

## INTEGRATING STRUCTURE AND LIGAND-BASED APPROACHES FOR MODELLING THE HISTONE DEACETYLASE INHIBITION ACTIVITY OF HYDROXAMIC ACID DERIVATIVES

HAI PHAM-THE<sup>1</sup>, HUONG LE-THI-THU<sup>2\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, Hanoi University of Pharmacy, 13–15 Le Thanh Tong, Hoan Kiem, Hanoi, Vietnam. <sup>2</sup>Department of Pharmacy and Pharmacology of Traditional Materials, School of Medicine and Pharmacy, Vietnam National University, Hanoi 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam. \*Email: Ltthuong1017@gmail.com

Received: 09 October 2017, Revised and Accepted: 10 November 2017

### ABSTRACT

**Objective:** Structure and ligand-based drug design approaches have been integrated to accurately predict the inhibition activity of hydroxamic acid (HA) derivatives against the histone deacetylase-2 enzyme (HDAC2).

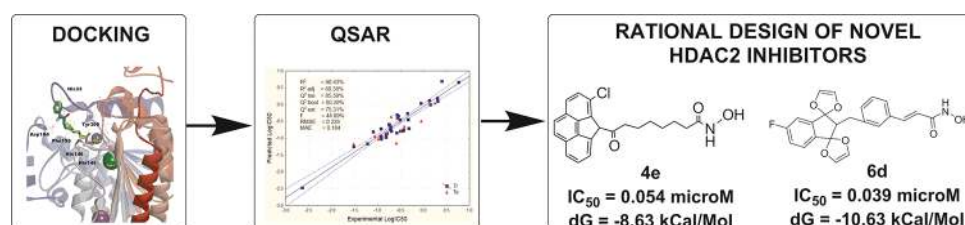
**Methods:** The “active conformations” of the ligands in the binding site of the enzyme were determined by docking assays. More than 1000 0–3 dimensional molecular descriptors included in Dragon package were calculated and utilized for developing quantitative structure-activity relationship (QSAR) models through a multiple linear regression approach coupled with the genetic algorithm (GA-MLR).

**Results:** The final model obtained showed suitable robustness and stability, with low correlation between descriptors and good predictive power. QSAR model was then used for screening bioactivity from a series of 36 novel HAs and found five candidates with very good bioactivity (half maximal inhibitory concentration < 0.1 μM). Docking experiment revealed the binding mode of these compounds into the active site of HDAC2. Drug-likeness and toxicity profiles of the compounds were checked through cheminformatics tools.

**Conclusion:** The results from this study can lead to rational design and synthesis of highly selective and potent HDAC2 inhibitors.

**Keywords:** Docking, Quantitative structure-activity relationship, Histone deacetylase, Rational drug design, Hydroxamic acid.

### GRAPHICAL ABSTRACT



© 2018 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2018.v11i2.22995>

### INTRODUCTION

Hydroxamic acids (HAs) belong to a very unique family of chemicals that show a broad spectrum of biological activities [1]. Their CONHOH moiety has been identified as a key functional group responsible to numerous therapeutic benefits, including anticancer, antiviral, and antioxidant, for treatment of cardiovascular diseases and Alzheimer's disease, to name but a few. The polypharmacology of HA moiety mainly relies on its ability to chelate with metal ions at the active site of numerous metalloenzymes, such as matrix metalloproteinases, peroxidases, and histone deacetylases [1]. In addition, the presence of nitrogen and oxygen lone pairs as well as hydroxyl and amine hydrogens enables HA to form multiple hydrogen bond interactions with its target [2].

One of the most widely investigated bioactivities of HAs is the potential inhibition of histone deacetylase (HDAC) [3]. HDACs comprise a family of 18 enzymes that control the *N*-ε-lysine deacetylation of histone to reestablish the positive charge in the *N*-terminus of the histone tails. Since then the histone-DNA interaction becomes so tight that the transcriptional machinery cannot access to the DNA template. Therefore, HDAC silencing or inhibition has been shown to

have an impact on cell cycle, cell growth, chromatin decondensation, cell differentiation, apoptosis, and angiogenesis in several cancer cell types [3]. Beyond cancer, modulators of HDAC activity are also of interest in other therapeutic areas, including cardiovascular diseases, inflammation, metabolic disorders, and neurological diseases [4,5]. In man, HDACs are commonly divided into four classes based on their homology to yeast HDAC proteins (Class I and IV vs. yeast Rpd3, Class II vs. yeast Hda1, and Class III vs. yeast Sir2). Among them, Class I HDACs, which comprise four isoforms (HDAC1-3 and 8) mainly found in the nucleus, have received extremely attention for the development of the novel anticancer targeted agents [6].

Up to date, two HA analogs have been approved for cancer therapy by the US food and drug administration (FDA), such as vorinostat (also known as suberoylanilide HA, [SAHA]) developed by Merck and Co., Inc., for use in patients with cutaneous T-cell lymphoma (CTCL), and belinostat developed by Spectrum Pharmaceuticals for the treatment of relapsed or refractory peripheral T-cell lymphoma [7]. Besides vorinostat and belinostat, several other HAs targeting HDACs have been subjected to different stages of clinical studies, such as abexinostat (PCI-24781), pracinostat (SB939), resminostat (RAS2410, 4SC-201), givinostat (ITF2357), quisinostat

(JNJ-26481585), panobinostat (LBH589), and CUDC-101 [7]. All of them share common pharmacophore features so that they can appropriately interact with HDACs in the active site region of the enzymes. In general, they include three parts: (1) A capping group (cap) generally consists of complex aromatic systems, (2) a zinc-binding group (ZBG) to chelate the catalytic Zn<sup>2+</sup> ion, and (3) a linker preferably bearing a n-alkyl, vinylic or aryl substituents that connect the Cap and ZBG groups [8].

The search for promising HDAC inhibitor candidates containing HA moiety remains one of the most active areas in medicinal chemistry today [3]. For the rational design of novel HAs, quantitative structure-activity relationship (QSAR) approaches have been extensively explored over the past 10 years [9]. QSAR models are generally developed based on the mathematical functions that relate the biological activity of a series of compounds with a set of appropriate molecular descriptors (MDs). Once constructed, QSAR models can be used to predict the bioactivity of newly designed compounds based on the descriptors calculated. Furthermore, the information translated from MDs is useful to rationalize the drug design process [10]. So far, several QSAR models predicting HDAC inhibitory activity have been developed for HA derivatives. The most relevant QSAR studies are summarized in Table 1. Note that, our survey only included the regression models developed for predicting HDAC inhibition activity of HA derivatives (compounds with CONHOH moiety).

As can be appreciated from Table 1, numerous two dimensional (2D) and 3D QSAR studies have been carried out to accurately predict HDAC inhibitory activity of novel HAs, as well as to identify key structural elements governing the bioactivity of analyzed compounds. However, previous studies on HAs have not focused on the inhibition activity toward specified HDAC isoform. In this aspect, QSAR model developed based on a specific HDAC isoform might be of pivotal importance to design potent, isoform-selective HDAC inhibitors [9,21].

In addition, various models were developed with very small dataset (<10 compounds) [18]. Consequently, the statistical significance and applicability of developed model become meaningless.

According to the report of Katritzky *et al.* [14], knowledge about the 3D geometrical conformation of compounds, including the important electrostatic and steric interactions, remained a key factor determining the accuracy of QSAR models. Pseudoreceptor methods, mainly 3D-QSAR techniques such as COMFA, COMSIA, and GOLPE, have been employed and showed higher performance in comparison with 2D-QSAR [14,18]. In this regard, structure-based methods, such as docking or molecular dynamics [22], are very useful to identify the "active conformation" of the ligands in the active sites of HDAC enzymes.

Taking all above issues into consideration, this study attempts to develop QSAR models able to accurately predict the inhibitory activity toward histone deacetylase-2 enzyme (HDAC2), which is an important isoform within Class I HDAC family [23]. To identify the active conformation of the ligands, HAs were firstly docked into the X-ray crystal structure of HDAC2 protein. Theoretical MDs were calculated for the 3D conformers and were used to find the correlation with bioactivity (half maximal inhibitory concentration [IC<sub>50</sub>]) through a multiple linear regression approach coupled with the genetic algorithm (GA-MLR). Finally, based on structural information extracted from QSAR models, several novel HA derivatives were designed and assessed the HDAC-2 inhibition activity.

## METHODS

### Dataset

A dataset of 46 HA derivatives was used to model the HDAC2 inhibitory activity (Fig. 1). These compounds were recollected from published

**Table 1: List of some QSAR models previously developed to predict HDAC inhibitory activity of hydroxamic acids**

HDAC targets	Data set size	Method	Model performance	Application	Reference
HDAC mixture	47 SAHA-like HAs	2D-QSAR	Best model R <sup>2</sup> =0.61, F=17.18	To determine structural characteristics essential for HDAC inhibition	Xie <i>et al.</i> [11]
Not specified HDAC	29 HAs	CoMFA CoMSIA	R <sup>2</sup> =0.98, R <sup>2</sup> <sub>cv</sub> =0.60 R <sup>2</sup> <sub>ext</sub> =0.95, R <sup>2</sup> <sub>cv</sub> =0.59	To predict HDAC inhibition activity	Guo <i>et al.</i> [12]
Not specified HDAC	57 HAs	CoMFA CoMSIA	R <sup>2</sup> <sub>cv</sub> =0.91, R <sup>2</sup> <sub>cv</sub> =0.50 R <sup>2</sup> =0.98, R <sup>2</sup> <sub>cv</sub> =0.53	To predict HDAC inhibition activity	Juvale <i>et al.</i> [13]
HDAC1/2 mixture	36 HAs	2D-QSAR Docking 3D-QSAR	R <sup>2</sup> =0.77, R <sup>2</sup> <sub>cv</sub> =0.72 R <sup>2</sup> =0.81, Q <sup>2</sup> =0.62, R <sup>2</sup> <sub>ext</sub> =0.72	To predict HDAC inhibition activity To determine structural characteristics essential for HDAC inhibition	Katritzky <i>et al.</i> [14]
Maize HDAC2	25 HAs	GOLPE 3D-QSAR	R <sup>2</sup> =0.96-98 Q <sup>2</sup> =0.81-85	To predict HDAC inhibition activity	Ragno <i>et al.</i> [15]
Not specified HDAC	55 HAs	Quantitative pharmacophore	Q <sup>2</sup> <sub>test</sub> =0.89	To determine structural characteristics essential for HDAC inhibition	Chen <i>et al.</i> [16]
Not specified HDAC	45 HAs	2D-QSAR	R <sup>2</sup> =0.78, R <sup>2</sup> <sub>adj</sub> =0.72, Q <sup>2</sup> =0.68	Virtual screening of HDAC inhibitors	Melagraki <i>et al.</i> [17]
Not specified HDAC	Varied data size (from 5-32 HAs)	2D-QSAR	R <sup>2</sup> =0.72-0.85 Q <sup>2</sup> =0.50-0.78	To predict HDAC inhibition activity	Pontiki and Hadjipavlou-Litina [18]
HDAC2	50 HAs	CoMFA CoMSIA	R <sup>2</sup> =0.98, R <sup>2</sup> <sub>cv</sub> =0.61 R <sup>2</sup> =0.99, R <sup>2</sup> <sub>cv</sub> =0.62	To design novel HDAC2 inhibitors	Xiang <i>et al.</i> [19]
Maize HDAC2	34 HAs	2D-QSAR 3D-QSAR	R <sup>2</sup> =0.85, Q <sup>2</sup> <sub>ext</sub> =0.79	To predict HDAC inhibition activity	Sharma and Sharma [20]

QSAR: Quantitative structure-activity relationship, HDAC: Histone deacetylase

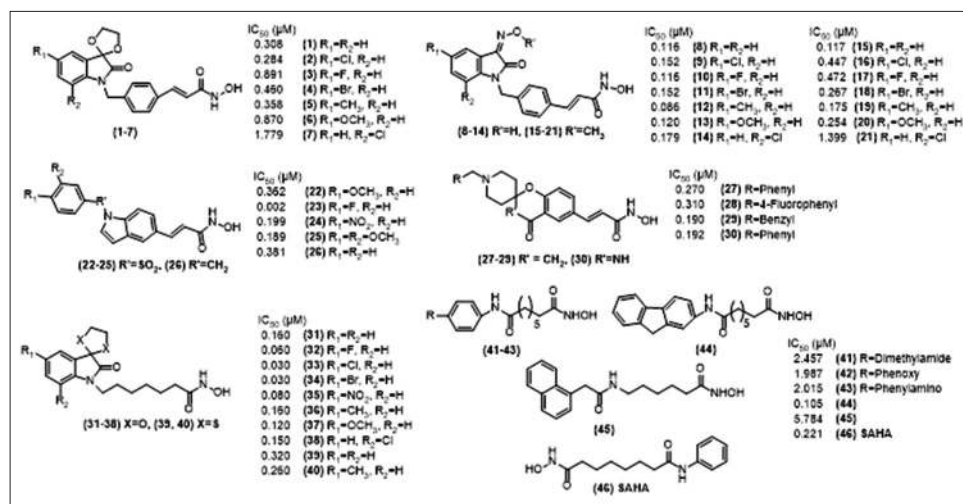


Fig. 1: Structures and bioactivity of hydroxamic acids included in the dataset

papers [24-28], taking into account the homogeneity of experimental conditions. Accordingly, the half maximal inhibitory concentration (IC<sub>50</sub>, µM) was determined using fluorogenic HDAC assay kit (BPS bioscience) with the same protocol [29]. In some experiments, SAHA was used as reference compound. Its IC<sub>50</sub> values showed slight variability ranging from 0.197 to 0.265 µM [24,26,27]. Therefore, we took the mean value, discarding those out of the (mean±SD) range. All the data were transformed into logarithmic form (pIC<sub>50</sub>=lgIC<sub>50</sub>).

#### Preparation of the ligands

The structures of 46 molecules were generated using MarvinSketch version 15.11.9 [30]. The 3D structures of the compounds were created and subsequently subjected to energy minimized to within an rms gradient of 0.1 kcal/mol/Å using MOE 2009.10 package, setting the force field to the 94s variant of the Merck molecular force field 94s [31]. Various studies have suggested negative hydroxamate-Zn<sup>2+</sup> coordination in the active site of HDACs [32] and linked with several experimental results. Therefore, the HA inhibitors were deprotonated after importing the 3D structures to MOE database.

#### Selection and preparation of the protein structure

It has been shown that histone-H3 and histone-H4 deacetylation is regulated principally by HDAC2 and HDAC3. Recently, the crystal structure of HDAC2 in complex with SAHA (PDB ID: 4LXZ) has been reported by Lauffer *et al.* [23], so we decided to select this structure as a docking template to study the interaction between HAs and HDAC2. After removing SAHA from the complex structure, the enzyme was prepared using ICM-pro version 3.8-3 [33]. All water molecules were removed, and the polar hydrogen atoms were aided to the protein. The active site of the enzyme includes key residues His145, His146, Asp181, His183, and Asp269, which were appropriately protonated according to the general catalytic transformations of zinc metalloenzymes described by Pottel *et al.* [34]. Afterward, the binding site was energy minimized applying AMBER99 force field based on truncated Newton method in MOE (convergence criteria of 0.01 Å root-mean-square deviation [RMSD]) [31].

#### Docking and binding affinity estimation procedures

For the docking experiments, the flexible-ligand rigid-protein simulations were performed using MOE [31], where flexible ligand conformations were generated using the Monte Carlo algorithm. The finally selected poses were rescored using London scoring and then by affinity scoring functions to estimate the free energy of binding of the ligands from the given poses (dG, kcal/mol) [35,36]. In the refinement stage, the energy minimization of the system was carried out using the molecular mechanics force-field method [31]. The pose with the best refining score of each ligand was retained and saved as MOL2 format

using Chimera [37]. All the other parameters were kept as default. This approach allows identifying suitable conformation for HA derivatives, which were further used in the MLR computation.

#### QSAR model development

In the present study, the parameters corresponding to 0-1D, 2D, and 3D molecular descriptors (MDs) were computed by Dragon 6.0 software [38]. In summary, a total of more than 1400 MDs, belonging to 18 classes, were calculated. By disregarding descriptors with constant or near constant values inside each class, a subset of 985 descriptors was retained. Besides, to decrease the redundancy existing in the descriptor data matrix, the descriptor correlation with each other was examined, and the collinear descriptors (r>0.9) were detected. Finally, only 617 descriptors were retained for each compound.

After calculating MDs, dataset was randomly separated into two parts: Training and test sets, which correspond to 75% (34 compounds) and 25% (12 compounds), respectively. Training set is for QSAR model development, and the test set was set aside to assess the predictive power of the final model.

MLR technique was used for model construction. Basically, MLR analysis consists of deriving a mathematical function (1), which best describes the activity Y (in this case is proteasome inhibitory potency LogIC<sub>50</sub>), as a linear combination of a subset of X-variables (the selected molecular descriptors) with the regression coefficients b<sub>n</sub> and the intercept b<sub>0</sub>.

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

In this study, a MLR approach coupled with the GA-MLR was performed using Mobydigs package versus 1.1 [39]. Accordingly, a random population of binary chromosomes (descriptor subsets) was generated. At each chromosome, the number of gene is equal to the number of descriptors. Values 1 or 0 is used to assign the gene in which corresponding descriptor was included or not, respectively. Subsequently, the next generation of the population is reproduced based on three operations: Selection, cross-over, and mutation [40]. The fitness criterion of squared predictive correlation coefficient (Q<sub>Loo</sub><sup>2</sup>) provided by the leave-one-out cross-validation (LOO-CV) procedure was used for population selection. In the crossover step, each pair of chromosomes is individually divided, mutually exchanged, and merged. The probability of crossover (pr) was randomly set to be 30, 50, 70, and 90%. In the third operation, a binary bit string randomly selected from chromosomes will be mutated. Hence, the probability of mutation (pm) was set at low level (0.4, 0.6, and 0.8%) to improve the overall fitness in the population of chromosomes. Finally, a population of 100 regression models was retained taking into account the Q<sub>Loo</sub><sup>2</sup>



coefficients. The GA was stopped when 90% of generation took the same fitness.

The principle of maximal parsimony (Occam's razor) was taken into account as a strategy for model selection [41]. Thus, we selected the models with the highest performance but having a few parameters ( $b_j$ ) as possible.

#### Model validation and applicability domain

Goodness of fit was assessed by examining the determination coefficient,  $R^2$ , the adjusted determination coefficient,  $R_{Adj}^2$ , the standard deviation (s), Fisher's statistic, F, as well as the ratio between the number of compounds and the number of adjustable parameters in the model, known as the  $\rho$  statistics. Besides these classic regression parameters, other important indices were adopted, i.e., the Kubinyi function and Akaike's information criterion [42], which give enough criteria for comparing models with different parameters, number of variables, and number of points. The robustness and predictive ability of the model were evaluated by LOO-CV and bootstrapping analyses. LOO-CV procedure consists of removing one data point from the training set, constructing the model only on the basis of the remaining training data, and then testing on the removed point. In this way, all of the training data points were tested and  $Q_{LOO}^2$  then calculated. In the bootstrapping procedure, K n-dimensional groups are generated by a randomly repeated selection of n-objects from the original dataset [43]. The model obtained on the first selected objects is used to predict the values for the excluded sample, and then,  $Q_{Boots}^2$  is calculated for each model. The bootstrapping was repeated 5000 times for each validated model.

For further application of QSAR models, the applicability domain (AD) was taken into account. The leverage approach was used to establish model AD [44]. Data outside the domain will be considered as extrapolated and predicted with low certainty.

The leverage ( $h_i$ ) [44] measures the distance from one molecule to the whole training set (TS) in the descriptors space and is calculated as follows:

$$h_i = \mathbf{t}_i(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{t}_i^T \quad (i=1, \dots, N) \quad (2)$$

Where  $\mathbf{t}_i$  is the descriptor vector of that compound and  $\mathbf{T}$  is the model matrix derived from the TS descriptor values. A "critical leverage" value ( $h^*$ ) is generally fixed at  $3 \times p/N$ , where  $p$  is the number of model variables plus one, and  $N$  is the number of the objects used to calculate the model.

Finally, to visualize the AD of a QSAR model, the plot of standardized cross-validated residuals (Stand. R) versus Leverage (Hat diagonal) values ( $h$ ) also known as the Williams plots can be used for an immediate and simple graphical detection of both the response outliers (i.e., compounds with standardized residuals greater than three standard deviation units) and structurally influential chemicals in mode ( $h > h^*$ ) [45].

#### Drug-likeness and toxicity profiling

Taking into account the importance of drug-likeness evaluation in early drug discovery phases, some end points related to the absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) processes were calculated using ADMETsar server (<http://lmmd.ecust.edu.cn/admetsar1/>) [46] and Volsurf 1.0.4+ software [47]. The important questions analyzed were as follows: Could the new compounds be applied through oral administration? Which are the main metabolism pathways of the compounds through intestinal and hepatic cells? and Do the molecules present any risk factor?

## RESULTS AND DISCUSSION

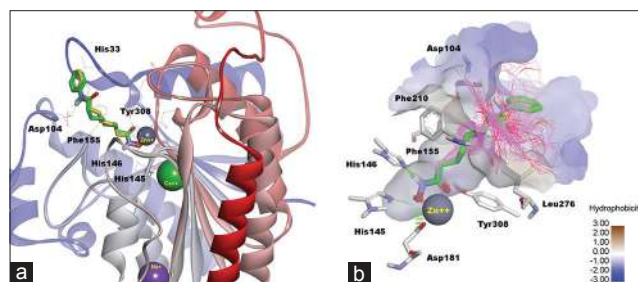
#### Docking results

In this study, the structure coordinates of human HDAC2 protein reported by the study of Lauffer *et al.* (PDB ID: 4LXZ) that showed a

good stereochemical quality were used as docking protein model [23]. In detail, the reported resolution of the entry is 1.85 Å and the reliability factor, which presents the agreement between the crystallographic model and the experimental X-ray diffraction data, is 0.161. According to the Protein Data Bank, the protein structure comprises 3 chains (A, B, and C) with the same length (368 amino acids). In particular, chain A coincides with the core domain of UniProt sequence Q92769 at residues number 8 up to 376. This chain was used for docking assays.

The first step was to demonstrate the suitability of ligand preparation procedures and docking protocol by redocking the X-ray cocrystal hydroxamate HDAC2 inhibitor (SAHA) into the active site. The accuracy of pose coordinates between the native and redocked ligand was evaluated by the RMSD. The results showed that the difference between cocrystallized and docked SAHA compounds is insignificant, displaying quite low value of RMSD (0.672 for coordinates of atoms involved in hydrogen bonding and 1.127 for coordinates of all heavy atoms). It is important to note that RMSD value  $< 2.0 \text{ \AA}$  between the docked ligand and the X-ray pose has been widely taken into account as acceptance for an accurate docked structure, according to Gohlke *et al.*'s criterion for examining the errors in crystallographic data [48]. In addition, the conservation of hydrogen bond interactions with the residues Asp104, His145, His146, Asp181, and Tyr308, as well as hydrophobic interactions with Phe155 and Phe210, can be observed (Fig. 2a). Especially, the similarly close distances ( $< 2.5 \text{ \AA}$ ) between hydroxamate groups and zinc ion, which play an important role in the zinc-binding motif of HDAC inhibitors [34], are conserved. Considering these results, our protocol is suitable for using in further docking experiments.

In the next steps, the remaining hydroxamic acids in the database were docked into the HDAC2 binding site following the protocol described above. The results highlighted the key interactions, which include: Non-covalent interactions of hydrogen bonding and electrostatic and van der Waals interactions. The active site comprises a narrow pocket located deep within the protein and contains the zinc metal cofactor. The metal was pentacoordinated by His183, Asp181, and Asp269 [32]. Docking results also showed a bidentate coordination through the carbonyl and hydroxyl groups of the ligand hydroxamate moieties (Fig. 2b). In addition, the carbonyl oxygen of the HA group formed an H bonding with the hydroxyl group of the Tyr308 residue, while the hydroxyl of the HA group formed another H bond with His146. In case of SAHA and some other compounds, this H bond was instead donated by His145. According to the catalytic mechanism of HDAC inhibition, His145, His146, Asp181, His183, Asp269, and Tyr308 are responsible for the stabilization of the substrate in the binding site and form part of the charge relay system necessary for the zinc-dependent hydrolysis of the acetylated lysine substrates [49]. Moreover, hydrophobic interactions involved in holding ligands within the active site include interactions between the linker moieties and aromatic rings of Phe155 and Phe210. In



**Fig. 2: (a) Superposition of co-crystal (yellow) and redocked suberoylanilide hydroxamic acid (SAHA) structures (green) at the histone deacetylase2 (HDAC2's) binding site. (b) Ligand alignment inside the binding pocket of HDAC2. Compounds are present in line and SAHA (green) in stick styles. Common hydrogen bond interactions are shown in green dash-line**

addition, bulky aromatic systems of capping groups formed both van der Waals and H bond interactions with numerous residues at the rim of the pocket, such as Pro34, Glu208, Tyr209, and Arg275. Finally, it is important to note that the docking affinity dG scores of hydroxamic inhibitors ranked ligands in accordance with the experimental  $IC_{50}$  data.

#### Development and selection of GA-MLR QSAR-models

The results from docking assays revealed the preferred orientation of HA derivatives in the active site of HDAC2. Such the "active conformations" were used for further development of QSAR models.

Herein, MLR models were obtained employing GA for the feature selection process. The fitting and predictive criteria for the best seven models after  $10^5$  iterations are given in Table 2. Note that, the maximum number of variables to be included in the final model was set to 6 as we have 34 observations in the training set, taking into account the Austin's criteria ( $5 \times \text{number of variables} \leq \text{number of observations}$ ) to avoid overfitting problem [50].

The best models were further analyzed and compared based on the quality of the statistical parameters. These parameters include correlation between the variables and the response ( $R^2$ ,  $K_{xy}$ ), robustness of the models ( $Q^2_{LOO}$ ,  $Q^2_{Bootstrap}$ ), significance of the model ( $p$ ), and low standard deviation ( $s$ ), among others.

As can be revealed from Table 2, model M1 displayed the highest performance and was selected as the final model for further application. The mathematical formula is given as follows:

$$pIC_{50} = 1.527 + 1.099 \text{ ve1\_B(m)} - 0.053 \text{ rdf105m} - 0.190 \text{ Mor02m} + 14.166 \text{ R7m}^+ - 96.459 \text{ R5v}^+ + 0.492 \text{ B07[N-O]} \quad (3)$$

The model presented very good fit with a  $R^2$  of 0.90 and adjusted  $R^2_{adj}$  of 0.88, which is significantly higher than the other models. In addition, Friedman index (LOF) computed for this model is quite low (0.078), indicating a lack of overfitting problem [51].

The autocorrelation between descriptors was checked through Pearson correlation matrix and the correlation among descriptors ( $K_{xx}$ ). As the results, no significant correlation between MDs was detected, as the highest pairwise correlation is between VE1\_B(m) and Mor02m (0.54) and the  $K_{xx} = 0.34$  (Table 3). In addition, variance inflation factors (VIF) and tolerance values were computed for the descriptors included in the model M1. Based on the results, we confirm the absence of multicollinearity in the finally selected model [52].

#### Validation of QSAR model

The stability of the models was evaluated using internal validation procedures. At first, the statistical parameters of the LOO-CV were checked. According to the obtained results, it is remarked that the internal predictions are good because the variance explained in the prediction by LOO ( $Q^2_{LOO} = 0.856$ ) is comparable with an  $R^2$  coefficient.

Another validation technique included in Mobydig is bootstrapping [39], which used the bootstrap resample approach to evaluate the behavior of the model in different training data distribution. Unsurprisingly, the average bootstrapping accuracy  $Q^2_{Bootstrap}$  was 0.803, quite similar to  $R^2$ , which referred to a very stable model.

At last, to demonstrate that the model was not the result of a chance in fitting given data, the Y-scrambling procedure was employed, placing the answers (dependent variable) at random, keeping all descriptors (independent variables) in the model, and then, performing whole model building procedure  $n$  times again [53]. If the model performance decays dramatically, it is a sign of a lack of chance. Mobydig provides the option to process Y-scrambling with extent iterations ( $n > 1000$ ) and output the average fitting ( $R^2_{Y-scr}$ ) and cross-validation ( $Q^2_{Y-scr}$ ) parameters. As can be observed in Table 2, the results of Y-scrambling procedure allow us to confirm the quality of correlation found, as the parameters are much lower than the original values of the model ( $R^2_{Y-scr} = 0.116$  and  $Q^2_{Y-scr} = -0.433$ ).

Taking all above results into consideration, we confirmed that the model M1 has a good fit, robustness, and stability. Finally, the predictive ability of the model was checked using an external set of 11 remaining HA derivatives. Before external validation of this model, applicability domain was defined. As can be seen in Fig. 3b, the critical leverage was determined as 0.6, and no molecules from the external test set were detected out of the AD of the model. The external prediction accuracy of the model M1 was more than 0.76, similar to the cross-validation and bootstrapping validation, suggesting a suitable goodness of prediction ability of the model.

#### Mechanistic interpretation

As can be seen in the mathematical equation (3), the model M1 includes 6 variables. Table 4 summarized the meaning of these MDs. At first sight, there existed a predominant role of 3D descriptors, which partly explained the importance of finding the active conformations through docking. Given the difficulty of interpreting 3D descriptors involved in QSAR model, a mechanistic interpretation of these variables was carried out as well.

Table 2: Performance of the best model series constructed by GA

Model	No var.*	$R^2_{Training}$	$R^2_{adj}$	$Q^2_{LOO}$	$Q^2_{Bootstrap}$	$Q^2_{ext}$	$R^2_{Y-scr}$	$Q^2_{Y-scr}$	$K_{xx}$	$s$	F
M1	6	0.904	0.884	0.856	0.803	0.763	0.116	-0.433	0.337	0.206	0.441
M2	6	0.866	0.838	0.819	0.744	0.758	0.109	-0.416	0.413	0.243	0.302
M3	6	0.845	0.811	0.765	0.676	0.751	0.141	-0.396	0.390	0.262	0.254
M4	5	0.844	0.817	0.783	0.723	0.764	0.099	-0.316	0.398	0.258	0.313
M5	5	0.849	0.823	0.788	0.726	0.704	0.087	-0.379	0.352	0.254	0.326
M6	5	0.788	0.778	0.705	0.592	0.695	0.092	-0.344	0.406	0.301	0.216
M9	4	0.787	0.758	0.716	0.643	0.709	0.075	-0.297	0.399	0.297	0.277

\*Number of variables selected for MLR models. MLR: Multiple linear regression, GA: Genetic algorithm

Table 3: Correlation matrix, tolerance, and VIF of the selected descriptors included in M1 model

	VE1_B(m)	RDF105m	Mor02m	R7m <sup>+</sup>	R5v <sup>+</sup>	B07[N-O]	Tolerance	VIF
VE1_B(m)	1						0.557	1.796
RDF105m	0.176	1					0.812	1.232
Mor02m	0.541	0.139	1				0.591	1.692
R7m <sup>+</sup>	-0.149	0.309	-0.074	1			0.664	1.506
R5v <sup>+</sup>	-0.422	0.029	-0.474	0.456	1		0.499	2.006
B07[N-O]	0.030	-0.046	-0.069	-0.099	-0.258	1	0.873	1.146

VIF: Variance inflation factor

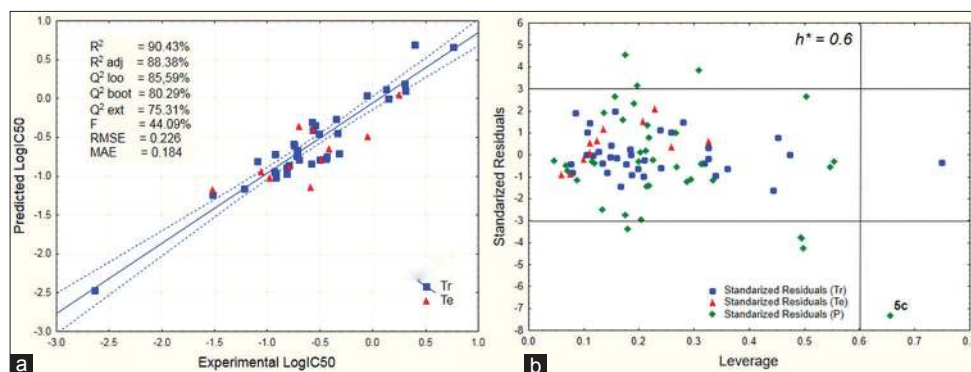


Fig. 3: (a) Experimental versus predicted p half maximal inhibitory concentration values for the M1 model, (b) Williams plot predicted by the final M1 model. Tr: Training set, Te: Test set, and P: Prediction set

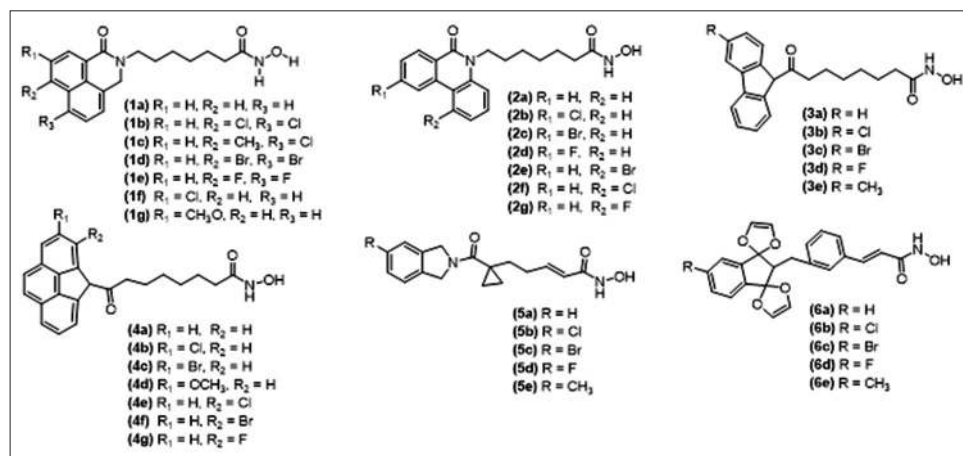


Fig. 4: Chemical structures of newly designed 36 hydroxamic acids

Table 4: Symbols and meaning of all the MD variables selected by the best model

Symbol	Descriptor family	Meaning
VE1_B(m)	2D (burden) matrix-based descriptors	Coefficient sum of the last Eigen vector from burden matrix weighted by mass
RDF105m	3D RDF descriptors	Radial distribution function - 105/weighted by mass
Mor02m	3D-MORSE descriptors	Signal 02/weighted by mass
R7m <sup>+</sup>	3D GETAWAY descriptors	R maximal autocorrelation of lag 7/weighted by mass
R5v <sup>+</sup>	3D GETAWAY descriptors	R maximal autocorrelation of lag 5/weighted by van der Waals volume
B07[N-O]	2D atom pairs	Presence/absence of N - O at topological distance 7

RDF105m is a radial distribution function index, which is the probability distribution to find an atom in a spherical volume of radius  $R$  around a group of atoms  $A$ . RDF105m indicates that, within a radius of 10.5 Å, the RDF value increases with the atomic masses and number of atoms. Likewise, Mor02m belongs to the group of geometrical descriptors that provide 3D information from the 3D coordinates using the electron diffraction transformation to prepare theoretical scattering curves [54]. This descriptor is calculated based on the atomic mass in 4<sup>th</sup> electron diffraction signal. Both RDF105m and Mor02m contributed positively to the activity.

R7m<sup>+</sup> and R5v<sup>+</sup> are 3D GETAWAY descriptors, which are derived from the molecular influence matrix (MIM) and atoms using atomic mass and van der Waals volume weighting schemes, respectively. Both descriptors selected belong to R-GETAWAY subfamily where the MIM is combined with the geometry matrix in a influence/distance matrix,  $[R]$ , defined as:  $[(h_{ii}h_{jj})/r_{ij}]_{ij}$  where the diagonal elements  $h_{ii}$  and  $h_{jj}$  are termed leverages and represent the influence of each atom in determining the shape of the molecule. Each off-diagonal element  $h_{ij}$  represents the degree of accessibility of the  $j^{\text{th}}$  atom to

interactions with the  $i^{\text{th}}$  atom [54]. Interestingly, R7m<sup>+</sup> and R5v<sup>+</sup> are associated with large and opposite coefficient signals. R5v<sup>+</sup> contributes positively to the activity, while R7m<sup>+</sup> has a negative contribution to the HDAC2 inhibitory bioactivity.

There are two topological descriptors negatively contributed to the activity, including VE1\_B(m) and B07[N-O]. VE1\_B(m) is a 2D matrix-based descriptor, derived from a Burden matrix weighted by mass, and depends in a complex way from molecular size, shape, and presence of heavy heteroatoms and multiple bonds. B07[N-O] in turn is a 2D atom pairs descriptor, which is equal to 1, if there is at least one pair of N and O atoms separated by 7 bonds [54].

#### Rational design of new HA derivatives

Based on the previous discussion of the role of the molecular descriptors involved in the QSAR model, six new hydroxamate series were designed based on the structures of compounds in the training set (Fig. 4).

The design strategy was mainly focused on the Cap group. It is postulated that Cap modifications could lead to the isoform selective inhibitory activity of HAs and that the complexity of aromatic



capping group might positively contribute to the activity [21]. Therefore, 6 new aromatic systems were incorporated into the Cap in this study, including 2,3-dihydro-1*H*-benzo[*de*]isoquinoline (1a-g), 6-(5*H*)-phenanthridinone (2a-g), 9*H*-fluorene (3a-e), 4*H*-cyclopenta[*def*]phenanthrene (4a-g), 2-substituted isoindoline (5a-e), and 2'*H*-dispiro[[1,3]dioxole-2,1'-indene-3';2''-[1,3]dioxole] (6a-e).

Before evaluating the bioactivity of newly designed HAs, we calculated their leverage and standardized residuals to detect possible compounds outside the AD. As can be seen in Fig.3b, compound 5c was identified as outlier, so it was excluded from the bioactivity screening assays by

QSAR model to avoid unreliable extrapolation. The predicted  $IC_{50}$  of the remaining 35 compounds was shown in (Fig. 5). In addition, to select the best inhibitors, the energy of binding affinity calculated was considered.

Five compounds (2f, 4e, 6b, 6e, and 6f) were identified as potent inhibitors as their predicted  $IC_{50}$  values were all below 0.1  $\mu$ M. Especially, compounds 4e and 6d displayed a five-to seven-fold more potent than reference drug SAHA (experimental  $IC_{50}$ =0.2–0.3  $\mu$ M). The docking affinity values computed for these 5 compounds ranged from  $-8.3$  to  $-9.9$  kCal/mol, significantly higher than that of SAHA ( $dG$ = $-7.4$  kCal/mol).

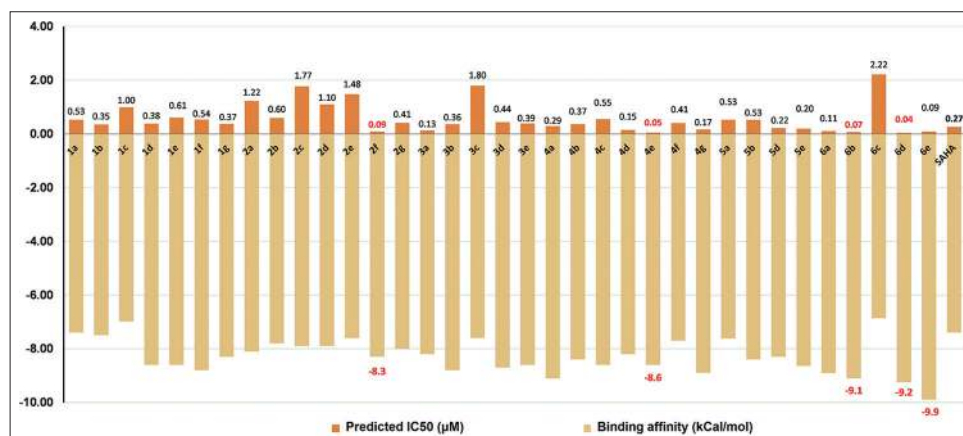


Fig. 5: Predicted half maximal inhibitory concentration ( $\mu$ M) by quantitative structure-activity relationship model and binding affinity  $dG$  (kCal/mol) by MOE docking experiments for 35 hydroxamic acids designed

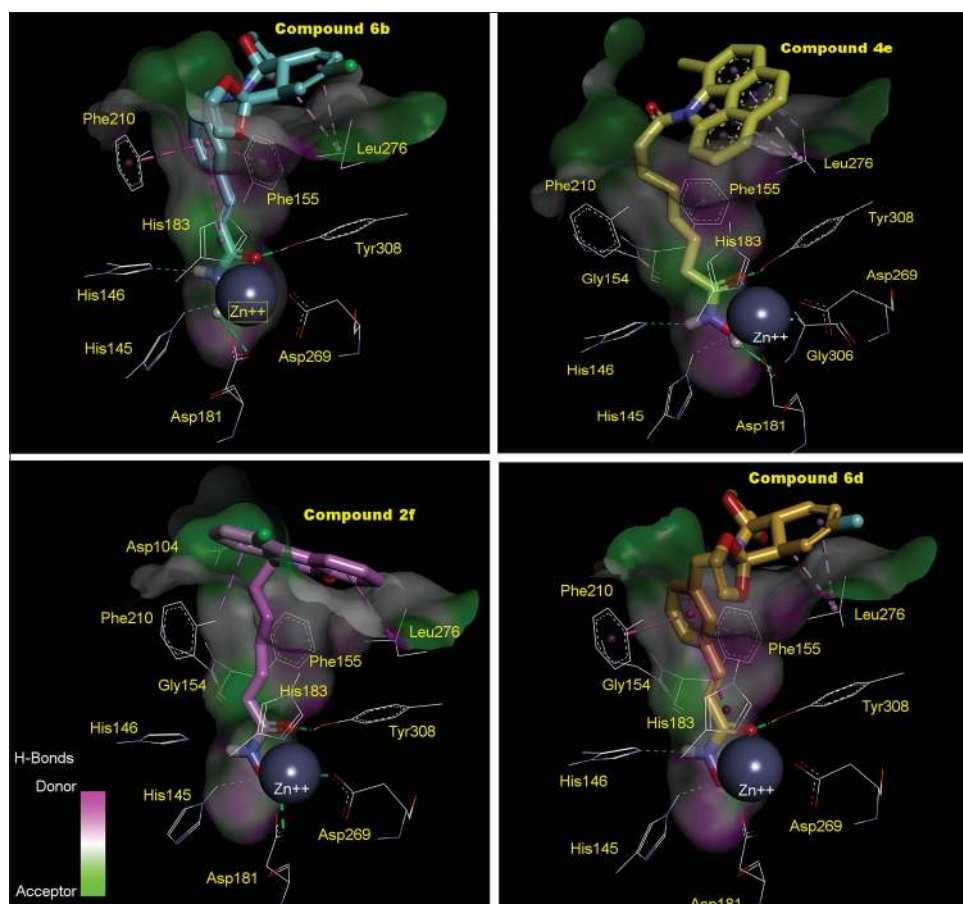


Fig. 6: Interactions of four compounds (2f, 4e, 6b, and 6e) with the active site residues of histone deacetylase2

Table 5: ADME-Tox profiling of selected hit compounds

Cpd.	DL <sup>a</sup>	LL <sup>b</sup>	Solubility <sup>c</sup> (mg/ml)	LogP <sub>app</sub> <sup>d</sup> (cm/s)	HIA <sup>e</sup>	BBB <sup>f</sup> permeant	P-gp subs. <sup>g</sup>	P450 CYP subs. <sup>h</sup>	Toxicity
2f	0/5	3/3	1.08×10 <sup>-2</sup>	0.112	High	Yes	No	3A4	Tetrahymena pyriformis
4e	0/5	3/3	2.64×10 <sup>-3</sup>	0.583	High	No	Yes	3A4	AMES toxic Fish toxicity Tetrahymena pyriformis
6b	0/5	1/3	1.20×10 <sup>-2</sup>	0.611	High	No	Yes	No	Fish toxicity Tetrahymena pyriformis
6d	0/5	1/3	3.10×10 <sup>-3</sup>	0.573	High	No	Yes	No	
6e	0/5	1/3	2.23×10 <sup>-2</sup>	0.529	High	No	Yes	No	

<sup>a</sup>Drug-likeness: Number of Lipinski's Rule of Five violations (over 5), <sup>b</sup>lead-likeness: Number of Teague's lead-like rules (over 3), <sup>c</sup>intrinsic solubility at 25°C calculated by volsurf+version 1.0.4 distributed by Molecular Discovery Ltd., London, U.K. (<http://www.moldiscovery.com>) [47], <sup>d</sup>Caco-2 cell permeability, <sup>e</sup>human intestinal absorption class, <sup>f</sup>blood-brain barrier permeation ability, <sup>g</sup>p-glycoprotein efflux substrate state, <sup>h</sup>substrate status of different P450 cytochrome enzymes, ADME-Tox: Absorption, distribution, metabolism, excretion, and toxicity

Docking results showed that all the compounds formed numerous H bonds with the residues in the active site of HDAC2. Commonly, 4 H-bonds were generated between HA group and His145, His146, Asp181, and Tyr308. This group also interacted with zinc ion in the same manner that SAHA did. A different situation was observed when analyzing hydrophobic interaction. The designed compounds have bulky hydrophobic Cap group that could form extent pi-alkyl stacking interactions with Asp104, Leu276, and Phe210 (Fig. 6). Especially, compounds 6b, 6d, and 6e exhibited a very high binding affinity and intensive hydrophobic interactions. The substitution of styrene moiety at the linker clearly enhanced interactions with the residues in the narrow tunnel of the binding pocket, including Phe155 and Phe210. Numerous studies have indicated that these pi-pi stacking interactions play significant role for increasing the bioactivity of HAs [3,8,27].

At last, we selected five most bioactive compounds to further investigate the ADME-Tox properties (Table 5). All the compounds exhibited acceptable intestinal absorption profile. However, their solubility appears to be a limiting factor. Intestinal membrane permeability was predicted to be moderate-to-high, according to 3Prule developed by Pham-The *et al.* [55]. Compound 2f showed good disposition ability in the central nervous system and was predicted as non-substrate characteristics of P-glycoprotein, a member of the MDR/TAP subfamily involved in multidrug resistance [56]. However, compounds 2f and 4e were identified as potential substrates of cytochrome CYP3A4 enzyme [46]. According to the three lead-likeness rules of Teague [57], they were identified as non-lead compounds.

On the other hand, compound 4e presented numerous toxic risks, including genotoxicity against Ames test, reproductive toxicity against fish fertilizer test, and bioaccumulation in *Tetrahymena pyriformis* [46]. Compounds 6b, 6d, and 6e displayed similar toxicity to 4e. However, they were predicted as possible lead compounds and non-substrates of cytochrome enzymes using Teague's rules and ADMESar server [46,57].

Taking all together, we concluded that 2f, 4e, 6b, 6e, and 6f displayed certain advantages and limitations (related to possible first-pass metabolism and toxicity) for being developed as drug candidates. In particular, compounds 6b, 6d, and 6e could be considered as a good starting point for further design and synthesis of novel HDAC2 inhibitors.

## CONCLUSION

The present study has demonstrated a versatile strategy for combining structure and ligand-based approaches to model the inhibition activity of HDAC2 enzyme. As the main results, a QSAR model was obtained on the basis of the active conformations of hydroxamates on the active site of HDAC2 identified by docking experiments. The model showed very encouraging performance according to the numerous internal and

external validation procedures. Another important result was the design of several novel hydroxamates targeting HDAC2. From the structures design, we identified five hit compounds with bioactivity significantly more potent than SAHA, a commercial drug currently used for cancer treatment. Molecular modelling and ADME-Tox profiling revealed that these compounds should be considered for further development of new potent HDAC inhibitor.

## ACKNOWLEDGMENT

The authors acknowledge the financial support of Vietnam National University, Hanoi-VNU (Grant No. QG.16.24).

## AUTHOR CONTRIBUTION

H L-T-T and H P-T designed the study. H P-T developed QSAR models and wrote the manuscript. H L-T-T validated the molecular docking studies and revised the manuscript.

## CONFLICTS OF INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

## REFERENCES

- Gupta SP, Sharma A. The chemistry of hydroxamic acids. In: Gupta SP, editor. Hydroxamic Acids: A Unique Family of Chemicals with Multiple Biological Activities. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 1-17.
- Politzer P, Murray JS. Some intrinsic features of hydroxylamines, oximes and hydroxamic acids: Integration of theory and experiment. PATAI'S Chemistry of Functional Groups. Chichester, West Sussex PO19 8SQ, UK: John Wiley & Sons, Ltd; 2009.
- Thaler F, Patil VM, Gupta SP. Hydroxamic acids as histone deacetylase inhibitors. In: Gupta SP, editor. Hydroxamic Acids: A Unique Family of Chemicals with Multiple Biological Activities. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 99-151.
- Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: Implications for disease and therapy. Nat Rev Genet 2009;10:32-42.
- Islam T. Crucial challenges in epigenetic cancer therapeutic strategy yet to be resolved. Int J Pharm Pharm Sci 2016;7:326-31.
- Falkenberg KJ, Johnstone RW. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. Nat Rev Drug Discov 2014;13:673-91.
- Mottamal M, Zheng S, Huang TL, Wang G. Histone deacetylase inhibitors in clinical studies as templates for new anticancer agents. Molecules 2015;20:3898-941.
- Whitehead L, Dobler MR, Radetich B, Zhu Y, Atadja PW, Claiborne T, *et al.* Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors. Bioorg Med Chem 2011;19:4626-34.
- Pham-The H, Casañola-Martin G, Diéguez-Santana K, Nguyen-Hai N,



- Ngoc NT, Vu-Duc L, *et al.* Quantitative structure-activity relationship analysis and virtual screening studies for identifying HDAC2 inhibitors from known HDAC bioactive chemical libraries. SAR QSAR Environ Res 2017;28:199-220.
10. Humberto GD. Editorial (Hot Topic: Bioinformatics and Quantitative Structure-Property Relationship (QSPR) Models). Curr Bioinform 2013;8:387-9.
  11. Xie A, Liao C, Li Z, Ning Z, Hu W, Lu X, *et al.* Quantitative structure-activity relationship study of histone deacetylase inhibitors. Curr Med Chem Anticancer Agents 2004;4:273-99.
  12. Guo Y, Xiao J, Guo Z, Chu F, Cheng Y, Wu S, *et al.* Exploration of a binding mode of indole amide analogues as potent histone deacetylase inhibitors and 3D-QSAR analyses. Bioorg Med Chem 2005;13:5424-34.
  13. Juvale DC, Kulkarni VV, Deokar HS, Wagh NK, Padhye SB, Kulkarni VM, *et al.* 3D-QSAR of histone deacetylase inhibitors: Hydroxamate analogues. Org Biomol Chem 2006;4:2858-68.
  14. Katritzky AR, Slavov SH, Dobchev DA, Karelson M. Comparison between 2D and 3D-QSAR approaches to correlate inhibitor activity for a series of indole amide hydroxamic acids. QSAR Comb Sci 2007;26:333-45.
  15. Ragno R, Simeoni S, Valente S, Massa S, Mai A. 3-D QSAR studies on histone deacetylase inhibitors. A GOLPE/GRID approach on different series of compounds. J Chem Inf Model 2006;46:1420-30.
  16. Chen YD, Jiang YJ, Zhou JW, Yu QS, You QD. Identification of ligand features essential for HDACs inhibitors by pharmacophore modeling. J Mol Graph Model 2008;26:1160-8.
  17. Melagraki G, Afantitis A, Sarimveis H, Koutentis PA, Kollias G, Igglessi-Markopoulou O, *et al.* Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors. Mol Divers 2009;13:301-11.
  18. Pontiki E, Hadjipavlou-Litina D. Histone deacetylase inhibitors (HDACIs). Structure-activity relationships: History and new QSAR perspectives. Med Res Rev 2012;32:1-165.
  19. Xiang Y, Hou Z, Zhang Z. Pharmacophore and QSAR studies to design novel histone deacetylase 2 inhibitors. Chem Biol Drug Des 2012;79:760-70.
  20. Sharma MC, Sharma S. Molecular modeling study of uracil-based hydroxamic acids-containing histone deacetylase inhibitors. Arab J Chem 2015;[Article In Press].
  21. Bieliauskas AV, Pflum MK. Isoform-selective histone deacetylase inhibitors. Chem Soc Rev 2008;37:1402-13.
  22. Rajaganapathy K, Sathiyasundar R, Ramesh GK, Kalaichelvan VK. Designing of anti-cancerous histone deacetylase inhibitors through mimicking of protein-protein interfaces. Int J Pharm Pharm Sci. 2014;6:208-12.
  23. Lauffer BE, Mintzer R, Fong R, Mukund S, Tam C, Zilberlyb I, *et al.* Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability. J Biol Chem 2013;288:26926-43.
  24. Mai DD, Phuong DP, Kim OD, Khac VT, Hyunggu H, Woo HP, *et al.* Exploration of novel 5'(7)-substituted-2'-oxospiro[1,3]dioxolane-2,3'-indoline-based N-hydroxypropenamides as histone deacetylase inhibitors and antitumor agents. Arab J Chem 2015;10:465-72.
  25. Chun PS, Kim WH, Kim JS, Kang JA, Lee HJ, Park JY, *et al.* Synthesis and importance of bulky aromatic cap of novel SAHA analogs for HDAC inhibition and anticancer activity. Bull Korean Chem Soc 2011;32:1891-6.
  26. Dung do TM, Dung PT, Oanh DT, Hai PT, Huong le TT, Loi VD, *et al.* Novel 3-substituted-2-oxoindoline-based N-hydroxypropenamides as histone deacetylase inhibitors and antitumor agents. Med Chem 2015;11:725-35.
  27. Lan HT, Mai DD, Phuong DP, Thanh HP, Khac VT, Hyunggu H, *et al.* Novel 2-oxoindoline-based hydroxamic acids: Synthesis, cytotoxicity, and inhibition of histone deacetylation. Tetrahedron Lett 2015;56:6425-9.
  28. Lai MJ, Huang HL, Pan SL, Liu YM, Peng CY, Lee HY, *et al.* Synthesis and biological evaluation of 1-arylsulfonyl-5-(N-hydroxyacrylamide) indoles as potent histone deacetylase inhibitors with antitumor activity *in vivo*. J Med Chem 2012;55:3777-91.
  29. Ito A, Kawaguchi Y, Lai CH, Kovacs JJ, Higashimoto Y, Appella E, *et al.* MDM2-HDAC1-mediated deacetylation of p53 is required for its degradation. EMBO J 2002;21:6236-45.
  30. Marvin Sketch. 6.5.1 ed. LLC 14<sup>th</sup> Floor, Cambridge Innovation Center, One Broadway Cambridge, MA 02142: Chem Axon; 2016.
  31. Vilar S, Cozza G, Moro S. Medicinal chemistry and the molecular operating environment (MOE): Application of QSAR and molecular docking to drug discovery. Curr Top Med Chem 2008;8:1555-72.
  32. Wu R, Lu Z, Cao Z, Zhang Y. Zinc chelation with hydroxamate in histone deacetylases modulated by water access to the linker binding channel. J Am Chem Soc 2011;133:6110-3.
  33. Abagyan RA, Totrov MM, Kuznetsov DA. ICM—a new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. J Comput Chem 1994;15:488-506.
  34. Pottel J, Therrien E, Gleason JL, Moitessier N. Docking ligands into flexible and solvated macromolecules 6. Development and application to the docking of HDACs and other zinc metalloenzymes inhibitors. J Chem Inf Model 2014;54:254-65.
  35. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Accurate calculation of the absolute free energy of binding for drug molecules. Chem Sci 2016;7:207-18.
  36. Huong TT, Dung DT, Huan NV, Cuong LV, Hai PT, Huong LT, *et al.* Novel N-hydroxybenzamides incorporating 2-oxoindoline with unexpected potent histone deacetylase inhibitory effects and antitumor cytotoxicity. Bioorg Chem 2017;71:160-9.
  37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, *et al.* UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605-12.
  38. Milano Chemometrics and QSAR Research Group. DRAGON for Windows (Software for Molecular Descriptor Calculator). 6.0 ed. Milano, Italy: Talet srl.
  39. Todeschini R, Ballabio D, Consonni V, Mauri A, Pavan M. Mobydigs: Computer Software. 1.1 ed. New York: WILEY – VCH; 2009.
  40. González MP, Terán C, Saiz-Urra L, Teijeira M. Variable selection methods in QSAR: An overview. Curr Top Med Chem 2008;8:1606-27.
  41. Hachamovitch R, Shufelt C. Statistical analysis of medical data. Part III: Multivariable analysis. J Nucl Cardiol 2000;7:484-95.
  42. Todeschini R, Consonni V, Mauri A, Pavan M. Detecting “bad” regression models: Multicriteria fitness functions in regression analysis. Anal Chim Acta 2004;515:199-208.
  43. Cruz-Monteagudo M, Borges F, Perez González M, Cordeiro MN. Computational modeling tools for the design of potent antimalarial bisbenzamidines: Overcoming the antimalarial potential of pentamidine. Bioorg Med Chem 2007;15:5322-39.
  44. Atkinson AC. Plots, Transformations and Regression. Oxford: Clarendon Press; 1985.
  45. Gramatica P. Principles of QSAR models validation: Internal and external. QSAR Comb Sci 2007;26:694-701.
  46. Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, *et al.* AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. J Chem Inf Model 2012;52:3099-105.
  47. Cruciani G, Pastor M, Guba W. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. Eur J Pharm Sci 2000;11 Suppl 2:S29-39.
  48. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 2000;295:337-56.
  49. Lombardi PM, Cole KE, Dowling DP, Christianson DW. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. Curr Opin Struct Biol 2011;21:735-43.
  50. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol 2015;68:627-36.
  51. Izenman AJ. Model Assessment and selection in multiple regression. In: Izenman AJ, editor. Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. New York, NY: Springer New York; 2008. p. 107-58.
  52. Tropsha A. Best practices for QSAR model development, validation, and exploitation. Mol Inform 2010;29:476-88.
  53. Rücker C, Rücker G, Meringer M. Y-Randomization and its variants in QSPR/QSAR. J Chem Inf Model 2007;47:2345-57.
  54. Todeschini R, Consonni V. Molecular Descriptors for Chemoinformatics. Mannhold R, Kubinyi H, Folkers G, editors. Weinheim, Germany: Wiley VCH Verlag GmbH & Co.; 2009.
  55. Pham-The H, González-Álvarez I, Bermejo M, Garrigues T, Le-Thi-Thu H, Cabrera-Pérez MÁ, *et al.* The use of rule-based and QSPR approaches in ADME profiling: A Case study on caco-2 permeability. Mol Inform 2013;32:459-79.
  56. Seelig A, Blatter XL, Wohnsland F. Substrate recognition by P-glycoprotein and the multidrug resistance-associated protein MRP1: A comparison. Int J Clin Pharmacol Ther 2000;38:111-21.
  57. Teague SJ, Davis AM, Leeson PD, Oprea T. The design of leadlike combinatorial libraries. Angew Chem Int Ed Engl 1999;38:3743-8.