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

HAI PHAM-THE¹, HUONG LE-THI-THU²*

¹Department of Pharmaceutical Chemistry, Hanoi University of Pharmacy, 13–15 Le Thanh Tong, Hoan Kiem, Hanoi, Vietnam. ²Department 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 chemoinformatics 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.

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, antiinflammatory, and antioxidant, for the 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 (RA2410, 4SC-201), givinostat (ITF2357), quisinostat...
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 CONH₂ 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. Pseudo receptor 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 (IC₅₀) 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 references.

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        | To determine structural characteristics essential for HDAC inhibition | Xie et al. [11] |
|                       |               |                     | R²=0.61, F=17.18  |                                                   |            |
| Not specified HDAC    | 29 HAs        | CoMFA               | R²=0.98, R²adj=0.60 | To predict HDAC inhibition activity              | Guo et al. [12] |
|                       |               |                     | R²=0.95, R²=0.59 |                                                   |            |
| Not specified HDAC    | 57 HAs        | CoMFA               | R²=0.91, R²adj=0.50 | To predict HDAC inhibition activity              | Juvalle et al. [13] |
|                       |               |                     | R²=0.98, R²=0.53 |                                                   |            |
| HDAC1/2 mixture       | 36 HAs        | 2D-QSAR, Docking    | R²=0.77, R²=0.72 | To determine structural characteristics essential for HDAC inhibition | Katritzky et al. [14] |
|                       |               | 3D-QSAR             | R²=0.81, Q²=0.62 |                                                   |            |
| Maize HDAC2           | 25 HAs        | GOLPE               | R²=0.96-98, Q²=0.81-85 | To predict HDAC inhibition activity | Rago et al. [15] |
| Not specified HDAC    | 55 HAs        | Quantitativepharmacophore | Q²=0.89 | To determine structural characteristics essential for HDAC inhibition | Chen et al. [16] |
|                       |               |                     |                   |                                                   |            |
| Not specified HDAC    | 45 HAs        | 2D-QSAR             | R²=0.78, R²=0.72, Q²=0.68 | Virtual screening of HDAC inhibitors | Melagnski et al. [17] |
|                       |               |                     |                   |                                                   |            |
| Not specified HDAC    | Varied data size (from 5-32 HAs) | 2D-QSAR             | R²=0.72-0.85, Q²=0.50-0.78 | To predict HDAC inhibition activity | Pontiki and Hadjipavlou-Litina [18] |
| HDAC2                 | 50 HAs        | CoMFA               | R²=0.98, R²=0.61 | To design novel HDAC2 inhibitors                  | Xiang et al. [19] |
|                       |               |                     | R²=0.99, R²=0.62 |                                                   |            |
| Maize HDAC2           | 34 HAs        | 2D-QSAR, 3D-QSAR     | R²=0.85, Q²=0.79  | To predict HDAC inhibition activity              | Sharma and Sharma [20] |

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

NOTES
papers [24-28], taking into account the homogeneity of experimental conditions. Accordingly, the half maximal inhibitory concentration (IC\(_{50}\), 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\(_{50}\) 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\(_{50}\)=lgIC\(_{50}\)).

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\(^{2+}\) 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 Lauffier 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 remained. 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 as 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\(_{50}\)) as a linear combination of a subset of X-variables (the selected molecular descriptors) with the regression coefficients b\(_j\) and the intercept b\(_0\).

\[ Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n \]  

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 gen 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\(_{loo}^2\)) 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 a 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\(_{loo}^2\) value.
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 \(k\) as possible.

**Model validation and applicability domain**

Goodness of fit was assessed by examining the determination coefficient, \(R^2\), the adjusted determination coefficient, \(R^2_{\text{adj}}\), 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 \(p\) statistic. 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 \(Q^2_{\text{LOO}}\) then calculated. In the bootstrapping procedure, \(k\)-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, \(Q^2_{\text{boot}}\) 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 \(h\) [44] measures the distance from one molecule to the whole training set (TS) in the descriptors space and is calculated as follows:

\[
h_i(t^T T T) = t_i t (i = 1, \ldots, N)
\]

Where \(t_i\) is the descriptor vector of that compound and \(T\) is the model matrix derived from the TS descriptor values. A “critical leverage” value \(h^*\) is generally fixed at \(3p/N\), 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 \(\text{Stand. R} \) 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 (ADMET-Tox) processes were calculated using ADMETsar server (http://1mmd.eucst.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 Q9V769 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 co-crystal 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 co-crystallized 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 <2.0Å 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 Å) 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 bond 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
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 IC50 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 103 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 x number of variables ≤ 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 (R2, Kxy), robustness of the models (Q2(samp), Q2(cross)), 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:

\[
p(\text{C50}) = 1.527 + 1.099 \, \text{VE1}_B(m) - 0.053 \, \text{RDF105m} - 0.190 \, \text{Mor02m} + 14.166 \, R7m' - 96.459 \, R5v' + 0.492 \, B07[N-O] \]

The model presented very good fit with a R2 of 0.90 and adjusted R2 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 (Kxy). 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 Kxy = 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 L00-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 L00 (Q2(samp)=0.856) is comparable with an R2 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 Q2(Y-scr) was 0.803, quite similar to R2, 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 (R2(Y-scr)) and cross-validation (Q2(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 (R2(Y-scr=0.116 and Q2(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.* | R2 Training | R2 Adj | Q2(samp) | Q2(cross) | Q2(Y-scr) | R2_Y-scr | Q2_Y-scr | KXY | s | F |
|-------|----------|-------------|--------|----------|-----------|-----------|-----------|----------|-----|----|---|
| M1    | 6        | 0.904       | 0.884  | 0.856    | 0.803     | 0.763     | 0.116     | 0.337    | 0.206 | 0.841 |
| M2    | 6        | 0.866       | 0.838  | 0.819    | 0.744     | 0.758     | 0.109     | 0.416    | 0.413 | 0.302 |
| M3    | 6        | 0.845       | 0.811  | 0.765    | 0.676     | 0.751     | 0.141     | 0.396    | 0.390 | 0.262 |
| M4    | 5        | 0.844       | 0.817  | 0.783    | 0.723     | 0.764     | 0.099     | 0.316    | 0.398 | 0.258 |
| M5    | 5        | 0.849       | 0.823  | 0.708    | 0.726     | 0.704     | 0.087     | 0.379    | 0.352 | 0.254 |
| M6    | 5        | 0.788       | 0.779  | 0.705    | 0.592     | 0.695     | 0.092     | 0.344    | 0.406 | 0.301 |
| M9    | 4        | 0.878       | 0.758  | 0.716    | 0.643     | 0.709     | 0.075     | 0.297    | 0.399 | 0.297 |

*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' | R5v' | B07[N-O] | Tolerance | VIF |
|----------|---------|--------|------|------|----------|-----------|-----|
| 1        |         |        |      |      |          | 0.557     | 1.796 |
| 0.176    | 1       |        |      |      |          | 0.812     | 1.232 |
| 0.541    | 0.139   | 1      |      |      |          | 0.591     | 1.692 |
| -0.149   | 0.309   | -0.074 | 1    |      |          | 0.664     | 1.506 |
| -0.422   | 0.029   | -0.474 | 0.456| 1    |          | 0.499     | 2.006 |
| 0.030    | -0.046  | -0.069 | -0.099| -0.258| 1        | 0.873     | 1.146 |

VIF: Variance inflation factor
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 4th electron diffraction signal. Both RDF105m and Mor02m contributed positively to the activity.

R7m$^+$ and R5v$^+$ 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: $[(h_i h_j)/r_{ij}]$, where the diagonal elements $h_i$ and $h_j$ 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 $j$th atom to interactions with the $i$th atom [54]. Interestingly, R7m$^+$ and R5v$^+$ are associated with large and opposite coefficient signals. R5v$^+$ contributes positively to the activity, while R7m$^+$ 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 [54-6].

**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...
caping 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-1H-benzo[de]isoquinoline (1a-g), 6-(5H)-phenanthridinone (2a-g), 9H-fluorene (3a-e), 4H-cyclopenta[d]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 Q SAR 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 μM. Especially, compounds 4e and 6d displayed a five-to seven-fold more potent than reference drug SAHA (experimental $IC_{50}$=0.2–0.3 μM). 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).
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 conformation 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

1. 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.
2. Politzer P, Murray JS. Some intrinsic features of hydroxylamines, oximes and hydroxamic acids: Integration of theory and experiment. PATAF’S Chemistry of Functional Groups.Chichester, West Sussex: John Wiley & Sons, Ltd; 2009.
3. 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.
4. 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.
5. Islam T. Crucial challenges in epigenetic cancer therapeutic strategy yet to be resolved. Int J Pharm Pharm Sci 2016;7:326-31.
6. Falkenberg KJ, Johnstone RW. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. Nat Rev Drug Discov 2014;13:673-91.
7. Moffatt M, Zheng S, Huang TL, Wang G. Histone deacetylase inhibitors in clinical studies as templates for new anticancer agents. Molecules 2015;20:3896-941.
8. Whitehead L, Dobler MR, Radeich 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.

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<sup>d</sup> (cm/s) | HA<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 AMES toxic |
| 4e   | 0/5    | 3/3    | 2.64×10<sup>-2</sup> | 0.583  | High   | No              | Yes            | 3A4            | Fish toxicity | Tetrahymena pyriformis Fish toxicity |
| 6b   | 0/5    | 1/3    | 1.20×10<sup>-2</sup> | 0.611  | High   | No              | Yes            | No             | Tetrahymena pyriformis |
| 6d   | 0/5    | 1/3    | 3.10×10<sup>-3</sup> | 0.573  | High   | No              | Yes            | No             | Tetrahymena pyriformis |
| 6e   | 0/5    | 1/3    | 2.23×10<sup>-3</sup> | 0.529  | High   | No              | Yes            | No             | Tetrahymena pyriformis |

<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. [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.
Predictive QSAR workflow for the AdmetSAR: The use of rule-based Novel N-hydroxybenzamides incorporating 2-oxoindoline with 28. Lai MJ, Huang HL, Pan SL, Liu YM, Peng CY, Lee HY, et al. Marvin Sketch. 6.5.1 ed. LLC 14.
27. Lan HT, Mai DD, Phuong DP, Thanh HP, Khac VT, Hyunggu H, et al. Integrated QSAR workflow for the in silico identification and screening of novel HDAC inhibitors. Mol Divers 2009;13:301-11.
25. Chun PS, Kim WH, Kim JS, Kang JA, Lee HJ, Park JY, et al. Designing of anti-cancerous histone deacetylase inhibitors through mimicking of protein-protein interfaces. Int J Pharm Pharm Sci. 2011;3:189-12.
22. Rajaganapathy K, Sathiyasundar R, Ramesh GK, Kalaichelvan VK. 19. Xiang Y, Hou Z, Zhang Z. Pharmacophore and QSAR studies to identify novel histone deacetylase inhibitors. J Biol Chem 2013;288:26926-43.
16. Chen YD, Jiang YJ, Zhou JW, Yu QS, You QD. Identification of ligand metalloenzymes. Curr Opin Struct Biol 2011;21:735-43.
15. Ragno R, Simeoni S, Valente S, Massa S, Mai A. 3D-QSAR studies on histone deacetylase inhibitors. A GROLJE/GIRD approach on different series of compounds. J Chem Inf Model 2006;46:1420-30.
14. Katritzky AR, Slavov SH, Dobchev DA, Kuznetsova DF. 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.
13. Juvale DC, Kulkarni VV, Deokar HS, Wagh NK, Padhye SB, Wahlberg JW, et al. Characterization of the absolute free energy of binding for drug molecules. J Comput Chem 2011;32:1891-6.
12. Laufer BE, Mintzer R, Fong R, Mukund S, Tam C, Zilberleyb I, et al. Histone deacetylase inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability. J Biol Chem 2008;283:29262-43.
11. Lan HT, Mai DD, Phuong DP, Thanh HP, Khac VT, Hyunggu H, et al. Novel 2-oxoindoline-based hydroxamic acids: Synthesis and inhibition of histone deacetylase. Tetrahedron Lett 2015;56:6425-9.
10. Humberto GD. Editorial (Hot Topic: Bioinformatics and Quantitative Structure-Activity relationship study of histone deacetylase inhibitors. Curr Med Chem Anticancer Agents 2004;4:273-99.
9. Ngioco 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.
8. Guo Y, Xiao J, Guo Z, Chu F, Cheng Y, Wu S, et al. Exploration of a binding mode of indole amine analogues as potent histone deacetylase inhibitors and 3D-QSAR analyses. Bioorg Med Chem 2005;13:3542-34.
7. Huang HL, Pan SL, Liu YM, Peng CY, Lee HY, et al. Synthesis and evaluation of 1-arylsulfonyl-5-(N-hydroxyacrylamide) hydroxamic acids. Angew Biomed Chem 2006;4:2588-65.
6. Katritzky AR, Slavov SH, Dobchev DA, Kuznetsova DF. Quantitative structure-activity relationship study of histone deacetylase inhibitors and inhibition of histone deacetylation. J Am Chem Soc 2011;133:6110-3.
5. Abagyan RA, Totrov MM, Kuznetsova DF. 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.
4. 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.
3. Huang 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.
2. Pottel J, Terrien E, Glessen JL, Moetissei M. 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.
1. Atkinson AC. Plots, Transformations and Regression. Oxford: Clarendon Press; 1985.
57. Teague SJ, Davis AM, Leeson PD, Oprea T. The design of leadlike combinatorial libraries. Angew Chem Int Ed Engl 1999;38:3743-8.