Journal of Molecular Structure 1199 (2020) 126961

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: http://www.elsevier.com/locate/molstruc

QSAR analysis of coumarin-based benzamides as histone deacetylase inhibitors using CoMFA, CoMSIA and HQSAR methods

Rahman Abdizadeh [a], Farzin Hadizadeh [b, c], Tooba Abdizadeh [d, *]

[a] Department of Medical Parasitology and Mycology, Faculty of Medicine, Shahrekord University of Medical Sciences, Shahrekord, Iran
[b] Biotechnology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
[c] Department of Medical Chemistry, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran
[d] Clinical Biochemistry Research Center, Basic Health Sciences Institute, Sharekord University of Medical Sciences, Shahrekord, Iran

ARTICLE INFO

Article history:
Received 26 March 2019
Received in revised form
5 August 2019
Accepted 19 August 2019
Available online 23 August 2019

Keywords:
Histone deacetylase inhibitors
CoMFA
CoMFA-RF
CoMSIA
HQSAR
2-AminoBenzamide

ABSTRACT

Histone deacetylases (HDACs) as the promising therapeutic targets for the treatment of cancer and other diseases, modify chromatin structure and contribute to aberrant gene expression in cancer. Inhibition of HDACs is emerging as an important strategy in human cancer therapy and HDAC inhibitors (HDACIs) enable histone to maintain a high degree of acetylation. In this work, molecular modeling studies, including CoMFA, CoMFA-RF, CoMSIA and HQSAR and molecular docking were performed on a series of coumarin-based benzamides as HDAC inhibitors. The statistical qualities of generated models were justified by internal and external validation, i.e., cross-validated correlation coefficient (q^2), non-cross-validated correlation coefficient (r^2_{ncv}) and predicted correlation coefficient (r^2_{pred}), respectively. The CoMFA (q^2, 0.728; r^2_{ncv}, 0.982; r^2_{pred}, 0.685), CoMFA-RF (q^2, 0.764; r^2_{ncv}, 0.960; r^2_{pred}, 0.552), CoMSIA (q^2, 0.671; r^2_{ncv}, 0.977; r^2_{pred}, 0.721) and HQSAR models (q^2, 0.811; r^2_{ncv}, 0.986; r^2_{pred}, 0.613) for training and test set of HDAC inhibition of HCT116 cell line yielded significant statistical results. Therefore, these QSAR models were excellent, robust and had better predictive capability. Contour maps of the QSAR models were generated and validated by molecular docking study. The final QSAR models could be useful for the design and development of novel potent HDAC inhibitors in cancer treatment. The amido and amine groups of benzamide part as scaffold and the bulk groups as a hydrophobic part were key factors to improve inhibitory activity of HDACIs.

1. Introduction

In cancer, the uncontrolled, rapid and pathological proliferation of abnormal cells, epigenetics has been found to play an important role in the origin, development, and metastasis [1]. Epigenetic writer, eraser, and reader enzymes as well as histone deacetylases (HDACs), DNA methyl transferases (DNMTs), and hidtone methyl transferases (HMTs) are being increasingly used as targets for drug design and discovery in cancer and other diseases including diabetes and neuro-degenerative and inflammatory disorders [2,3].

Among the epigentic modifications, acetylation/deacetylation of histones is the most common mechanism which used by cells in regulating the normal cellular processes like cell differentiation, proliferation, angiogenesis, and apoptosis [4]. Dysregulation of the acetylation has been associated with diverse cellular events in cancer pathologies. Global hypoacetylation of the H4 is one such common feature of human tumors [5].

Acetylation of histones and non-histones proteins is regulated by two antagonistic families of enzymes, histone acetyl transferases (HATs) and histone deacetylases (HDACs) [6].

HATs neutralized the positive charge of lysine residues of histones by adding an acetyl group and leads to the relaxation of chromatin and activates transcription [7].

HDACs are a family of ubiquitous enzymes found in bacteria, fungi, plants, and animals [8,9] that HDACs remove acetyl groups from ε-nitrogens of lysine residues within core histones and non-

Abbreviations: HDAC, Histone Deacetylases; HATs, Histone Acetyl Transferases; HSP 90, Heat Shock Protein 90; SAHA, suberoylanilide hydroxamic acid; TSA, Trichostatin A; ZBG, Zinc Binding Group; CU, Connect Unit; QSAR, Quantitative structure-activity relationship; 3D-QSAR, Three-dimensional QSAR; CoMFA, Comparative molecular field analysis; CoMFA-RF, CoMFA region focusing; CoMSIA, Comparative molecular similarity index analysis; HQSAR, Hologram QSAR; AD, Application Domain.

* Corresponding author.
E-mail address: t.abdizadeh@gmail.com (T. Abdizadeh).

https://doi.org/10.1016/j.molstruc.2019.126961
0022-2860/

histone proteins (e.g. tubulin, p53, ER_{α}, HSP 90, NF-γ and GATA-1) during post-translational protein modification [10].

The human HDAC enzyme consists of 18 isoforms, which are divided into four classes according to their homology to yeast models, subcellular distribution and enzymatic activity [11]. Class 1 (HDAC 1, 2, 3 and 8), class IIa (HDAC 4, 5, 7 and 9), class IIb (HDAC 6 and 10) and class IV (HDAC 11) enzymes are zinc-dependent HDACs, whereas class III HDACs (sirtuins 1–7) are NAD^{+} dependent as a cofactor foe activity [12].

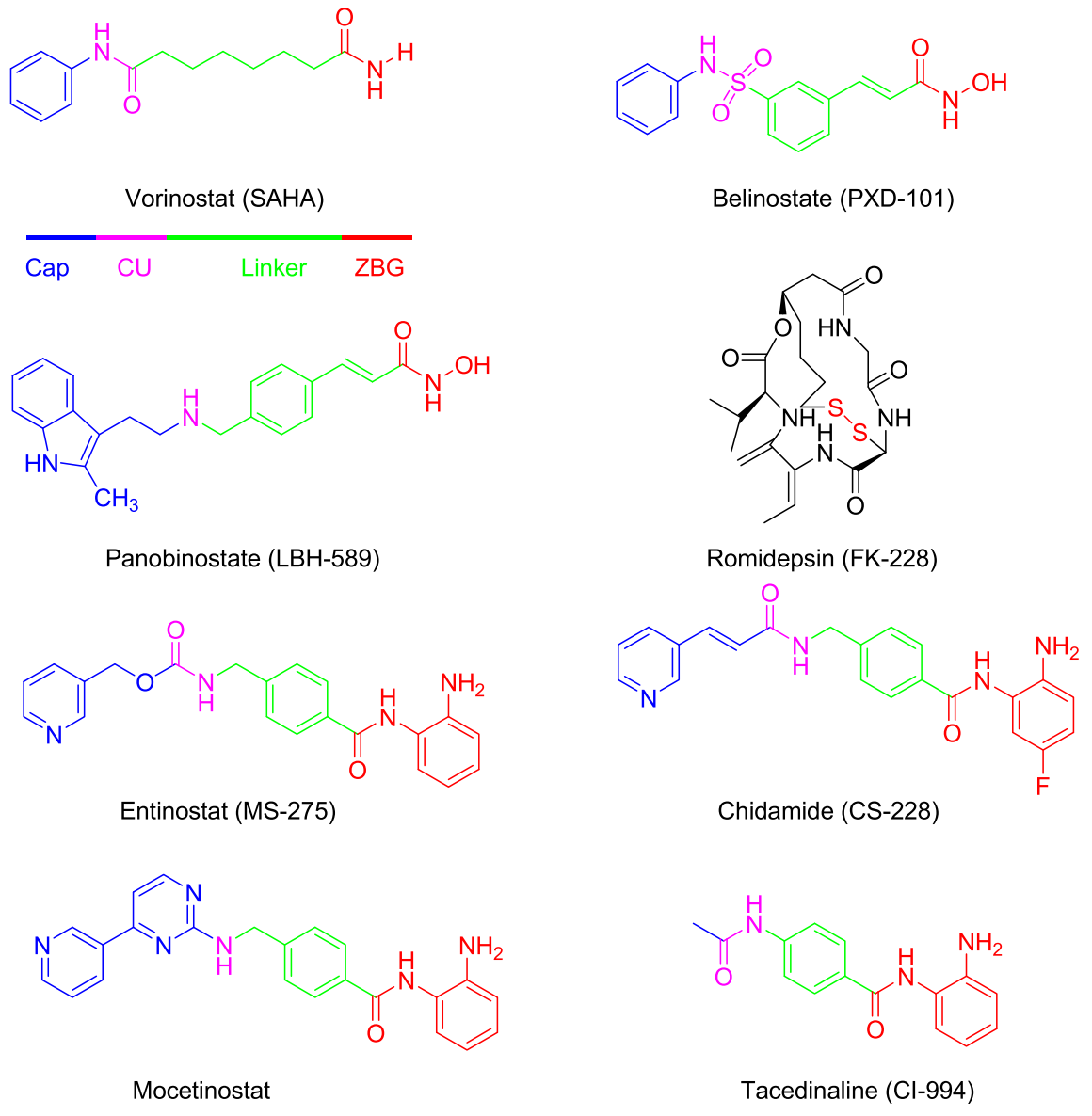
Class I HDACs are homologus to yest Rpd3 protein and are mainly located in the nucleus, while class II HDACs are structurally related to yeast Hda2 and shuttle between nucleus and cytoplasm. Class III have sequence homology to yeast sir2 and class IV, also shuttle between nucleus and cytoplasm and shows the characteristics of both class I and II HDACs [13].

Zn^{2+}-dependent HDACs, especially class I enzymes are involved with cell cycle progression, metastasis, angiogenesis and invasion [14]. Class I HDAC isoforms show a key role in the development of cancer and overexpressed in various human cancers such as prostate, breast, ovarian, leukemia, colon and pancreas [15,16].

HDAC enzyme inhibition can abolish abnormal epigenetic changes associated with different diseases such as cancer [17]. The classification of HDAC inhibitors (HDACIs) in clinical studies depends on their chemical structures and categorized into four class: hydroxamic acid (Trichostatine A (TSA) [18], Vorinostat (SAHA) [19], Panobinostat (LBH-5890 [20], Belinostat (PXD-101) [21]), benzamides (Entinostate (MS-275) [22,23], Mocetinostat (MGCD-0103) [24], Chidamide (CS-055) [25], Tacedinaline (CI-994) [26]), short-chain fatty acids (Valproic acid) [27] and depsipeptides (Romidepsin (FK-228) [28].

To date, four HDACIs have been approved by the FDA: vorinostate (cutaneous T-cell lymphoma; CTCL), romidepsin (CTCL and peripheral T-cell lymphoma; PTCL), belinostat (relapsed or refractory PTCL), and panobinostat (multiple myeloma; MM) [29]. The molecular views of several approved and clinical HDACIs are shown in Scheme 1.

Most hydeoxamate HDACIs are pan inhibitors that target broad-spectrum of HDAC isoforms with unwanted side effects while 2-



Scheme 1. Pharmacophore model and structures of approved and in clinical trials HDAC inhibitors.

amino benzamides display specific inhibiting effects on HDAC 1, 2, 3 enzymes [30,31].

Thus, understanding the structural regions of HDACIs is essential to design and discover potent and selective inhibitors. Despite the huge variety of structural motifs, the HDACIs share a common pharmacophore model: a zinc binding group (ZBG) that chelates with the catalytically Zn^{2+} ion at the bottom of the active site, a hydrophobic linker region, plus a polar connecting unit (CU), interacting with hydrophobic residues of the narrow tunnel and connect the ZBG and the CAP group, surface recognition moiety (CAP) that interacts with residues at the entry point of the active site of HDAC [32,33].

The CAP moiety is essential for recognizing and binding to the residues in the enzyme active site [34]. Molecular views of entinostat and coumarin benzamides as HDACIs are shown in Scheme 2.

Quantitative structure-activity relationship (QSAR) is a technique that is used in computer-assisted rational drug design and predicts the protein-ligand interaction and to explore the correlation between biological activity and molecular structure [35–37]. Three-dimensional QSAR (3D-QSAR) is a broad term encompassing all those QSAR methods which are utilized to calculate the highly specific interactions and a molecule, how far and with how much power can be connected to the active site of an enzyme or protein [38–40]. Recently, comparative molecular field analysis (CoMFA), CoMFA region focusing (CoMFA-RF), comparative molecular similarity index analysis (CoMSIA) and hologram QSAR (HQSAR) are especially effective methods of QSAR based on statistical techniques [34–36]. The CoMFA model proposed by Cramer et al. describes the molecular properties by steric (Lennard-Jones) and electrostatic (Coulomb) energy fields of important regions of a set of aligned compounds that predict their biological activity over a lattice of point [41,42]. In CoMFA-RF model, steric and electrostatic fields are calculated for aligned fragments by creating specific grid space at the specific lattice points [43]. In CoMSIA model, proposed by Klebe et al., a probe atom is used to calculate similarity indices, at regularly placed grid points for the aligned molecules. Compared to CoMFA, CoMSIA uses a Gaussian-type distance-dependent function to assess five fields of different physicochemical properties (i.e., steric, electrostatic, hydrophobic, hydrogen bonding donor and acceptor [44]. Also, CoMSIA is differentiated by Gaussian functions and no arbitrary definitions of cut off limits should be used.

HQSAR study is a comparatively new 2D-QSAR method which employs the fragment fingerprints of molecular holograms and other molecular descriptors to predict the biological activity of a series of molecules [45–47]. In these models, all regression analyses performed in two steps using the partial least squares (PLS) method [48–51].

In the present study, we performed a molecular modeling study by combined 2D- and 3D-QSAR and molecular docking techniques. 2D-QSAR, using HQSAR method, and 3D-QSAR, using CoMFA, CoMFA-RF and CoMSIA methods, were used to identify the key structural factors influencing on inhibitory activity. Molecular docking was used to identify some key amino acid residues at the active site of HDAC protein and investigate the binding modes between HDAC and the selected inhibitors. The obtained results can apply to the further structural modification, design and development new and more potent anti-cancer drugs.

2. Materials and methods

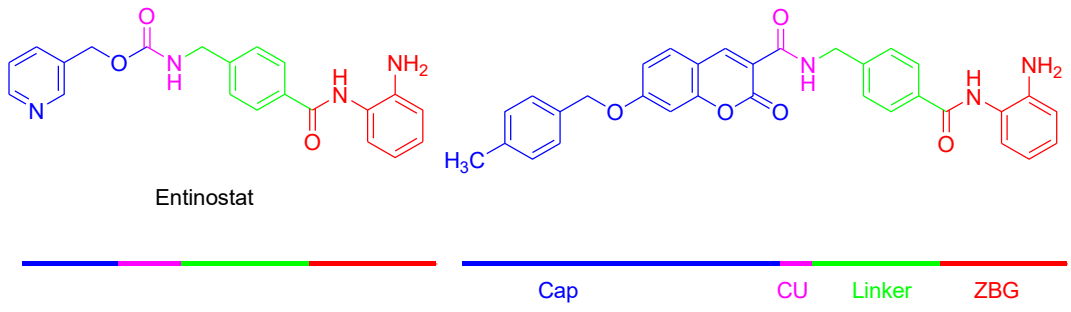
2.1. Data set

QSAR studies were performed on a set of 21 coumarin-based benzamides as a new class of anti-cancer agents and HDAC inhibitors with their biological activities (IC_{50} values) that recently reported by our group [52].

These activity values (IC_{50} in μM) were converted to corresponding pIC_{50} (-log IC_{50}) values and used as a dependent variable in CoMFA, CoMFA-RF, CoMSIA and HQSAR models. The data set was randomly divided into a training set (15 compounds, 70%) for QSAR model generation and a test set (6 compounds, 30%) for external validation of the models (Fig. 1.).

2.2. Molecular modeling and alignment

The QSAR models including CoMFA, CoMFA-RF, CoMSIA and HQSAR were performed using the SYBYL-X 1.2. molecular modeling software (Tripos, Inc, St. Louis, MO). Before modeling with these primary methods, the 3D structures of compounds were drawn using Chemoffice Bio 3D Ultra (version 12.0, Cambridge Soft Corporation, Cambridge, UK, 2010). All the compounds were energy minimized using the standard molecular mechanics force field with a distance dependent dielectric and the powell conjugate gradient algorithm with a convergence criterion of 0.05 kcal/molÅ using the maximum iteration set to 5000 [53]. Partial atomic charges of the compounds for electrostatic interactions were calculated by the Gasteiger-Hückel method. Structure alignment was one of the most important input variables in 3D-QSAR analysis and the accuracy of the prediction power of the models was reliability dependent on contour maps according to the structural alignment of the molecules. In this study, rigid body alignment of molecules in a Mol2 database was performed using maximum common substructures defined by Distill alignment. Compound **11** was selected as template because the most active compound of the data set and other compounds were aligned according to the common structure. The molecular view and molecular structure of compound **11** with bold red common substructure and final super imposition of compounds are shown in Fig. 2a and b, respectively.



Scheme 2. Chemical structures of Entinostat and a coumarin-based benzamide as HDACIs.

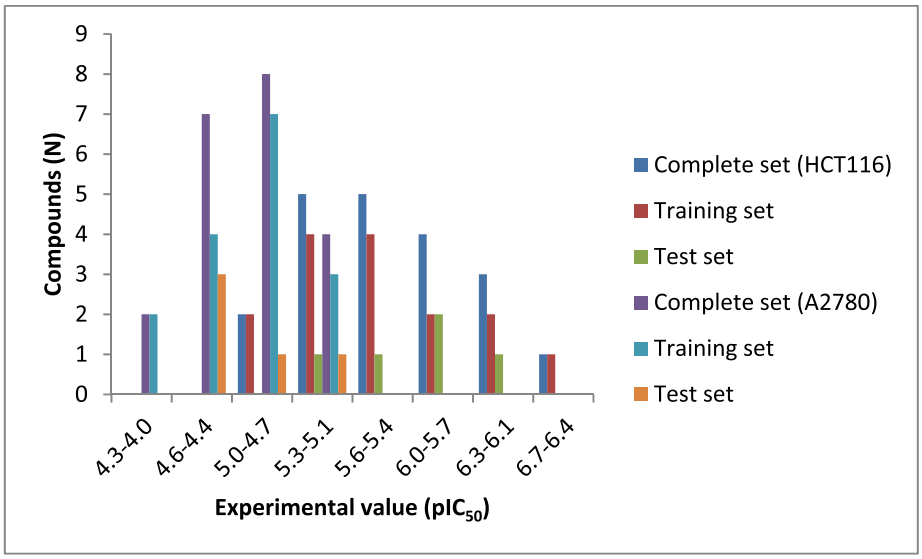


Fig. 1. Distribution of experimental inhibitory activities (pIC_{50}) for the training and test sets compounds in the QSAR models.

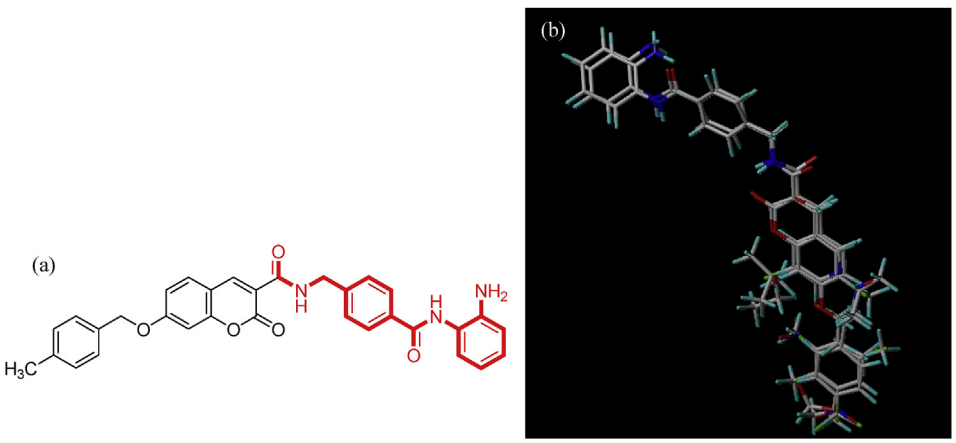


Fig. 2. Compound **11** used as the template molecule for database alignment and Common substructure in Distill alignment shown in the bold red (a) and aligned compounds in the training and test sets (b).

2.3. CoMFA and CoMSIA analysis

In CoMFA method, the aligned molecules in optimal orientation were located in a 3D cubic lattice with grid spacing of 2 Å in the x, y, and z direction which extended 4.0 Å around the align molecules in all Cartesian directions. The CoMFA steric and electrostatic fields were calculated for each molecule using a hybridized sp^3 carbon probe atom with a Van Der Waals radius of 1.52 Å and a charge of +1.0. The Coulomb and Lennard-Jones potential functions were used to estimate the electrostatic and steric interactions, respectively. The energy cut off values for both steric and electrostatic fields were set at 30 kcal/mol. In order to reduce noise and improve efficiency, column filtering was tested in the range of 0.0−2.0 kcal/mol and a threshold column filtering value of 2.0 kcal/mol. CoMFA-RF in the "Advanced CoMFA" module is a technique of application of weight to the lattice point in a CoMFA region to increase or decrease the contribution of these points to subsequent analysis. "StDev*-Coefficients" values as different weighting factors were employed in addition to grid spacing for getting the better models. This increases the resolution and predictive capability (q^2, cross validated r^2) of a followed PLS analysis.

The CoMSIA method calculates the similarity indices descriptors with the same lattice box used in CoMFA. Five physicochemical properties of steric, electrostatic, hydrophobic, hydrogen bonding donor and acceptor fields were evaluated using a probe atom to charge +1.0, radius 1 Å, hydrophobicity +1.0, hydrogen bonding donor +1.0, hydrogen bonding acceptor +1.0, attenuation factor α of 0.3 and grid spacing 2.0 Å. A distance-dependent Gaussian type was used between the probe atom and each molecule atom [53,54].

2.4. HQSAR analysis

Hologram QSAR study is a 2D-QSAR technique which certain the relationship between the biological activity with the structural fragments. This method eliminates the need for 3D structure, the ability to achieve molecular alignment and conformational specification [55,56] by transforming the chemical representation of a molecule into its corresponding molecular hologram. 2D chemical database storage and searching technologies rely on linear notations that define chemical structures [Wiswesser line-formula notation (WLN), simplified molecular input line entry system (SMILES); SLN-SYBYL line notation]. The process involves the

generation of fragments that are hashed into the array is called molecular hologram and bin occupancies are the descriptor variable [57–59].

The HQSAR method employs different parameters for the molecular hologram generation, such as hologram length (HL) values (53, 59, 61, 72, 83, 97, 151, 199, 257, 307, 353 and 401), a fragment distinction (atom (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA), and the fragment size (2–5, 3–6, 4–7, 5–8, 6–9, 7–10).

2.5. Partial least-square (PLS) analysis

In 3D-QSAR studies, PLS method [56] an extension of multiple regression analysis was used for the model building. Calculated CoMFA and CoMSIA descriptors as independent variables were used with the pIC_{50} values as dependent variables in the PLS regression analysis, respectively. Before the PLS analysis, the CoMFA and CoMSIA columns were filtered by using column filtering value equal to 2.0 kcal/mol. The predictive ability of the models was evaluated by leave-one-out (LOO) and leave-ten-out (L-10-O) methods. LOO cross-validation method was used as an internal validation to generate the optimal number of components (ONC) with the lowest standard error of prediction (SEP) and the highest cross-validated coefficient q^2 (r^2_{cv}) that was calculated by Equation (1):

$$q^2 = 1 - \frac{\sum_{n=1}^{\infty}(\hat{y}_i - y_i)^2}{\sum_{n=1}^{\infty}(y_i - \bar{y})^2} \quad (1)$$

whereas, $\hat{y}_i$ and y_i are predicted, observed activity values, and $\bar{y}$ and $\hat{y}$ are observed and predicted mean activity values of the training set, respectively [60]. The $\sum_{n=1}^{\infty}(\hat{y}_i - y_i)^2$ is the predictive residual sum of squares (PRESS).

After cross validation, the final PLS analysis was carried out using the optimal number of components with no validation to generate the final QSAR model. The non-cross-validated analysis performed by the conventional correlation coefficient $r^2(r^2_{ncv})$ (Equation (2)), standard error of estimation (SEE) and F values calculated with the same column filtering set. High q^2 and r^2 ($q^2 > 0.5$, $r^2 > 0.6$) values are regarded as a proof of high predictive ability of the built model and also $r^2 - q^2$ for a good model should not be more than 0.3 [56].

$$r^2 = \frac{[\sum(y_i - \bar{y}_i)(\hat{y}_i - \hat{y})]^2}{\sum(y_i - \bar{y}_i)^2 \times \sum(\hat{y}_i - \hat{y})^2} \quad (2)$$

Bootstrapping analysis was performed for 100 runs to assess the statistical confidence of the derived models [42,61–63]. Contour maps were generated graphically after models were developed in CoMFA/CoMFA-RF and CoMSIA using the field type "StDev*Coeff" and the contour levels were set to default values.

In HQSAR, LOO cross-validation was applied to determine the number of components that yields a good predictive model. PLS then yields a mathematical equation that related the molecular hologram bin values to the inhibitory activity of the compounds in the database.

2.6. Validation of the QSAR model

A good internal validation showed only a high q^2 in the training set of compounds, but it did not indicate the high predictive ability of the established models, therefore external validation was essential. The predictive ability of 3D-QSAR models was validated by calculating biological activities of the compounds which were not included in the training set and used as a test set. Test set was marked with * in Tables 1 and 2.

The predictive correlation coefficient r^2_{pred} (r^2_{pred}>0.6) [64], based on the test set was calculated using Equation (3):

$$r^2_{pred} = \left(\frac{SD - PRESS}{SD}\right) \quad (3)$$

SD is the sum of squared deviation between the biological activities of the test set molecules and the mean activity of the training set molecules. PRESS is the sum of squared derivations between the predicted and actual activities of the test set molecules.

The performance of the regression models constructed here was evaluated using the root mean squared error (RMSE), mean absolute error (MAE) (RMSE and MAE close to zero), residual sum of squares (RSS) and concordance correlation coefficient (CCC; CCC ≥ 0.85) of the training and validation sets [65]. The RMSE and the MAE are calculated for the data set as Equations (4)–(7):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad (5)$$

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (6)$$

$$\mathrm{CCC} = \frac{2\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \hat{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2 + \sum_{i=1}^{n}(\hat{y}_i - \hat{y})^2 + n(\bar{y} - \hat{y})^2} \quad (7)$$

To obtain the best predictive model for the test set, additional validation of model, the following Parameters [64] were used (Eqs. (8)–(12)):

$$r^2_o = 1 - \frac{\sum(y_i - k \times \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (8)$$

$$r'^2_o = 1 - \frac{\sum(\hat{y}_i - k' \times y_i)^2}{\sum(\hat{y}_i - \hat{y})^2} \quad (9)$$

$$k = \frac{\sum(y_i \times \hat{y}_i)}{\sum(\hat{y}_i)^2} \quad (10)$$

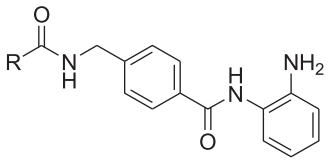
$$k' = \frac{\sum(y_i \times \hat{y}_i)}{\sum(y_i)^2} \quad (11)$$

$$\frac{(r^2 - r^2_o)}{r^2} < 0.1 \quad \text{or} \quad \frac{(r^2 - r'^2_o)}{r^2} < 0.1 \quad (12)$$

$0.85 \le k \le 1.15$ or $0.85 \le k' \le 1.15$ r^2_0 and r'^2_0 are squared correlation coefficients of determination for regression lines through the origin between predicted (y) and observed (x) activities and vice versa. The values of k and k′ are the slopes of their models, respectively.

To further assess the models, another validation statistical parameters r^2_m and Δr^2_m were determined by following Equations (13) and (14):

Table 1
Molecular view and the Corresponding Experimental and Predicted pIC_{50} Values by QSAR models (HCT116).



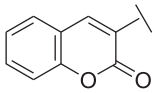
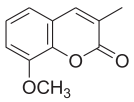
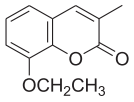
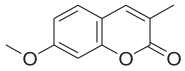
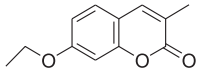
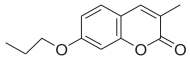
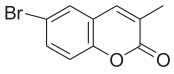
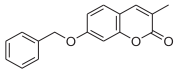
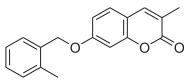
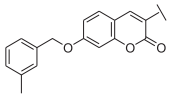
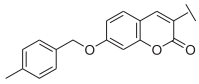
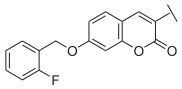
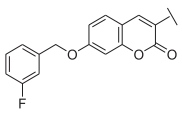
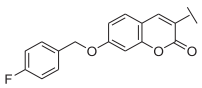
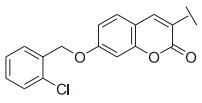
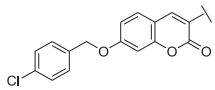
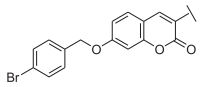
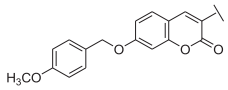
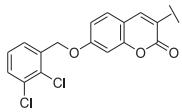
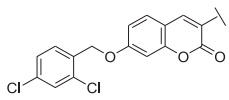
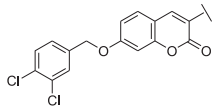
Predicted pIC_{50}						
Compd	R	Experimental pIC_{50}	CoMFA	CoMFA-RF	CoMSIA	HQSAR
1		4.942	4.969	5.061	4.952	4.933
2		5.286	5.295	5.227	5.284	5.251
3[a]		5.690	5.407	5.690	5.524	5.552
4		5.511	5.495	5.49	5.508	5.463
5		5.636	5.642	5.585	5.646	5.662
6[a]		5.962	5.842	5.629	5.684	5.915
7		5.177	5.152	5.148	5.169	5.248
8		5.603	5.617	5.594	5.639	5.547
9[a]		5.853	5.647	5.723	5.760	5.847
10		5.882	5.893	5.808	5.86	5.864
11		6.602	6.592	6.583	6.587	6.623
12		5.030	5.072	5.07	5.011	5.027
13		5.428	5.426	5.364	5.431	5.343
14[a]		5.609	5.513	5.533	5.526	5.6
15		5.322	5.278	5.268	5.288	5.37

Table 1 (*continued*)

Compd	R	Experimental pIC_{50}	Predicted pIC_{50} CoMFA	CoMFA-RF	CoMSIA	HQSAR
16		5.752	5.938	6.011	5.965	5.833
17		6.376	6.286	6.362	6.252	6.284
18[a]		6.00	6.011	5.937	6.014	6.06
19		5.212	5.195	5.334	5.242	5.268
20		5.371	5.409	5.175	5.414	5.125
21[a]		6.097	5.928	5.854	5.917	5.745

[a] Test set.

$$r_m^2 = r^2\left(1 - \sqrt{|r^2 - r_o^2|}\right) \quad (13)$$

$$\Delta r_m^2 = \left|r_m^2 - r_m^{\acute{2}}\right| \quad (14)$$

r_m^2 value more than 0.5 (r_m^2>0.5) and Δr_m^2<0.2 show good external predictability of the models.

2.7. Molecular docking study

Molecular docking as one of the most frequently methods in drug design was used to investigate the mode of interaction of small molecules with the appropriate target binding sites. The docking study was performed using Operation Environment (MOE) software (www.chemcomp.com) between the most and least active compounds with HDAC1 enzyme. For the preparation of ligands prior to docking, the 2D structures of ligands were prepared by Chemoffice ultra (version 12.0, Cambridge Soft Corporation, Cambridge, UK, 2010) and converted to 3D format by Hyper Chem7 (Hyper cube Inc, USA) using AM1 semi-empirical method. The ligands in our data set were docked in the active site of HDAC1 (PDB ID: 4BKX) by MOE software. The docking was performed by the triangle matcher placement algorithm in combination with London dG scoring function and force field as refinement method and the conformation of compounds were further analyzed by LigX module in MOE software.

3. Results and discussion

3.1. CoMFA and CoMFA-RF statistical results

The statistical results of CoMFA and CoMFA-RF models of HDAC inhibitory activity of HCT116 and A2780 are summarized in Tables 3 and 4. The CoMFA analysis was carried out with steric and electrostatic fields at column filtering of 2.0 kcal/mol.

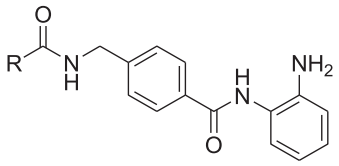
PLS analysis of CoMFA of training set of HTC-116 and A2780 cell lines including leave-one-out (LOO) and leave-ten-out (L-10-O) cross validation with ONC 6 and 4 showed q^2 values of 0.728 and 0.721, r_{cv}^2 (L-10-O) values of 0.750 and 0.711 and SEP of 0.309 and 0.224, respectively. These statistical results showed that the model had a better predictive capability.

The non-cross-validated PLS analysis gave r_{ncv}^2 of 0.982 and 0.961 with standard error of estimate (SEE) of 0.081 and 0.084, F values of 79.930 and 67.786, $r^2 - q^2$ of 0.254 and 0.240 and $R_{pearson}$ of 0.954 and 0.936 for HCT116 and A2780 cell lines, respectively which supported the statistical validity of the developed model. The contributions from steric and electrostatic field descriptors explained 0.557 and 0.414; 0.586 and 0.443 of the total variance, respectively, that indicated steric effect was more important than the electrostatic fraction.

After using region focusing, a new model of CoMFA-RF was created in the statistical parameters. The cross-validation and non-cross-validated PLS calculation results were similar to CoMFA model. This approach showed the q^2 values of 0.764 and 0.665 with ONC of 4 and r_{cv}^2 (L-10-O) of 0.758 and 0.679 and SEP of 0.245 and 0.261 for HCT116 and A2780 cell lines, respectively. The non-cross-validated PLS analysis resulted in high r_{ncv}^2 values of 0.960 and 0.842 with low SEE values of 0.108 and 0.144, F values of 65.274 and 21.082, $r^2 - q^2$ values of 0.196 and 0.220 and $R_{pearson}$ values of 0.925 and 0.912 for HCT116 and A2780 cell lines, respectively. The contribution of steric and electrostatic field descriptors was 0.563 and 0.437; 0.730 and 0.270, respectively in CoMFA-RF.

The bootstrapped results were shown in r_{bs}^2 and SEE_{bs} values of 1.00 and 0.998; 0.00 and 0.002 (CoMFA) and 0.997 and 0.995; 0.004 and 0.006 (CoMFA-RF), for HCT116 and A2780 cell lines, respectively, that suggesting a good internal consistency and the absence of systematic errors of the models within the training data set.

Table 2
Molecular view and the Corresponding Experimental and Predicted pIC_{50} Values by QSAR models (A2780).



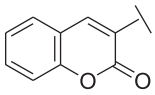
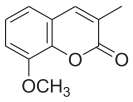
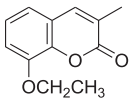
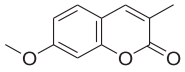
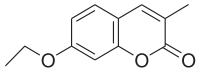
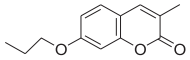
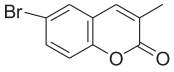
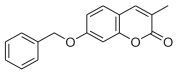
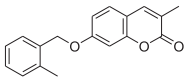
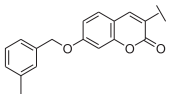
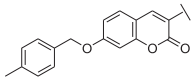
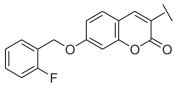
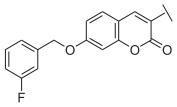
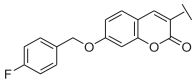
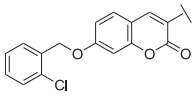
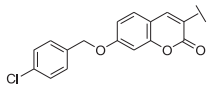
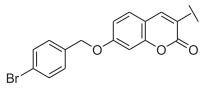
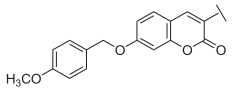
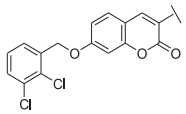
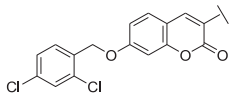
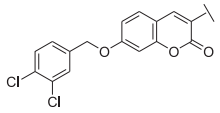
Predicted pIC_{50}						
Compd	R	Experimental pIC_{50}	CoMFA	CoMFA-RF	CoMSIA	HQSAR
1		4.260	4.338	4.385	4.438	4.364
2		4.397	4.407	4.339	4.444	4.374
3[a]		4.483	4.367	4.304	4.411	4.567
4		4.665	4.637	4.714	4.565	4.62
5		4.768	4.75	4.783	4.631	4.823
6[a]		4.829	4.864	4.839	4.729	4.977
7		4.560	4.494	4.448	4.476	4.477
8		4.892	5.02	4.913	5.097	5.00
9[a]		4.681	4.836	4.754	4.895	4.803
10		4.996	4.957	4.841	5.146	4.987
11		5.686	5.602	5.478	5.334	5.609
12		4.761	4.764	4.877	4.825	4.69
13		4.802	4.768	4.896	4.785	4.803
14[a]		5.012	4.958	5.02	5.046	5.031
15		4.996	4.858	4.754	4.814	4.833

Table 2 (*continued*)

Predicted pIC_{50}						
Compd	R	Experimental pIC_{50}	CoMFA	CoMFA-RF	CoMSIA	HQSAR
16		5.219	5.245	5.257	5.22	5.172
17		5.267	5.416	5.423	5.324	5.287
18[a]		5.186	5.152	5.205	5.29	5.303
19		4.533	4.636	4.669	4.566	4.629
20		4.622	4.879	4.907	4.837	4.426
21[a]		5.310	5.197	5.153	5.271	5.157

[a] Test set.

Table 3
Statistical parameters of QSAR models (HCT116).

Parameters	CoMFA	CoMFA-RF	CoMSIA	HQSAR
PLS analysis				
q^2	0.728	0.764	0.678	0.811
r^2_{ncv}(L-10-O)	0.750	0.758	0.678	0.808
ONC	6	4	6	5
SEP	0.309	0.261	0.386	0.245
r^2_{ncv}	0.982	0.960	0.977	0.986
SEE	0.081	0.108	0.102	0.067
R pearson	0.954	0.925	0.931	0.967
F	79.930	65.274	37.118	138.870
r^2_{bs}	1.00	0.997	0.997	0.999
SEE_{bs}	0.00	0.004	0.006	0.001
Contribution				
Steric	0.586	0.563	0.206	–
Electrostatic	0.414	0.437	0.311	–
Hydrophobic	–	–	0.383	–
Donor	–	–	0.043	–
Acceptor	–	–	0.057	–

q^2: cross-validated correlation coefficient after the leave-one-out procedure; ONC: optimal number of principal components; r^2_{ncv} : non-cross-validated correlation coefficient; SEE: standard error of estimate; F: the value of F statistic; r^2_{bs}: the average r^2 value from a bootstrapping analysis for 100 runs; SEE_{bs}: the average SEE value from a bootstrapping analysis for 100 runs; r^2_{cv} :(mean) the average r_{cv} from ten times tenfold cross-validation.

Table 4
Statistical parameters of QSAR models (A2780).

Parameters	CoMFA	CoMFA-RF	CoMSIA	HQSAR
PLS analysis				
q^2	0.721	0.665	0.634	0.829
r^2_{cv}(L-10-O)	0.711	0.679	0.565	0.743
ONC	4	4	2	4
SEP	0.224	0.245	0.235	0.224
r^2_{ncv}	0.961	0.885	0.842	0.992
SEE	0.084	0.144	0.155	0.093
R pearson	0.936	0.912	0.910	0.940
F	67.786	21.082	34.520	54.800
r^2_{bs}	0.998	0.995	1.00	0.998
SEE_{bs}	0.002	0.006	0.00	0.004
Contribution				
Steric	0.557	0.730	0.227	–
Electrostatic	0.443	0.270	0.355	–
Hydrophobic	–	–	0.318	–
Donor	–	–	0.041	–
Acceptor	–	–	0.059	–

q^2: cross-validated correlation coefficient after the leave-one-out procedure; ONC: optimal number of principal components; r^2_{ncv}: non-cross-validated correlation coefficient; SEE: standard error of estimate; F: the value of F statistic; r^2_{bs}: the average r^2 value from a bootstrapping analysis for 100 runs; SEE_{bs}: the average SEE value from a bootstrapping analysis for 100 runs; r^2_{cv} :(mean) the average r_{cv} from ten times tenfold cross-validation.

3.2. CoMSIA statistical results

The CoMSIA technique deals with direct correlation of ligand affinities to changes in molecular properties [66]. The CoMSIA model was generated using combinations of five steric (S), electrostatic (E), hydrophobic (H), hydrogen bonding acceptor (A) and hydrogen bonding donor (D) fields. The statistical parameters of CoMSIA model were summarized in Tables 3 and 4. In PLS analysis, the q^2 values of 0.678 and 0.634 with ONC of 6 and 2, SEP of 0.386 and 0.235 and r^2_{cv} (L-10-O) of 0.678 and 0.586 was obtained with column filtering of 2.0 kcal/mol for HCT116 and A2780 cell lines, respectively. The non-cross-validated PLS analysis gave r^2_{ncv} values of 0.977 and 0.842 with SEE values of 0.102 and 0.155, F values of 37.118 and 34.520, $r^2 - q^2$ values of 0.306 and 0.208 and $R_{pearson}$ values of 0.931 and 0.910 for HCT116 and A2780 cell lines, respectively.

High bootstrapped r^2 values of 0.997 and 1.00 and SEE_{bs} of 0.006 and 0.00 suggest a high degree of confidence in the analysis. For

CoMSIA, the contribution of the steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor field descriptors were 0.206 and 0.227; 0.311 and 0.355; 0.383 and 0.318; 0.0.043 and 0.041; 0.057 and 0.0.059, for HCT116 and A2780 cell lines, respectively. These molecular fields were not completely independent of each other and could form 31 combinations (Fig. 3.).

Among the first five models, hydrophobic field with high q^2 values ($q^2 = 0.615$ and 0.739 for HCT116 and A2780) was more important than the other four fields.

In CoMSIA model, the combination of hydrophobic and hydrogen bond donor (HD) was found to be the best. CoMSIA (HD) combination gave q^2 values of 0.675 and 0.779 for HCT116 and A2780 cell lines, respectively. In the model CoMSIA, this combination shared the large part and indicated that internal prediction of HD combination was good.

In the CoMFA model, the steric field contribution and in the CoMSIA model, the hydrophobic and H-donor and hydrogen bond donor contributions shared the large part. In docking studies, it was also recognized that the steric and hydrophobic effects and H-bond network around the key residues in the active site played a significant role in the binding of ligand to HDAC1. It was also demonstrated that the hydrophobic and steric properties were important in the design of new HADC1 inhibitors.

3.3. HQSAR statistical results

The HQSAR is a technique for QSAR analysis that is useful in exploring the combination of each molecule under study to the biological activity and eliminates the need of alignment, generation of 3D structures and putative binding conformation. The performance of the HQSAR model was affected by three parameters, including the fragment size, the fragment type (fragment distinction) and hologram length. The HQSAR models with statistical parameters are shown in Tables 3 and 4.

The best statistical results of HQSAR model were obtained with q^2 values of 0.811 and 0.829, ONC of 5 and 4, SEP of 0.245 and 0.224 and r^2_{cv} (L-10-O) of 0.808 and 0.743, r^2_{ncv} of 0.986 and 0.992 with SEE of 0.067 and 0.093, F values of 138.870 and 54.800, $r^2 - q^2$ of 0.175 and 0.163, r^2_{bs} of 0.999 and 0.998 with SEE_{bs} of 0.001 and 0.004 and $R_{pearson}$ of 0.967 and 0.940 using a relevant hologram length (HL) of 61 and 353, fragment distinction (atom (A), bonds (B), hydrogen atoms (H) and chirality (Ch), and atoms (A), hydrogen atoms (H), chirality (Ch) and donor and acceptor (DA), and the fragment size of 4–7 (Tables 5–8) for HDAC inhibitory activity of HCT116 and A2780 cell lines, respectively. All the results demonstrated that the HQSAR model was also highly predictive.

3.4. Validation of QSAR models

The predictive abilities of the QSAR models were externally

Table 5
HQSAR analysis of various fragment distinctions on the key statistical parameters using fragment size (4–7) (HCT116).

model	Fragment distinction	q^2	SEP	r^2	SEE	HL	N
1–1	A	0.390	0.385	0.739	0.252	307	2
1–2	A/B	0.468	0.375	0.890	0.170	257	3
1–3	A/C	0.573	0.336	0.914	0.151	199	3
1–4	A/H	0.779	0.241	0.965	0.096	307	3
1–5	A/Ch	0.390	0.385	0.739	0.252	307	2
1–6	A/DA	0.386	0.403	0.876	0.181	83	3
1–7	A/B/C	0.522	0.355	0.903	0.160	401	3
1–8	A/B/H	0.760	0.242	0.890	0.164	61	2
1–9	A/B/Ch	0.466	0.375	0.890	0.171	257	3
1–10	A/B/DA	0.522	0.371	0.933	0.138	257	4
1–11	A/C/H	0.745	0.259	0.959	0.104	199	3
1–12	A/C/Ch	0.573	0.336	0.914	0.151	199	3
1–13	A/C/DA	0.493	0.382	0.938	0.134	199	4
1–14	A/H/Ch	0.771	0.246	0.961	0.101	307	3
1–15	A/H/DA	0.705	0.291	0.959	0.109	199	4
1–16	A/Ch/DA	0.361	0.411	0.868	0.187	83	3
1–17	A/B/C/H	0.764	0.250	0.964	0.098	151	3
1–18	A//C/Ch	0.522	0.355	0.903	0.160	401	3
1–19	A/B/C/DA	0.478	0.388	0.934	0.137	401	4
1–20	A/B/H/Ch	0.811	0.245	0.986	0.067	61	3
1–21	A/B/H/DA	0.725	0.295	0.977	0.085	61	5
1–22	A/B/Ch/DA	0.507	0.377	0.933	0.139	257	4
1–23	A/C/H/Ch	0.742	0.261	0.958	0.105	199	3
1–24	A/H/Ch/DA	0.715	0.286	0.968	0.101	97	4
1–25	A/C/H/DA	0.793	0.270	0.991	0.055	53	6
1–26	A/C/H/Ch/DA	0.696	0.310	0.978	0.084	53	5
1–27	A/B/H/Ch/DA	0.772	0.268	0.979	0.081	61	5
1–28	A/B/C/Ch/DA	0.521	0.371	0.921	0.151	307	4
1–29	A/B/C/H/DA	0.661	0.320	0.941	0.130	151	4
1–30	A/B/C/H/Ch	0.765	0.249	0.964	0.097	151	3
1–31	A/B/C/H/Ch/DA	0.664	0.311	0.938	0.134	151	4

q^2, cross-validated correlation coefficient; r^2, non-cross-validated correlation coefficient; SEE, standard estimated error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atom; B, bond; C, connections; H, hydrogen atom; Ch, chirality; D, donor and acceptor.
The model chosen for analysis is highlighted in bold fonts.

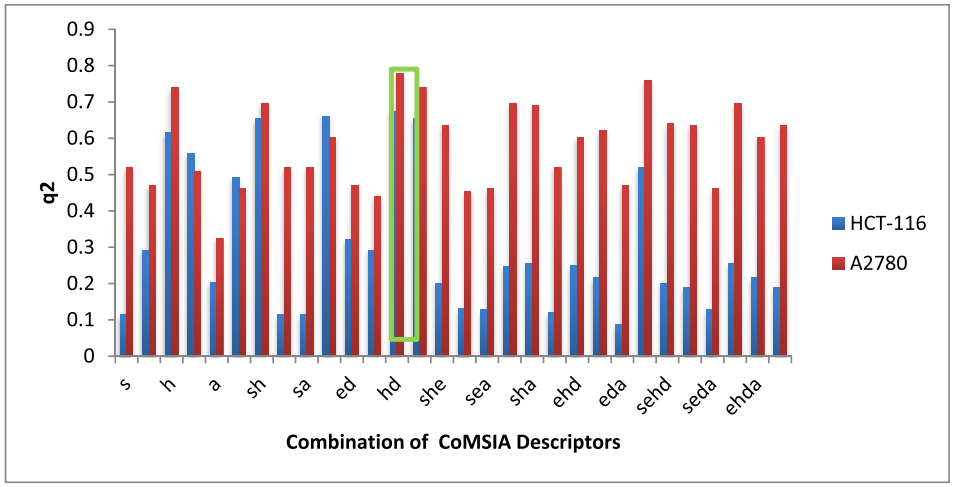


Fig. 3. The results of the distribution of q^2 values that were obtained from 31 combinations of CoMSIA fields of HCT116 and A2780. s, steric; e, electrostatic; h, hydrophobic; d, H-bond donor; a, H-bond acceptor.

Table 6
The statistical results of QSAR model using the model 1−20 (including fragments A/B/H/Ch) with different fragment sizes (HCT116).

model	Fragment size	q^2	SEP	r^2	SEE	HL	N
2−1	1−4	0.382	0.338	0.639	0.297	353	2
2−2	2−5	0.694	0.297	0.955	0.114	61	4
2−3	3−6	0.573	0.336	0.914	0.151	199	3
2−4	4−7	0.807	0.226	0.954	0.110	61	3
2−5	5−8	0.776	0.243	0.948	0.117	71	3
2−6	6−9	0.741	0.261	0.940	0.126	71	3
2−7	7−10	0.758	0.264	0.982	0.071	257	4
2−8	8−11	0.682	0.278	0.821	0.209	199	2
2−9	9−12	0.571	0.323	0.821	0.209	199	2

q^2, cross-validated correlation coefficient; r^2, non-cross-validated correlation coefficient; SEE, standard estimated error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atom; B, bond; C, connections; H, hydrogen atom; Ch, chirality; D, donor and acceptor.
The model chosen for analysis is highlighted in bold fonts.

Table 7
HQSAR analysis of various fragment distinctions on the key statistical parameters using fragment size (4−7) (A2780).

model	Fragment distinction	q^2	SEP	r^2	SEE	HL	N
1−1	A	0.399	0.291	0.576	0.244	401	1
1−2	A/B	0.525	0.279	0.913	0.119	53	3
1−3	A/C	0.398	0.302	0.761	0.190	61	2
1−4	A/H	0.792	0.185	0.965	0.076	61	3
1−5	A/Ch	0.399	0.291	0.576	0.244	401	1
1−6	A/DA	0.404	0.290	0.566	0.247	307	1
1−7	A/B/C	0.375	0.297	0.571	0.246	353	1
1−8	A/B/H	0.737	0.217	0.967	0.077	61	4
1−9	A/B/Ch	0.598	0.268	0.974	0.068	53	4
1−10	A/B/DA	0.572	0.306	0.979	0.068	53	6
1−11	A/C/H	0.755	0.210	0.975	0.067	97	4
1−12	A/C/Ch	0.398	0302	0.761	0.190	61	2
1−13	A/C/DA	0.460	0.298	0.760	0.199	61	3
1−14	A/H/Ch	0.760	0.198	0.966	0.075	257	3
1−15	A/H/DA	0.821	0.198	0.991	0.044	353	6
1−16	A/Ch/DA	0.581	0.287	0.956	0.094	151	5
1−17	A/B/C/H	0.758	0.218	0.978	0.066	151	5
1−18	A//C/Ch	0.374	0.297	0.571	0.246	353	1
1−19	A/B/C/DA	0.524	0306	0.960	0.089	401	5
1−20	A/B/H/Ch	0.743	0.206	0.950	0.091	61	3
1−21	A/B/H/DA	0.815	0.201	0.986	0.055	61	6
1−22	A/B/Ch/DA	0.329	0.307	0.524	0.259	353	1
1−23	A/C/H/Ch	0.754	0.210	0.975	0.068	97	4
1−24	A/H/Ch/DA	0.829	0.224	0.992	0.093	353	6
1−25	A/C/H/DA	0.750	0.234	0.985	0.058	71	6
1−26	A/C/H/Ch/DA	0.785	0.217	0.971	0.079	53	6
1−27	A/B/H/Ch/DA	0.783	0.207	0.976	0.069	61	5
1−28	A/B/C/Ch/DA	0.528	0.305	0.970	0.077	401	5
1−29	A/B/C/H/DA	0.730	0.231	0.982	0.059	151	5
1−30	A/B/C/H/Ch	0.762	0.217	0.979	0.065	151	5
1−31	A/B/C/H/Ch/DA	0.528	0.305	0.970	0.077	401	5

q^2, cross-validated correlation coefficient; r^2, non-cross-validated correlation coefficient; SEE, standard estimated error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atom; B, bond; C, connections; H, hydrogen atom; Ch, chirality; D, donor and acceptor.
The model chosen for analysis is highlighted in bold fonts.

validated using the independent test set that was not used for the model generation [64]. q^2 and r^2 parameters, obtained from internal validation, were used for confirming the stability and the predictive ability of the models. The results of external validation parameters are listed in Tables 9 and 10. The QSAR models for the whole test set includes six compounds gave the r^2_{pred} and r^2_m values of 0.685 and 0.737; 0.629 and 0.574 (CoMFA), 0.552 and 0.680; 0.540 and 0.532 (CoMFA-RF), 0.721 and 0.756; 0.640 and 0.720 (CoMSIA), and 0.613 and 0.730; 0.612 and 0.552 (HQSAR) and high slope regression lines with k and k' values of 0.947 and 1.080; 1.031

Table 8
The statistical results of QSAR model using the model 1−24 (including fragments A/H/Ch/DA) with different fragment sizes (A2780).

model	Fragment size	q^2	SEP	r^2	SEE	HL	N
2−1	1−4	0.471	0.273	0.604	0.236	59	1
2−2	2−5	0.468	0.274	0.603	0.236	61	1
2−3	3−6	0.820	0.198	0.993	0.040	97	6
2−4	4−7	0.829	0.194	0.992	0.042	353	6
2−5	5−8	0.816	0.182	0.964	0.080	307	4
2−6	6−9	0.819	0.180	0.954	0.090	53	4
2−7	7−10	0.797	0.191	0.937	0.106	83	4
2−8	8−11	0.748	0.212	0.960	0.084	307	4
2−9	9−12	0.743	0.225	0.973	0.073	83	5

q^2, cross-validated correlation coefficient; r^2, non-cross-validated correlation coefficient; SEE, standard estimated error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atom; B, bond; C, connections; H, hydrogen atom; Ch, chirality; D, donor and acceptor.
The model chosen for analysis is highlighted in bold fonts.

Table 9
Statistical parameters of validation method for QSAR models (HCT116).

Parameters	CoMFA	CoMFA-RF	CoMSIA	HQSAR
r^2_{pred}	0.685	0.552	0.721	0.613
r^2_o	0.765	0.629	0.804	0.798
r'^2_o	0.817	0.631	0.890	0.726
$(r^2\text{-}r^2_o)/r^2$	0.064	0.053	0.082	0.050
$(r^2\text{-}r'^2_o)/r^2$	0.001	0.051	0.052	0.045
k	0.974	0.972	0.979	0.972
k'	1.031	1.027	1.026	1.027
r^2_m	0.629	0.540	0.640	0.612
r'^2_m	0.792	0.543	0.772	0.620
Δr^2_m	0.163	0.003	0.132	0.008
$RMSE_{train}$	0.097	0.104	0.094	0.107
$RMSE_{test}$	0.035	0.101	0.091	0.024
MAE_{train}	0.053	0.068	0.062	0.060
MAE_{test}	0.014	0.063	0.044	0.011
RSS_{train}	0.199	0.227	0.134	0.239
RSS_{test}	0.026	0.213	0.175	0.011
CCC_{train}	0.948	0.945	0.966	0.937
CCC_{test}	0.979	0.850	0.985	0.993

r^2_{pred}: predicted correlation coefficient for the test set of compounds.
r^2_o: correlation coefficient of regression through the origin for predicted versus observed activities (test set).
r'^2_o: correlation coefficient of regression through the origin for observed versus predicted activities (test set).
r^2_m: modified squared correlation coefficient (test set); RMSE: root mean squared error; MAE: mean absolute error; RSS: residual sum of squares; CCC: concordance correlation coefficient.

and 0.990 (CoMFA), 0.972 and 1.00; 1.027 and 0.998 (CoMFA-RF), 0979 and 1.008; 1.026 and 0.998 (CoMSIA), and 0.972 and 1.001; 1.027 and 0.997 (HQSAR) for HDAC inhibitory activity of HCT116 and A2780, respectively. r^2_o and r'^2_o values of 0.765 and 0.689; 0.817 and 0.735 (CoMFA), 0.629 and 0.628; 0.631 and 0.648 (CoMFA-RF), 0.804 and 0.739; 0.890 and 0.757 (CoMSIA), and 0.798 and 0.662; 0.726 and 0.705 (HQSAR) for HDAC inhibitory activity of HCT116 and A2780 cell lines, respectively, were used to calculate the relationship between r^2, r^2_o and r'^2_o that $(r^2\text{-}r^2_o)/r^2$ and $(r^2\text{-}r'^2_o)/r^2$ values of 0.064 and 0.068; 0.001 and 0.005 (CoMFA), 0.053 and 0.064; 0.051 and 0.034 (CoMFA-RF), 0.082 and 0.034; 0.052 and 0.010 (CoMSIA), and 0.050 and 0.072; 0.045 and 0.011 (HQSAR), respectively were obtained.

The QSAR models yielded RMSE, MAE and CCC values of 0.097 and 0.089, 0.053 and 0.058, 0.948 and 0.940; 0.035 and 0.126, 0.014 and 0.045, 0.979 and 0.959 (CoMFA); 0.104 and 0.109, 0.068 and 0.077, 0.945 and 0.913; 0.101 and 0.076, 0.063 and 0.030, 0.850 and

Table 10
Statistical parameters of validation method for QSAR models (A2780).

Parameters	CoMFA	CoMFA-RF	CoMSIA	HQSAR
r^2_{pred}	0.737	0.680	0.756	0.730
r^2_o	0.689	0.628	0.739	0.662
$r^{2'}_o$	0.735	0.648	0.757	0.705
$(r^2\text{-}r^2_o)/r^2$	0.068	0.064	0.034	0.072
$(r^2\text{-}r^{2'}_o)/r^2$	0.005	0.034	0.010	0.011
k	1.080	1.00	1.008	1.001
k'	0.990	0.998	0.990	0.997
r^2_m	0.574	0.532	0.720	0.552
$r^{2'}_m$	0.692	0.570	0.640	0.649
Δr^2_m	0.118	0.038	0.080	0.097
$RMSE_{train}$	0.089	0.109	0.107	0.078
$RMSE_{test}$	0.126	0.076	0.093	0.061
MAE_{train}	0.058	0.077	0.075	0.053
MAE_{test}	0.045	0.030	0.038	0.029
RSS_{train}	0.170	0.250	0.241	0.126
RSS_{test}	0.334	0.121	0.180	0.078
CCC_{train}	0.940	0.913	0.913	0.952
CCC_{test}	0.959	0.878	0.857	0.937

r^2_{pred}: predicted correlation coefficient for the test set of compounds.
r^2_o: correlation coefficient of regression through the origin for predicted versus observed activities (test set).
$r^{2'}_o$: correlation coefficient of regression through the origin for observed versus predicted activities (test set).
r^2_m: modified squared correlation coefficient (test set); RMSE: root mean squared error; MAE: mean absolute error; RSS: residual sum of squares; CCC: concordance correlation coefficient.

0.878 (CoMFA-RF); 0.094 and 0.107, 0.062 and 0.075, 0.966 and 0.913; 0.091 and 0.093, 0.044 and 0.038, 0.985 and 0.857 (CoMSIA); 0.107 and 0.078, 0.060 and 0.053, 0.937 and 0.952; 0.024 and 0.061, 0.011 and 0.029, 0.993 and 0.937 (HQSAR) for Training and test set of HCT116 and A2780 HDAC inhibitory activity, respectively.

From the values of the performance criteria parameters yielded by the QSARs in training and test data (Tables 9 and 10), it is evident that all of the models yielded considerably low RMSE and MAE values and high CCC values which show that models built by training set could be used for the prediction of these chemotypes.

These results confirm that the QSAR models, especially QSAR model results from HDAC inhibitory activity of HCT116 could be used to predict the biological activities of new compounds and their derivatives.

The correlation plots between the predicted and experimental activities for HCT116 and A2780 are shown in Figs. 4 and 5, respectively. Most of the compounds were located on or near to the trend line in the QSAR models and these results confirm that these models had good predictive ability for new compounds.

The residual values of the QSAR models for HCT116 and A2780 cell lines are shown in Figs. 6 and 7, respectively. The CoMSIA and HQSAR models showed smaller residuals than the CoMFA and CoMFA-RF models and were the better models.

3.5. Evaluation of the Y -randomization test and application domain (AD) of model

The QSAR models were further validated by applying the Y-

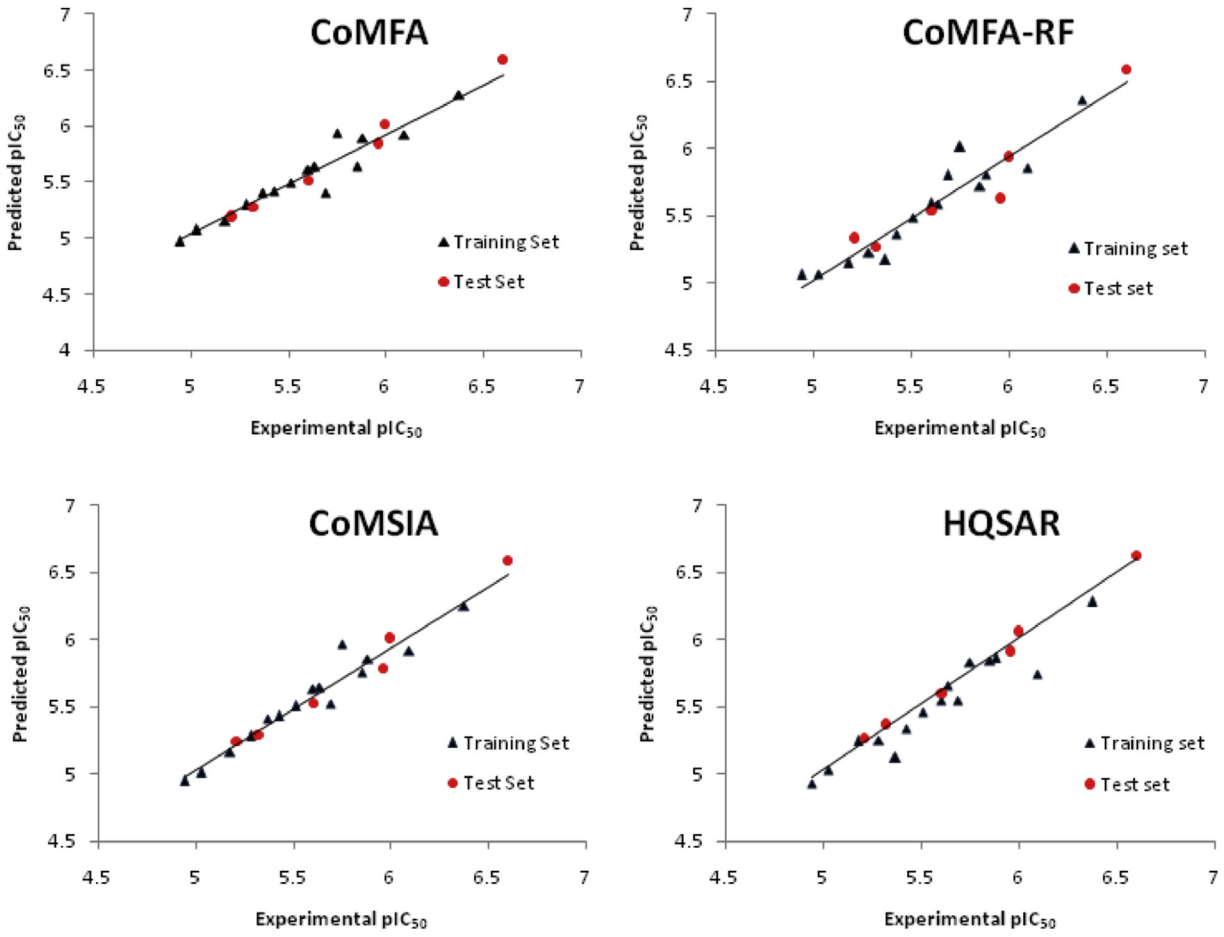


Fig. 4. The plot of predicted pIC_{50} versus experimental pIC_{50} values for training and test sets compounds by QSAR models (HCT116).

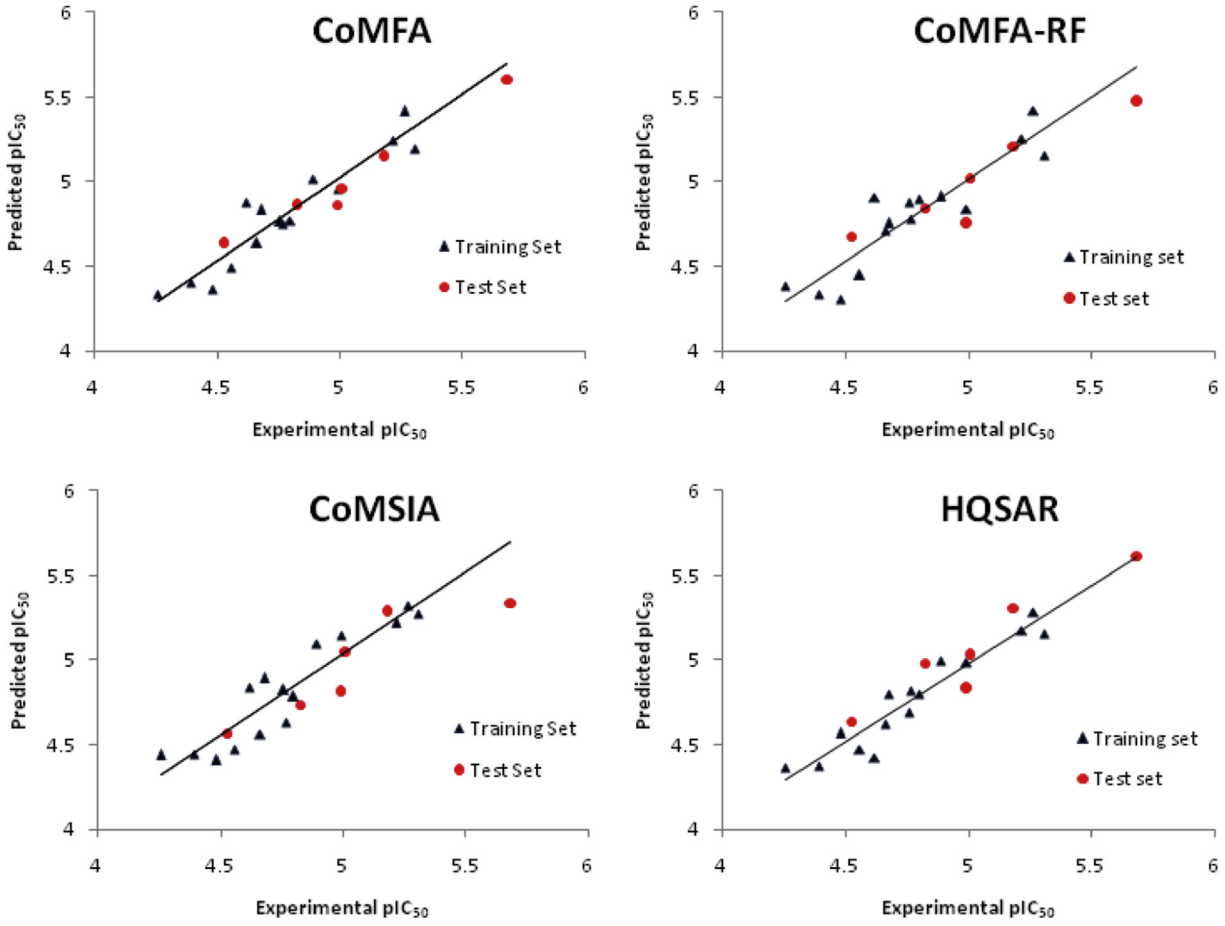


Fig. 5. The plot of predicted pIC_{50} versus experimental pIC_{50} values for training and test sets compounds by QSAR models (A2780).

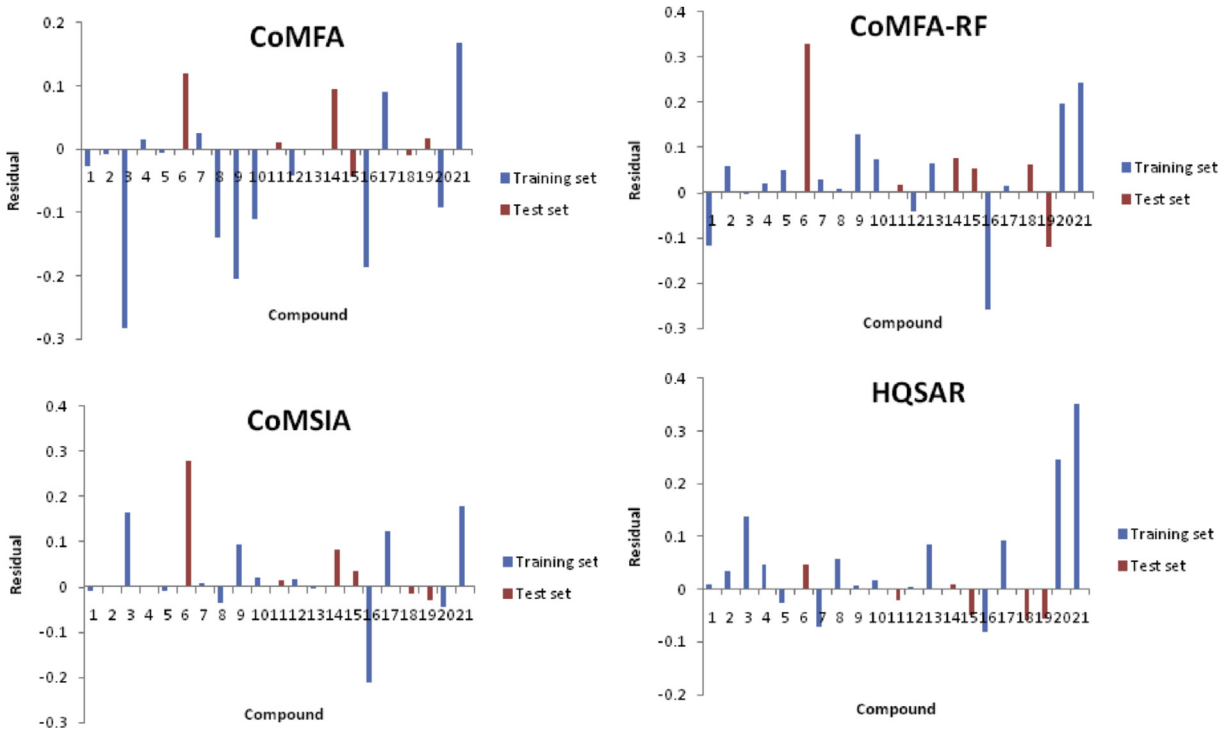


Fig. 6. Residual plots between experimental and predicted values for QSAR models (HCT116).

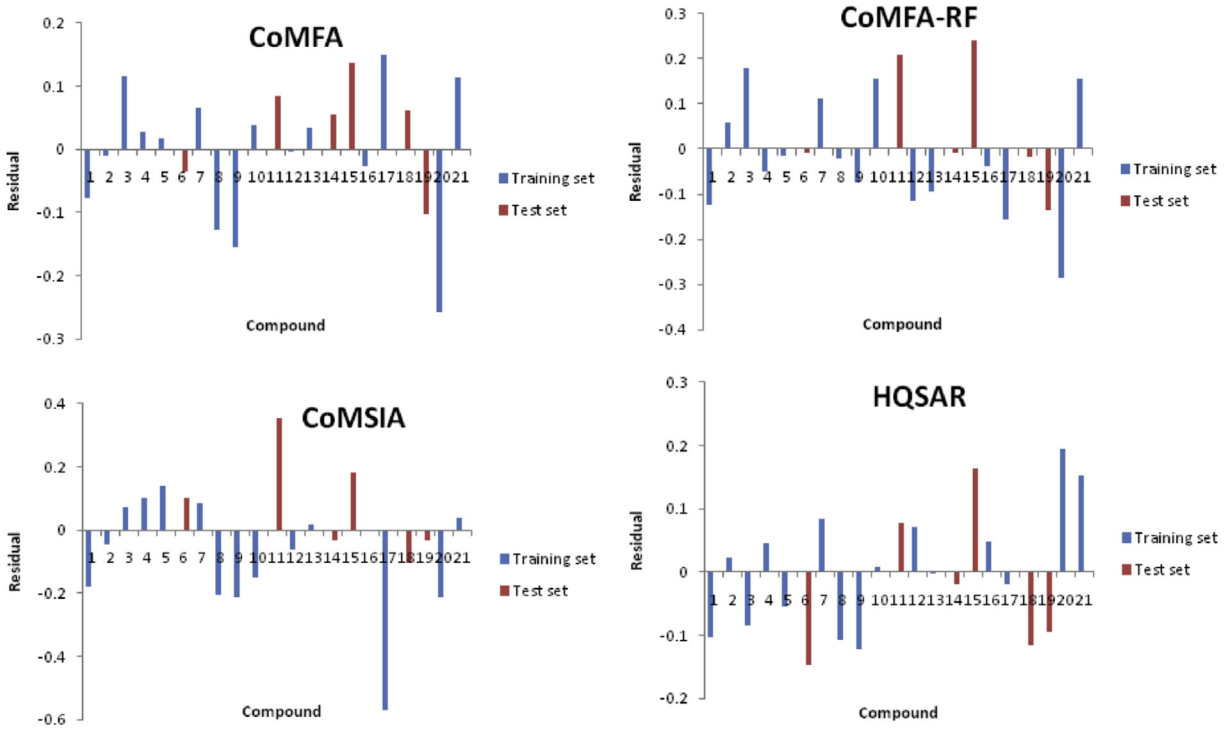


Fig. 7. Residual plots between experimental and predicted values for QSAR models (A2780).

randomization test to assess the robustness of the models and to avoid chance correlation [67,68]. Thus, for every original model, several random shuffles of the dependent variable (biological activity) were performed and a new QSAR model was developed using the original independent variable matrix and the results are shown in Table 11. The low q^2 and r^2_{ncv} values ($q^2 < 0.5$ and $r^2_{ncv} < 0.6$) show that the good results obtained in the formulation of the final models were not by chance.

For a new compound with no experimental data, a predicted value of QSAR models without an idea of reliability of the value is not useful. Therefore, for evaluating new compounds, a very important step in QSAR model development is the definition of the applicability domain of regression or classification models [69].

The Williams plot, the plot of the standardized residuals (δ) vs. leverage values (h_i), was used to illustrate the predictive and express the applicability domain of the models for each chemical compound [70,71].

The standardized residuals (δ) value is calculated by Equation (15) [72]:

Table 11
q^2 and r^2_{ncv} values after several Y-randomization tests.

Y-random Iteration	HCT116				A2780			
	CoMFA		CoMSIA		CoMFA		CoMSIA	
	q^2	r^2_{ncv}	q^2	r^2_{ncv}	q^2	r^2_{ncv}	q^2	r^2_{ncv}
1	−0.210	0.371	−0.353	0.228	−0.221	0.375	−0.352	0.237
2	−0.321	0.430	−0.320	0.344	−0.342	0.441	−0.226	0.350
3	−0.172	0.332	−0.182	0.275	−0.201	0.334	−0.164	0.289
4	−0.175	0.307	−0.186	0.330	−0.189	0.321	−0.144	0.342
5	−0.092	0.341	−0.155	0.346	−0.111	0.352	−0.189	0.355
6	−0.120	0.307	−0.137	0.166	−0.127	0.320	−0.162	0.172
7	−0.117	0.330	−0.401	0.380	−0.120	0.332	−0.435	0.390
8	−0.240	0.362	−0.256	0.271	−0.251	0.369	−0.227	0.283
9	−0.190	0.343	−0.193	0.332	−0.187	0.347	−0.184	0.341
10	−0.111	0.321	−0.117	0.173	−0.105	0.335	−0.117	0.180
11	−0.237	0.353	−0.248	0.348	−0.224	0.372	−0.262	0.368
12	−0.260	0.380	−0.248	0.282	−0.235	0.383	−0.195	0.276
13	−0.193	0.334	−0.198	0.328	−0.175	0.335	−0.203	0.348
14	−0.172	0.320	−0.183	0.338	−0.169	0.316	−0.190	0.357
15	−0.203	0.332	−0.220	0.359	−0.184	0.352	−0.255	0.373
16	−0.281	0.370	−0.273	0.372	−0.286	0.374	−0.249	0.381
17	−0.227	0.361	−0.237	0.330	−0.253	0.368	−0.230	0.346
18	−0.250	0.374	−0.260	0.352	−0.236	0.373	−0.268	0.369
19	−0.220	0.335	−0.240	0.381	−0.221	0.333	−0.241	0.392
20	−0.245	0.373	−0.242	0.334	−0.258	0.380	−0.122	0.340
Non-Random	0.728	0.982	0.678	0.977	0.721	0.961	0.634	0.842

$$\delta = \frac{y_i - \widehat{y_i}}{\sqrt{\sum_{i=1}^{n} \frac{(y_i - \widehat{y_i})^2}{(n-A-1)}}} \quad (15)$$

where $y_i, \widehat{y_i}$ are the observed and predicted values for *i-th* the compound, respectively, n is the number of compounds and A is the number of descriptors. Also, the leverage value (h_i) is defined by Equation (16):

$$h_i = X_i^T (X^T X)^{-1} X_i \quad (i = 1, \quad n) \quad (16)$$

where x_i is the descriptor-row vector of the *i-th* the compound, X_i^T is the transpose of x_i, X is the descriptor matrix of the training set compounds and X^T is the transpose of X.

The warning leverage value (h^*), as a prediction tool, is expressed as:

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

where k is the number of model descriptors and n is the number of training compounds.

The Williams plot illustrates the distribution of data and its restricting rang termed cutoff lines which all data should be between ±3 units (horizontal dotted line) for standardized residuals and the leverage value (h_i) should be less than warning leverage ($h_i < h^*$). The Williams plot for the training set is used to identify molecules with the greatest structural influence ($h_i < h^*$) in developing the QSAR models. Molecules with $h_i > h^*$ are evaluated to be unreliably predicted by the models due to substantial extrapolate.

Cook's distance is used to estimate the influence of a single observation of the model [73], and is defined by Equation (17):

$$D_i = \frac{e_i^2}{p+1} * \frac{h_i}{1-h_i} \quad (17)$$

where e_i^2 is the standard residual of the *i-th* the compound, p is the number of descriptors, and h_i is the leverage value of the *i-th* the compound. The cutoff of the Cook's distance is defined as $\frac{4}{(n-p-1)}$, and the compounds with Cook's distance higher than the cutoff value are marked as highly influential points of the model.

In this work, for CoMFA and CoMSIA models, most of the compounds fall into their corresponding application domain. These results indicated that our QSAR models had achieved a reliable activity prediction for the compounds.

As shown in the Williams plot of CoMFA model for data set (Fig. 8a), only one compound (**8**) of the training set for HCT116 had greater value than the warning leverage (h^*) value of 0.567. This compound had low standard residual value and could be considered as influential in fitting the model performance, but not necessarily outlier to be deleted from the training set. The test compounds were within the applicability domain (AD) of HCT116 and A2780 indicating that their predicted activity values were reliable. Also, at the Cook's plot (cutoff = 0.333) of CoMFA model for HCT116 (Fig. 8b); there was not any highly influential compound for training and test set and one highly influential compound for A2780 may slightly distort the regression. In addition, the histograms of the residuals distribution for HCT116 and A2780 were confirmed with histogram plots as shown in Fig. 9c and c*.

Also, at the Williams plots of CoMSIA model for a data set of HCT116 and A2780 (Fig. 10a and a*); there was not any outlier compound for training and test set ($h^* = 1.153$). Otherwise, according to the Cook's distances (cutoff = 0.444) of the compounds in the data set, two highly influential compounds for HCT116 and

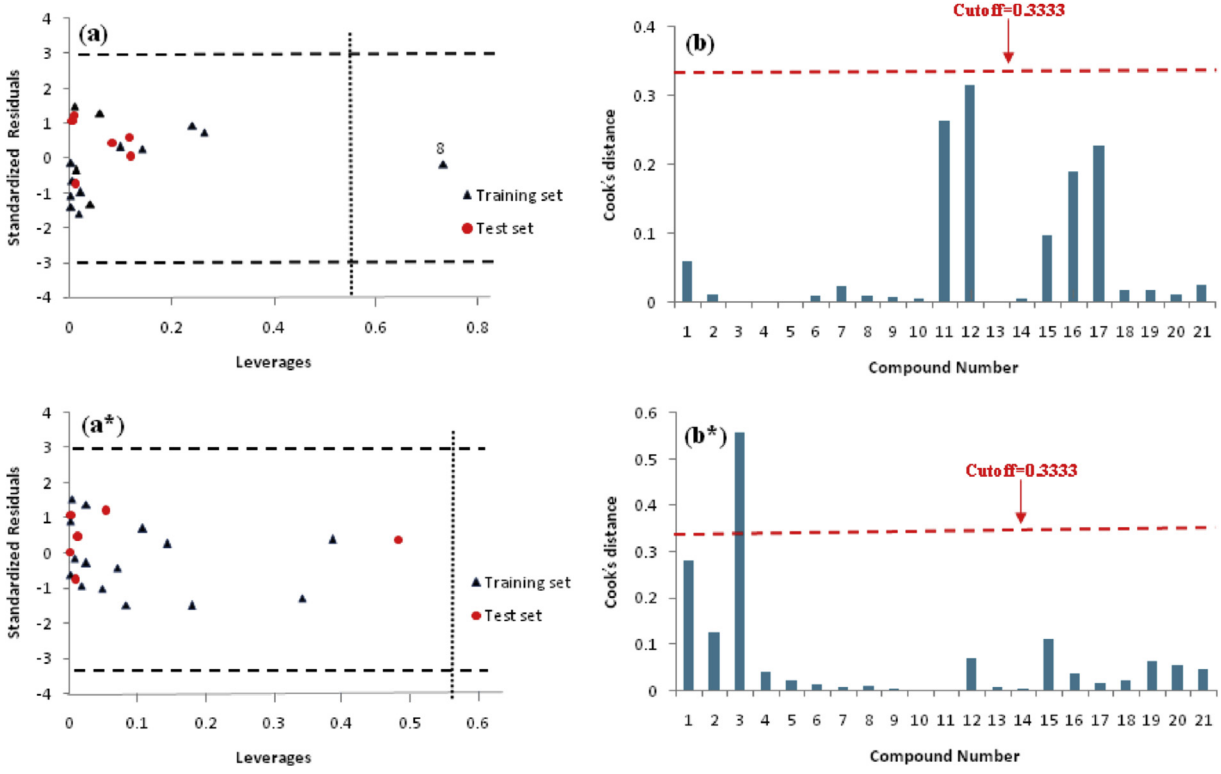


Fig. 8. Williams plot describing the applicability domain of the CoMFA model for the training and test sets of HCT116 and A2780 ($h^* = 0.567$) (a, a*); Cook's distance plot (b, b*).

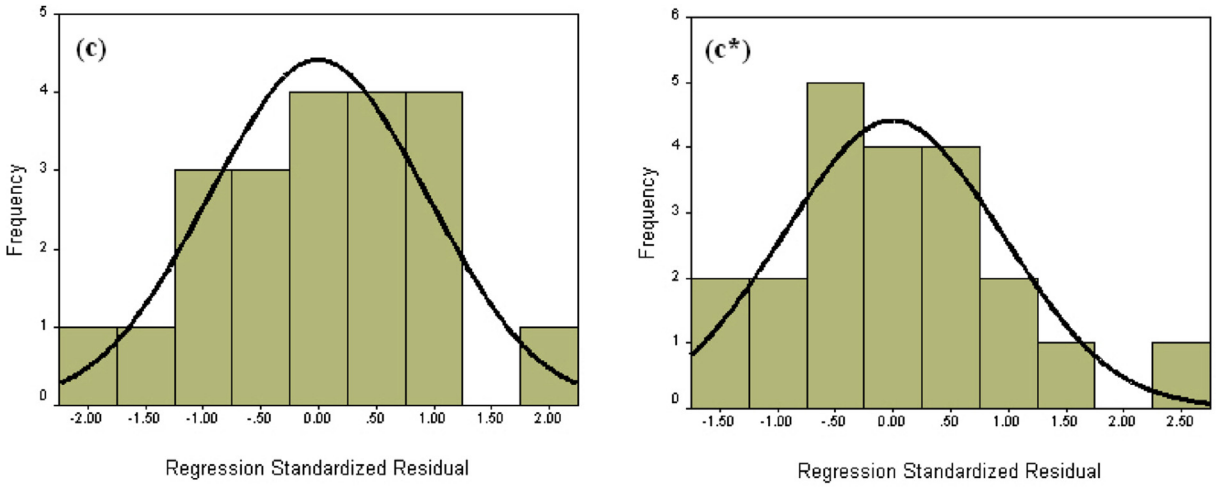


Fig. 9. Histogram of model CoMFA residuals (c, c*).

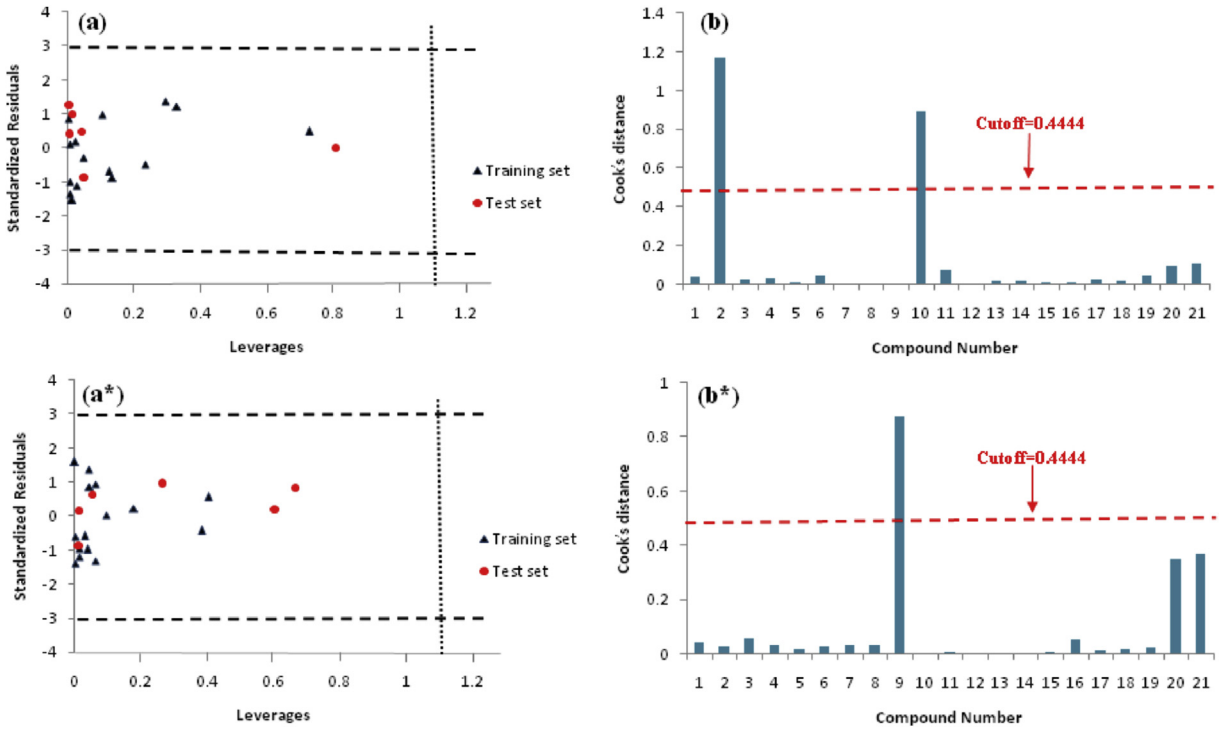


Fig. 10. Williams plot describing the applicability domain of the CoMFA model for the training and test sets of HCT116 and A2780 (h* = 0.567) (a, a*); Cook's distance plot (b, b*).

one highly influential compound for A2780 may slightly distort the regression (Fig. 10b and b*), also, the histograms of the residuals distribution were confirmed with histogram plots as shown in Fig. 11c and c* and prediction of CoMSIA models is reliable.

3.6. Interpretation of CoMFA and CoMSIA contour maps

The QSAR contour maps were used as an informative tool to visualize the effects of the different fields on the target compound 3D grid orientation of the models. The CoMFA and CoMSIA results were graphically interpreted by field contribution maps using the standard deviation (StDev) at each grid point and the coefficient from the PLS analysis (StDev*Coefficients).

The CoMFA contour maps of the steric and electrostatic fields for the best HDAC inhibitors (compound **11**) are shown in Fig. 12a and b. The field steric is shown by favorable groups (80% contribution) in green color and unfavorable ones (20% contribution) in yellow where the introduction of bulky groups may enhance or diminish the activity.

In the CoMFA steric maps, there was a green contour covering the methylene of benzyloxy group at C-7 position of coumarin ring. The bulky groups in this position of compound improved activity and had the highest activity. The compounds **8**–**21**and **4**–**6** with bulky substituents (e.g. methoxy, ethoxy, propoxy) at this region exhibited more potency, while compounds **1**–**2** and **7** due to the absence of this group had relatively low activity. Substituting the electron-donating or electron-withdrawing at the *ortho* and *meta* positions of benzyloxy ring in compounds **9**–**10** and **12**–**13, 15** and

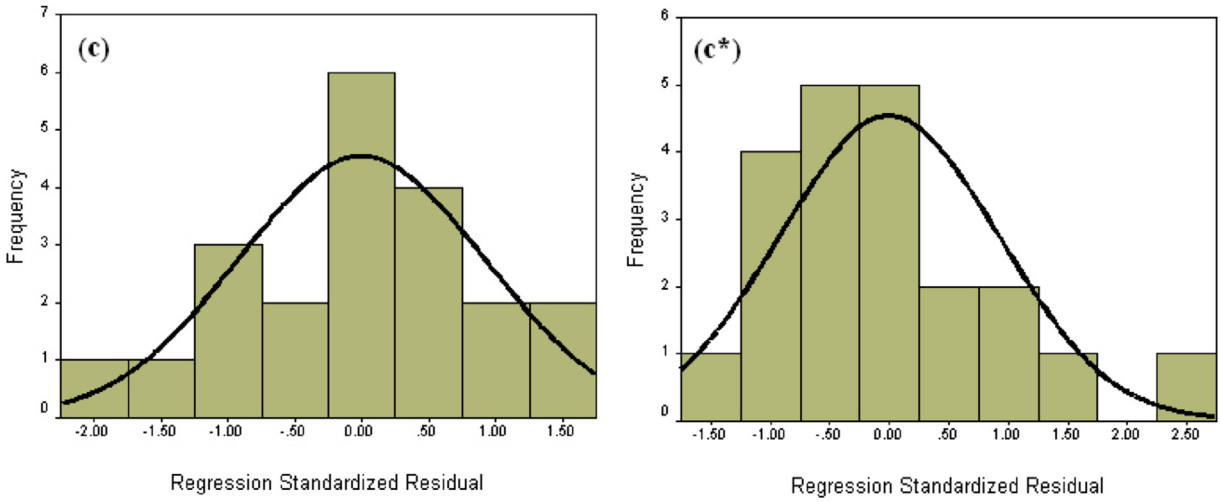


Fig. 11. Histogram of model CoMSIA residuals of HCT116 and A2780 (c, c*).

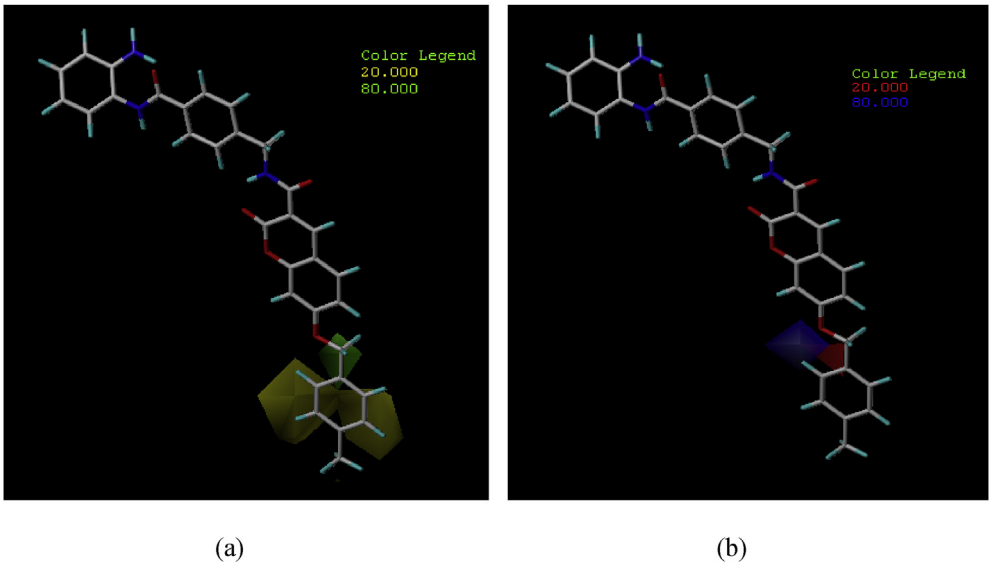


Fig. 12. CoMFAStDev*coeff. Contour plots with the combination of compound **11**. (a) *Steric contour maps*: Green contours indicate regions where bulky groups increase activity and yellow contours indicate regions where bulky groups decrease activity. (b) *Electrostatic contour maps*: Blue contours indicate regions where positive charges increase activity and red contours indicate regions where negative charges increase activity.

19 decreased activity because these substituents were located at disfavored yellow contours. Therefore, these positions of benzyloxy ring should be occupied by the steric moderate and low crowed substituents.

In CoMFA electrostatic contour maps (Fig. 12b), the blue region (80% contribution) is favorable for electropositive groups and red region (20% contribution) is favorable for electronegative groups. The blue contour near the 2 position of benzyloxy ring of compound **11** indicated the introduction of electropositive groups, e.g. methyl in this position could improve the biological activity (compounds **12** < **15**<**9**). Besides, a red contour near to the oxygen atom of benzyloxy ring showed that the electronegative substituent was beneficial to activity (e.g. compounds **1** < **7**<**2** < **4**<**8** < **5**<**3** < **6**<**11**).

In CoMSIA model, the steric and electrostatic, hydrophobic, hydrogen bonding (H-bond) donor and acceptor contour maps of compound **11** are shown in Fig. 15. In the CoMSIA steric maps, the substituents at green region were next to the yellow contours (Fig. 13a). This position should not be occupied by very large groups, so that the substituted methyl in the *para* position of benzyloxy ring was acceptable (compounds **14** < **16**<**18** < **17**<**11**). The blue contours near benzyloxy ring of compound **11** indicated the introduction of electropositive groups, e.g. methyl and methoxy in these positions could improve the biological activity (compounds **9**–**11 and 18**). Besides, a red contour near to the oxygen atom of benzyloxy ring showed that the electronegative substituent was beneficial to activity (e.g. compounds **1** < **7**<**2** < **4**<**8** < **5**<**3** < **6**<**11**) (Fig. 13b).

Therefore, the CoMSIA steric and electrostatic contours were nearly similar to that of CoMFA contours, the hydrophobic interaction and hydrogen bond fields were described here. In the hydrophobic contour map, the yellow region is favorable (80% contribution) for the hydrophobic group while white region (20% contribution) is favorable for the hydrophilic group.

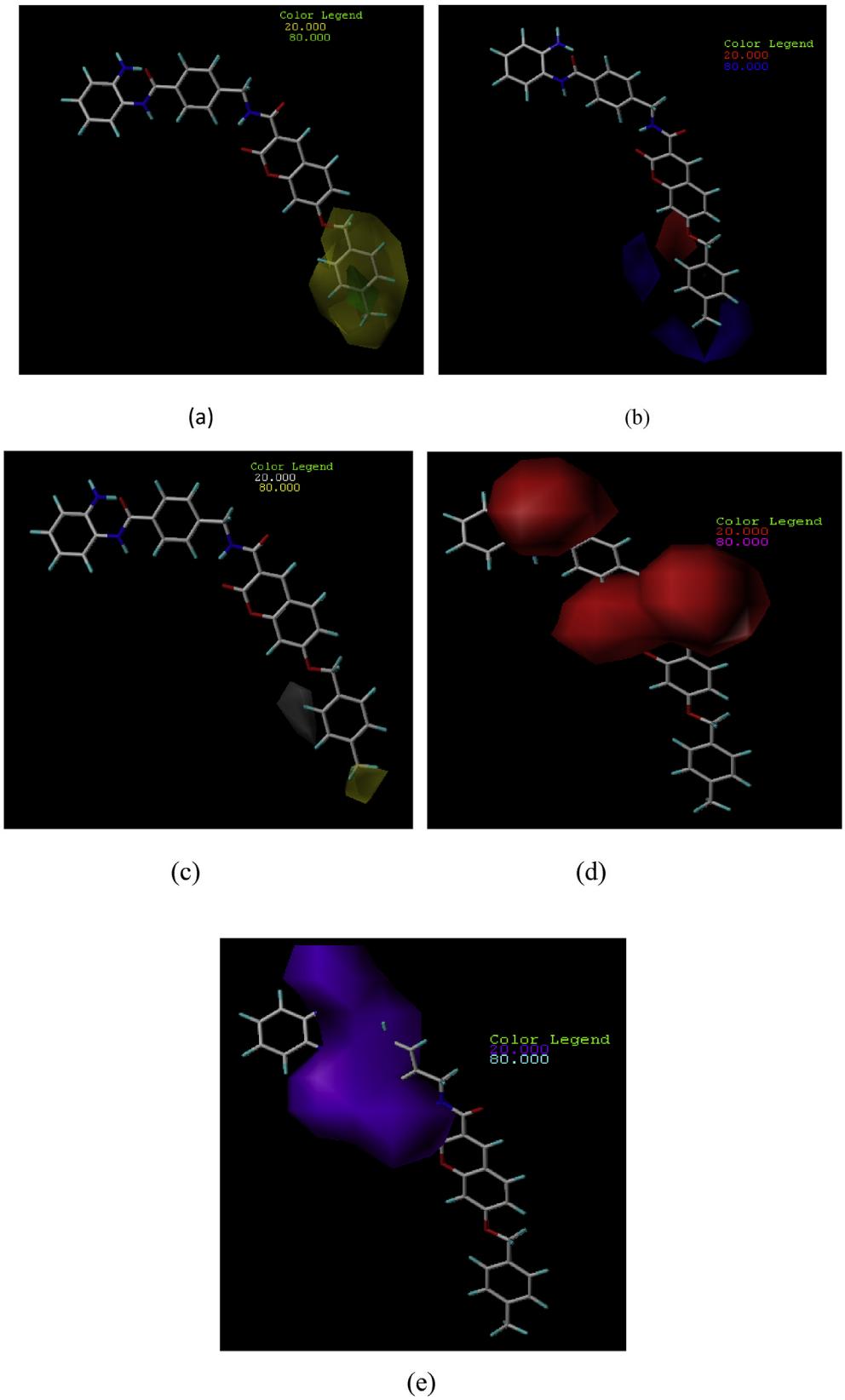


Fig. 13. CoMSIAStDev*Coeff contour plots with the combination of compound **11**. (a) *Steric contour maps*: Green contours indicate regions where bulky groups increase activity; yellow contours indicate regions where bulky groups decrease activity; (b) *Electrostatic contour maps*: Blue contours indicate regions where positive charges increase activity; red contours indicate regions where negative charges increase activity; (c) *Hydrophobic contour maps*: yellow contours indicate regions where hydrophobic substituents enhance activity; white contours indicate regions where hydrophobic groups decrease activity; (d) *Hydrogen bond donor contour maps*: Cyan contours indicate regions where H-bond donor groups increase activity and purple contours indicate the unfavorable regions for hydrogen bond donor substituents; (e) *H-bond acceptor contour maps*: Magenta contours indicate regions where an H-bond acceptor substituents increase activity; red contours indicate the disfavor regions for H-bond acceptor groups.

A white region near *ortho* and *meta* positions of benzyloxy ring showed that the introduction of hydrophilic groups into these positions might be beneficial for inhibitory activity (Fig. 13c). The yellow contour in the *para* position of benzyloxy ring indicated that hydrophobic groups such as methyl in this region could be increasing the activity of the compounds. These results confirm that the yellow contour of hydrophobic map was in agreement with green contour of steric map.

The CoMSIA H-bond donor and acceptor contour maps correlated with hydrogen bond interactions of the ligand with target. The cyan and purple contour maps of H-bond donor indicated favorable (80% contribution) and unfavorable (20% contribution) interactions and the magenta and red contour maps indicated favorable (80% contribution) and unfavorable (20% contribution) H-bond acceptor groups (Fig. 13d and e). However, no favorable cyan and magenta contours were observed for H-bond donor and acceptor contour maps. There were purple and red contours near to amide and amine groups of CU and ZBG regions of coumarin-based benzamides that changing the amide and amine groups and introducing other functional groups significantly decreased HDAC inhibition activity because these groups are pharmacophore of HDAC inhibitors [52].

3.7. Interpretation of HQSAR contribution map

HQSAR calculations are based on the contributions of molecular fragments to the biological activity for each molecule. The results of the HQSAR contribution maps can be graphically shown as a color-coded structure diagram which the color of each atom reflects its contribution to the molecule's overall activity. The red end of the spectrum (red, red orange and orange) reflects negative contribution to the activity, while the green end (yellow, blue, green-blue and green) represents positive effect and intermediate contributions are colored in white. The individual atomic contributions of the most active HDAC inhibitory analogues (compounds **11** and **1**) were displayed in Fig. 14.

The (4-(2-aminophenyl) benzamide) benzyl) carbamoyl scaffold as maximal common structural fragment represented by green and yellow color codes in the most active compound (**11**) because it was a common fragment to all molecules and contributed in the same way to all inhibitors. The coumarin portion was highlighted in green and yellow colors, indicating the importance of this fragment to enhance the biological activity. The C-2 position of benzyloxy ring was colored in yellow that may contribute moderately to the inhibitory activity. The atoms of methyl at position 4 of benzyloxy ring was colored in yellow and made a positive contribution to increase activity. Compound (**1**) (compound with the lowest activity) has green and yellow colors in the coumarin ring that indicating favorable contribution to the inhibitory activity. Moreover, phenyl ring and amine group of compound (**1**) were colored red orange, demonstrating their intermediate or negative contribution to the inhibitory activity. Finally, the structure-activity relationship and binding features obtained by present QSAR models and molecular docking analysis are summarized in Scheme 3.

3.8. Molecular docking studies

In order to gain functional and structural insight into the binding mode of all of the compounds and the HDAC1 enzyme (PDB ID: 4BKX) and also, to validate the results of QSAR contour maps, docking studies were carried out using MOE software (Fig. 15).

Analysis of docking results revealed high docking scores (−15.43-12.75 kcal/mol) for all of the compounds. All of the compounds were well stabilized in the active site of HDAC1 and had significant interactions with the key amino acid residues of the enzyme (Fig. 16).

HDAC1 is a zinc dependent enzyme and its active site consisted of a long, narrow tunnel leading to a cavity that contains the catalytic Zn ion. Studying interaction mode of all of the compounds with HDAC1 by LigX module of MOE software revealed that the zinc atom was *penta*-coordinated with Asp176, His178 and Asp264 as well as the carbonyl and NH_2 groups of benzamide region. The amidic NH, NH_2 and carbonyl groups of all of the compounds could also form hydrogen bonding with His140, His141, Gly149 and Tyr303 (Fig. 16.).

The aryl linker region of compounds was located in lipophilic tube including hydrophobic amino acids of Leu271, Phe150, Phe205 and Met30. This linker had hydrophobic and π-π staking interactions with benzene group of Phe150 which improved the binding affinity of ligands to HDAC1.

The surface recognition and hydrophobic amino acids of the lipophilic tube are shown in Figs. (15 and 16), that hydrophobic residues (e.g. Pro29, Leu271) in active site were significant and should be considered in developing potent HDAC inhibitors.

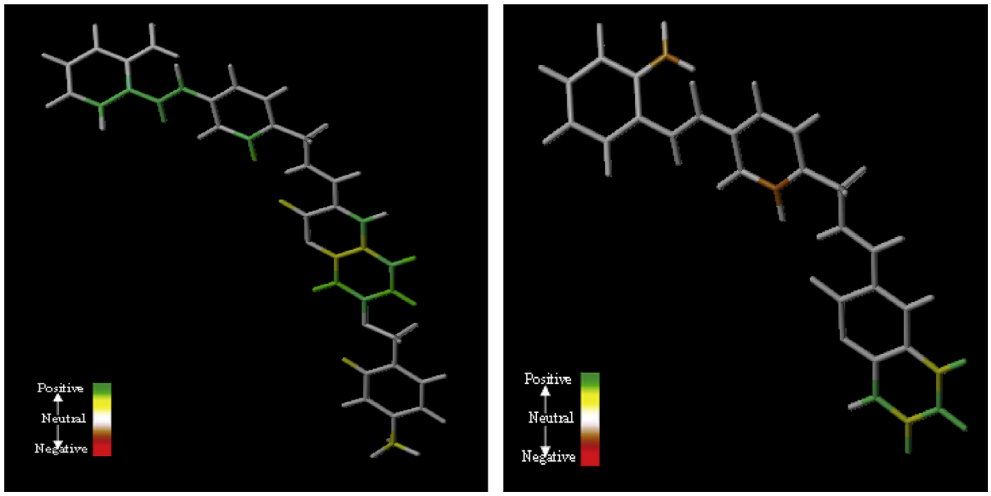


Fig. 14. The HQSAR contribution map of the most active compound (**11**) and the less active compound (**1**). The colors in yellow, blue, green-blue or green indicate positive contributions, while colors with red, red-orange or orange represent negative contributions and intermediate contributions are colored in white.

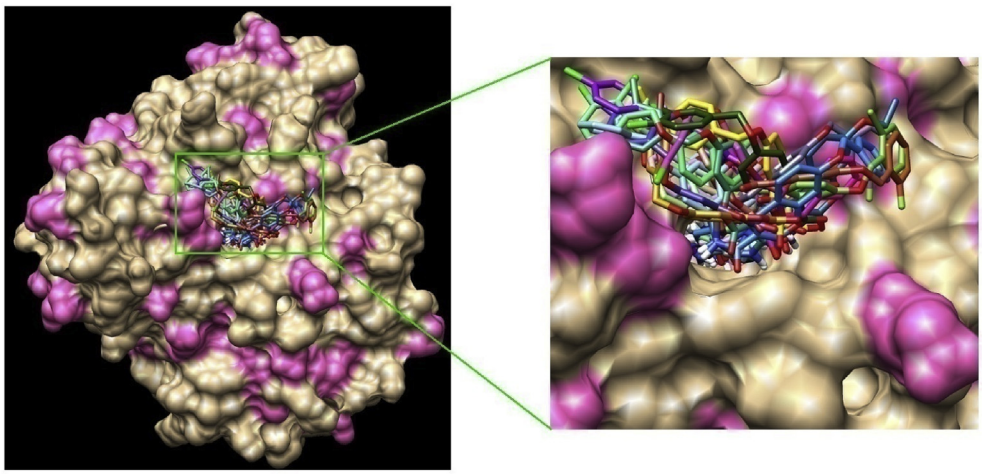
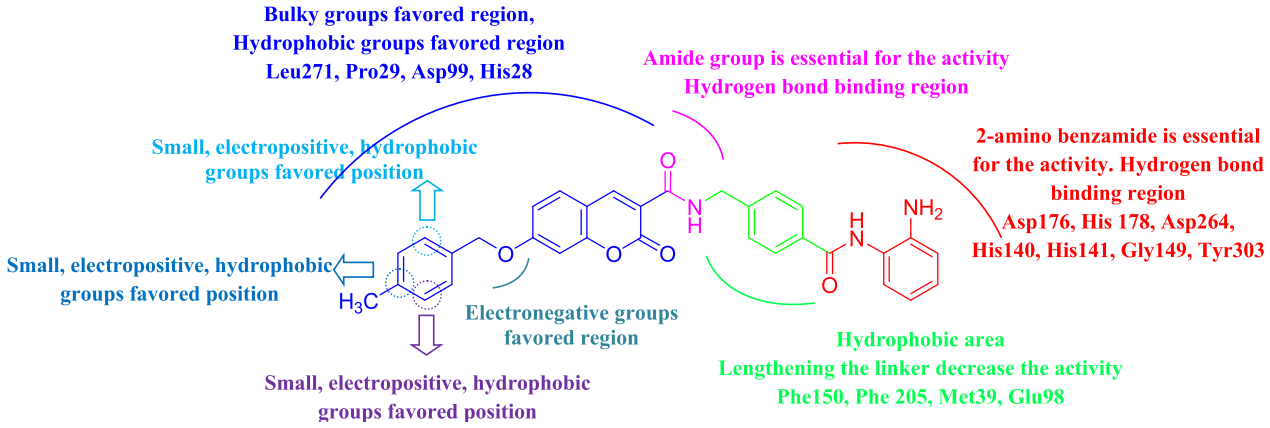


Fig. 15. 3D representation of docked ligands into binding site of HDAC1.



Scheme 3. Structure-activity relationship revealed by QSAR and molecular docking studies.

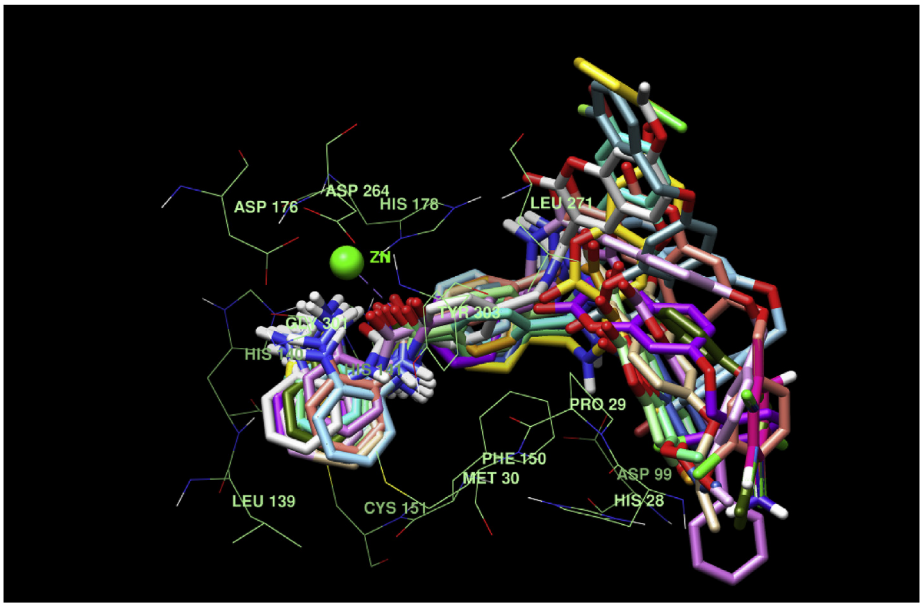


Fig. 16. The best pose of the compounds obtained from docking study in the active site of HDAC1.

4. Conclusion

The 2D- (HQSAR) and 3D-QSAR (CoMFA, CoMFA-RF and CoMSIA) methods were employed to study coumarin-based benzamides as histone deacetylase inhibitors. The CoMFA, CoMFA-RF, CoMSIA and HQSAR models for HDAC inhibitory activity of HCT116 cell line provided statistically significant results for internal and external validations including q^2 Values of 0.728, 0.764, 0.671, and 0.811, r^2_{ncv} values of 0.982, 0.960, 0.977, and 0.986, r^2_{pred} values of 0.685, 0.552, 0.721, and 0.613 and r^2_m values of 0.629, 0.540, 0.640, and 0.612, respectively. The CoMFA and CoMSIA contour maps and the HQSAR fragment contribution map were explained structure-activity relationship of this series of HDAC inhibitors. Also, molecular docking studies were carried out to confirm the rationality of the derived models. The amido and amine groups of benzamide part as scaffold and the bulky groups in the heterocyclic moiety as a hydrophobic part were key factors to improve inhibitory activity of HDAC. These results showed good predictive models for the rational design of novel HDAC inhibitors for the treatment of cancer disease.

Disclosure statement

No potential conflict of interest was reported by the authors.

Acknowledgements

We are grateful to Clinical Biochemistry Research Center, Basic Health Sciences Institute, Sharekord University of Medical Sciences for financial support of this research.

References

[1] K. Nepali, S. Sharma, M. Sharma, P. Bedi, K. Dhar, Rational approaches, design strategies, structure activity relationship and mechanistic insights for anticancer hybrids, Eur. J. Med. Chem. 77 (2014) 422–487.
[2] M. Ahmad, M.A. Aga, J.A. Bhat, B. Kumar, A. Rouf, N. Capalash, M.J. Mintoo, A. Kumar, P. Mahajan, D.M. Mondhe, Exploring derivatives of quinazoline alkaloid L-vasicine as cap groups in the design and biological mechanistic evaluation of novel antitumor histone deacetylase inhibitors, J. Med. Chem. 60 (2017) 3484–3497.
[3] L. Zhang, Y. Han, Q. Jiang, C. Wang, X. Chen, X. Li, F. Xu, Y. Jiang, Q. Wang, W. Xu, Trend of histone deacetylase inhibitors in cancer therapy: isoform selectivity or multitargeted strategy, Med. Res. Rev. 35 (2015) 63–84.
[4] A.M. Aboeldahab, E.A. Beshr, M.E. Shoman, S.M. Rabea, O.M. Aly, Spirohydantoins and 1, 2, 4-triazole-3-carboxamide derivatives as inhibitors of histone deacetylase: design, synthesis, and biological evaluation, Eur. J. Med. Chem. 146 (2018) 79–92.
[5] Y. Luan, J. Li, J.A. Bernatchez, R. Li, Kinase and histone deacetylase hybrid inhibitors for cancer therapy, J. Med. Chem. (2018), https://doi.org/10.1021/acs.jmedchem.8b00189.
[6] K. Stenzel, A. Hamacher, F.K. Hansen, C.G. Gertzen, J. Senger, V. Marquardt, L. Marek, M. Marek, C. Romier, M. Remke, Alkoxyurea-based histone deacetylase inhibitors increase cisplatin potency in chemoresistant cancer cell lines, J. Med. Chem. 60 (2017) 5334–5348.
[7] R. Sangwan, R. Rajan, P.K. Mandal, HDAC as onco target: reviewing the synthetic approaches with SAR study of their inhibitors, Eur. J. Med. Chem. 158 (2018) 620–706.
[8] T. Heimburg, A. Chakrabarti, J. Lancelot, M. Marek, J. Melesina, A.T. Hauser, T.B. Shaik, S. Duclaud, D. Robaa, F. Erdmann, M. Schmidt, C. Romier, R.J. Pierce, M. Jung, W. Sippl, Structure-based design and synthesis of novel inhibitors targeting HDAC8 from *Schistosoma mansoni* for the treatment of schistosomiasis, J. Med. Chem. 59 (2016) 2423–2435.
[9] H. Fu, L. Han, X. Hou, Y. Dun, L. Wang, X. Gong, H. Fang, Design, synthesis and biological evaluation of saccharin-based N-hydroxybenzamides as histone deacetylases (HDACs) inhibitors, Bioorg. Med. Chem. 23 (2015) 5774–5781.
[10] R. De Vreese, M. D'hooghe, Synthesis and applications of benzohydroxamic acid-based histone deacetylase inhibitors, Eur. J. Med. Chem. 135 (2017) 174–195.
[11] R. Xie, Y. Li, P. Tang, Q. Yuan, Design, synthesis and biological evaluation of novel 2-aminobenzamides containing dithiocarbamate moiety as histone deacetylase inhibitors and potent antitumor agents, Eur. J. Med. Chem. 143 (2018) 320–333.
[12] S.-W. Chao, L.-C. Chen, C.-C. Yu, C.-Y. Liu, T.E. Lin, J.-H. Guh, C.-Y. Wang, C.-Y. Chen, K.-C. Hsu, W.-J. Huang, Discovery of aliphatic-chain hydroxamates containing indole derivatives with potent class I histone deacetylase inhibitory activities, Eur. J. Med. Chem. 143 (2018) 792–805.
[13] V. Krieger, A. Hamacher, C.G. Gertzen, J. Senger, M.R. Zwinderman, M. Marek, C. Romier, F.J. Dekker, T. Kurz, M. Jung, Design, multicomponent synthesis, and anticancer activity of a focused histone deacetylase (HDAC) inhibitor library with peptoid-based cap groups, J. Med. Chem. 60 (2017) 5493–5506.
[14] S. Banerjee, N. Adhikari, S.A. Amin, T. Jha, Histone deacetylase 8 (HDAC8) and its inhibitors with selectivity to other isoforms: an overview, Eur. J. Med. Chem. 164 (2018) 214–240.
[15] M.K. Wambua, D.A. Nalawansha, A.T. Negmeldin, M.K.H. Pflum, Mutagenesis studies of the 14 Å internal cavity of histone deacetylase 1: insights toward the acetate-escape hypothesis and selective inhibitor design, J. Med. Chem. 57 (2014) 642–650.
[16] D.R. Walkinshaw, X.-J. Yang, Histone deacetylase inhibitors as novel anticancer therapeutics, Curr. Oncol. 15 (2008) 237.
[17] K. Keller, M. Jung, Histone deacetylase (HDAC) inhibitors in recent clinical trials for cancer therapy, in: Epigenetic Therapy of Cancer, Springer, 2014, pp. 227–255.
[18] M. Yoshida, Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A, Tanpakushitsu kakusan koso. Protein. nucleic acid. Enzyme. 52 (2007) 1788–1789.
[19] S. Grant, C. Easley, P. Kirkpatrick, Vorinostat, Nature Publishing Group, 2007.
[20] K.P. Garnock-Jones, Panobinostat: first global approval, Drugs 75 (2015) 695–704.
[21] H.-Z. Lee, V.E. Kwitkowski, P.L. Del Valle, M.S. Ricci, H. Saber, B.A. Habtemariam, J. Bullock, E. Bloomquist, Y.L. Shen, X.-H. Chen, FDA approval: belinostat for the treatment of patients with relapsed or refractory peripheral T-cell lymphoma, Clin. Cancer Res. 21 (2015) 2666–2670.
[22] T. Suzuki, T. Ando, K. Tsuchiya, N. Fukazawa, A. Saito, Y. Mariko, T. Yamashita, O. Nakanishi, Synthesis and histone deacetylase inhibitory activity of new benzamide derivatives, J. Med. Chem. 42 (1999) 3001–3003.
[23] C.L. Smith, Conference scene: 2nd cancer epigenetics conference, Epigenomics 5 (2013) 123–130.
[24] M. Fournel, C. Bonfils, Y. Hou, P.T. Yan, M.-C. Trachy-Bourget, A. Kalita, J. Liu, A.-H. Lu, N.Z. Zhou, M.-F. Robert, MGCD0103, a novel isotype-selective histone deacetylase inhibitor, has broad spectrum antitumor activity in vitro and in vivo, Mol. Cancer Ther. 7 (2008) 759–768.
[25] O. Moradei, A. Vaisburg, R.E. Martell, Histone deacetylase inhibitors in cancer therapy: new compounds and clinical update of benzamide-type agents, Curr. Top. Med. Chem. 8 (2008) 841–858.
[26] L.K. Gediya, A. Belosay, A. Khandelwal, P. Purushottamachar, V.C. Njar, Improved synthesis of histone deacetylase inhibitors (HDIs)(MS-275 and CI-994) and inhibitory effects of HDIs alone or in combination with RAMBAs or retinoids on growth of human LNCaP prostate cancer cells and tumor xenografts, Bioorg. Med. Chem. 16 (2008) 3352–3360.
[27] S. Chateauvieux, F. Morceau, M. Dicato, M. Diederich, Molecular and therapeutic potential and toxicity of valproic acid, BioMed Res. Int. 2010 (2010).
[28] R. Furumai, A. Matsuyama, N. Kobashi, K.-H. Lee, M. Nishiyama, H. Nakajima, A. Tanaka, Y. Komatsu, N. Nishino, M. Yoshida, FK228 (depsipeptide) as a natural prodrug that inhibits class I histone deacetylases, Cancer Res. 62 (2002) 4916–4921.
[29] R. Liu, J. Wang, W. Tang, H. Fang, Design and synthesis of a new generation of substituted purine hydroxamate analogs as histone deacetylase inhibitors, Bioorg. Med. Chem. 24 (2016) 1446–1454.
[30] J. Cai, H. Wei, K.H. Hong, X. Wu, M. Cao, X. Zong, L. Li, C. Sun, J. Chen, M. Ji, Discovery and preliminary evaluation of 2-aminobenzamide and hydroxamate derivatives containing 1, 2, 4-oxadiazole moiety as potent histone deacetylase inhibitors, Eur. J. Med. Chem. 96 (2015) 1–13.
[31] N. Steele, J. Plumb, L. Vidal, J. Tjørnelund, P. Knoblauch, P. Buhl-Jensen, R. Molife, R. Brown, J. De Bono, T. Evans, Pharmacokinetic and pharmacodynamic properties of an oral formulation of the histone deacetylase inhibitor Belinostat (PXD101), Cancer Chemother. Pharmacol. 67 (2011) 1273–1279.
[32] M.S. Finnin, J.R. Donigian, A. Cohen, V.M. Richon, R.A. Rifkind, P.A. Marks, R. Breslow, N.P. Pavletich, Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors, Nature 401 (1999) 188.
[33] P. Bertrand, Inside HDAC with HDAC inhibitors, Eur. J. Med. Chem. 45 (2010) 2095–2116.
[34] C. Seidel, M. Schnekenburger, C. Zwergel, F. Gaascht, A. Mai, M. Dicato, G. Kirsch, S. Valente, M. Diederich, Novel inhibitors of human histone deacetylases: design, synthesis and bioactivity of 3-alkenoylcoumarines, Bioorg. Med. Chem. Lett 24 (2014) 3797–3801.
[35] X. Zheng, M. He, X. Tan, J. Zheng, F. Wang, S. Liu, 3D-quantitative structure-activity relationship and docking studies of coumarin derivatives as tissue kallikrein 7 inhibitors, J. Pharm. Pharmacol. 69 (2017) 1136–1144.
[36] H. Kubinyi, QSAR and 3D QSAR in drug design Part 1: methodology, Drug Discov. Today 2 (1997) 457–467.
[37] H. Kubinyi, QSAR and 3D QSAR in drug design Part 2: applications and problems, Drug Discov. Today 2 (1997) 538–546.
[38] T. Abdizadeh, R. Ghodsi, F. Hadizadeh, 3D-QSAR (CoMFA, CoMSIA) and molecular docking studies on histone deacetylase 1 selective inhibitors, Recent Pat. Anticancer Drug Discov. 12 (2017) 365–383.
[39] M. Akamatsu, Current state and perspectives of 3D-QSAR, Curr. Top. Med. Chem. 2 (2002) 1381–1394.
[40] J. Verma, V.M. Khedkar, E.C. Coutinho, 3D-QSAR in drug design–a review, Curr.

Top. Med. Chem. 10 (2010) 95–115.
[41] G.E. Kellogg, S.F. Semus, D.J. Abraham, HINT: a new method of empirical hydrophobic field calculation for CoMFA, J. Comput. Aided Mol. Des. 5 (1991) 545–552.
[42] R.D. Cramer III, J.D. Bunce, D.E. Patterson, I.E. Frank, Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies, Quantit. Struct. Act. Relat. 7 (1988) 18–25.
[43] A. Borisa, H. Bhatt, 3D-QSAR (CoMFA, CoMFA-RG, CoMSIA) and molecular docking study of thienopyrimidine and thienopyridine derivatives to explore structural requirements for aurora-B kinase inhibition, Eur. J. Pharm. Sci. 79 (2015) 1–12.
[44] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem. 37 (1994) 4130–4146.
[45] T.L. Moda, C.A. Montanari, A.D. Andricopulo, Hologram QSAR model for the prediction of human oral bioavailability, Bioorg. Med. Chem. 15 (2007) 7738–7745.
[46] M.S. Castilho, M.P. Postigo, C.B.V. de Paula, C.A. Montanari, G. Oliva, A.D. Andricopulo, Two- and three-dimensional quantitative structure–activity relationships for a series of purine nucleoside phosphorylase inhibitors, Bioorg. Med. Chem. 14 (2006) 516–527.
[47] D.R. Lowis, HQSAR: a new, highly predictive QSAR technique, Tripos. Technical. Notes. 1 (1997) 17.
[48] B.L. Bush, R.B. Nachbar Jr., Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA, J. Comput. Aided Mol. Des. 7 (1993) 587–619.
[49] W. Dunn Iii, S. Wold, U. Edlund, S. Hellberg, J. Gasteiger, Multivariate structure-activity relationships between data from a battery of biological tests and an ensemble of structure descriptors: the PLS method, Quantit. Struct. Act. Relat. 3 (1984) 131–137.
[50] P. Geladi, Notes on the history and nature of partial least squares (PLS) modelling, J. Chemom. 2 (1988) 231–246.
[51] H. Kubinyi, Y.C. Martin, G. Folkers, 3D QSAR in Drug Design: Volume 1: Theory Methods and Applications, Springer Science & Business Media, 1993.
[52] T. Abdizadeh, M.R. Kalani, K. Abnous, Z. Tayarani-Najaran, B.Z. Khashyarmanesh, R. Abdizadeh, R. Ghodsi, F. Hadizadeh, Design, synthesis and biological evaluation of novel coumarin-based benzamides as potent histone deacetylase inhibitors and anticancer agents, Eur. J. Med. Chem. 132 (2017) 42–62.
[53] M. Clark, R.D. Cramer, N. Van Opdenbosch, Validation of the general purpose Tripos 5.2 force field, J. Comput. Chem. 10 (1989) 982–1012.
[54] A. Politi, S. Durdagi, P. Moutevelis-Minakakis, G. Kokotos, M.G. Papadopoulos, T. Mavromoustakos, Application of 3D QSAR CoMFA/CoMSIA and in silico docking studies on novel renin inhibitors against cardiovascular diseases, Eur. J. Med. Chem. 44 (2009) 3703–3711.
[55] J. Sainy, R. Sharma, QSAR analysis of thiolactone derivatives using HQSAR, CoMFA and CoMSIA. SAR. QSAR, Environ. Res. 26 (2015) 873–892.
[56] C.L. Waller, A comparative QSAR study using CoMFA, HQSAR, and FRED/SKEYS paradigms for estrogen receptor binding affinities of structurally diverse compounds, J. Chem. Inf. Comput. Sci. 44 (2004) 758–765.
[57] H. Zhang, H. Li, C. Liu, CoMFA, CoMSIA, and molecular hologram QSAR studies of novel neuronal nAChRs ligands-open ring analogues of 3-pyridyl ether, J. Chem. Inf. Model. 45 (2005) 440–448.
[58] L. Jiao, X. Zhang, Y. Qin, X. Wang, H. Li, Hologram QSAR study on the electrophoretic mobility of aromatic acids, Chemometr. Intell. Lab. Syst. 157 (2016) 202–207.
[59] J. Sun, H. Mei, QSAR and molecular mechanism analysis of N-substituted oseltamivir derivatives as potent avian influenza H5N1 neuraminidase inhibitors, Chemometr. Intell. Lab. Syst. 146 (2015) 485–493.
[60] F.J. Pérez-Areales, N. Betari, A. Viayna, C. Pont, A. Espargaró, M. Bartolini, A. De Simone, J.F. Rinaldi Alvarenga, B. Pérez, R. Sabate, Design, synthesis and multitarget biological profiling of second-generation anti-Alzheimer rhein–huprine hybrids, Future Med. Chem. 9 (2017) 965–981.
[61] L. Ståhle, S. Wold, Partial least squares analysis with cross-validation for the two-class problem: a Monte Carlo study, J. Chemom. 1 (1987) 185–196.
[62] S. Wold, Cross-validatory estimation of the number of components in factor and principal components models, Technometrics 20 (1978) 397–405.
[63] M. Kearns, D. Ron, Algorithmic stability and sanity-check bounds for leave-one-out cross-validation, Neural Comput. 11 (1999) 1427–1453.
[64] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.
[65] A. Rácz, D. Bajusz, K. Héberger, Consistency of QSAR models: correct split of training and test sets, ranking of models and performance parameters, SAR QSAR Environ. Res. 26 (2015) 683–700.
[66] S. Zhang, Z. Lin, Y. Pu, Y. Zhang, L. Zhang, Z. Zuo, Comparative QSAR studies using HQSAR, CoMFA, and CoMSIA methods on cyclic sulfone hydroxyethylamines as BACE1 inhibitors, Comput. Biol. Chem. 67 (2017) 38–47.
[67] M. Lorca, C. Morales-Verdejo, D. Vásquez-Velásquez, J. Andrades-Lagos, J. Campanini-Salinas, J. Soto-Delgado, G. Recabarren-Gajardo, J. Mella, Structure-activity relationships based on 3d-QSAR CoMFA/CoMSIA and design of aryloxypropanol-amine agonists with selectivity for the human β3-adrenergic receptor and anti-obesity and anti-diabetic profiles, Molecules 23 (2018) 1191.
[68] C. Rücker, G. Rücker, M. Meringer, y-Randomization and its variants in QSPR/QSAR, J. Chem. Inf. Model. 47 (2007) 2345–2357.
[69] I.M. Fawzy, K.M. Youssef, D.S. Lasheen, N.S. Ismail, K.A. Abouzid, Design, synthesis and 3D QSAR based pharmacophore study of novel imatinib analogs as antitumor-apoptotic agents, Future Med. Chem. 10 (2018) 1421–1433.
[70] S. Weaver, M.P. Gleeson, The importance of the domain of applicability in QSAR modeling, J. Mol. Graph. Model. 26 (2008) 1315–1326.
[71] H. Kaneko, K. Funatsu, Applicability domain based on ensemble learning in classification and regression analyses, J. Chem. Inf. Model. 54 (2014) 2469–2482.
[72] X. Yang, H. Liu, Q. Yang, J. Liu, J. Chen, L. Shi, Predicting anti-androgenic activity of bisphenols using molecular docking and quantitative structure-activity relationships, Chemosphere 163 (2016) 373–381.
[73] T. Lei, F. Chen, H. Liu, H. Sun, Y. Kang, D. Li, Y. Li, T. Hou, ADMET evaluation in drug discovery. Part 17: development of quantitative and qualitative prediction models for chemical-induced respiratory toxicity, Mol. Pharm. 14 (2017) 2407–2421.