ELSEVIER

Exploration of a binding mode of indole amide analogues as potent histone deacetylase inhibitors and 3D-QSAR analyses

Yanshen Guo, Jingfa Xiao, Zongru Guo,* Fengming Chu, Yonghao Cheng and Song Wu

Department of Medicinal Chemistry, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China

Abstract—Docking simulations and three-dimensional quantitative structure–activity relationship (3D-QSAR) analyses were conducted on a series of indole amide analogues as potent histone deacetylase inhibitors. The studies include comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Selected ligands were docked into the active site of human HDAC1. Based on the docking results, a novel binding mode of indole amide analogues in the human HDAC1 catalytic core is presented, and enzyme/inhibitor interactions are discussed. The indole amide group is located in the open pocket, and anchored to the protein through a pair of hydrogen bonds with Asp99 O-atom and amide NH group on ligand. Based on the binding mode, predictive 3D-QSAR models were established, which had conventional r^2 and cross-validated coefficient values (r_{cv}^2) up to 0.982 and 0.601 for CoMFA and 0.954 and 0.598 for CoMSIA, respectively. A comparison of the 3D-QSAR field contributions with the structural features of the binding site showed good correlation between the two analyses. The results of 3D-QSAR and docking studies validate each other and provided insight into the structural requirements for activity of this class of molecules as HDAC inhibitors. The CoMFA and CoMSIA PLS contour maps and MOLCAD-generated active site electrostatic, lipophilicity, and hydrogen-bonding potential surface maps, as well as the docking studies, provided good insights into inhibitor–HDAC interactions at the molecular level. Based on these results, novel molecules with improved activity can be designed.
© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Histone deacetylases (HDACs) mediate changes in nucleosome conformation and are important in the regulation of gene expression.[1] HDACs are involved in cell-cycle progression and differentiation, and their deregulation is associated with several cancers.[2] HDACs catalyze the removal of acetyl groups for the ε-amino groups of lysine residues clustered near the amino terminus of nucleosomal histones. Acetylation and deacetylation of the specific lysines within histones play a crucial role in the transcriptional process.[3] Two families of enzymes, acetylases and deacetylases, are involved in controlling the acetylation state of histones. Recent studies show that inhibition of histone deacetylases elicits anticancer effects in several tumor cells by inhibition of cell growth and induction of cell differentiation. The development of HDAC inhibitors as anticancer drugs has

been initiated, and compounds such as the hydroxamic acid Trichostatin A (TSA),[4] suberanilohydroxamic acid (SAHA),[5] the cyclic tetrapeptides apicidin[6] and trapoxin,[7] as well as synthetic inhibitors have been studied for this purpose in cancer cell lines. The inhibition of HDACs is a rapidly growing and very promising area for cancer chemotherapy.

Dai et al. designed and synthesized a series of hydroxamic acid-based HDAC inhibitors with an indole amide residue at the terminus. The introduction of substituents on the indole ring generated a series of potent inhibitors with significant antiproliferative activity. No previous effort has been carried out to seek new insight into the relationship between the structure information and the inhibitory potency of indole amide hydroxamic acid employing combined computational methods of molecular docking with 3D-QSAR approaches. In this paper, we studied the binding mode of indole amide hydroxamic acid derivatives[8] against HDAC using a molecular docking approach. Following the docking results, 3D-QSAR models were constructed by using approaches of comparative molecular field analysis (CoMFA) and comparative

molecular similarity analysis (CoMSIA).[9] The aim of this research includes two parts. One is to demonstrate the common binding mode of indole amide hydroxamic acid inhibitors with HDAC, and the other is to obtain QSAR models involving the main intermolecular interactions between inhibitors and HDAC, which can be used not only in rapidly and accurately predicting the activities of newly designed inhibitors, but also offer some beneficial clues in structural modification for designing new inhibitors with much higher inhibitory activities against HDAC.

2. Methods

2.1. Data sets

A series of hydroxamic acid-based HDAC inhibitors with an indole amide residue at the terminus were chosen in this study (Table 1). The HDAC inhibition activity of the compounds reported was assayed using a mixture of HDACs, predominantly HDAC1 and HDAC2. These inhibition activities were converted into the corresponding pIC_{50} ($-\log IC_{50}$) values, which were used as dependent variables in the CoMFA and CoMSIA analyses. The pIC_{50} values span a range of 3 log units. The total set of HDAC inhibitors (29 compounds) was divided into training (25 compounds) and test (4 compounds) sets, which were selected randomly and structurally diverse molecules possessing activities of wide range.

The sequence of HDAC1 is the same as that of HDAC2 at the active sites, and both the sequences are highly homological with HDAC8. So the homology structure of HDAC1 based on human HDAC8 in complex with TSA and SAHA (PDB entry 1T64, 1T67) is applied in this research. The binding mode of indole amide hydroxamic acid inhibitors with HDAC1 was studied using FlexX docking program. The molecular modeling software InsightII was used to build the structure of human HDAC1, Discover3 was used to minimize the residues of the binding site, SYBYL 6.9[10] was employed for CoMFA, CoMSIA, FlexX docking and visualization. All calculations were performed on SGI Fuel workstation with the IRIX 6.5 operating system.

2.2. Protonation state assignment

The crystal structure of HDAC8 in complex with TSA was recovered from Brookhaven Protein Database (PDB, http://www.rcsb.org/pdb). The ligand binding mode and the binding affinity can be strongly pH-dependent. The receptor structures was generated by using default values to determine the protonation states of the titratable groups. As such, amines were protonated, carboxylate groups were negatively charged, and hydroxyl groups are considered to be neutral. Imidazole rings were considered neutral, with the hydrogen on the δ-nitrogen, except when a hydrogen bond involving the N ε as a donor could be formed. Then, the hydrogen was placed on the ε-nitrogen. The homology structure of HDAC1 was put into PDB file and used for docking,

and hydrogen was added and minimized. The active site of HDAC was defined based on TSA, and all amino acid residues within a 6.5 Å radius to any of the inhibitor atoms were specified and a core sub-pocket was mentioned in order to obtain the appropriate bioactive conformation of the ligand within the active site. The default SYBYL FlexX parameters were used. The orientation of indole amide hydroxamic acid-based inhibitor in the active site of HDAC was used to superimpose the contours of CoMFA to find out the interaction with active site residue.

2.3. Molecular docking

The initial structure of small molecular inhibitors was constructed by SYBYL 6.9, and the geometries of these compounds were subsequently optimized using the TRIPOS force field, Gasteiger–Hückel charges, and Powell method; a nonbond cutoff of 8 Å[11] was adopted to consider the intramolecular interaction. For the purpose of investigating (tackling) the interacting mode of indole amide hydroxamic acid inhibitors with HDAC, the FlexX program[12] interfaced with SYBYL 6.9 was used to dock the compounds to the active site of HDAC. FlexX is a fast flexible automated docking program that considers ligand conformational flexibility by an incremental fragment placing technique.[13] Scoring calculations according to the FlexX and CSCORE modules were performed. The scoring methods available included empirical methods such as ChemScore.[14] FlexX score,[15] and G Score[16] and knowledge-based methods such as PMF score[17] and DrugScore.[18] In general, empirical and knowledge-based scoring functions do not require a protonation model.

2.4. Alignment

All the molecules in a database containing both the training and test sets were docked into the active site, and among the 100 possible docking structures generated by the FlexX docking program, the conformation that was indicated as the most tightly bound by FlexX scoring was overlapped with the cocrystallized inhibitor of HDAC8. Figure 1 shows the cocrystallized compound in magenta and ball-and-stick rendering, and the highest ranked conformations of the training and test sets are shown in black lines.

2.5. Calculations of the atomic charges

Five different kinds of partial atomic charges were considered: (1) Gasteiger–Marsili charges, (2) Gasteiger–Hückel charges, (3) MMFF94 charges, (4) the electrostatic potential fit charge at the AM1 level (ESP-AM1), and (5) the Mulliken population analysis atomic charges at the AM1 level (MPA-AM1). Both Gasteiger–Marsili and Gasteiger–Hückel methods calculate atomic charges based on the information of the atoms and the connectivity within the molecule. The MMFF94 atomic charges are simply calculated based on the bond increment parameters in MMFF94 force field. The calculation of Gasteiger–Marsili, Gasteiger–Hückel, and MMFF94 charges in SYBYL was

Table 1. Comparison of experimentally observed and CoMFA, CoMSIA predicted activities (pIC$_{50}$) for 29 indole amide hydroxamic acid-based HDAC inhibitors
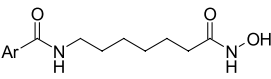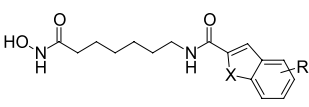
Compound		HDAC pIC$_{50}$	CoMFA predicted	CoMFA residue	CoMSIA predicted	CoMSIA residue
7		7.84	7.85	−0.01	8	−0.16
12[a]		7.43	7.53	−0.10	7.57	−0.14
13		7.1	7.83	−0.03	7.07	0.03
14		7.5	7.51	−0.01	7.63	−0.13
15		7.25	7.17	0.08	7.03	0.22
16[a]		7.4	7.46	−0.06	7.59	−0.19
29		7.33	7.6	−0.27	7.7	−0.37
30		7.25	7.19	0.06	7.08	0.17
31		7.92	7.91	0.01	7.97	−0.05
32[a]		8.46	8.18	0.28	7.74	0.72
33		7.93	7.77	0.16	7.7	0.23
34		7.72	7.77	−0.05	7.78	−0.06
5		5.77	5.65	0.12	5.59	0.18
6		6.28	6.43	−0.15	6.53	−0.25
9		6	6.04	−0.04	6.25	−0.25
10		7.9	8.01	−0.11	7.82	0.08

Table 1 (*continued*)

Compound	R	X	HDAC pIC$_{50}$	CoMFA predicted	CoMFA residue	CoMSIA predicted	CoMSIA residue
17	4-CH$_3$O	NH	8.51	8.45	0.06	8.34	0.17
18	5-CH$_3$O	NH	8.19	8.02	0.17	8.13	0.06
19	6-CH$_3$O	NH	8	8.04	−0.04	8.11	−0.11
20	5-Br	NH	8.46	8.29	0.17	8.21	0.25
21	5-CH$_3$	NH	8.2	8.14	0.06	8.27	−0.07
22	5-F	NH	8.27	8.33	−0.06	8.04	0.23
23	4-CH$_3$O, 6-CH$_3$O	NH	8.51	8.51	−0.0029	8.71	−0.2
24	5-BnO	NH	7.87	7.83	0.04	7.92	−0.05
25	5-CF$_3$O	NH	7.8	7.83	−0.03	7.72	0.08
8[a]	H	O	7.53	7.39	0.14	7.21	0.32
26	5-CH$_3$O	O	7.82	7.88	−0.06	7.87	−0.05
27	7-CH$_3$O	O	7.77	7.87	−0.1	7.79	−0.02
28	5-Cl	O	7.82	7.91	−0.09	7.76	0.06

[a] The compounds of test set.
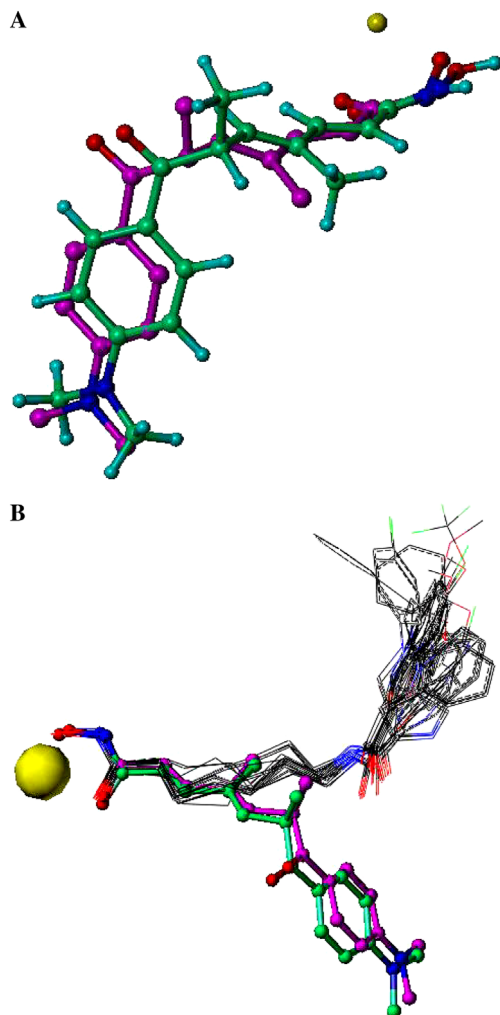
Figure 1. FlexX docked results. (A) Compounds in ball-and-stick rendering is TSA. The nitrogen atoms are colored blue, oxygen atoms are colored red, the carbon atoms are colored magenta or green. The magenta compound is the cocrystallized HDAC inhibitor TSA, and the green represents the FlexX docking result of TSA; (B) representative docked structures of the training and test sets (black lines).

automated with SPL scripts. The calculations of ESP-AM1 and MPA-AM1 charges were calculated using MOPAC module in SYBYL.

2.6. CoMFA and CoMSIA studies

The overlapped molecules were surrounded by a 3D grid of points in the three dimensions extending at least 4 Å beyond the union volume occupied by the superimposed molecules. The CoMFA descriptors, steric (Lennard–Jones 6–12 potential) and electrostatic (coulombic potential) field energies, were calculated using the SYBYL default parameters: 2 Å grid points spacing, an sp^3 carbon probe atom with +1|e| charge and a van der Waals radius of 1.52 Å, and energy cutoff of 30 kcal/mol except for those had been indicated.

The five similarity indices in CoMSIA, that is, steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor descriptors, were calculated using a C^{1+} probe atom with a radius of 1.0 Å placed at regular grid spacing of 2 Å. CoMSIA similarity indices $A_{F,K}^q(j)$ for a molecule j with atoms i at a grid point q are calculated by Eq. 1:

$$A_{F,K}^q(j) = \sum \omega_{\text{probe},k} \omega_{ik} e^{-\alpha r_{iq}^2}, \qquad (1)$$

in which k represents the following physicochemical properties: steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor. A Gaussian type distance dependence was used between the grid point q and each atom i of the molecule. The default value of 0.3 was used as the attenuation factor (α). Here, steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from atomic partial charges, hydrophobic fields are derived from atom-based parameters,[19] and H-bond donor and acceptor indices are obtained by a rule-based method based on experimental results.[20]

The CoMFA and CoMSIA descriptors derived above were used as explanatory variables, and pIC_{50} values were used as the target variable in PLS analyses to derive 3D-QSAR models using the implementation in the SYBYL package. The predictive value of the models was evaluated by leave-one-out (LOO) cross-validation. The cross-validated coefficient (r_{cv}^2) was calculated using Eq. 2:

$$r_{cv}^2 = 1 - \frac{\sum(Y_{predicted} - Y_{actual})^2}{\sum(Y_{actual} - Y_{mean})^2}, \qquad (2)$$

where $Y_{predicted}$, Y_{actual}, and Y_{mean} are predicted, actual, and mean values of the target property (pIC_{50}), respectively. $\sum(Y_{predicted} - Y_{actual})^2$ is the predictive sum of squares (PRESS). To maintain the optimum number of PLS components and minimize overdetermination, the number of components giving the lowest PRESS value was used for deriving the final PLS regression models. The conventional correlation coefficient, r^2, and its standard error, S, were also computed for the final PLS models. CoMFA and CoMSIA coefficient maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding CoMFA or CoMSIA descriptor values.

3. Results and discussion

3.1. Binding mode of indole amide hydroxamic acid-based inhibitors

Crystal structures of HDAC8 complex with TSA and SAHA revealed that hydroxamic acid-based inhibitors bind the deacetylase core by inserting their aliphatic chains into the HDAC pocket and by making multiple contacts to its tubelike hydrophobic portion. Particularly, their hydroxamic acid group reaches the polar bottom of the pocket, where it coordinates the zinc ion in a bidentate fashion (through CO and OH groups) and also contacts active site residues (forming two hydrogen bonds between its NH and OH groups and the two charge-relay systems His131/Asp166 and His132/Asp173). Moreover, the hydroxamic acid function replaces the zinc-bound water molecule of the active structure with its OH group.[21] The FlexX dock-predicted conformation of TSA with HDAC8 is shown in Figure 1 with the X-ray crystallographic-obtained conformation. The root mean square deviation (RMSD) between these two conformations is about 0.64 Å, indicating that the FlexX docking simulation is reasonable to reproduce the X-ray structure and could be extended to search the enzyme binding conformations for other inhibitors.

The human HDAC1 and HDAC2 is highly homologous with HDAC8, especially on the binding site of Zn^{2+}. So a reliable homology structure can be obtained based on the crystal structure of HDAC8. All 29 indole amide hydroxamic acid-based inhibitors were docked to the active site of HDAC1 based on the modeling structure, 100 docked structures of every compound were saved, the highest score of the docked results is in a similar orientation and the binding conformations of indole amide hydroxamic acids could be aligned quite well overall (Fig. 1). Two classes of orientation were found in the 100 docked results. In the first orientation, indole amide is near to Asp99, most docked results are in this orientation and have the highest rank in original FlexX score, only very few of docked results are in the second orientation, on which the indole amide located on the site of dimethyl amino benzyl of TSA in HDAC8.
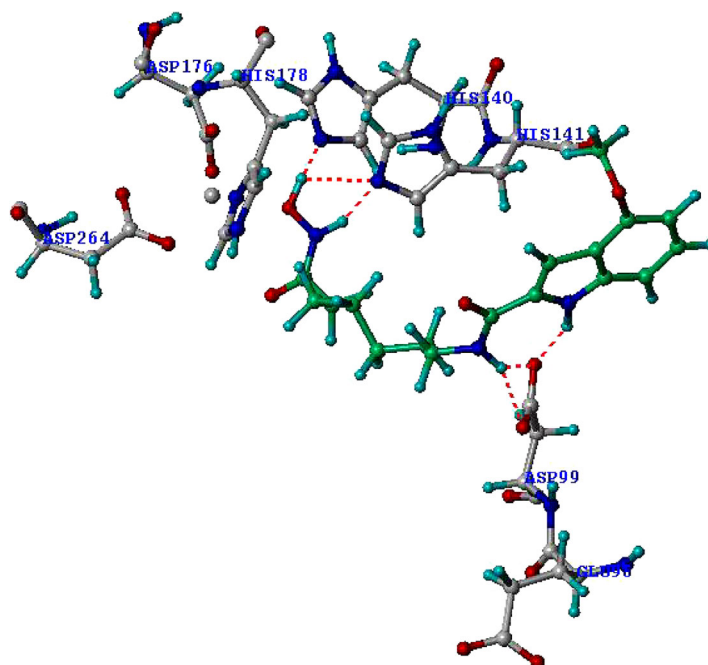
Figure 2. The highest score docking result in the FlexX docking of compound **17** docked into the HDAC1 catalytic core. Compound **17** is in green, and some residues are shown in the tube/ribbon graphic.

Figure 2 generally represents the interacting mode of compound **17** (the most potent inhibitor among the 29 indole amide hydroxamic acid-based inhibitors) with HDAC. An inspection of the HDAC1/17 complex shows that the hydroxamic acid moiety of the inhibitor is located in the active site and interacts with the catalytic zinc ion (CO–Zn, 2.18Å; OH–Zn, 2.39 Å) for producing efficient metal chelation, just as the related groups of SAHA and TSA were positioned at the optimal bond distance (CO_{SAHA}–Zn, 1.98 Å; OH_{SAHA}–Zn, 1.95 Å; CO_{TSA}–Zn, 2.22 Å; OH_{TSA}–Zn, 2.00 Å) for zinc ion complexing. Both NH on amide and the indole ring can potentially form hydrogen bonds with COO of Asp99, which perhaps affected the orientation of all the indole amide hydroxamic acid-based inhibitors.

Surface physicochemical property maps (Fig. 3), that is, electrostatic potential, hydrophobicity (lipophilicity) potential, and hydrogen bonding (donor/acceptor) potential maps of the active site were generated on the solvent accessible (Connolly) surface using the MOL-CAD program in SYBYL. Figure 3 shows that there are two grooves at the placket of the active site, and that one is near Asp99, which can be occupied by the moiety of indole amide, as both NH on the amide and NH on indole ring can potentially form hydrogen bonds with COO of Asp99. The other is bigger and more opening, located on the site of the dimethyl aniline moiety of TSA on HDAC8. The Tyr100 is mutated into Glu98 in HDAC1, which cannot form an effective π–π interaction between 4-hydroxy benzyl of Tyr100 and dimethyl amino benzyl of TSA.

The result of molecular modeling suggests that the NH on amide is important for an optimal orientation of the indole amide group in the enzyme. In Figure 1, those which are shown in black lines are the highest ranked conformations of the training and test sets, and illustrate the probable binding conformational alignment for the 29 indole amide hydroxamic acid-based HDAC inhibitors.

3.2. CoMFA Using the inhibition of HDAC

CoMFA and CoMSIA 3D-QSAR models were derived using a series of indole amide hydroxamic acids HDAC inhibitors based on the FlexX docking results. The chemical structures of the molecules and their actual pIC_{50} values are shown in Table 1. The predictive power of the 3D-QSAR model which was derived from the training set was assessed by predicting biological activities of the test set molecules. The CoMFA and CoMSIA 3D-QSAR methods are based on the assumption that the changes in binding affinities of ligands are related to changes in molecular properties represented by fields. The alignment rule and the bioactive conformation are crucial variables in any 3D-QSAR analysis, as both will affect outcome of statistical analysis. In this study the FlexX docked conformations are constrained and aligned by the active pocket of histone deacetylases.

3.2.1. Partial charges. Steric and electrostatic CoMFA fields were generated using standard procedures. The different sets of partial charges were used in building the CoMFA models, and all of them exhibit good statistical quality between the predicted and experimentally determined values of pIC_{50}. The statistical details are summarized in Table 2. The r_{cv}^2 for ESP-AM1, MMFF94, and Gasteiger–Hückel are similar (0.550, 0.520, and 0.505); it is lowest in the case of Gasteiger–Marsili (0.342) and highest in the case of MPA-AM1 (0.589) partial charges, respectively. The conventional r^2 for MPA-AM1 and ESP-AM1 are comparable (0.980 and 0.962) with four components, and that for MMFF94 and Gasteiger–Hückel are comparable (0.966 and 0.964) too, while it is lowest (0.948) in the case of Gasteiger–Marsili partial charge. The standard error of estimation value (SEE) for MPA-AM1 partial
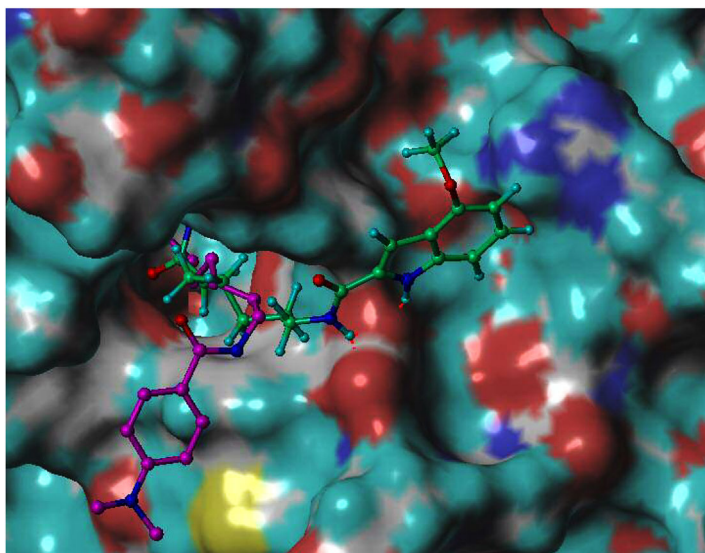

Figure 3. Surface representation of the HDAC-TSA and compound **17** interface in a different orientation. The protein surface is colored according to atom types. Magenta indicates TSA.

Table 2. Influence of different partial charges on the CoMFA models

	CoMFA1	CoMFA2	CoMFA3	CoMFA4	CoMFA5
Partial charge	MPA-AM1	ESP-AM1	MMFF94	Gasteiger–Marsili	Gasteiger–Hückel
r_{cv}^2 [a]	0.589	0.550	0.520	0.342	0.505
r_{conv}^2 [b]	0.980	0.962	0.966	0.958	0.964
SEE[c]	0.115	0.130	0.150	0.183	0.153
F-value[d]	229.6	176.2	151.2	90.48	129.2
N[e]	4	4	4	4	4
Contributions					
Steric	0.573	0.580	0.601	0.535	0.722
Electronic	0.427	0.420	0.399	0.465	0.278
Grid spacing: 2.0 Å					

[a] Cross-validated correlation coefficient.
[b] Non-cross-validated correlation coefficient.
[c] Standard error of estimate.
[d] F-test value.
[e] Optimum number of components obtained from cross-validated PLS analysis and same used final non-cross-validated analysis.

charge is better (0.115) than that of the other four partial charges. The steric and electrostatic field contributions of the three models for MPA-AM1, ESP-AM1, and Gasteiger–Marsili are almost similar (~50:50) indicating equal requirement of these two fields on ligand–receptor interactions. Based on the above observations, the best CoMFA model obtained by using MPA-AM1 partial charge was chosen for further analysis.

3.2.2. Grid spacing. Often for QSARs developed with CoMFA, a shift in the r_{cv}^2 values is observed as the grid spacing is altered.[22] To examine this possibility with these data, the different grid boxes with 1.0, 1.5, 2.0, 2.5, and 3.0 Å grid spacing, respectively, were used for the CoMFA calculations. The influence of the different grid spacing on the CoMFA model is obvious

(Table 3). Only from the r_{cv}^2 value after leave-one-out cross-validation, the model with the grid spacing of 2.0 Å was selected as the best model. A lower grid spacing (1.0 or 1.5 Å) may generate more noise in the PLS calculations and require a greater computational effort. When the grid spacing is defined as a larger value such as 2.5 or 3.0 Å, some important information about field properties in some regions may be lost. The standard error of 2.0 Å grid spacing is the smallest in the five results. The following discussion concerned with CoMFA will only refer to the model generated from 2.0 Å grid spacing.

3.2.3. Column filtering. To explore the effect of column filtering on the CoMFA model with 2.0 Å grid spacing and MPA-AM1 partial charge, different column filtering values were used.[23] The minimum-sigma (column

Table 3. Influence of different grid spacings on the CoMFA models

Model	CoMFA1	CoMFA2	CoMFA3	CoMFA4	CoMFA5
Grid spacing (Å)	1.0	1.5	2.0	2.5	3.0
r_{cv}^2	0.544	0.540	0.589	0.430	0.520
r^2	0.968	0.960	0.980	0.962	0.969
Standard error	0.139	0.153	0.115	0.153	0.161
F	156.3	128.6	229.6	128.5	117.7
N	4	4	4	3	4
Contributions					
Steric	0.516	0.531	0.573	0.575	0.487
Electronic	0.484	0.469	0.427	0.425	0.513

Table 4. Influence of column filtering value on the CoMFA models

Model	CoMFA1	CoMFA2	CoMFA3	CoMFA4	CoMFA5	CoMFA6
Column filtering (kcal/mol)	3.0	2.5	2.0	1.5	1.0	0
r_{cv}^2	0.568	0.576	0.589	0.591	0.592	0.594
r^2	0.961	0.976	0.980	0.979	0.978	0.980
Standard error	0.134	0.127	0.115	0.118	0.116	0.115
F	165.9	192.5	229.6	204.3	225.3	228.6
N	4	4	4	4	4	4
Contributions						
Steric	0.505	0.518	0.573	0.508	0.559	0.574
Electronic	0.495	0.482	0.427	0.492	0.441	0.426

filtering) was set to 0, 1.0, 1.5, 2.0, 2.5, and 3.0 kcal/mol (Table 4), respectively, to improve the signal-to-noise ratio by omitting those lattice points on which the energy variation was below this threshold. Any region point column having a variation less than this value is omitted for PLS. The results show that when the value of column filtering is bigger than 2.0 kcal/mol, the r_{cv}^2 decreases from 0.594 to 0.568 and the standard error is also bigger than that of 2.0 kcal/mol. The model of both 2.5 and 3.0 kcal/mol is not better than that of 2.0 kcal/mol, but they need less computational effort. When the column filtering is less than 2.0 kcal/mol, it needs more time to run. A typical value of 2.0 decreases the analysis time about an order of magnitude, with typically small effects on the r_{cv}^2 values obtained.

3.3. CoMSIA using the inhibition of HDAC

3.3.1. Different field combinations. The CoMSIA method defines explicit hydrophobic (H) and hydrogen bond donor (D) and acceptor (A) descriptors in addition to the steric (S) and electrostatic (E) fields used in CoMFA. PLS analyses of various CoMSIA models with different combinations of fields were obtained. Figure 4 shows the distributing of r_{cv}^2 that resulted from all of the 31 field combinations. Considering the steric, electronic, hydrophobic, H-bond donor and H-bond accepter fields alone, electronic and steric fields are more important, which can be seen by nothing that in using steric and electrostatic fields only the quality of the CoMSIA model ($r_{cv}^2 = 0.47$) is poorer than the CoMFA model ($r_{cv}^2 = 0.589$). The CoMSIA model that included S, E, and D fields performed better than the other field combinations. The cross-validated r^2 for these three higher models are 0.534, 0.517, and 0.512, respectively. Hence, the optimal field combinational CoMSIA model that included the steric (S) and electrostatic (E) and hydrogen bond donor (D) was chosen for further analysis.

3.3.2. Grid spacing. Using the steric (S) and electrostatic (E) and hydrogen bond donor (D) field combination, the sensitivity of the CoMSIA models to different grid spacing was also investigated. The difference of the r_{cv}^2 values for grid spacing from 1.0 to 2.5 is about 0.015 units (Table 5), which is obviously smaller than that of CoMFA models. The instabilities of CoMFA can be attributed to

Figure 4. Result of cross-validated r_{cv}^2 for different field combinations.

Table 5. Influence of different grids on the CoMSIA models

Model	CoMSIA1	CoMSIA2	CoMSIA3	CoMSIA4	CoMSIA5
Grid spacing (Å)	1.0	1.5	2.0	2.5	3.0
r_{cv}^2	0.560	0.565	0.551	0.555	0.366
r^2	0.951	0.966	0.966	0.961	0.827
Standard error	0.166	0.163	0.170	0.163	0.333
F	93.5	93.9	82.7	91.7	31.2
N	5	5	5	5	3
Contributions					
Steric	0.266	0.266	0.275	0.284	0.230
Electronic	0.531	0.537	0.522	0.499	0.498
H-bond donor	0.203	0.197	0.203	0.217	0.272

Figure 5. Comparison of different potential functions used in CoMFA and CoMSIA field.

the shape and steepness of the Lennard–Jones potential and in consequence to the required arbitrary fixation of cutoff values in CoMFA. CoMFA calculates steric field using a Lennard–Jones potential (Fig. 5). When some atoms of the aligned molecules are near the grid points, the Lennard–Jones potentials will increase to larger values, and generally the large potentials will be truncated to the cutoff value. The contributions of those grids using cutoff values will introduce certain unpredicted instability to the final models. In CoMSIA, the distance dependence between the probe atom and the molecular atoms a Gaussian function is used, which prevents the abrupt change of the potential on the grid points near the molecule surface. When the grid spacing is 3.0 Å, the columns for PLS analysis decrease to so few that a good model cannot be obtained, perhaps because some important information has been lost. From the results of r_{cv}^2 and the standard error the model of grid spacing 1.5 is optimum, and the following discussion concerned with CoMSIA will only refer to the model generated from 1.5 Å grid spacing.

3.3.3. Attenuation factor. In CoMSIA, a Gaussian-type distance dependence function is applied. In the preliminary parameter study, we calibrated the attenuation factor α to 0.3. To decide whether this is an appropriate value, α was varied in a parameter study within the range from 0.1 to 0.5 in steps of 0.1 (Table 6), and subsequently similarity indices and r_{cv}^2 values were computed each time. Reducing α to smaller values means that a probe placed at a particular lattice point detects molecular similarity in its neighborhood more globally. On the other hand, larger values of α imply a more localized

evaluation of similarity. Our systematic parameter study for α shows that $\alpha = 0.2$ is optimum for these data sets.[24]

3.4. CoMFA contour maps

The CoMFA steric and electrostatic fields from the final non-cross-validated analysis were plotted as three-dimensional colored contour maps in Figure 6. The field
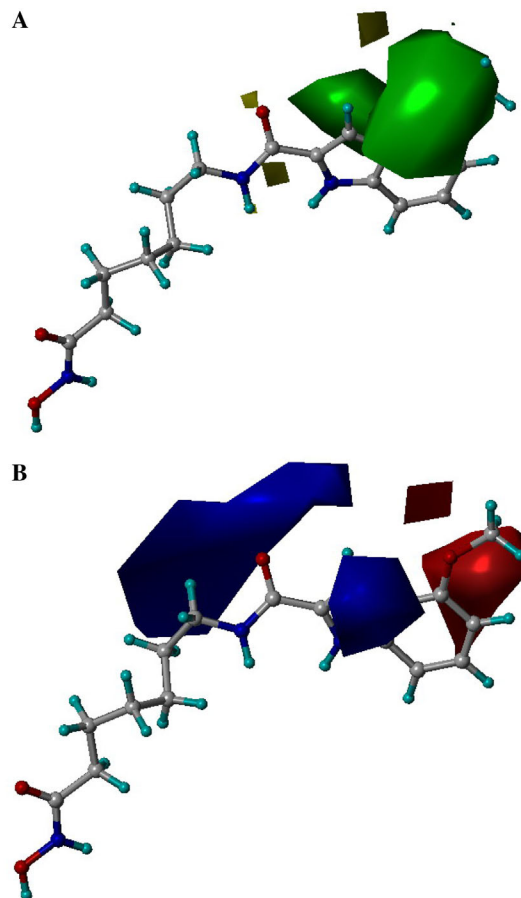
Figure 6. Stereoview of the contour plots (SD*coeff) of the CoMFA (A) steric fields; green contours indicate regions where bulky groups increase activity, whereas yellow contours indicate regions where bulky groups decrease activity and (B) electrostatic fields; blue contours indicate regions where electropositive groups increase activity, whereas red contours indicate regions where electronegative groups will increase activity. Potent HDAC inhibitor 17 is displayed in the background for reference.

Table 6. Influence of different attenuation factor on the CoMSIA models

Model	CoMSIA1	CoMSIA2	CoMSIA3	CoMSIA4	CoMSIA5
Attenuation factor	0.1	0.2	0.3	0.4	0.5
r_{cv}^2	0.581	0.591	0.565	0.529	0.497
r^2	0.962	0.964	0.966	0.963	0.958
Standard error	0.167	0.141	0.163	0.170	0.192
F	72.2	65.5	93.9	122.5	244.0
N	6	5	5	5	6
Contributions					
Steric	0.263	0.267	0.266	0.274	0.301
Electronic	0.585	0.548	0.537	0.519	0.479
H-bond donor	0.152	0.185	0.197	0.207	0.220

energies at each lattice point were calculated as the scalar results of the coefficient and the standard deviation associated with a particular column of the data table (SD*coeff), always plotted as the percentage of contribution to the CoMFA equation. These maps show regions where differences in molecular fields are associated with differences in biological activity. The maps do not show what is common to all molecules of a set, and hence one cannot expect to generate a complete image of the receptor. The steric interactions are represented by green- and yellow-colored contours (Fig. 6A), while electrostatic interactions are represented by red- and blue-colored contours. In the green regions of steric contour plot, bulky substituents enhance biological activity, while bulky substituents in the yellow regions are likely to decrease activity. The green steric contour near the 4-position of the indole ring indicates that any bulkier substituents are preferred at this position (Fig. 6A). Thus, compounds **17** and **23** with bulkier substituents such as OCH_3 at this position are more active than compound **7** which has H at the 4-position. The enhancement of the biological activities may be caused by the greater hydrophobic interactions induced by the substituted methoxyl group. Blue-colored contours represent regions where electropositive groups increase activity, whereas red-colored regions represent areas where electronegative groups enhance activity (Fig. 6B). The electrostatic contour plot on the set of 29 compounds shows that there is a red-colored region situated close to the 4- and 5-positions of the indole ring, that is to say, the negative charges in these two regions are very important to ligand binding, and a charge withdrawing group linked to this position will enhance the biological activity. For example, the substitution of the hydrogen atom at the 5-position in compound **7** with Br and F gives compounds **20** and **22** increased inhibitory potency from 14.6 to 3.5 and 5.4 nM.

3.5. CoMSIA contour maps

The CoMSIA steric and electrostatic field contour plots employing S, E and D fields are shown in Figure 7. These plots (Figs. 7A and B) are more or less similar to the corresponding CoMFA plots (Figs. 6A and B), except that there is one red contour near the 4-position on the indole ring (Fig. 6B). Interestingly, the blue CoMSIA electrostatic contour is less than that of CoMFA near the amide group, which is difficult to interpret directly in CoMFA contour. The hydrogen bond donor contours are shown in Figure 7c. In principle, it should highlight the areas near which H-bonding donor on the ligand can form H-bonds with the receptor to influence binding affinity. The presence of the cyan-colored contour near the NH of amide indicates that hydrogen bond donor substituted at this position enhances the HDAC inhibitory activity (Fig. 7C). The atoms near this position may act as H-bonding donor and produce H-bonding interactions with H-bonding acceptor in protein. That is to say, the NH on amide may form stable H-bonds with some residues in the protein, which is consistent with the residue of ASP101 located on this region.
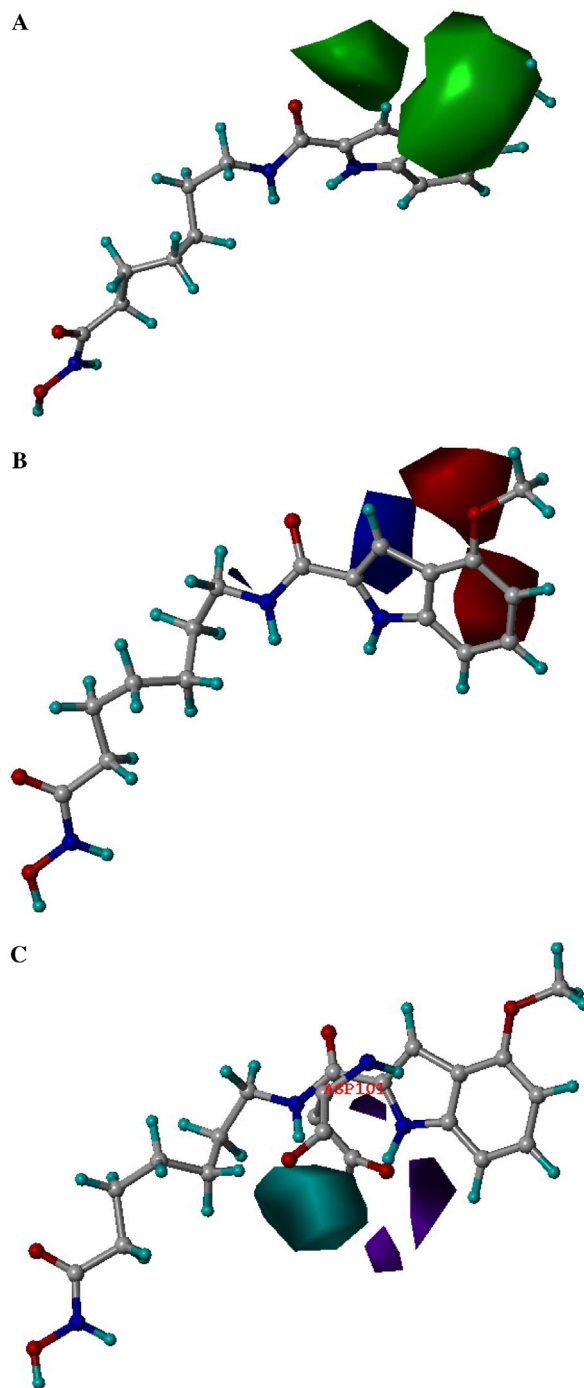
Figure 7. Stereoview of the contour plots of the CoMSIA (A) steric fields (green, bulky substitution favored; yellow, bulky substitution disfavored), (B) electrostatic fields (blue, electropositive group favored; red, electronegative group favored), and (C) hydrogen bond donor fields (cyan, favored; purple, disfavored). Potent HDAC inhibitor 17 is displayed in the background for reference.

3.6. CoMFA versus CoMSIA

The predictive power of CoMFA and CoMSIA 3D-QSAR models was evaluated by using the same test set of four molecules. The PLS statistics of both CoMFA and CoMSIA 3D-QSAR models indicate that CoMFA is somewhat better than CoMSIA. In both models, the predictive values fall close to the actual pIC_{50} values,
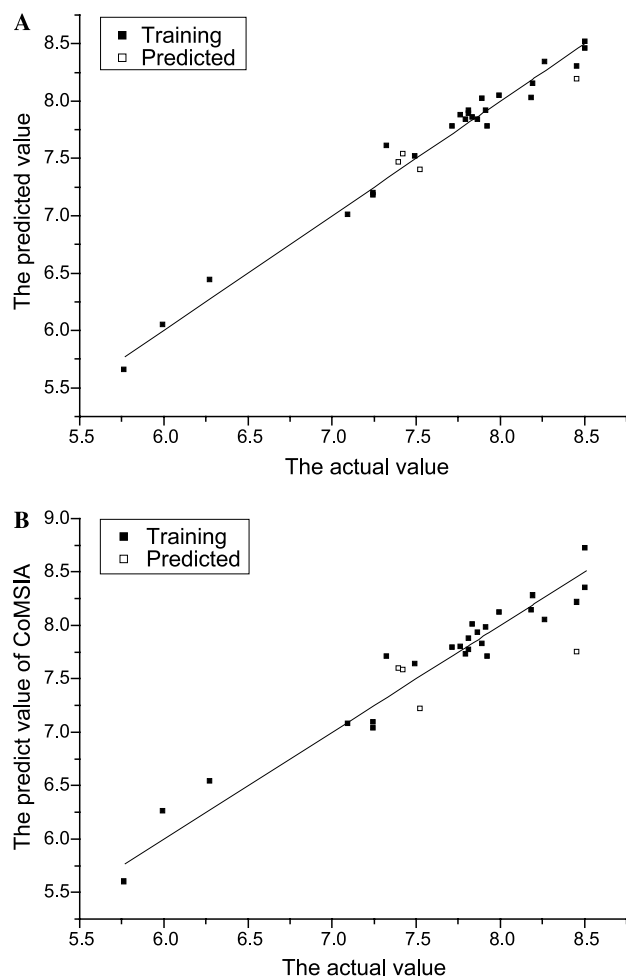
Figure 8. CoMFA (A) and CoMSIA (B) predicted versus experimental pIC$_{50}$ values of training and test sets, filled squares represent predictions for the training set, while open squares represent predictions for the test set.

not deviating by more than 1 logarithmic unit (Fig. 8, Table 1) in the case of CoMSIA. Both models predicted lower activity for the test set of four molecules. In summary, the differences between CoMFA and CoMSIA are not striking and both models demonstrated good predictive ability.

4. Conclusions

A series of hydroxamic acid-based HDAC inhibitors with an indole amide residue at the terminus were obtained from the literature. The docking studies revealed that the orientation and hydrogen bonding interactions of indole amide hydroxamic acid-based HDAC inhibitors inside the active site of HDAC are similar to those in the crystal structure of TSA–HDAC complex. CoMFA and CoMSIA 3D-QSAR models were developed for 29 substituted hydroxamic acid-based HDAC inhibitors with an indole amide residue at the terminus. Both models showed similar predictive capabilities. To our knowledge, this is the first study aimed at deriving predictive 3D-QSAR models for indole amide hydroxamic acid-based HDAC inhibitors. In addition, the dock-

ing studies provided good insights into inhibitor–HDAC interactions at the molecular level. This information will be useful in the design of novel broad-spectrum anticancer drug candidates. A comparison of the 3D-QSAR field contributions with the structural features of the binding site showed good correlation between the two analyses.

References and notes

1. Davie, J. R.; Chadee, D. N. *J. Cell Biochem. (Suppl.)* **1998**, *30–31*, 203.
2. Kouzarides, T. *Curr. Opin. Genet. Dev.* **1999**, *9*, 40.
3. Pazin, M. J.; Kadonaga, J. T. *Cell* **1997**, *89*, 325.
4. Yoshida, M.; Kijima, M.; Akita, M.; Beppu, T. *J. Biol. Chem.* **1990**, *265*, 17174.
5. Richon, V. M.; Emiliani, S.; Verdin, E.; Webb, Y.; Breslow, R.; Rifkind, R. A.; Marks, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3003.
6. Han, J. W.; Ahn, S. H.; Park, S. H.; Wang, S. Y.; Bae, G. U.; Seo, D. W.; Known, H. K.; Hong, S.; Lee, Y. W.; Lee, H. W. *Cancer Res.* **2000**, *60*, 6068.
7. Kijima, M.; Yoshida, M.; Suguta, K.; Horinouchi, S.; Beppu, T. *J. Biol. Chem.* **1993**, *268*, 22429.
8. Dai, Y.; Guo, Y.; Guo, J.; Pease, L. J.; Li, J.; Marcotte, P. A.; Glaser, K. B.; Tapang, P.; Albert, D. H.; Richardson, P. L.; Davidsen, S. K.; Michaelides, M. R. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1897.
9. Cramer, M.; Cramer, R. D., III; Jones, D. M. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
10. Sybyl, Version 6.9; Tripos Associates: St. Louis, MO, 2000.
11. Huang, X.; Liu, T.; Gu, J.; Luo, X.; Ji, R.; Cao, Y.; Xue, H.; Wong, J. T.; Wong, B. L.; Pei, G.; Jiang, H.; Chen, K. *J. Med. Chem.* **2001**, *44*, 1883.
12. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
13. Kramer, B.; Rarey, M.; Lengauer, T. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 228.
14. Eldridge, M. D.; Murray, C. W.; Auton, R. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425.
15. Stahl, M. *Perspect. Drug Discovery Des.* **2000**, *20*, 83.
16. Jones, G.; Willett, P.; Glen, R.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
17. Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791.
18. Gohlke, H.; Hendlich, M.; Klebe, G. *Perspect. Drug Discov. Des.* **2000**, *20*, 115.
19. Viswanadhan, V. N.; Ghose, A. K.; Revenkar, G. R.; Robins, R. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.
20. Klebe, G. *J. Mol. Biol.* **1994**, *237*, 212.
21. Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. *Nature* **1999**, *401*, 188.
22. Hou, T. J.; Li, Y. Y.; Xu, X. J. *J. Mol. Model.* **2000**, *6*, 438.
23. Singh, J.; van Vlijmen, H.; Lee, W. C.; Liao, Y.; Lin, K. C.; Ateeq, H.; Cuervo, J.; Zimmerman, C.; Hammond, C.; Karpusas, M.; Palmer, R.; Chattopadhyay, T.; Adams, S. P. *J. Comput. Aided Mol. Des.* **2002**, *16*, 201.
24. Markus, B.; Jörg, S.; Gerhard, K. *J. Med. Chem.* **1999**, *42*, 458.