Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 935–946
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Pharmacophore modeling and virtual screening studies to design some potential histone deacetylase inhibitors as new leads

S. Vadivelan [a,b,*], B.N. Sinha [b], G. Rambabu [a], Kiran Boppana [a], Sarma A.R.P. Jagarlapudi [a]

[a] GVK Biosciences Pvt. Ltd., S-1, Phase-1, T.I.E. Balanagar, Hyderabad 500037, India
[b] Department of Pharmaceutical Sciences, Birla Institute of Technology, Mesra 835215, Ranchi, India

Received 31 January 2007; received in revised form 7 July 2007; accepted 8 July 2007
Available online 12 July 2007

Abstract

Histone deacetylase is one of the important targets in the treatment of solid tumors and hematological cancers. A total of 20 well-defined inhibitors were used to generate Pharmacophore models using and HypoGen module of Catalyst. These 20 molecules broadly represent 3 different chemotypes. The best HypoGen model consists of four-pharmacophore features—one hydrogen bond acceptor, one hydrophobic aliphatic and two ring aromatic centers. This model was validated against 378 known HDAC inhibitors with a correlation of 0.897 as well as enrichment factor of 2.68 against a maximum value of 3. This model was further used to retrieve molecules from NCI database with 238,819 molecules. A total of 4638 molecules from a pool of 238,819 molecules were identified as hits while 297 molecules were indicated as highly active. Also, a Similarity analysis has been carried out for set of 4638 hits with respect to most active molecule of each chemotypes which validated not only the Virtual Screening potential of the model but also identified the possible new Chemotypes. This type of Similarity analysis would prove to be efficient not only for lead generation but also for lead optimization.

Keywords: Histone deacetylase; Pharmacophore; HypoGen model; Virtual screening; Similarity analysis

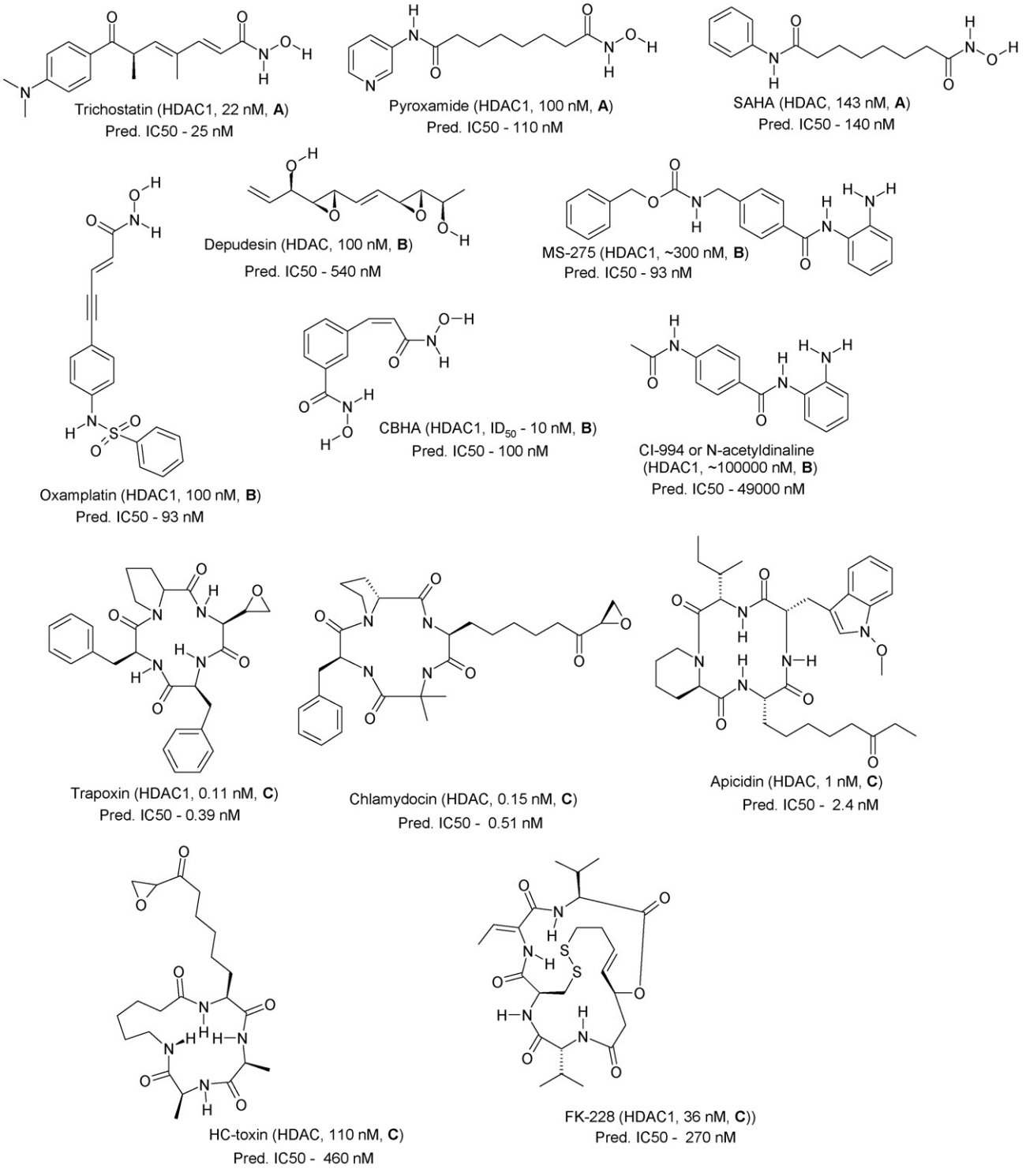
1. Introduction

Histone can be in one of the two antagonist forms, acetylated or deacetylated, equilibrium regulated by corresponding enzymes, histone acetylases and histone deacetylases [1]. Histone deacetylases (HDACs) represent a family of enzymes that compete with histone acetyltransferases (HATs) to modulate chromatin structure and transcriptional activity via change in acetylation status of nucleosomal histones [2]. Histone deacetylase enzymes remove acetyl groups from epsilon-*N*-acetyl-lysine histone tails and restore a positive charge to the Lysine residues. Such chemical modifications result in a tightening of nucleosome structure and gene silencing [3]. HDAC have been recognized as attractive therapeutic targets for anticancer, antifungal, antiviral and anti-inflammatory treatment [4].

Comprehensive reviews on histone deacetylase as a new target for cancer chemotherapy and several HDAC inhibitors of various structural families have been now advanced to Phase I and II of clinical trials (Scheme 1) [5,6]. Mammalian HDACs can be divided into three distinct classes. Class I HDACs corresponds to the analogues of Rpd3 yeast protein, isoforms HDACs 1, 2, 3 and 8 share a homology in their catalytic sites and exclusively expressed in nucleus [1,7]. Class II deacetylases includes the isoforms HDACs 4, 5, 6, 7, 9 and 10 which is shuttled between the cytoplasm and the nucleus and correspond to Hda1 yeast protein. HDACs 4, 5, 7 and 9 share homology in two regions, the C-terminal catalytic domain and a N-terminal regulatory domain. The third class of deacetylases is the conserved Sir2 family of proteins which are dependent on NAD+ for activity whereas Class I and II HDACs operate by zinc-dependent mechanisms [1,7]. In spite of the low sequence or structure similarity among different HDAC class proteins [8], selectivity in HDAC activity has been seldom observed and many diverse HDAC inhibitors have been included in the Pharmacophore model and virtual screening studies.

* Corresponding author at: GVK Biosciences Pvt. Ltd., S-1, Phase-1, T.I.E. Balanagar, Hyderabad 500037, India. Tel.: +91 40 65180632; fax: +91 40 23721010.
E-mail address: vadivelan@gvkbio.com (S. Vadivelan).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2007.07.002



Scheme 1. Clinical HDAC inhibitors from different scaffolds.

Crystal structure of HDAC8 (1T64) is available in literature [6], however, it does not share any good structural or sequence similarity with other HDAC isoforms despite being the Class I [5,6]. As the active residues are highly conserved in Class I HDAC proteins, molecular dynamics and free energy studies have been well carried out on a few selective HDAC inhibitors earlier [2,3,8–10].

HDAC inhibitors (Scheme 1) can be broadly characterized by a common pharmacophore that includes key elements of inhibitor–enzyme interactions [11]. The aim of this study was to construct a Pharmacophore model based on common chemical features of molecule with inhibitory activity towards histone deacetylase using HipHop and HypoGen modules implemented in the Catalyst software package [12], a leading

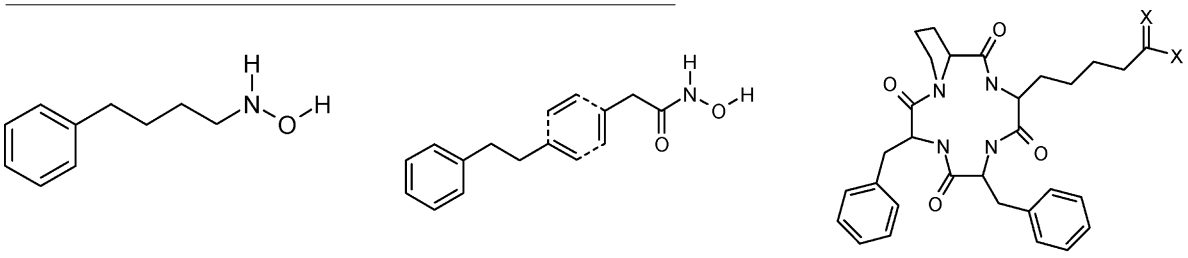
software product for the automated generation of Pharmacophore models and 3D database searching, as a large number of successful application in Medicinal Chemistry already demonstrated [13–20]. This Pharmacophore model have been used for virtual screening to identify more hits/leads from virtual library of molecules with high varied chemical nature.

2. Materials and methods

2.1. Database mining and training set

Medichem database [21] produced 638 HDAC inhibitors from 33 references while 398 molecules have IC_{50} values for human HDACs. While selectivity for different isoforms is not available for over 50% of these molecules, HDAC1 values are available for majority of the cases where such selectivity information is available and has been used. The most critical aspect of pharmacophore hypothesis generation is the selection of the training set. There should not be any redundancy in information content in terms of both structural features and activity ranges. The 398 molecules are arranged in the decreasing order activities and were clustered into 80 bins. Based on atom–atom pair distance descriptors, similarity between most active molecules to all other molecules in 80 bins has been performed based on ISIS 960 keys and Tanimoto analysis. The most diverse molecules from 20 bins have been selected as the training set and are given in Fig. 1. In general, HDAC inhibitors have been analyzed to possess (a) a functional group serving to chelate the metal ion in the active site (anchor), (b) a hydrophobic cap which interacts with amino acid residues at the entrance of the *N*-acetyl lysine binding channel and (c) a linker which is often a 5–6 hydrocarbon chain, that connects both hydrophobic cap and anchor [4,6]. For all the 398 molecules used for model development, the IC_{50} values are either in nM or μM. Assay conditions were also taken into consideration. The molecules with Ki, ED_{50} and EC_{50} values were ignored for modeling.

HDAC activities for the training set of 20 molecules [11,22–28] covers 5 orders of magnitude ($0.02\ nM \leq IC_{50} \leq 1500\ nM$). The molecules in the training set can be broadly classified into three slightly different scaffolds with 14, 3, and 3 molecules respectively in scaffolds **A**, **B**, and **C**. Molecules with scaffold **A** contain one or two aromatic rings connected to a very flexible long alkyl chain. The other end of the alkyl chain contains some hydrogen bond acceptor groups. Molecules with scaffold **B**, also contains both aromatic rings and hydrogen bond acceptors, however, the long alkyl chain is replaced with cyclic aliphatic, alkenyl or aromatic rings, which are somewhat less flexible. On the other hand, in the scaffold **C**, the aromatic rings are replaced with a substituted macrocycle. The flexibility associated with the macrocycle may be significant enough to preclude a large number of conformational changes. Incidentally scaffolds A, B and C represent three class of compounds, namely, hydroxamic acids/short chain fatty acids, sulfa/benzamides and cyclic tetrapeptide or epoxides, respectively. These set of class of compounds represent major HDAC inhibitors that have been studied so far.



A **B** **C**

This training set is used for HypoGen to generate Pharmacophore models while some of the molecules are used for the HipHop Pharmacophore models [12]. The mol files of all molecules from the database were exported and minimized using modified CHARMM force field in Catalyst package installed on a SGI Octane 275 MHz MIPS R12000 processor, and conformational analysis was carried out with Confirm module. Poling algorithm of Confirm module reduces considerably the probability of reappearance of nearly similar conformers by usage of penalty function. This feature could be very useful for the conformational search in the case of cyclic ring systems with scaffolds **B** and **C**. For each molecule, a maximum of 250 conformers that lie within 10 kcal/mol from the observed global minimum was considered for the model generation. This method ensures an exhaustive conformational mapping even for most complex molecules. Default values are used for all other parameters in the conformational analysis.

2.2. HipHop

HipHop hypotheses are produced by comparing a set of conformational models and a number of 3D configurations of chemical features shared among the training set molecules [12,29]. This results in a qualitative model wherein important chemical features can easily be identified. It is very important to identify such chemical features before one proceeds to the quantitative model generation such that results can be easily understood in terms of chemically meaningful patterns. However to confirm these features, 10 common feature hypotheses were generated using 3 most active molecules

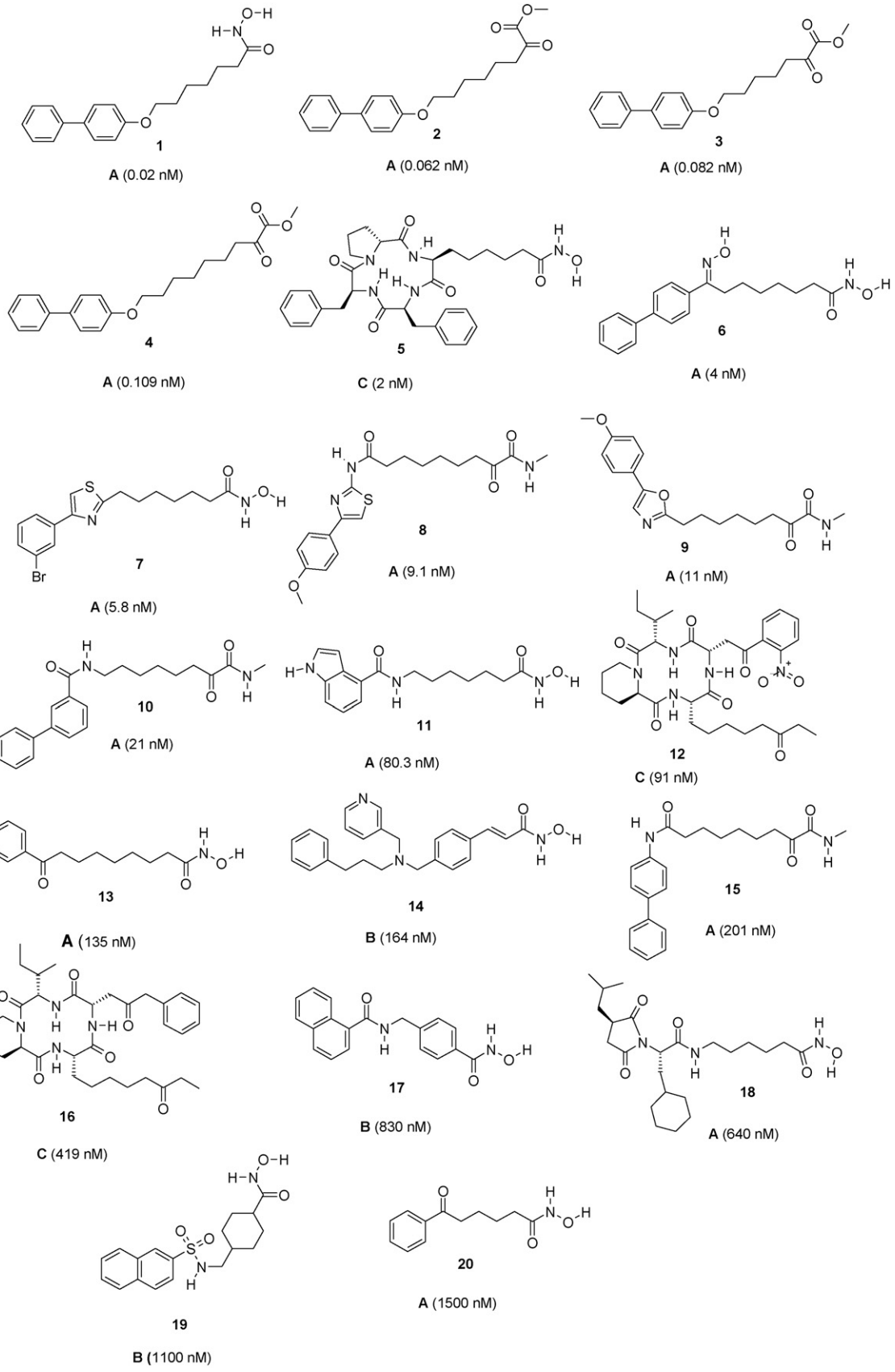


Fig. 1. Structures of 20 training set molecules with its experimental IC_{50} values.

Table 1
Ten Pharmacophore models generated by the HypoGen algorithm

Hypo no.	Total cost	Cost difference[a]	Error cost	R.M.S. deviation	Training set (*r*)	Features[b]
1	86.05	59.04	70.27	0.547	0.976	AHRR
2	89.94	55.15	74.18	0.831	0.964	AHRR
3	90.21	54.88	72.75	0.740	0.973	AHRR
4	90.87	54.22	75.67	0.916	0.956	AHRR
5	91.11	53.98	75.80	0.923	0.955	AHRR
6	91.16	53.93	75.49	0.906	0.957	AHRR
7	91.84	53.75	76.63	0.967	0.950	AHRR
8	91.96	53.13	75.58	0.911	0.957	AHRR
9	92.10	52.99	76.38	0.954	0.952	AHRR
10	92.42	52.67	76.28	0.949	0.933	AHRR

[a] (Null cost − total cost), null cost = 145.09, fixed cost = 82.32, for the Hypo-1 weight = 1.824, configuration = 13.931, all cost units are in bits.
[b] A, hydrogen bond acceptor; H, hydrophobic aliphatic; R, ring aromatic.

from each of the three scaffolds. In all these hypotheses hydrogen bond acceptors, hydrophobic aliphatic and ring aromatic groups are common features. However these models cannot be directly used for quantitative estimation of activities for the molecules screened from other databases or sets.

2.3. HypoGen

The training set of 20 molecules defined earlier has been used in HypoGen to generate Pharmacophore models, which could be used for quantitative estimation of activities while screening large virtual compound libraries [12]. While generating HypoGen model, a minimum of 0 to a maximum of 5 features involving hydrogen bond acceptor, hydrophobic aliphatic and ring aromatic systems have been specified based on the HDAC inhibitory activities, the molecules in the training set are broadly classified into three categories namely highly active (<20 nM), moderately active (20–200 nM) and low active (>200 nM). The quality of HypoGen models are best described in terms of fixed cost, null cost and total cost and these terms are well defined by Debnath [30]. The cost for each hypothesis is the summation of the three cost components (error (*E*), weight (*W*), and configuration (*C*)) multiplied by a coefficient (default coefficient is 1.0 for each). The fixed cost represents the simplest model that fits the data perfectly. The null cost represents the cost of a hypothesis with no features that estimates every activity to be the average activity [12]. In simple terms there should be a large difference between fixed cost and null cost with a value of 40–60 bits for the unit of cost, which would imply a 75–90% probability for correlating the experimental and predicted activity data. The total cost of any hypothesis should be close to the fixed cost for a good model. The best model Hypo-1 has been given in the Fig. 2 while the parameters that describe the Hypo-1 are given in Table 1.

2.4. Model validation and NCI 3D-database screening

The best pharmacophore hypothesis was used initially to screen 378 HDAC inhibitors from the Medichem database. All queries were performed using the Best Flexible search databases/Spreadsheet method. Enrichment factors and goodness of hit has been calculated in order to validate the model [31]. Subsequently the same model has also been used to retrieve potent molecules from NCI databases of 238,819 molecules that resulted in 4638 hits.

2.5. Similarity index

Molecular similarity index has been performed for 4638 molecules that were identified as hits from the pharmacophore based virtual screening of NCI database with most active molecules, **1**, **14** and **5** from chemotypes **A**, **B** and **C**, respectively. ISIS 960 keys were used to calculate the Tanimoto similarity coefficients [32]. The predicted activities are correlated with compounds similar to each scaffold.

3. Results and discussion

3.1. HipHop and HypoGen models

The best HipHop Pharmacophore model indicated the importance of hydrogen bond acceptors and hydrophobic

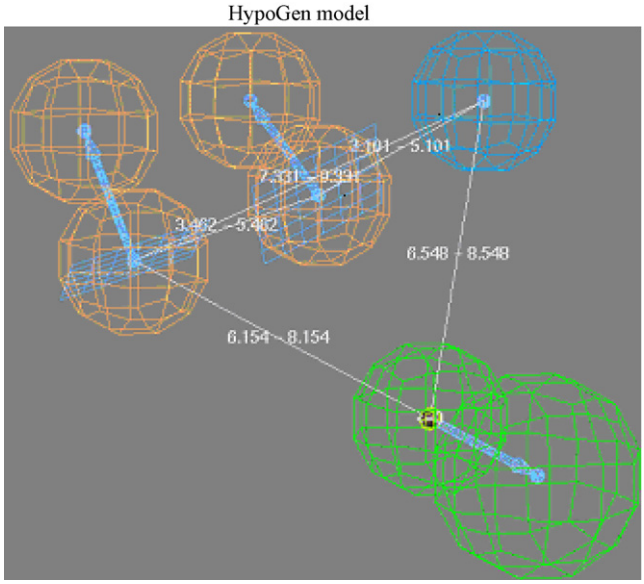


Fig. 2. Histone deacetylase inhibitor Pharmacophore model with its distance constraints. Features are color coded with green: one hydrogen bond acceptor, light blue: one hydrophobic aliphatic, orange: two aromatic rings.

aliphatic and ring aromatic groups, which were further, confirmed in the HypoGen generated models. In the HypoGen studies, a couple of sets of 10 hypotheses were generated using the most diverse 20 molecules in the training set molecules. The best 10 hypotheses consists again hydrogen bond acceptor, one hydrophobic aliphatic and two ring aromatic features and their relative distances, angles and other geometric parameters. This study also resulted in very good cost values, correlation (r) and root mean square deviations (R.M.S.D.) which are given in Table 1. The best hypothesis, Hypo-1, is characterized by the highest cost difference (59 bits), lowest RMSD value error (0.547) and with correlation 0.976. The fixed cost, pharmacophore (total) cost and null cost are 82, 86 and 145 bits, respectively. It is evident that as error, weight and configuration component are very low and not deterministic to the model; the total pharmacophore cost is also low and close to the fixed cost. Also, as total cost is less than the null cost, this model accounts for all the pharmacophore features and has a good predictability power.

Ragno et al. [33] shows the hydrogen bond interactions between the ZBG with zinc ion or active site residues and also hydrophobic interactions contribute positively to the activity. The hydrophobic aliphatic feature shows the necessity of a linker leading to enhanced metal chelation and higher anti-HDAC activity. One of the most active compounds has two hydrophobic groups in the cap portion that could contemporaneously interact with both the side chains of Phe337 and Tyr90 which matches with two ring aromatic pharmacophore features. Hence the pharmacophore features—one hydrogen bond acceptor, one hydrophobic aliphatic and two ring aromatic centers correlates well with the previously developed 3D QSAR model and proved to be highly predictive model.

Fig. 2 shows the Hypo-1 pharmacophore features with their geometric parameters between. Fig. 3a and b represents the Hypo-1 aligned with the most active and inactive molecules (**1** and **20**; IC_{50} values are 0.02 and 1500 nM, respectively) and shows a nice fit with all the features of Hypo-1. In this case of **1**, the hydrogen bond acceptor seems to be mapped to the hydroxyl amine group and the hydrophobic aliphatic fit well with the aliphatic chain while the biphenyl group matches with the two ring aromatic ring features. On the other hand, for molecule **20**, the two features hydrophobic aliphatic group and one of the ring aromatic groups could not fit well (Fig. 3b). Based on the Hypo-1 model, Table 2 shows experimental and predicted IC_{50} values for 20 training set molecules, along with other details such as, error values and fitness scores.

In the training set, all nine highly active molecules are correctly predicted as highly active, in the five medium active molecules, four were predicted well while one molecule was indicated as low active and all six low active molecules were also predicted as low active. From Table 2, it is noted that molecules are not only picked right as high, medium and low active but also the difference between the experimental and predicted activities are also minimal. Thus the error values are very less, and the fit value gives a good measure of how well the defined features in Hypo-1 fit well with the pharmacophore of each molecule.

3.2. Model validation and database screening

3.2.1. Validation and mining using known HDAC inhibitors

The validity of any Pharmacophore model needs to be ascertained by screening some known inhibitors (test set) that were retrieved from the Medichem databases [21] in order to check how many active molecules are picked in the screening process, how their predicted activities are correlated with the experimental activities and the efficiency in reducing the false positives or negatives. Hypo-1 was used to screen 378 HDAC of known high, medium and low active inhibitors of the test set. Database mining was performed in Catalyst software using the BEST flexible searching technique. A number of parameters such as hit list (H_t), number of active percent of yields ($\%Y$), percent ratio of actives in the hit list ($\%A$), enrichment factor of 2.68 (E), false negatives, false positives and goodness of hit score of 0.65 (GH) are calculated (Table 3) while carrying out the Pharmacophore model and virtual screening of test set

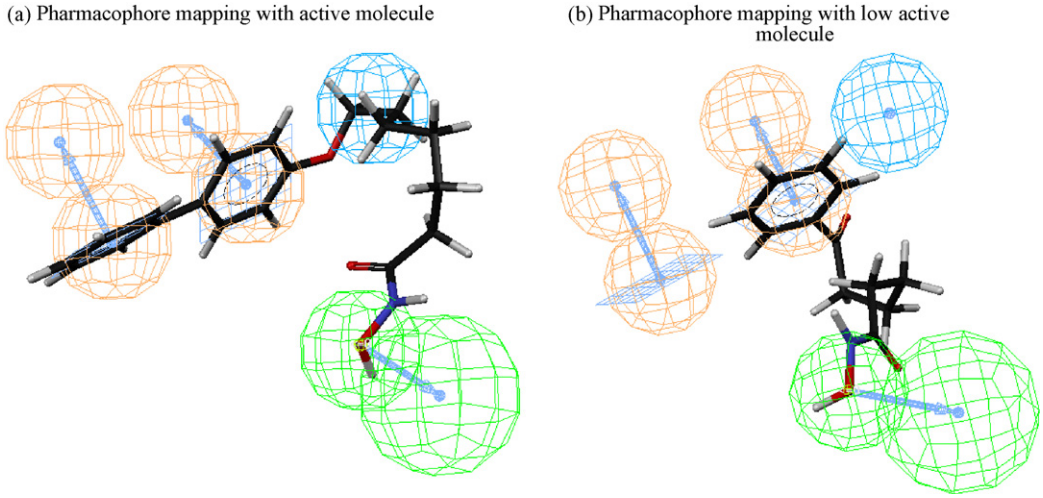


Fig. 3. Hypo-1 mapped to the most active molecule **1** (IC_{50} = 0.02 nM, a) and also mapped to low active molecule **20** (IC_{50} = 1500 nM, b) in the training set. Pharmacophore features are color coded with green: hydrogen bond acceptor, light blue: hydrophobic aliphatic, orange: aromatic rings.

Table 2
Experimental IC_{50} and predicted IC_{50} data of 20 training set molecules

Molecule	Scaffold	Exp. IC_{50} (nM)	Predicted IC_{50} (nM)	Error[a]	Fit value[b]	Experimental scale[c]	Predicted scale[c]
1	**A**	0.02	0.02	+1	9.09	+++	+++
2	**A**	0.062	0.15	+2.4	8.21	+++	+++
3	**A**	0.082	0.17	+2.1	8.14	+++	+++
4	**A**	0.11	0.077	−1.4	8.50	+++	+++
5	**C**	2	5.5	+2.7	6.65	+++	+++
6	**A**	4	3.6	−1.1	6.83	+++	+++
7	**A**	5.8	3	−2	6.91	+++	+++
8	**A**	9.1	6.2	−1.5	6.59	+++	+++
9	**A**	11	4.6	−2.4	6.72	+++	+++
10	**A**	21	26	+1.2	5.96	++	++
11	**A**	80.3	86	+1.1	5.45	++	++
12	**C**	91	83	−1.1	5.46	++	++
13	**A**	135	160	+1.2	5.17	++	++
14	**B**	164	700	+3.1	4.68	++	+
15	**A**	201	250	+1.3	4.98	+	+
16	**C**	419	840	+2	4.72	+	+
17	**B**	640	480	−1.4	4.73	+	+
18	**A**	830	1600	+1.9	4.39	+	+
19	**B**	1100	270	+4	4.75	+	+
20	**A**	1500	400	+3.7	4.62	+	+

[a] '+' indicates that the predicted IC_{50} is higher than the experimental IC_{50}; '−' indicates that the predicted IC_{50} is lower than the experimental IC_{50}; a value of 1 indicates that the predicted IC_{50} is equal to the experimental IC_{50}.
[b] Fit value [12] indicates how well the features in the pharmacophore overlap the chemical features in the molecule. fit = weight × [max(0, 1 − SSE)] where SSE = $(D/T)^2$, D = displacement of the feature from the center of the location constraint and T = the radius of the location constraint sphere for the feature (tolerance).
[c] Activity scale: IC_{50} <20 nM = +++ (highly active); IC_{50} 20–200 nM = ++ (moderately active); IC_{50} >200 nM = + (low active).

molecules. While the false positives and negatives, 36 and 12, respectively, are minimal, enrichment factor of 2.68 against a maximum value of 3.0 is a very good indication on the highly efficiency of the screening. In 100 molecules predicted to be active, 88 molecules were correctly picked thus missing only 12 false negatives with 36 false positives overall. GH score assessment of hit lists should be used to optimize the working Pharmacophore model as databases with known biological activities, note that the technique can also be used to focus a list of active molecules as a post-HTS processing, or to prioritize a virtual library as a pre-HTS screening.

In case of score Hypothesis against 378 molecules, from 109 high active molecules (<20 nM), 95 were predicted correctly while the remaining 14 molecules were predicted as moderately active. Similarly, from 108 moderately active molecules (20–200 nM), 88 were predicted correctly, 7 were predicted as highly active and the remaining 13 molecules were predicted as low active. From 161 low active molecules (>200 nM), 143 were predicted correctly, but 18 molecules were predicted as moderately active. The prediction of several clinical HDAC inhibitors had been done using the validated model and the predictions were pretty good for most of the molecules except a few like FK-228 and HC-toxin. The reason may be conformational flexibility of cyclic peptides and absence of aromatic features.

Three molecules representing each of **A**, **B** and **C** types and three activity ranges are given in Fig. 4. A fairly good agreement between the predicted and experimental activities is observed not only these nine molecules (Fig. 4). In fact, such correlation is very good for all the molecules in the training (blue) as well as test set molecules as represented by the 'r' values in Fig. 5. In the test set, correlation (r) for scaffolds **A**, **B** and **C** are 0.894 (black), 0.922 (red) and 0.877 (green), respectively, which indicates that the prediction may be slightly better in the case of **B** molecules in relation to **A** and **C** owing to some false positives and negatives respectively in these two cases.

Table 4 gives the molecule and scaffold-wise distribution from the hits, active hits, false positives and negatives. From Fig. 5 as well as Table 4, it is observed that scaffold A has more false positives and a few negatives. Scaffold **A** has a highly flexible long alkyl chain of 'n' atoms with an aryl and hydrogen bond acceptor groups connected on either end. In many cases, some simple groups are present in the 2, $n-2$ and/or $n-1$ positions of the alkyls chain however without affecting the conformational flexibility. The steric and other interaction effects that could result from these additional

Table 3
Statistical parameters from screening test set molecules

S. no.	Parameter	Value
1	Total molecules in database (D)	378
2	Total number of actives in database (A)	100
3	Total hits (H_t)	124
4	Active hits (H_a)	88
5	% Yield of actives $[(H_a/H_t) \times 100]$	70.97
6	% Ratio of actives $[(H_a/A) \times 100]$	88.0
7	Enrichment factor (E) $[(H_a \times D)/(H_t \times A)]$	2.68
8	False negatives $[A - H_a]$	12
9	False positives $[H_t - H_a]$	36
10	Goodness of hit score[a]	0.65

[a] $[(H_a/4H_tA)(3A + H_t) \times (1 - ((H_t - H_a)/(D - A))]$; GH score of 0.6–0.7 indicates a very good model.

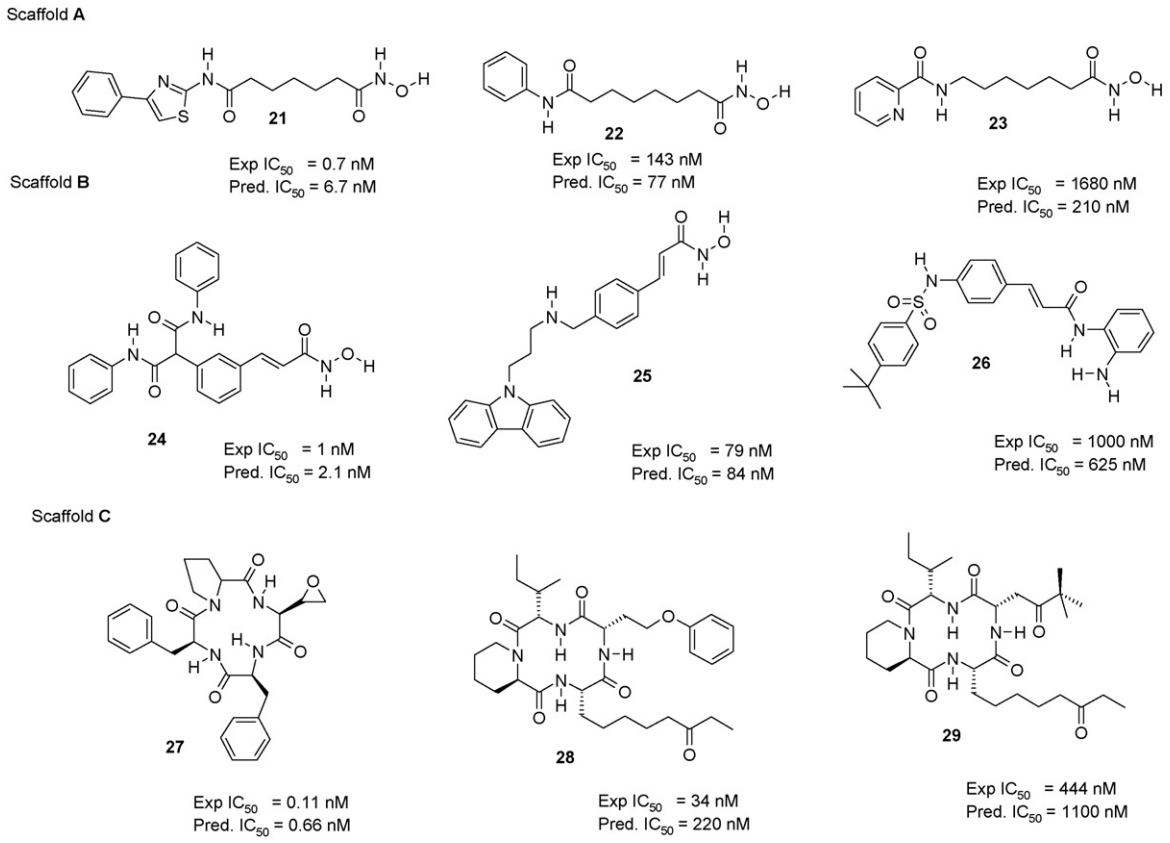


Fig. 4. Some test set molecules from scaffolds **A**, **B** and **C**.

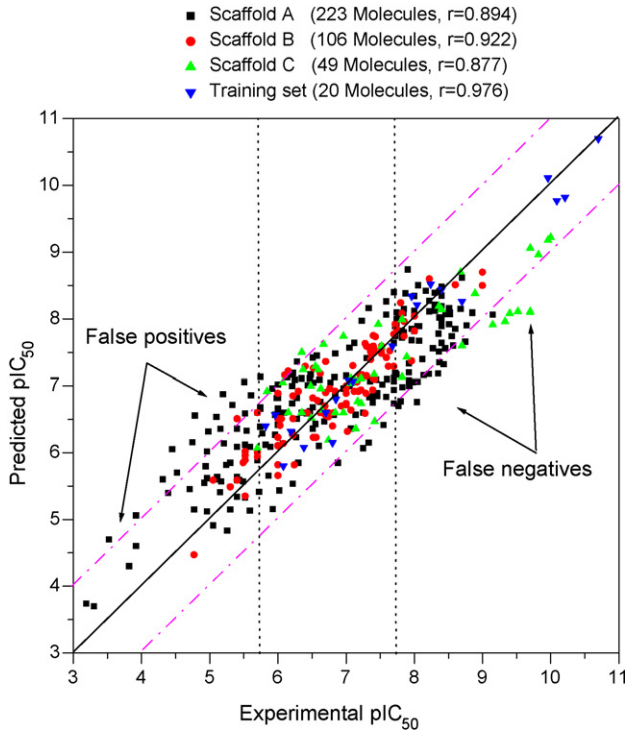


Fig. 5. Graph showing the chemotype-wise correlation (*r*) between experimental and predicted activities for the 378 test set molecules against Hypo-1 model along with 20 training set molecules.

substituents with amino acid residues in the active site might have a subtle, yet crucial role on the predicted activity. While these additional groups may not prevent in identifying many low energy conformers or add any penalty for the total cost, but could be detrimental to fit these conformers in the active site. Thus the features of Hypo-1 are relatively well optimized with some of the low energy conformers of medium active molecules of **A** without contributing much difference between the fixed and total costs. However, it may not be the same in the case of either low actives or high actives as too many false positives or a few false negatives are shown respectively, wherein the pharmacophore features are not best optimized due to the effect of additional substituents in the case of high active molecules and over optimized in the case of low active molecules.

In **C**, conformational flexibility of cyclic peptides is highly influenced by solvent and amino acid residues in the active site, while such effects have not been considered in the conformation generation while the difficulty in predicting the conformation in cyclic systems are many especially when the system contains the amino acid residues. For example N–H···O bonds might play a crucial role on the conformation energy as well as in the total cost, while the effect of solvent or protein environment on such conformational energy or total cost is very high which has not been completely taken into consideration. Thus these molecules have greater tendency to show more false negatives. In **B**, wherein these two aberrations are almost absent, predictivity is very good

Table 4
Hypo-1 validation with the test set molecules

Database details	Total molecules	Molecules in scaffold **A**	Molecules in scaffold **B**	Molecules in scaffold **C**
Total molecules	378	223	106	49
Active molecules	100	71	11	18
Hits	124	95	18	11
Active hits	88	67	10	11
False positives	36	28	8	0
False negatives	12	4	1	7

and almost in many molecules a very good correlation is observed.

3.2.2. NCI database search

The Hypo-1 model was used to screen the NCI database consisting of 238,819 molecules, which yielded 4638 hits comprising 297 high, 1433 medium and 2988 low active molecules. Some of the molecules with the good predicted activities were shown in Fig. 6. Nearly 56% of the hits (2608) fall in the activity range of 200–1000 nM, while the hits from medium and high activity groups show nearly a Gaussian distribution. Only 86 hits are below the 10 nM. In fact one can consider all the 297 hits that are predicted as highly active could become a good source for further evaluation, while some others in medium actives cannot be completely neglected.

The predicted activity of the 4638 hits ranges from 0.28 to 8900 nM. Details on the further distribution of these hits into the chemotypes **A**, **B**, **C** and Others are also shown in Table 5 and such classification is based on chemical intuition rather than any automated methods. Irrespective of the activity range, nearly 62% of hits cannot be easily categorized to any of the three chemotypes defined here, but some may prove to be useful. Also similar consistency is maintained in the hits when viewed with respect to scaffold or activity. The % of molecules of high, medium and low activities is nearly similar (6.4, 30.9 and 62.5%, respectively) in all three chemotypes. Also in the 38% hits represented three chemotypes, it is interesting to note that **B** molecules (26%) comprise a large set compared to those of either **A** (5.7%) or **C** (6.4%) in all activity ranges and it goes well that performance of **B** molecules is better even in the test set.

Similarity analysis of 4638 hits with respect to **1**, **14** and **5** representing scaffolds **A**, **B** and **C** has been performed based on ISIS 960 keys and Tanimoto analysis. Fig. 7 shows the correlation between the predicted activity and the similarity index for the 4638 hits. It is noticed that irrespective of the activity range, relatively higher similarity with respect to

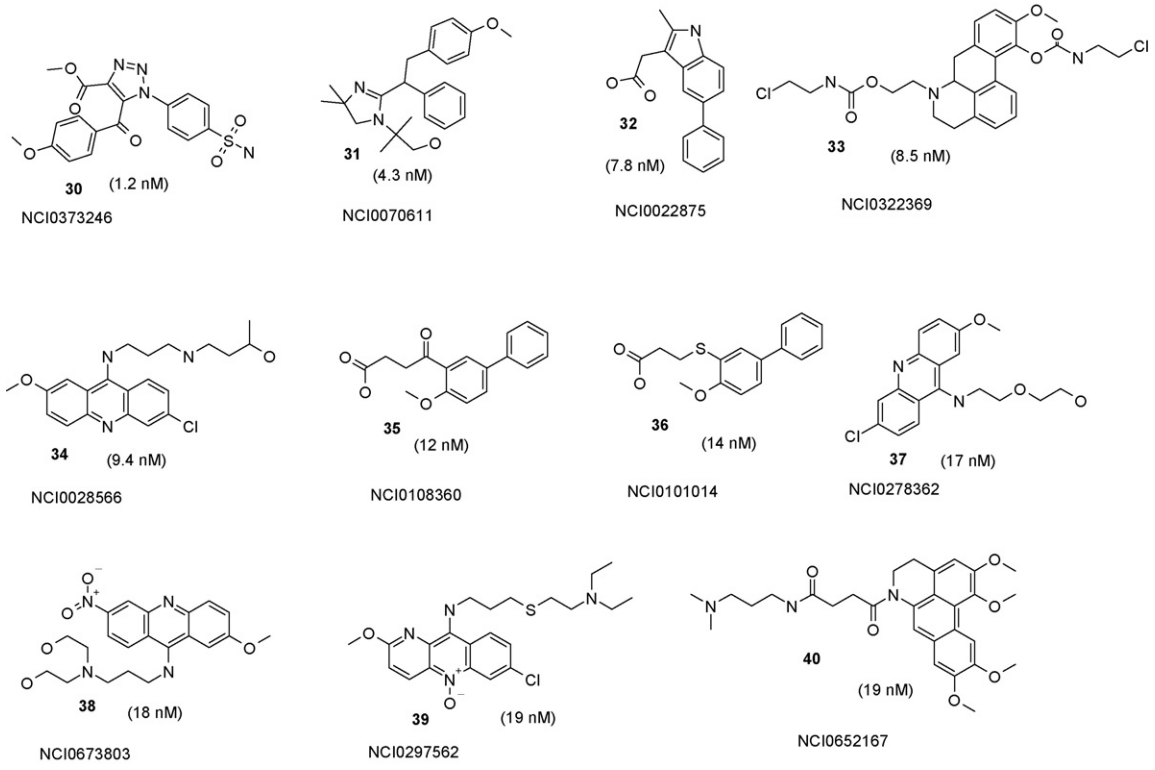


Fig. 6. Identified lead molecules through NCI database search along with predicted IC_{50}, nM.

Table 5
Predicted activities of 4638 hits that are retrieved from NCI database on virtual screening with Hypo-1 model

	Total hits	Hits			
		Scaffold **A**	Scaffold **B**	Scaffold **C**	Others
Total hits	4638	265	1202	297	2874
Highly active	297	17	77	19	184
Medium	1438	82	372	92	892
Low active	2903	166	753	186	1798
Variation with respect to scaffold (%)					
Total (%)	100.0	5.7	25.9	6.4	61.9
High (%)	6.4	5.7	25.9	6.4	62.0
Medium (%)	30.9	5.7	26.0	6.4	62.0
Low (%)	62.5	5.7	26.0	6.4	62.0
Variation with respect to activity (%)					
Total (%)	100.0	100.0	100.0	100.0	100.0
High (%)	6.4	6.4	6.4	6.4	6.4
Medium (%)	30.9	30.9	30.9	31.0	30.9
Low (%)	62.5	62.6	62.6	62.3	62.5

chemotype B is observed compared to **A** and **C**. Based on the earlier observations that model performs for this class of molecules than others, it is recognizable that this virtual screening effort produced relatively more hits which would be worthwhile to continue further experimental studies. Also, this would indicate the consistency of Pharmacophore model against scaffold **B** and produced large number of hits from NCI database.

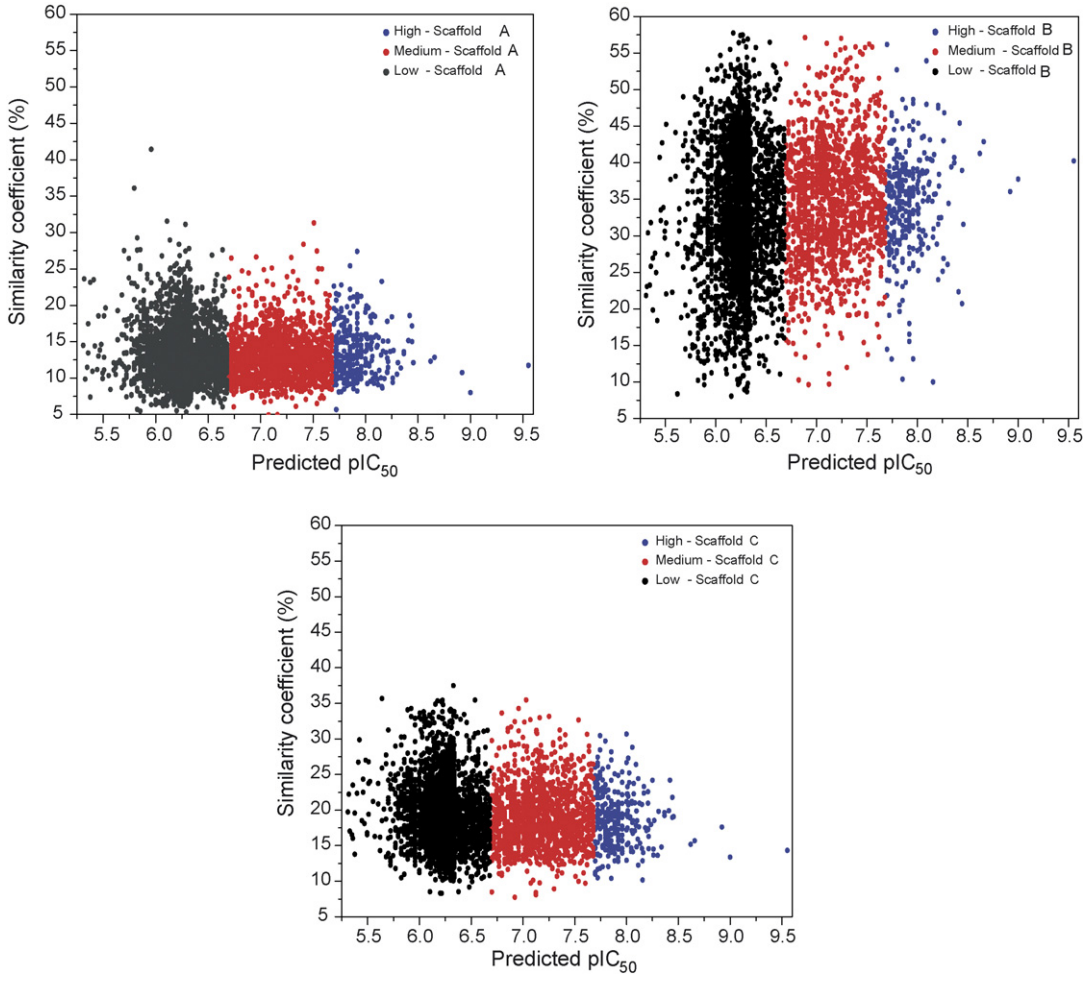


Fig. 7. Predicted pIC_{50} vs. similarity indices of scaffold A, B and C.

4. Conclusions

HDAC enzymes have proven to be exciting and promising novel targets for the treatment of solid tumors and hematological cancers. In-house build Medichem database is a powerful resource to identify many HDAC inhibitors with highly varied activities and chemotypes. These inhibitors have been retrieved from the resource and some of them have been used to general a Pharmacophore model while other inhibitors have been used for virtual screening to validate the model. The purpose of the study is useful for two purposes:

1. To generate Pharmacophore models as powerful search tool to be used as a 3D query to identify lead molecules from chemical databases as potential HDAC inhibitors.
2. To utilize the Pharmacophore model as a predictive tool for estimating biological activity of virtual molecules or molecules designed on the basis of structure activity analysis.

The best quantitative Pharmacophore model in terms of predictive value consisted of four features like one hydrogen bond acceptor, one hydrophobic aliphatic and two ring aromatic features, which is further validated by using an large set of 378 HDAC inhibitors and gives a r value of 0.897. The most active molecule 1 (IC_{50} = 0.02 nM) in the training set fits very well with this top scoring pharmacophore hypothesis. Virtual screening produced some false positives and a few false negatives. It is believed concurrent use or a consensus study, which readily minimizes these errors, could be an added tool for Pharmacophore model based virtual screening in order to produce reliable true positives and negatives.

This Pharmacophore model was further used to search the NCI database consisting of 238,819 structurally diversified molecules, which yielded 4638 molecules as hits that satisfied the 3D query. The activities of those molecules were predicted using the developed Pharmacophore model and the highly active molecules are further used to design more potent lead molecules against histone deacetylase inhibitors for the treatment of various types of cancer.

5. Experimental

All molecular modeling works were performed on a Silicon Graphics Octane R12000 computer running Irix 6.5.12 (SGI, 1600 Amphitheatre Parkway, Mountain View, CA 94043) Catalyst 4.10 software was used to generate Pharmacophore models.

Acknowledgements

Vadivelan would like to thank D.S. Brar, Chairman and G.V. Sanjay Reddy, CEO of GVK Biosciences Pvt. Ltd., for their continuous support, S. Rama Devi, T. Sunita, G. Madhavi Sastry, Vema Aparna and M. Anil Kumar for their guidance and technical support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.07.002.

References

[1] C. Monneret, Histone deacetylase inhibitors, Eur. J. Med. Chem. 40 (2005) 1–13.
[2] A. Mai, S. Massa, R. Ragno, M. Esposito, G. Sbardella, G. Nocca, R. Scatena, F. Jesacher, P. Loidl, G. Brosch, Binding mode analysis of 3-(4-benzoyl-1-methyl-1*H*-2-pyrrolyl)-*N*-hydroxy-2-propenamide: a new synthetic histone deacetylase inhibitor inducing histone hyperacetylation, growth inhibition, and terminal cell differentiation, J. Med. Chem. 45 (2002) 1778–1784.
[3] A. Mai, S. Massa, R. Ragno, I. Cerbara, F. Jesacher, P. Loidl, G. Brosch, 3-(4-Aroyl-1-methyl-1*H*-2-pyrrolyl)-*N*-hydroxy-2-alkylamides as a new class of synthetic histone deacetylase inhibitors. 1. Design, synthesis, biological evaluation, and binding mode studies performed through three different docking procedures, J. Med. Chem. 46 (2003) 512–524.
[4] G.V. Kapustin, G. Fejer, J.L. Gronlund, D.G. McCafferty, E. Seto, F.A. Etzkorn, Phosphorus-based SAHA analogues as histone deacetylase inhibitors, Org. Lett. 5 (2003) 3053–3056.
[5] A.J. De Ruijter, A.H. Van Gennip, H.N. Caron, S. Kemp, A.B. Van Kuilenburg, Histone deacetylases (HDACs): characterization of the classical HDAC family, Biochem. J. 370 (2003) 737–749.
[6] J.R. Somoza, R.J. Skene, B.A. Katz, C. Mol, J.D. Ho, A.J. Jennings, C. Luong, A. Arvai, J.J. Buggy, E. Chi, J. Tang, B.C. Sang, E. Verner, R. Wynands, E.M. Leahy, D.R. Dougan, G. Snell, M. Navre, M.W. Knuth, R.V. Swanson, D.E. McRee, L.W. Tari, Structural snapshots of human HDAC8 provide insights into the Class I histone deacetylases, Structure 12 (2004) 1325–1334.
[7] W. Hilmar, O. Eckhard, Recent advances in medicinal chemistry of histone deacetylase inhibitors, Annu. Rev. Med. Chem. 39 (2004) 185–196.
[8] D.F. Wang, P. Helquist, N.L. Wiech, O. Wiest, Toward selective histone deacetylase inhibitor design: homology modeling, docking studies, and molecular dynamics simulations of human Class I histone deacetylases, J. Med. Chem. 48 (2005) 6936–6947.
[9] R. Ragno, A. Mai, S. Massa, I. Cerbara, S. Valente, P. Bottoni, R. Scatena, F. Jesacher, P. Loidl, G. Brosch, 3-(4-Aroyl-1-methyl-1*H*-pyrrol-2-yl)-*N*-hydroxy-2-propenamides as a new class of synthetic histone deacetylase inhibitors. 3. Discovery of novel lead compounds through structure-based drug design and docking studies, J. Med. Chem. 47 (2004) 1351–1359.
[10] D.F. Wang, O. Wiest, P. Helquist, H.Y. Lan-Hargest, N.L. Wiech, On the function of the 14 Å long internal cavity of histone deacetylase-like protein: implications for the design of histone deacetylase inhibitors, J. Med. Chem. 47 (2004) 3409–3417.
[11] T.A. Miller, D.J. Witter, S. Belvedere, Histone deacetylase inhibitors, J. Med. Chem. 46 (2003) 5097–5116.
[12] Catalyst, Version 4.10, Accelrys Inc., San Diego, California, USA, 2005.
[13] P. Prathipati, A.K. Saxena, Characterization of β3-adrenergic receptor: determination of pharmacophore and 3D QSAR model for β3 adrenergic receptor agonism, J. Comput. Aided Mol. Des. 19 (2005) 93–110.
[14] C. Laggner, C. Schieferer, B. Fiechtner, G. Poles, R.D. Hoffmann, H. Glossmann, T. Langer, F.F. Moebius, Discovery of high-affinity ligands of ó1 receptor, ERG2, and emopamil binding protein by pharmacophore modeling and virtual screening, J. Med. Chem. 48 (2005) 4754–4764.
[15] M.L. Lopez-Rodriguez, B. Benhamu, T. de la Fuente, A. Sanz, L. Pardo, M. Campillo, A three-dimensional pharmacophore model for 5-hydroxytryptamine6 (5-HT6) receptor antagonists, J. Med. Chem. 48 (2005) 4216–4219.
[16] R. Kristam, V.J. Gillet, R.A. Lewis, D. Thorner, Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst, J. Chem. Inf. Model. 45 (2005) 461–476.
[17] M.Y. Li, K.C. Tsai, L. Xia, Pharmacophore identification of alpha 1A-adrenoceptor antagonists, Bioorg. Med. Chem. Lett. 15 (2005) 657–664.

[18] Y. Kurogi, O. Guner, Pharmacophore modeling and three-dimensional database searching for drug design using catalyst, Curr. Med. Chem. 8 (2001) 1035–1055.
[19] G.I. Mustata, A. Brigo, J.M. Briggs, HIV-1 integrase pharmacophore model derived from diverse classes of inhibitors, Bioorg. Med. Chem. Lett. 14 (2004) 1447–1454.
[20] E.A. Hecker, C. Duraiswami, T.A. Andrea, D.J. Diller, Use of catalyst pharmacophore models for screening of large combinatorial libraries, J. Chem. Inf. Comput. Sci. 42 (2002) 1204–1211.
[21] MediChem Database, Informatics, GVK Biosciences Pvt. Ltd., S-1, Phase-1, T.I.E. Balanagar, Hyderabad-500 037, India.
[22] S.L. Colletti, R.W. Myers, S.J. Darkin-Rattray, A.M. Gurnett, P.M. Dulski, S. Galuska, J.J. Allocco, M.B. Ayer, C. Li, J. Lim, T.M. Crumley, C. Cannova, D.M. Schmatz, M.J. Wyvratt, M.H. Fisher, P.T. Meinke, Broad spectrum antiprotozoal agents that inhibit histone deacetylase: structure activity relationships of apicidin. Part 2, Bioorg. Med. Chem. Lett. 11 (2001) 113–117.
[23] S. Uesato, M. Kitagawa, Y. Nagaoka, T. Maeda, H. Kuwajima, T. Yamori, Novel histone deacetylase inhibitors: *N*-hydroxycarboxamides possessing a terminal bicyclic aryl group, Bioorg. Med. Chem. Lett. 12 (2002) 1347–1349.
[24] M.L. Curtin, R.B. Garland, H.R. Heyman, R.R. Frey, M.R. Michaelides, J. Li, L.J. Pease, K.B. Glaser, P.A. Marcotte, S.K. Davidsen, Succinimide hydroxamic acids as potent inhibitors of histone deacetylase (HDAC), Bioorg. Med. Chem. Lett. 12 (2002) 2919–2923.
[25] Y. Dai, Y. Guo, J. Guo, L.J. Pease, J. Li, P.A. Marcotte, K.B. Glaser, P. Tapang, D.H. Albert, P.L. Richardson, S.K. Davidsen, M.R. Michaelides, Indole amide hydroxamic acids as potent inhibitors of histone deacetylases, Bioorg. Med. Chem. Lett. 13 (2003) 1897–1901.
[26] C.K. Wada, R.R. Frey, Z. Ji, M.L. Curtin, R.B. Garland, J.H. Holms, J. Li, L.J. Pease, J. Guo, K.B. Glaser, P.A. Marcotte, P.L. Richardson, S.S. Murphy, J.J. Bouska, P. Tapang, T.J. Magoc, D.H. Albert, S.K. Davidsen, M.R. Michaelides, α-Keto amides as inhibitors of histone deacetylase, Bioorg. Med. Chem. Lett. 13 (2003) 3331–3335.
[27] Y. Dai, Y. Guo, M.L. Curtin, J. Li, L.J. Pease, J. Guo, P.A. Marcotte, K.B. Glaser, S.K. Davidsen, M.R. Michaelides, A novel series of histone deacetylase inhibitors incorporating hetero aromatic ring systems as connection units, Bioorg. Med. Chem. Lett. 13 (2003) 3817–3820.
[28] S.W. Remiszewski, L.C. Sambucetti, K.W. Bair, J. Bontempo, D. Cesarz, N. Chandramouli, R. Chen, M. Cheung, S. Cornell-Kennon, K. Dean, G. Diamantidis, D. France, M.A. Green, K.L. Howell, R. Kashi, P. Kwon, P. Lassota, M.S. Martin, Y. Mou, L.B. Perez, S. Sharma, T. Smith, E. Sorensen, F. Taplin, N. Trogani, R. Versace, H. Walker, S. Weltchek-Engler, A. Wood, A. Wu, P. Atadja, *N*-Hydroxy-3-phenyl-2-propenamides as novel inhibitors of human histone deacetylase with in vivo antitumor activity: discovery of (2*E*)-*N*-hydroxy-3-[4-[[(2-hydroxyethyl)[2-(1*H*-indol-3-yl)ethyl]amino]methyl]phenyl]-2-propenamide (NVP-LAQ824), J. Med. Chem. 46 (2003) 4609–4624.
[29] A. Hirashima, M. Morimoto, E. Kuwano, E. Taniguchi, M. Eto, Three-dimensional common-feature hypotheses for octopamine agonist 2-(arylimino)imidazolidines, Bioorg. Med. Chem. 10 (2002) 117–123.
[30] A.K. Debnath, Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of mycobacterium avium complex dihydrofolate reductase, J. Med. Chem. 45 (2002) 41–53.
[31] F.G. Osman, R.H. Douglas, in: F. Osman, Guner (Eds.), Phamacophore Perception, Developmentand Use in Drug Design, International University Line, La Jolla, California, 2000 , pp. 193–210 (Chapter 11).
[32] ISIS Host, Version 5.1, Elsevier MDL., 2440 Camino Ramon, Suite 300, San Ramon, CA 94583.
[33] R. Ragno, S. Simeoni, S. Valente, S. Massa, A. Mai, 3D QSAR studies on histone deacetylase inhibitors. A GOLPE/GRID approach on different series of compounds, J. Chem. Inf. Model. 46 (2006) 1420–1430.