Journal of Enzyme Inhibition and Medicinal Chemistry, 2011; 26(4): 535–545

ISSN 1475-6366 print/ISSN 1475-6374 online
DOI: 10.3109/14756366.2010.535795

informa
healthcare

ORIGINAL ARTICLE

Discovery of potential pancreatic cholesterol esterase inhibitors using pharmacophore modelling, virtual screening, and optimization studies

Shalini John*, Sundarapandian Thangapandian*, Sugunadevi Sakkiah, and Keun Woo Lee

Department of Biochemistry and Division of Applied Life Science (BK21 Program), Environmental Biotechnology National Core Research Center (EB-NCRC), Gyeongsang National University (GNU), Jinju, Republic of Korea

Abstract

Pancreatic cholesterol esterase (CEase) is a serine hydrolase involved in the hydrolysis of variety of lipids and transport of free cholesterol. In this study, pharmacophore hypotheses based on known inhibitors were generated using common feature pharmacophore generation protocol available in Discovery Studio program. The best pharmacophore model containing two hydrogen bond acceptor and three hydrophobic features was selected and validated. It was further used in screening three diverse chemical databases. Hit compounds were subjected to drug-likeness and molecular docking studies. Four hits, namely SEW00846, NCI0040784, GK03167, and CD10645, were selected based on the GOLD fitness score and interaction with active site amino acids. All hit compounds were further optimized to improve their binding in the active site. The optimized compounds were found to have improved binding at the active site. Strongly binding optimized hits at the active site can act as virtual leads in potent CEase inhibitor designing.

Keywords: Pancreatic cholesterol esterase, common feature pharmacophore, virtual screening, molecular docking, optimization

Introduction

Pancreatic cholesterol esterase (CEase), an enzyme, which is also known as bile salt-activated lipase, is responsible for the hydrolysis of various substrates including dietary cholesterol esters, fat-soluble vitamins, triglycerides, and phospholipids.[1,2] CEase is secreted from vertebrate pancreas into the intestinal track and activated by primary bile salts. It is present in few mammals including humans. It is also present in the milk to assure efficient triglycerol utilization in breast-fed newborns as their pancreas develops.[3,4] CEase gene-knockout transgenic mice showed a reduced uptake of cholesteryl ester, and confirms that this enzyme is responsible for the intestinal fat absorption.[5] Lack of CEase activity can also lead to the incomplete digestion of milk fat and accumulation by enterocytes in the ileum of newborn mice.[3] Apart from its role in fat digestion, it is also involved directly in lipoprotein metabolism in which it catalyses the conversion of larger and less atherogenic low-density lipoprotein to the smaller and more atherogenic low-density lipoprotein subspecies and may also regulate serum cholesterol level.[6,7] Therefore, CEase may function in these roles and act as cholesterol transfer protein and have harmful effect in atherosclerosis processes.[7,8] The role of CEase in atherogenesis and the relationship of this enzyme to various pathological conditions are not completely established so far.[9]

CEase belongs to α/β-hydrolase fold family. The active site includes both the catalytic triad (Ser194–Asp320–H435) and oxyanion hole (Gly107–Ala108–Ala195) residues.[10] CEase, serine protease, and serine lipase enzymes share the same catalytic mechanism that is the formation of discrete acyl enzyme intermediate via serine hydroxyl group in the active site, by having a Ser–Asp–His catalytic

*These authors have equally contributed to this work.

Address for Correspondence: Keun Woo Lee, Department of Biochemistry and Division of Applied Life Science (BK21 Program), Environmental Biotechnology National Core Research Center (EB-NCRC), Gyeongsang National University (GNU), 900 Gazwa-dong, Jinju 660-701, Republic of Korea. E-mail: kwlee@gnu.ac.kr

(Received 01 July 2010; revised 22 October 2010; accepted 26 October 2010)

Abbreviations

CEase	cholesterol esterase
DS	Discovery Studio
HBA	hydrogen bond acceptor
HBD	hydrogen bond donor
PI	positive ionizable
RA	ring aromatic
HY	hydrophobic
E	enrichment factor
GH	goodness of hit score
ADMET	absorption, distribution, metabolism, elimination, and toxicity
GOLD	genetic optimization for ligand docking
PDB	protein data bank
LGA	Lamarckian genetic algorithm

triad as serine protease.[11] These three category of enzymes may also expected to be inhibited by the same class of mechanism-based inhibitors since they share this common mechanism of hydrolysis by possessing a Ser–Asp–His catalytic triad. Recently, CEase has been emerging as a potential target particularly for the development of hypocholesterolaemic agents. It has been demonstrated that aryl haloketones, diethyl-*p*-nitrophenol phosphate, carbamates, boronic acids, isocoumarins, β-lactones, β-lactams, aryl carbamates, and fluoroketones are the most studied class and mechanism-based inhibitors of CEase.[12–16] The characterization of enzyme inhibition by aryl carbamates provides information that this process occurs mainly because of fast carbamylation of the active site serine followed by slow decarbamylation. Designing mechanism-based inhibitors with a scissile CO–O or CO–N bond within a ring system for the inhibition of serine protease and serine lipase enzymes was frequently observed in various studies.[17–22] This concept has not been adopted during the development of early CEase inhibitors. However, 6-chloro-2-pyrones, representatives of a known class of mechanism-based inhibitors of serine proteases, have been described as potent CEase inhibitors later.[23] Though, there are many compounds available for the inhibition of CEase, still there is a need for a new class of compounds as drugs in the treatment of cholesterol-related diseases.

In the present study, we have developed a pharmacophore model based on the common chemical features present in the known CEase inhibitors. The obtained pharmacophore model was used to identify a set of new lead candidates through virtual screening. The screened compounds were considered in drug-likeness evaluation and molecular docking study. The retrieved final hits were further optimized to bind with high affinity at the active site of the enzyme. Synthetic accessibility of the final lead candidates that are listed as a result of this study was calculated using a recently developed program.

Materials and methods

Pharmacophore generation

All the studies were carried out using Discovery Studio (DS) 2.5 unless it is mentioned. A data set containing chemical compounds known for their *in vitro* inhibitory activity values against human CEase from the literature[8,11,23–25] was prepared. Of this data set, six most active compounds as shown in Figure 1 were selected as a training set to generate qualitative pharmacophore models to be used in database screening to identify new scaffolds for the future drug discovery. These compounds are diverse in terms of their structures and found to inhibit CEase in nanomolar level. The 2D chemical structures of the training set compounds were built using ChemSketch program version 12, and subsequently converted into 3D structures using DS. All compounds in the training set were given a *Principal* value of 2 and a *Maximum Omitted Feature* value of 0 to make sure that all the features of these compounds are considered during pharmacophore generation. Diverse conformational models for every training set compound were generated to cover the flexibility of their chemical nature using polling algorithm.[26] All the compounds were energetically minimized using CHARMM force field implemented in DS. *Diverse Conformation Generation* protocol with *BEST flexible search* option implemented in DS was employed with the default value of generating maximum of 250 conformers within the energy range of 20 kcal/mol, with respect to the global minimum. *Feature mapping* protocol was employed prior to the original pharmacophore generation calculation to identify the chemical features present in the training set compounds. The chemical features such as hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), positive ionizable (PI), ring aromatic (RA), and hydrophobic (HY) features were used during pharmacophore generation. These chemical features were selected based on the feature mapping results and the possible interaction points available at the active site of CEase. All the other parameters were maintained at their default settings. The six compounds in the training set along with the generated conformational models were used in pharmacophore model generation. Common feature pharmacophore models, generally, are developed by comparing a set of conformational models and a number of 3D configurations of chemical features shared among the training set compounds. *Common Feature Pharmacophore Model Generation* protocol implemented in DS was used to generate pharmacophore models. Ten pharmacophore models were generated based on the common chemical features present in the training set compounds. In order to generate a reliable and best pharmacophore model that can be used further in drug

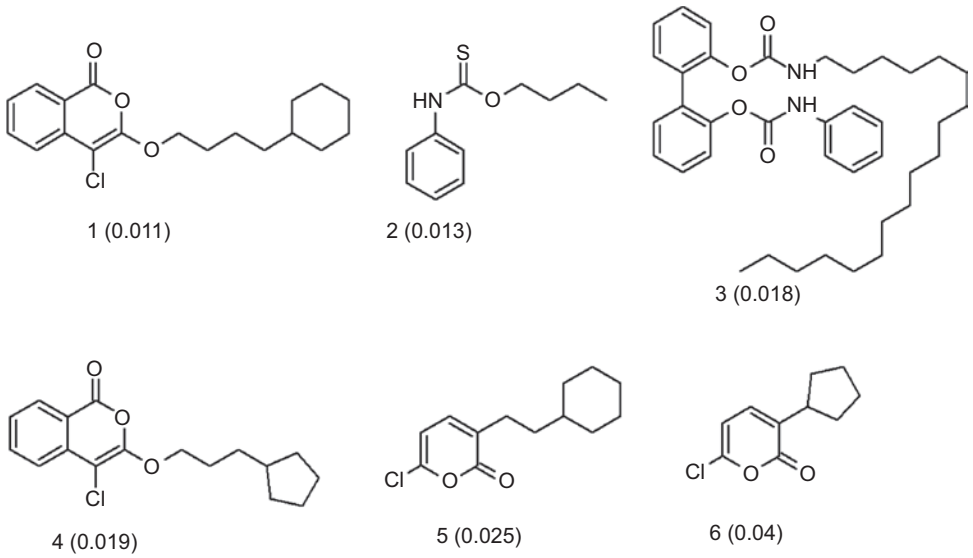


Figure 1. Chemical structures of the training set compounds together with their experimental K_i values in μM used for pharmacophore generation.

discovery process, number of pharmacophore models was generated by selecting different set of features and by varying the default values of other parameters.

Pharmacophore validation

The best pharmacophore model "Hypo 1" was validated based on (i) its potential in distinguishing active from the inactive compounds in a database, (ii) the best fit values of the training set compounds, and (iii) the presence of significant chemical features required for the interaction with key catalytic triad and oxyanion hole residues. For the first validation, an external dataset containing 90 compounds was developed with 35 active and 55 inactive compounds for CEase inhibition and used to validate the generated pharmacophore models. These active and inactive set of compounds in the database were obtained from published literature.[8,9,11,23–25,27–29] The chemical structures were sketched and their 3D structures were generated as explained in training set preparation. The best pharmacophore model, "Hypo 1", was used as a 3D query to screen this dataset and a set of statistical parameters[30,31] like hit list (H_t), number of active percent of yields (%Y), percent ratio of actives in the hit list (%A), enrichment factor (E), false negatives, false positives, and goodness of hit score (GH) were calculated.

Virtual screening

The pharmacophore model that performed better in all the validation procedures was considered as a best pharmacophore model. It was used further as a 3D query to search chemical databases like Maybridge, Chembridge, and NCI2000 containing 59,652, 50,000, and 238,819 compounds, respectively, in order to identify new scaffolds to be utilized in novel CEase inhibitor design. *Search 3D Database* protocol with *BEST search* option as available in DS was employed in virtual screening. Lipinski's rule of five and absorption, distribution, metabolism, elimination, and toxicity (ADMET) properties were used as primary and secondary filters to remove non-drug-like hits.[32–34] In order to further refine and also to reduce the false positives of the retrieved hit compounds, molecular docking study was carried out using well-validated two docking programs, namely, genetic optimization for ligand docking (GOLD) and AutoDock.

Molecular docking study

The binding orientation of the retrieved hit compounds in the active site of human CEase was analysed by molecular docking studies using the GOLD 4.1 program from Cambridge Crystallographic Data Center, UK.[35] GOLD uses genetic algorithm for docking flexible ligands into protein-binding site to explore the full range of ligand conformational flexibility with partial flexibility of the protein. The GOLD fitness score was calculated from the contribution of hydrogen bonds and van der Waals interactions between protein and ligand. GOLD has two scoring fitness functions, namely, GOLDScore and ChemScore. Among these two functions, GOLDScore was identified, in previous studies, as better scoring function than ChemScore in predicting the inhibitor-binding positions and affinity.[36,37] The GOLD fitness score was calculated using the following equation:

$$\text{GOLD Fitness} = S_{hb\_ext} + 1.3750 \times S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int}$$

S_{hb_ext} and S_{vdw_ext} indicate the contribution from hydrogen bonds and van der Waals interactions between protein and ligand, respectively; S_{hb_int} represents the contribution to the Fitness due to intramolecular hydrogen bonds in the ligand; and S_{vdw_int} is the contribution from intramolecular strain in the ligand.[37] Thereby, we used GOLD fitness score function to measure the binding positions and affinity of ligand to the protein-binding site. X-ray crystal structure of human CEase from protein data bank (PDB ID: 1F6W) was used to define the binding site for

molecular docking studies. Ester hydrolysis by CEase is catalysed by the operation of catalytic triad and oxyanion hole residues. Thus interaction with hydroxyl group of Ser194 that acts as nucleophile is essential for the ester hydrolysis reaction. Protein residues within the radius of 10 Å around the hydroxyl group of Ser194 were defined to form the active site of the enzyme during molecular docking. The early termination option of three was changed to five and 10 docking poses per ligand was set whereas the other parameters were kept at their default values. From the molecular docking results, the compounds were selected based on their GOLD fitness score, interaction with active site amino acids, and structural diversity. The hit compounds from the molecular docking study were optimized further to improve their binding affinity at the active site. Ligplot program[38] was used to observe the protein–ligand HY interactions.

Optimization

The selected compounds were used as lead for further optimization. Many different substitutions were made at their side chains. These optimized compounds were also subjected to molecular docking study using GOLD program with the same settings used previously. The final hit compounds were selected based on the GOLD fitness score and interaction with essential amino acids. These final hits were also docked using AutoDock 4.2 program to validate the GOLD prediction. AutoDock is a fully automated docking program that employs Lamarckian genetic algorithm (LGA) as a search engine and LUDI-type scoring function.[39] Three-dimensional energy scoring grids were calculated. The default parameters for LGA were used. AutoDock predicts the binding conformation of the small molecules efficiently but takes time. As no scoring function employed in currently available docking programs performs better for all macromolecular targets, a combination of scoring functions from various programs (GOLD and AutoDock in this study) may provide significant value in predicting favourable binding conformations. Synthetic accessibility scores for all the optimized compounds scoring a GOLD fitness score value >50 was used to validate the synthetic possibilities. SYLVIA v 1.0 program[40,41] from the *Molecular Networks* group was employed to calculate the synthetic accessibility of these optimized compounds. The estimation of synthetic accessibility using SYLVIA provides a number between 1 and 10 for compounds that are very easy to synthesize and compounds that are very difficult to synthesize, respectively. The method for calculating synthetic accessibility takes account of a variety of criteria such as complexity of the molecular structure, complexity of the ring system, number of stereo centres, similarity to commercially available compounds, and potential for using powerful synthetic reactions. These criteria have been individually weighted to provide a single value for synthetic accessibility. Binding energies of all the optimized compounds were calculated using AutoDock as a cross-validation to the GOLD predictions. The novelty of the final hits and the optimized structures were evaluated using *SciFinder Scholar* and *Pubchem* compound search.

Results and discussion

Pharmacophore generation

Common feature pharmacophore models were built with six active training set compounds using DS. The training set included derivatives of isocoumarin, binapthol, and pyrone (Figure 1) chemical scaffolds. Several pharmacophore runs were carried out by changing the control parameters to develop the best model. The common feature pharmacophore generation run resulted in 10 pharmacophore models. Interestingly, all the 10 models were generated with same five pharmacophoric features that include 3 HY and 2 HBA, along with good ranking scores range from 57.887 to 49.212 (Table 1). We observed from the distance constraints of all 10 pharmacophore models that the two HBA features available in all pharmacophore models are present adjacent to each other and the two HY features are present close to the HBA features, whereas the third HY feature is located away from them. All pharmacophore models do not have much difference in their 3D distance constraints. Hence, the ranking score of a pharmacophore model and the fit values of the training set compounds based on the generated pharmacophore models were analyzed to choose the best model. "Hypo 1" was selected as the best pharmacophore model based on the highest ranking score of 57.887 and good fit values from the mapping of the training set compounds upon the chemical features (all the training set compounds showed good fit value with respect to this model). The second generated pharmacophore model "Hypo 2" has also scored a ranking very close to "Hypo 1" but the best fit values of the training set compounds were not predicted well (Table 2). The best pharmacophore model "Hypo 1" containing three HY and two HBA features is shown in Figure 2 along with its inter-feature distance constraints.

Table 1. Summary of the generated common feature pharmacophore models.

Hypotheses	Molecular features[a]	Ranking score	Direct hit[b]	Partial hit[c]	Max. fit
Hypo 1	HHHAA	57.887	111111	000000	5
Hypo 2	HHHAA	57.587	111111	000000	5
Hypo 3	HHHAA	56.563	111111	000000	5
Hypo 4	HHHAA	55.823	111111	000000	5
Hypo 5	HHHAA	55.542	111111	000000	5
Hypo 6	HHHAA	54.082	111111	000000	5
Hypo 7	HHHAA	53.880	111111	000000	5
Hypo 8	HHHAA	53.792	111111	000000	5
Hypo 9	HHHAA	52.673	111111	000000	5
Hypo 10	HHHAA	49.212	111111	000000	5

[a]H, Hydrophobic; A, hydrogen bond acceptor.
[b]Direct hit indicates whether (1) or not (0) a molecule in the training set mapped every feature in the hypothesis.
[c]Partial hit indicates whether (1) or not (0) a particular molecule in the training set mapped all but one feature in the hypothesis.

In order to reduce the false positive and/or false negative rate, "Hypo 1" was subjected to further validation[42] and employed in virtual screening to retrieve new scaffolds that matched its functional and spatial constraints.

Pharmacophore validation and database screening

A three-stage validation procedure has been employed to validate the pharmacophore model "Hypo 1" for its reliability to be employed in database screening. As a first stage of validation, an external dataset containing active and inactive compounds was screened using "Hypo 1" as 3D query to evaluate the discriminative ability of this pharmacophore model identifying the actives from inactive compounds. The active and inactive compounds were selected based on their biological activities. This dataset comprised a total of 90 compounds including 35 active (<1 μM) and 55 inactive (≥1 μM) compounds for CEase inhibition. Various statistical parameters were calculated to analyse the results (Table 3). "Hypo 1" has retrieved 94% of active compounds from the database. In addition to it, the calculated *E* value of 2.06 signifies that the selected pharmacophore model "Hypo 1" is good enough for virtual screening. GH score of 0.72 also ensures the statistical significance of the pharmacophore model, "Hypo 1". A significant pharmacophore model should have the GH value between 0.6 and 0.8. This validation result confirmed that the quality of "Hypo 1" is reliable to be used in database screening. Second validation step is based on the scored best fit values of the training set compounds when mapped over the pharmacophore model "Hypo 1". The best fit values of the training set compounds derived by mapping onto "Hypo 1" were generally higher than the closely ranked "Hypo 2". The most active compound 1 in the training set (Figure 3A) overlaid on all the features of "Hypo 1". Its fused ring, HY side chain and the cyclohexane moieties overlaid on 3 HY features. The carbonyl group and oxygen atom of isocoumarin were overlaid on HBA features. Another active compound 6 (Figure 3B) also mapped effective on "Hypo 1". The cyclopentane group and the chlorine attached with the pyrone moiety were overlaid on two of three HY features. The carbonyl group and oxygen atom present in pyrone moiety were overlaid on HBA features. As a third and final validation step, the presence of required chemical features to interact with the key amino acid residues (catalytic triad and oxyanion hole-forming residues) in the active site of human CEase. "Hypo 1" model was made of five features including three HY and two HBA. Molecular docking results of the training set compounds revealed that the parts overlaid on HBA features of "Hypo 1" are interacting with Ala195 and His435, whereas the HY parts interact with Ala108 and other HY residues such as Ala436 in the active site. Based on these validation results, "Hypo 1" was selected and subsequently used as a 3D query in the screening of chemical databases such as Maybridge, Chembridge, and NCI2000. *Search 3D Database* protocol with *Best search* option implemented in DS was used in database screening. The virtual screening of these databases resulted in 22,548, 19,751, and 48,537 compounds from Maybridge, Chembridge, and NCI2000, respectively. Several filtration criteria were used to reduce the number of compounds for further analysis. Lipinski's rule of five, as a primary filter, could retrieve 1279 from Maybridge, 1253 from Chembridge, and 2984 from NCI2000, respectively. This step of filtration selects only the compounds characterized with the following properties: molecular weight <500, number of HBAs <10, HBDs <5, and calculated octanol–water partition coefficient <5. Secondary filter involves the calculation of ADMET properties to further refine the retrieved hits with favourable bioavailability. ADMET properties were calculated for all the hit compounds using *ADMET descriptors* protocol available in DS. The ADMET descriptors like solubility, intestinal absorption, blood–brain barrier penetration, cytochrome P450 2D6 inhibition, and hepatotoxicity models were used. Finally, the compounds with the calculated fit value of ≥4 were chosen based on the most active compound in the training set. A total of 353 compounds were obtained from virtual screening and were subjected to molecular docking study using GOLD program to investigate the binding pose and interactions with the active site components.

Molecular docking

The retrieved 353 hit compounds were docked in to the active site of CEase (PDB ID: 1F6W) using GOLD. From the GOLD solutions, 57 compounds scoring GOLD fitness score greater than any training set compound were selected. Most active compound in the training set has

Table 2. Best fit values and absolute energy for the training set compounds respective to "Hypo 1" and "Hypo 2".

		Fit value		Absolute energy (kcal/mol)	
Compound	K_i μM	Hypo 1	Hypo 2	Hypo 1	Hypo 2
1	0.011	4.263	2.708	37.411	44.775
2	0.013	1.614	2.592	24.981	24.981
3	0.018	4.999	1.905	68.057	68.057
4	0.019	3.110	2.791	33.584	47.717
5	0.025	3.028	5.000	13.881	12.766
6	0.04	1.471	3.202	17.044	15.805

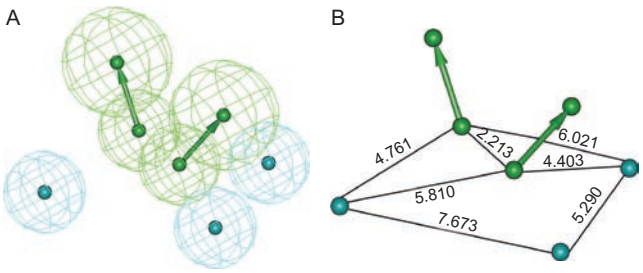


Figure 2. (A) Three-dimensional spatial arrangement of the best pharmacophore hypothesis "Hypo 1". (B) The distance constraints between the chemical features. Green colour represents hydrogen bond acceptor (HBA) and cyan colour represents hydrophobic (HY) features.

Table 3. Statistical parameters and scores from decoy set validation for "Hypo 1".

S. No.	Parameters	Results
1	Total molecules in database (D)	90
2	Total no. of actives in database (A)	35
3	Total hits (H_t)	41
4	Active hits (H_a)	33
5	% Yield of actives [$(H_a/H_t)\times 100$]	80.49
6	% Ratio of actives [$(H_a/A)\times 100$]	94.29
7	Enrichment factor (E) [$(H_a\times D)/(H_t\times A)$]	2.06
8	False negatives [$A - H_a$]	2
9	False positives [$H_t - H_a$]	8
10	Goodness of hit (GH)*	0.72

*$[(H_a/4H_tA)(3A + H_t)\times(1 - (H_t - H_a)/(D - A))]$; GH score of 0.6–0.8 indicates a very good model.

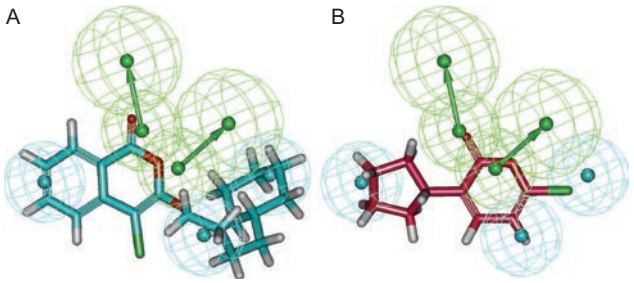


Figure 3. Compound 1 (A) and compound 6 (B) in the training set mapped on "Hypo 1". Green colour represents hydrogen bond acceptor (HBA) and cyan colour represents hydrophobic (HY) features.

scored a GOLD fitness score of 38.183. Ser194, Asp320, and His435 that form catalytic triad and Gly107, Ala108, and Ala195 are known as oxyanion hole amino acids responsible for the formation of tetrahedral intermediate in the active site of CEase and thereby essential for its catalytic function. Therefore, interactions with these amino acids are required for a potent CEase inhibitor. Finally four compounds, namely, SEW00846, NCI0040784, GK03167, and CD10645 were selected based on the GOLD fitness score, visual inspection of the binding modes, interaction with essential amino acids, and structural diversity. Binding mode and the molecular interactions for the most active compound in the training set is shown Supplementary Figure S1.

The binding mode of SEW00846 within the active site of CEase has been analysed. This compound was identified from Maybridge database. It has scored a GOLD fitness score of 56.127 and has shown interactions with Gly107, Ala108, Ala195, and His435. The oxazole ring of this compound has formed three hydrogen bonds with Gly107, Ala108, and Ala195, whereas the oxadiazole ring in the centre has formed a bidentate hydrogen bond with His435. The substituted phenyl ring and the methyl groups attached with the terminal oxazole ring, which were overlaid on the HY features of "Hypo 1" (Figure 4A) hydrophobically interact with Ala108, Ser194, and Ala195 (Figure 5A).

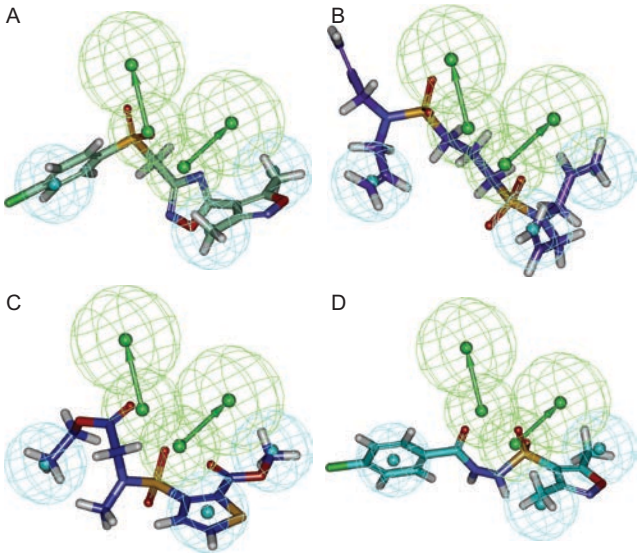


Figure 4. Pharmacophore mapping of final database hit compounds on the best pharmacophore hypothesis "Hypo 1", (A) SEW00846 represented in green colour, (B) NCI0040784 represented in violet colour, (C) GK03167 represented in blue colour, and (D) CD10645 represented in cyan colour. In the pharmacophore hypothesis, green represents hydrogen bond acceptor (HBA) and cyan represents hydrophobic (HY) features.

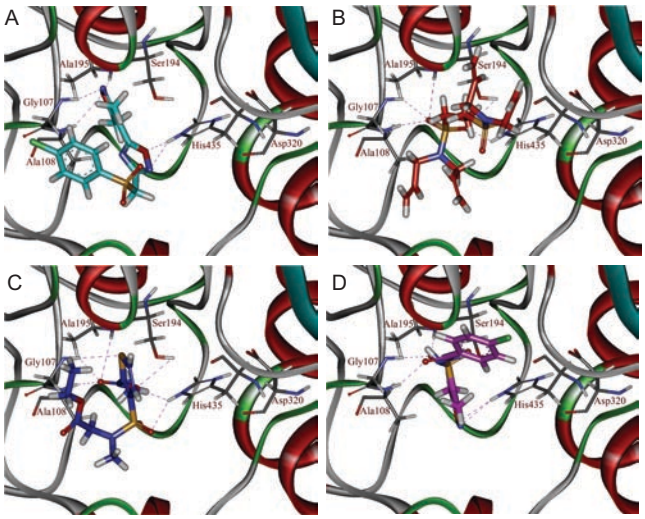


Figure 5. Binding orientations of database hit compounds: (A) SEW00846, (B) NCI0040784, (C) GK03167, and (D) CD10645 are shown in cyan, red, blue, and magenta colours, respectively. Hydrogen bonds are shown in dotted lines.

NCI0040784 has been retrieved from NCI2000 database and scored a GOLD fitness score of 51.904. This compound has formed hydrogen bond interactions with Gly107, Ala108, Ser194, Ala195, and His435. One of the SO_2 groups of the sulphonamides present in this compound that were overlaid on HBA features of "Hypo 1" (Figure 4B) has formed a hydrogen bond network with Gly107, Ala108, Ser194, Ala195, and His435 residues. The terminal propene substitutions that were mapped onto the HY features of "Hypo 1" interacted hydrophobically with Ala108, Ser194, and His435 at the active site (Figure 5B).

GK03167, a Maybridge compound, has scored a GOLD fitness score of 45.817 and also formed hydrogen bond interactions with Gly107, A108, Ser194, Ala195, and His 435. The SO_2 group of the central sulphonamide moiety that was mapped upon one of the HBA features of "Hypo 1" (Figure 4C) has formed a hydrogen bond with His435. Sulphur atom of the thiazole ring and ester group next to the thiazole ring formed hydrogen bond network with Gly107, Ala108, Ser194, and Ala195. The thiazole ring and the ethyl ester group away from it have shown HY interactions with His435 and Ala108, respectively (Figure 5C).

CD10645 from Maybridge database has scored a GOLD fitness score of 45.528. It formed hydrogen bond interactions with essential amino acids Gly107, Ala108, Ala195, and His435. The sulphonamide moiety and the carbonyl group next to it (mapped onto the HBA features of "Hypo 1", shown in Figure 4D) of CD10645 formed hydrogen bonds with Gly107, Ala108, and Ala195. The five-membered oxazole ring has formed bidentate hydrogen bond with His435. This oxazole ring also hydrophobically interacts with Ala108 and His435 residues in the active site (Figure 5D). The 2D structures of the final hits were shown in Figure 6. These compounds were further optimized to bind the active site of CEase with high affinity.

Optimization

The four hit compounds identified from the database screening were carried for further optimization. Different substitutions were added at their side chains in order to improve their binding affinity towards the catalytically active amino acids. For example, the database hit compound SEW00846 fits well into the active site of CEase but there is a possibility to extend its interaction towards Asp320 and other HY amino acids thereby improving its binding at the active site. Similarly, adding various substitutions to the hit compounds, totally 104 compounds were designed in this optimization study. The optimized compounds were docked into the active site of CEase using GOLD program with the same parameters used to dock the direct database hits. The results were analysed and 91 compounds that scored GOLD fitness score greater than their respective precursor compound were selected. The binding modes and molecular interactions with essential amino acids like Ser94, Ala107, Gly108, Ala195, Asp320, and His435 were analysed. AutoDock program has been employed to evaluate the GOLD predictions for the set of optimized compounds. We have presented the top four optimized compounds as a representative set to show their molecular interactions (Figure 7). OPT-SEW-5, the compound optimized from SEW00846, has scored a GOLD fitness score of 70.566, which was 56.127 before optimization and formed improved hydrogen bond interactions with essential amino acids compared with its precursor SEW00846. The SO_2 group of sulphonamide, which was not in interaction with any of the active site residues before optimization, formed hydrogen bond with His435. The optimization has also enabled extended interactions with Phe324 and Gly193 (data not shown). These interactions enhance the binding of the optimized SEW00846 (Figure 7A). The optimized compound based

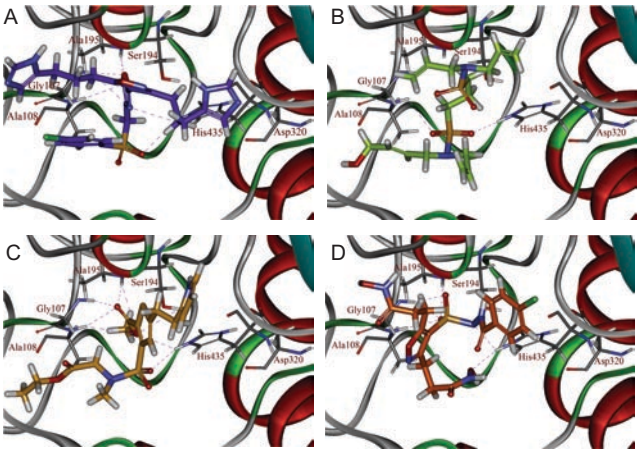


Figure 7. Binding orientations of the optimized compounds in the active site of Cease: (A) OPT-SEW-5, (B) OPT-NCI-1, (C) OPT-GK-1, and (D) OPT-CD-2 are shown in blue, green, yellow, and orange colours, respectively. Hydrogen bonds are shown in dotted lines.

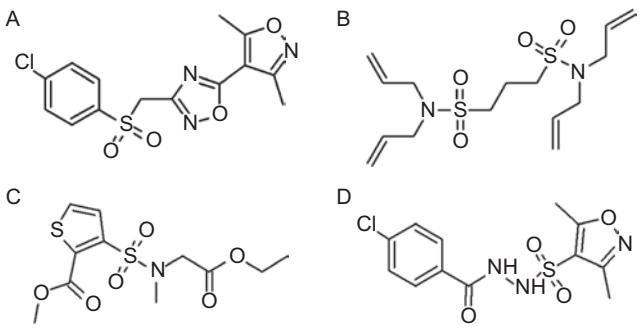


Figure 6. 2D representation of the database hit compounds: (A) SEW00846, (B) NCI0040784, (C) GK03167, and (D) CD10645.

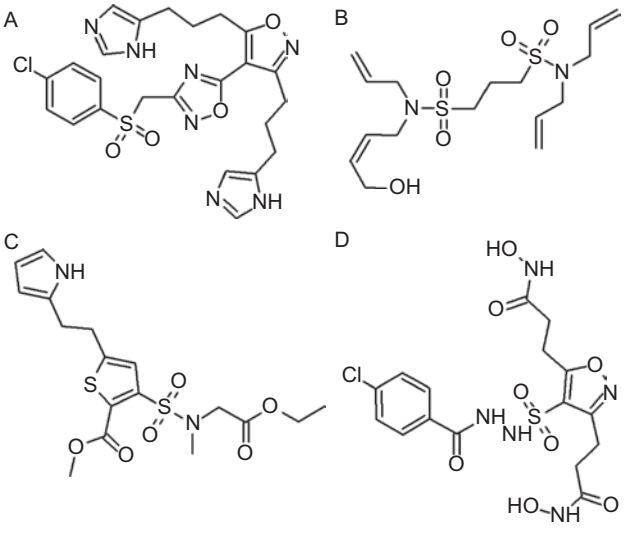


Figure 8. 2D representation of the final optimized compounds: (A) OPT-SEW-5, (B) OPT-NCI-1, (C) OPT-GK-1, and (D) OPT-CD-2.

Table 4. GOLD scores, AutoDock binding energies, and SYLVIA synthetic accessibility scores of top 10 optimized compounds.

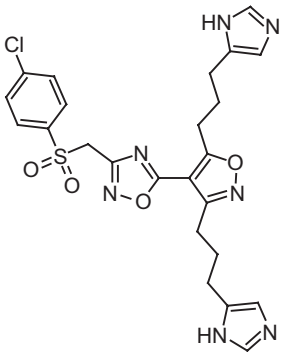
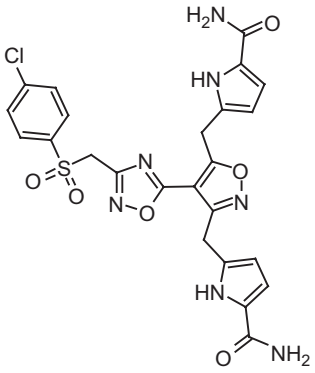
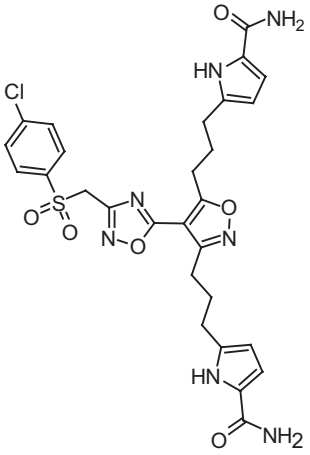
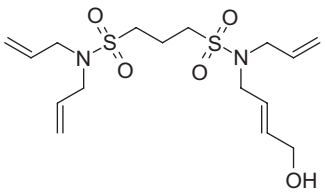
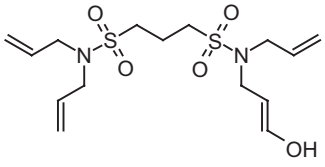
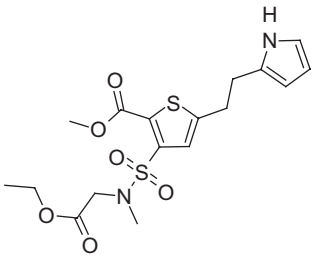
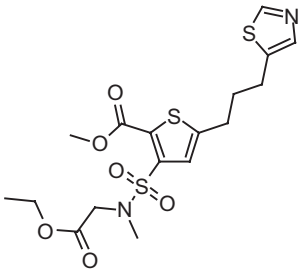
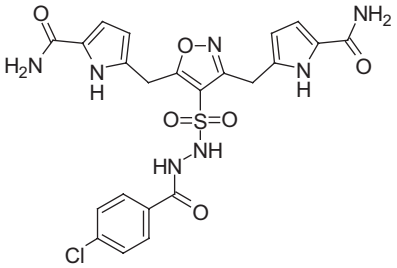
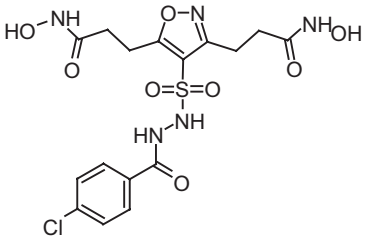
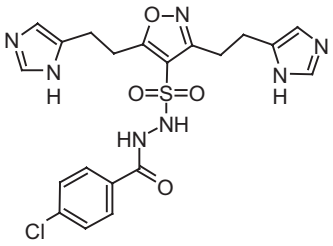
Compound	Structure	Precursor	GOLD fitness score	AutoDock binding energy	SYLVIA score[a]
OPT-SEW-1		SEW00846	70.566	−7.87	4.631
OPT-SEW-2		SEW00846	70.500	−9.70	5.064
OPT-SEW-3		SEW00846	64.473	−7.73	5.208
OPT-NCI-1		NCI0040784	58.110	−5.33	3.573
OPT-NCI-2		NCI0040784	54.520	−4.29	3.554

Table 4. continued on next page

Table 4. Continued.

Compound	Structure	Precursor	GOLD fitness score	AutoDock binding energy	SYLVIA score[a]
OPT-GK-1		GK03167	64.581	−5.48	4.555
OPT-GK-2		GK03167	64.397	−5.63	4.568
OPT-CD-1		CD10645	67.749	−9.77	5.107
OPT-CD-2		CD10645	66.979	−7.94	4.397
OPT-CD-3		CD10645	64.408	−7.17	4.471

[a]The calculated score for the synthetic accessibility. A number between 1 and 10 denoting the synthetic simplicity and the complexity of a particular chemical compound, respectively.

on NCI0040784 (OPT-NCI-1) scored an improved GOLD fitness score of 58.110 from 51.904 and shared hydrogen bond interactions with Ala108 and His435 and HY interaction with Gly107, Ser194, Glu193, and Leu124 (Figure 7B). Compound optimized form GK03167 (OPT-GK-1) with a GOLD fitness score of 64.581 from its precursor compound that scored 45.817 has formed hydrogen bond interactions with Ala195, Gly107, and His435 as well as HY interactions with Ser194, Phe324, and Ala275 (Figure 7C). Optimized compound from CD10645 (OPT-CD-2) has scored a GOLD fitness score of 66.979 from the previous score of 45.528. It has formed hydrogen bond interactions with Gly107, Ala108, Ala195, His435 as well as the HY interactions with Gly107, Ala108, Ser194, Phe324, and Ala436 (Figure 7D). The selected compounds were checked for their synthetic accessibility using SYLVIA 1.0 program. Two-dimensional representations of the top four optimized compounds were shown in Figure 8. The SYLVIA score for most of the final

optimized hits showed that these compounds are easy to be synthesized. Finally top 10 optimized compounds based on their GOLD fitness scores, AutoDock binding energies along with the SYLVIA scores were listed (Table 4) as possible virtual leads for CEase inhibitor designing. The novelty of the optimized compounds was confirmed by *SciFinder Scholar*[43] and *Pubchem compound* search.[44]

Conclusion

In summary, in this study we have developed a five-feature pharmacophore model "Hypo 1" containing 3 HY and 2 HBA features. "Hypo 1" was validated using three methods and used as 3D query in screening chemical databases like Maybridge, Chembridge, and NCI2000 followed by drug-likeness filtration and molecular docking studies. The four compounds, namely SEW00846, NCI0040784, GK03167, and CD10645, were selected based on their binding orientation, GOLD fitness score, and their interactions with active site amino acids. These final hits were further optimized with different substitutions at their side chains based on the interaction points observed from the active site cavity and subjected to molecular docking studies. Out of 104 optimized compounds, 91 compounds were selected based on the GOLD fitness score and interaction with essential amino acids. Optimization step has improved the binding and molecular interactions of the compounds. The binding energies of these compounds were calculated using AutoDock program to ensure the GOLD predictions. The synthetic accessibility of these optimized compounds was calculated using SYLVIA program. Combining the results from molecular docking, SYLVIA calculations and novelty search 10 compounds were listed as potent virtual leads for the designing of novel CEase inhibitors.

Acknowledgements

This research was supported by Basic Science Research Program (2009-0073267), Pioneer Research Center Program (2009-0081539), and Environmental Biotechnology National Core Research Center program (20090091489) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST). And all students were recipients of fellowship from the BK21 Program of MEST.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

1. Wang, C.S., Hartsuck, J.A. Bile salt-activated lipase. A multiple function lipolytic enzyme. Biochim. Biophys. Acta 1993, 1166, 1–19.
2. Lombardo, D. Bile salt-dependent lipase: its pathophysiological implications. Biochim. Biophys. Acta 2001, 1533, 1–28.
3. Wang, C.S., Martindale, M.E., King, M.M., Tang, J. Bile-salt-activated lipase: effect on kitten growth rate. Am. J. Clin. Nutr. 1989, 49, 457–463.
4. Howles, P.N., Stemmerman, G.N., Fenoglio-Preiser, C.M., Hui, D.Y. Carboxyl ester lipase activity in milk prevents fat-derived intestinal injury in neonatal mice. Am. J. Physiol. 1999, 277, G653–G661.
5. Howles, P.N., Carter, C.P., Hui, D.Y. Dietary free and esterified cholesterol absorption in cholesterol esterase (bile salt-stimulated lipase) gene-targeted mice. J. Biol. Chem. 1996, 271, 7196–7202.
6. Brodt-Eppley, J., White, P., Jenkins, S., Hui, D.Y. Plasma cholesterol esterase level is a determinant for an atherogenic lipoprotein profile in normolipidemic human subjects. Biochim. Biophys. Acta 1995, 1272, 69–72.
7. Fält, H., Hernell, O., Bläckberg, L. Does bile salt-stimulated lipase affect cholesterol uptake when bound to rat intestinal mucosa *in vitro*? Pediatr. Res. 2002, 52, 509–515.
8. Lin, G., Liao, W.C., Chiou, S.Y. Quantitative structure–activity relationships for the pre-steady-state inhibition of cholesterol esterase by 4-nitrophenyl-*N*-substituted carbamates. Bioorg. Med. Chem. 2000, 8, 2601–2607.
9. Pietsch, M., Gutschow, M. Alternate substrate inhibition of cholesterol esterase by thieno[2,3-*d*][1,3]oxazin-4-ones. J. Biol. Chem. 2002, 277, 24006–24013.
10. Terzyan, S., Wang, C.S., Downs, D., Hunter, B., Zhang, X.C. Crystal structure of the catalytic domain of human bile salt activated lipase. Protein Sci. 2000, 9, 1783–1790.
11. Lin, G., Shieh, C.T., Tsai, Y.C., Hwang, C.I., Lu, C.P., Chen, G.H. Structure–reactivity probes for active site shapes of cholesterol esterase by carbamate inhibitors. Biochim. Biophys. Acta 1999, 1431, 500–511.
12. Feaster, S.R., Quinn, D.M. Mechanism-based inhibitors of mammalian cholesterol esterase. Meth. Enzymol. 1997, 286, 231–252.
13. Feaster, S.R., Lee, K., Baker, N., Hui, D.Y., Quinn, D.M. Molecular recognition by cholesterol esterase of active site ligands: structure–reactivity effects for inhibition by aryl carbamates and subsequent carbamyl enzyme turnover. Biochemistry 1996, 35, 16723–16734.
14. Tew, D.G., Boyd, H.F., Ashman, S., Theobald, C., Leach, C.A. Mechanism of inhibition of LDL phospholipase A2 by monocyclic-beta-lactams. Burst kinetics and the effect of stereochemistry. Biochemistry 1998, 37, 10087–10093.
15. Chiou, S.Y., Lai, C.Y., Lin, L.Y., Lin, G. Probing stereoselective inhibition of the acyl binding site of cholesterol esterase with four diastereomers of 2′-*N*-alpha-methylbenzylcarbamyl-1,1′-bi-2-naphthol. BMC Biochem. 2005, 6, 17.
16. Krantz, A. A classification of enzyme inhibitors. Bioorg. Med. Chem. Lett. 1992, 2, 1327–1334.
17. Leung, D., Abbenante, G., Fairlie, D.P. Protease inhibitors: current status and future prospects. J. Med. Chem. 2000, 43, 305–341.
18. Martyn, D.C., Moore, M.J., Abell, A.D. Succinimide and saccharin-based enzyme-activated inhibitors of serine proteases. Curr. Pharm. Des. 1999, 5, 405–415.
19. Neumann, U., Gütschow, M. *N*-(Sulfonyloxy)phthalimides and analogues are potent inactivators of serine proteases. J. Biol. Chem. 1994, 269, 21561–21567.
20. Hlasta, D.J., Pagini, E.D. Human leukocyte elastase inhibitors. Annu. Rep. Med. Chem. 1994, 29, 195–204.
21. Powers, J.C., Odake, S., Oleksyszyn, J., Hori, H., Ueda, T., Boduszek, B., Kam, C. Proteases—structures, mechanism and inhibitors. Agents Actions Suppl. 1993, 42, 3–18.
22. Krantz, A., Spencer, R.W., Tam, T.F., Liak, T.J., Copp, L.J., Thomas, E.M., Rafferty, S.P. Design and synthesis of 4*H*-3,1-benzoxazin-4-ones as potent alternate substrate inhibitors of human leukocyte elastase. J. Med. Chem. 1990, 33, 464–479.
23. Deck, L.M., Baca, M.L., Salas, S.L., Hunsaker, L.A., Vander Jagt, D.L. 3-Alkyl-6-chloro-2-pyrones: selective inhibitors of pancreatic cholesterol esterase. J. Med. Chem. 1999, 42, 4250–4256.

24. Lin, G., Shieh, C.T., Ho, H.C., Chouhwang, J.Y., Lin, W.Y., Lu, C.P. Structure–reactivity relationships for the inhibition mechanism at the second alkyl-chain-binding site of cholesterol esterase and lipase. Biochemistry 1999, 38, 9971–9981.
25. Heynekamp, J.J., Hunsaker, L.A., Vander Jagt, T.A., Royer, R.E., Deck, L.M., Vander Jagt, D.L. Isocoumarin-based inhibitors of pancreatic cholesterol esterase. Bioorg. Med. Chem. 2008, 16, 5285–5294.
26. Bharatham, N., Bharatham, K., Lee, K.W. Pharmacophore identification and virtual screening for methionyl-tRNA synthetase inhibitors. J. Mol. Graph. Model. 2007, 25, 813–823.
27. Pietsch, M., Gütschow, M. Synthesis of tricyclic 1,3-oxazin-4-ones and kinetic analysis of cholesterol esterase and acetylcholinesterase inhibition. J. Med. Chem. 2005, 48, 8270–8288.
28. Sakai, K., Watanabe, K., Masuda, K., Tsuji, M., Hasumi, K., Endo, A. Isolation, characterization and biological activities of novel triprenyl phenols as pancreatic cholesterol esterase inhibitors produced by *Stachybotrys* sp. F-1839. J. Antibiot. 1995, 48, 447–456.
29. Eilfeld, A., González Tanarro, C.M., Frizler, M., Sieler, J., Schulze, B., Gütschow, M. Synthesis and elastase-inhibiting activity of 2-pyridinyl-isothiazol-3(2*H*)-one 1,1-dioxides. Bioorg. Med. Chem. 2008, 16, 8127–8135.
30. Guner, O.F., Henry, D.R. Phamacophore Perception, Development, and Use in Drug Design. La Jolla, CA: International University Line, 2000, pp. 193–210.
31. Thangapandian, S., John, S., Sakkiah, S., Lee, K.W. Ligand and structure based pharmacophore modeling to facilitate novel histone deacetylase 8 inhibitor design. Eur. J. Med. Chem. 2010, 45, 4409–4417.
32. Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 2001, 46, 3–26.
33. Walters, W.P., Murcko, M.A. Prediction of 'drug-likeness.' Adv. Drug Deliv. Rev. 2002, 54, 255–271.
34. John, S., Thangapandian, S., Sakkiah, S., Lee, K.W. Identification of potent virtual leads to design novel indoleamine 2,3-dioxygenase inhibitors: pharmacophore modeling and molecular docking studies. Eur. J. Med. Chem. 2010, 45, 4004–4012.
35. Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 1997, 267, 727–748.
36. Manoj, K.A., Krishna, K.I., Lalitha, G. Docking of phosphonate and trehalose analog inhibitors into *M. tuberculosis* mycolyltransferase Ag85C: comparison of the two scoring fitness functions GoldScore and ChemScore, in the GOLD software. Bioinformation 2007, 1, 339–350.
37. Marcel, L.V., Jason, C.C., Michael, J.H., Christopher, W.M., Richard, D.T. Improved protein–ligand docking using GOLD. Proteins: Struct. Funct. Genet. 2003, 52, 609–623.
38. Wallace, A.C., Laskowski, R.A., Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. Protein Eng. 1995, 8, 127–134.
39. Böhm, H.J. The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. J. Comput. Aided Mol. Des. 1994, 8, 243–256.
40. Krisztina, B., Thomas, S., Johann, G. Structure and reaction based evaluation of synthetic accessibility. J. Comput. Aided Mol. Des. 2007, 21, 311–325.
41. Andrea, Z., Krisztina, B., Thomas, S., Achim, H., Christof, H.S., Johann, G., Holger, C., Christian, L., Jorg, D., Juri, P., Matthias, R. Second-generation *de novo* design: a view from a medicinal chemist perspective. J. Comput. Aided Mol. Des. 2009, 23, 593–602.
42. Sheng, Y.Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov. Today 2010, 15, 444–450.
43. Wagner, A.B. SciFinder Scholar 2006: an empirical analysis of research topic query processing. J. Chem. Inf. Model. 2006, 46, 767–774.
44. Wang, Y., Bolton, E., Dracheva, S., Karapetyan, K., Shoemaker, B.A., Suzek, T.O., Wang, J., Xiao, J., Zhang, J., Bryant, S.H. An overview of the PubChem BioAssay resource. Nucleic Acids Res. 2010, 38, D255–D266.