Hybrid Structure-Based Virtual Screening Protocol for the Identification of Novel BACE1 Inhibitors

R. S. K Vijayan, M. Prabu, Nahren Manuel Mascarenhas, and Nanda Ghoshal*

Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology (A unit of CSIR), Kolkata 700032, India

Received October 17, 2008

BACE1, also called β-secretase or memapsin 2, is an extensively studied aspartic protease, involved in etiopathogenesis and progression of Alzheimer's disease (AD). We report herein a modified structure-based virtual screening protocol that augments the lead identification process against BACE1 during virtual screening endeavors. A hybrid structure-based virtual screening protocol that incorporates elements from both ligand-based and structure-based techniques was used for the identification of prospective small molecule inhibitors. Virtual screening, using an active-site-derived pharmacophore, followed by ROCS (rapid overlay of chemical structures)-based GOLD (genetic optimization in ligand docking) docking was used to identify a library of focused candidates. The efficacy of the ROCS-based GOLD docking method together with our customized weighted consensus scoring function was evaluated against conventional docking methods for its ability to discern true positives from a screening library. An in-depth structural analysis of the binding mode of the top-ranking molecules reveals that emulation of the curial interaction patterns deemed necessary for BACE1 inhibition. The results obtained from our validation study ensure the superiority of our docking methodology over conventional docking methods in yielding higher enrichment rates.

INTRODUCTION

Alzheimer's disease (AD) is a common age-related neurodegenerative disorder, first described by Alois Alzheimer[1] in 1906, that continues to increase worldwide. A major histopathological manifestation of Alzheimer's disease (AD) involves the accumulation of extra cellular β-amyloid (Aβ) neuritic plaques.[2] One of the two isoforms of proteases termed BACE1, involved in the cleavage of β amyloid precursor protein (APP), plays a central role in the etiology of Alzheimer's disease.[3,4] Reducing the production of Aβ or increasing its clearance offers an axiomatic strategy for the treatment of Alzheimer's disease.[5] Hence inhibition of the enzyme BACE1 offers a tractable target for drug development that would aid in the management and delay the progression of AD. Though considerable attention has been paid toward the development of aspartic protease inhibitors, the development of inhibitors targeting proteases of human origin is challenging because BACE1 is a cellular protein and presumably has some normal physiological functions. Further the existence of other homologous aspartlyl proteases like BACE2 and a ubiquitous protein termed cathepsin D obviously reminds one of the need to consider selectivity aspects as a part of the investigation. Therefore, identifying selective nonpeptidic BACE1 inhibitors with CNS penetration and good pharmacokinetic properties would be demanding.[6] Working toward this end, we carried out a hybrid structure-based virtual screening program that could provide greater strides for the identification of novel non-peptidic BACE1 inhibitors.

A drug discovery project conducted by Merck research laboratories and NeoGenesis pharmaceuticals identified a series of aminopentyl oxacetamides as potent and selective inhibitors.[7] The published crystal structure of BACE1 in complex with an aminopentyl oxacetamide (PDB ID 1TQF) served as a starting point to pursue our structure-based drug design program. This cocrystallized complex was chosen because it reveals a binding mode unconventional to aspartyl proteases and more importantly it characterizes the nontransition state of the enzyme complex. To begin with, an active-site-directed pharmacophore model was generated on the basis of the geometric disposition of the protein-inhibitor complex using the program LigandScout.[8] Hits identified by pharmacophores generated through direct approach are more reliable because they impose the necessary constraints required for interaction and selectivity.

In the next phase, scaffold hopping was performed against multiconformational databases using pharmacophore-derived queries. Molecular docking, a widely used approach in virtual screening was used for post processing hits that lack complementarities with the target protein outside the pharmacophoric definition.[9] Though the concept of protein−ligand docking brought about a paradigm shift in early stage drug discovery, it is apparent that there are certain lacunas that impede the identification of high quality leads during a virtual screen.[10] High quality lead identification is an important prerequisite in drug discovery research; hence optimization of docking applications represents the next step in attaining better enrichment rates during virtual screening. While other efforts center on addressing issues like receptor flexibility and development of improved scoring functions for protein−

* To whom correspondence should be addressed. Phone: +91-33-2473-3491 ext. 254/236. Fax: +91-33-2473-0284/5197. E-mail: nghoshal@iicb.res.in. Address: Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology, 4 Raja S.C. Mullick Road, Jadavpur, Kolkata 700032, India.

10.1021/ci800386v CCC: $40.75
Published on Web 02/13/2009

ligand docking, we decided to focus on an alternative philosophy that has the potential for enrichment during virtual screening. Generally structure-based virtual screening (SBVS) can be carried out using the protein coordinates (docking) or the ligand coordinates of the bioactive conformation (shape-based comparison).[11] Both the methods stated above have independently demonstrated their strength in virtual screening. Hence, to capitalize the benefit of these two independent techniques, we decided to devise a modified docking methodology that encodes the structural and similarity attributes of the bound ligand and the protein to enhance hit rate in a virtual screening program and move beyond the "protein structural information centric view", currently employed in docking and scoring. ROCS (Rapid Overlay of Chemical Structures),[12] a ligand-centric shape-based program used to identify conformations that overlap in shape and chemistry with respect to the reference ligand. ROCS-optimized conformations were rigidly docked in to the active site using GOLD (Genetic Optimization in Ligand Docking) docking[13] software. The steric and chemical complementarity of the pregenerated conformations for the protein active site was quantified using different scoring functions. Although a myriad of scoring functions has been developed, a reliable scoring function still proves to be elusive.[14] A usual approach employed to compensate the imperfections of individual scoring functions is to employ consensus scoring schemes.[14] Generally, scoring functions are biased toward high molecular weight (MW) compounds[15] because most scoring functions treats Van der Waals (VDW) interaction energy as the sum over all pairs of ligand and target protein atoms within a specified cutoff distance. This assumption leads to a significant size-dependent bias in the scoring scheme. Hence, a modified consensus scoring scheme, which effectively handles the bias of the score toward molecular weight, was incorporated in the study. Traditionally, scoring functions are aimed at providing an estimate of the binding strength and rarely penalize unintended binding modes.[16] In the course of deriving a penalty function, a weighted term was incorporated for penalizing binding modes that deviate from what is to be expected. As a proof of concept, the said protocol was extensively validated for its enrichment ability using receiver operator characteristic (ROC) curves and enrichment factor plots.

A comparative assessment was also carried out against the routine docking methods implemented in some highly regarded commercial packages like Ligandfit[17] and GOLD.[13] The results obtained from our preliminary study suggest that incorporation of ligand-based information in docking provides alternate avenues to attain better enrichment rates. The approach implemented in the study is general and transferable to any protein system in complex with inhibitors. The top ranked hits produced by weighted C_{score} also retain the requisite binding mode in terms of hydrogen-bonding pattern with the catalytic aspartic diad[18] and further stabilized by hydrophobic interactions with the adjacent subpockets. Because there is no unique solution to a drug design problem, we emphasize that a knowledge-driven approach that seamlessly integrates structure based and ligand based aspects in docking holds promise to augment and accelerate early stage drug discovery programs.

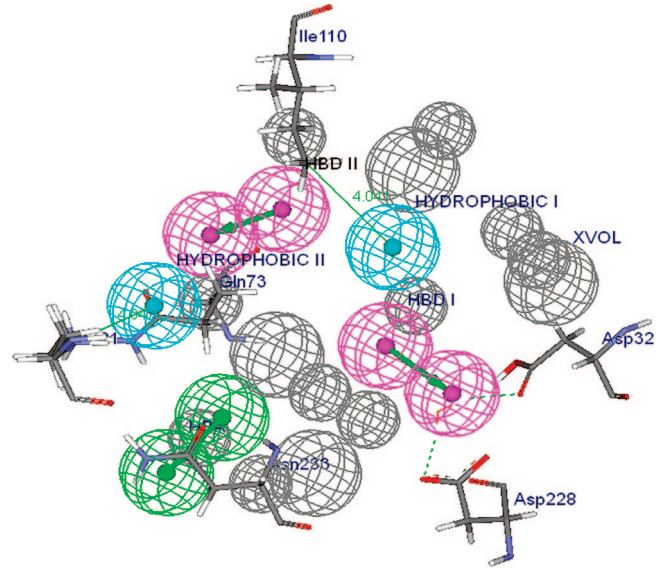


Figure 1. Structure-based pharmacophore generated using LigandScout showing the pharmacophoric features and their corresponding amino acids involved in interaction. A few excluded volume spheres, shown in gray, have been hidden for clarity.

MATERIALS AND METHODS

The computational studies outlined here were performed using the following software packages. Discovery studio 2.0,[19] GOLD 3.2,[20] ROCS 2.3.1,[21] LigandScout 1.03,[22] and TSAR 3.0[23] were ran on a Pentium 4 core2 Duo workstation using a Windows XP operating system. Catalyst 4.1,[24] Cerius^2 4.10,[25] and Insight II[26] studies were performed on SGI Fuel workstations running on IRIX 6.5 operating system.

Pharmacophore-Based Scaffold Hopping. LigandScout,[8] a tool that constructs structure-based pharmacophore model, was used to translate the structural data of macromolecule/ligand complexes (1TQF)[7] into a 3D pharmacophoric model. Excluded volume (EV) spheres that define areas, which are sterically forbidden, were automatically added to the pharmacophore model.[27] Inclusion of EV to the pharmacophore models helps in pruning the hits, improves enrichment rate, and also imparts target selectivity implicitly. The pharmacophore model generated by LigandScout was translated by a script into a Catalyst compatible pharmacophore model and exported to Catalyst. The structure-based pharmacophore model that incorporates features from both the ligand and the protein was queried across Maybridge database supplied with Catalyst (56 000 compounds) and the downloaded LeadQuest database (50 000 compounds) to identify compounds with newer scaffolds similar to known *active principle*. Initial filtering of the hits was based on Lipinsiki's druglikeness principle[28] (MW $<$ 500, Log P $<$ 5, HBA $\leq$ 10, HBD $\leq$ 5), and those hits that satisfy at least three of the said parameters were subsequently considered for further studies.

Preparation of Ligand and Protein System for Docking. The necessity of providing a correct representation of the ligand is a prerequisite for carrying out a successful docking study. Hence to ensure quality, the following aspects were considered. The initial 3D structure was generated using CORINA;[29] charge calculation and further energy optimization were carried out using the COSMIC module of TSAR.[23] The assignment of ionization states, removal of salts, and

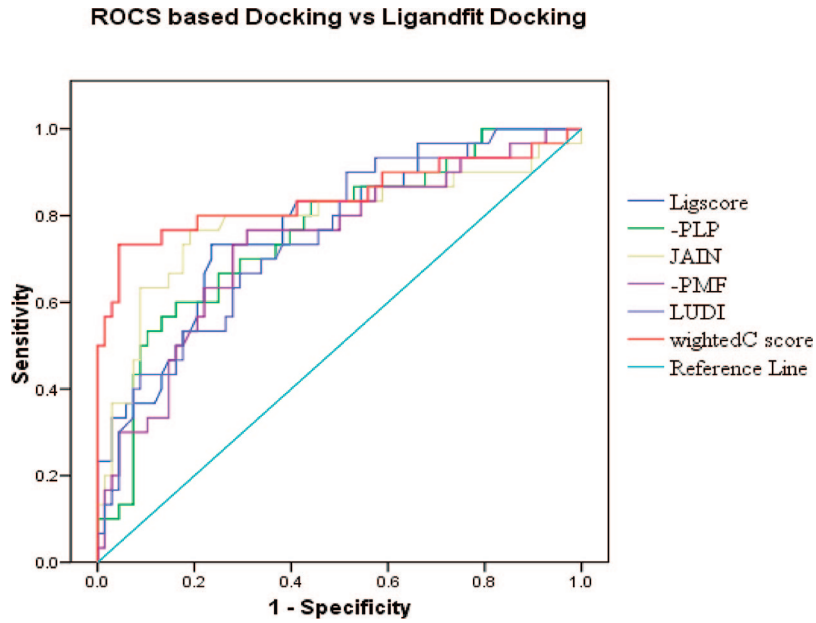


Figure 2. ROC curves showing the tradeoff between sensitivity and specificity calculated using SPSS. The scores are color-coded and the reference line indicates random prediction.

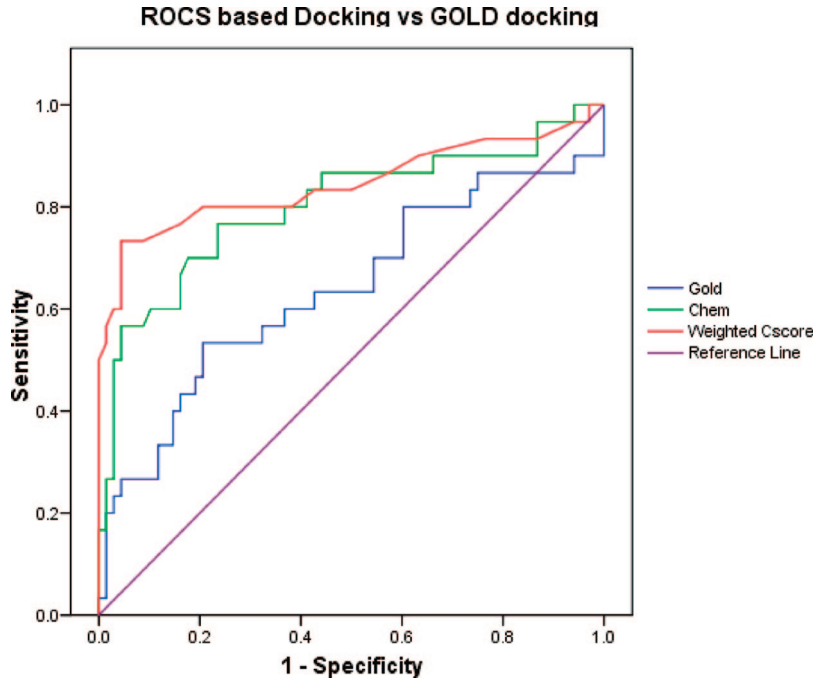


Figure 3. ROC curve plotted for ROCS-based docking versus GOLD generated using SPSS. The scores are color-coded and the reference line indicates random prediction.

Table 1. Distance Matrix Plot Showing the Interfeature Distance between the Pharmacophoric Points

distance matrix (Å)	hydrophobic 1	hydrophobic 2	HBD1	HBD2	HBA
hydrophobic 1	0	8.158	5.873	5.699	8.244
hydrophobic 2	8.158	0	10.159	8.251	4.468
HBD1	5.873	10.159	0	5.81	8.727
HBD2	5.699	8.251	5.81	0	9.701
HBA	8.244	4.468	8.727	9.628	0

removal of duplicate structures were performed using the "prepare ligand" module of Discovery Studio. Protein preparation was carried out using the BIOPOLYMER module of Insight II. The assignment of protonation states are very crucial because the definitions of H-bond donor/acceptors are bound to change if appropriate protonation states are not assigned correctly. Hence hydrogens were added assuming a pH of 4.5 reported to be optimal for BACE1 activity.[18] The side chain of the catalytic residues Asp 228 and Asp 32 were kept neutralized and ionized, respectively.[30] Hydrogens, added to the protein system, were relaxed keeping the rest of the system static, using 2000 steps

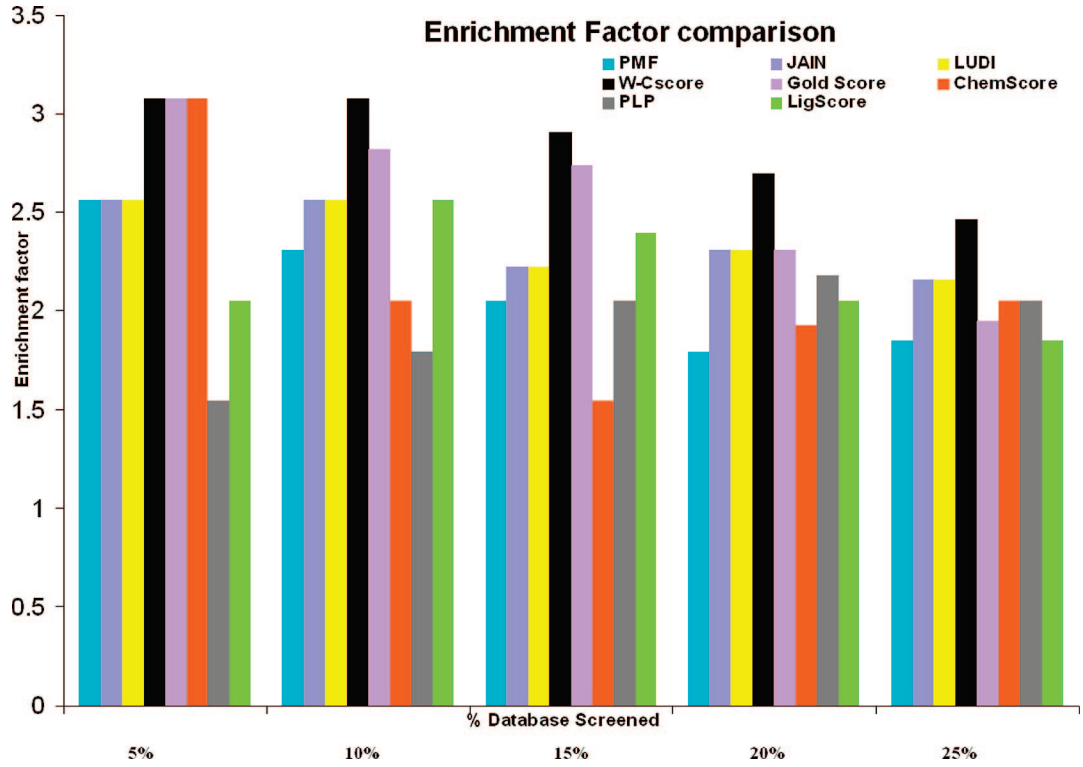


Figure 4. Enrichment factor plot obtained for ROCS-based GOLD docking is shown in context with other molecular docking methods together with their respective scoring functions.

Table 2. Enrichment Factor Obtained for Different Docking Methods and Their Scoring Functions at Various Levels of Percentage Database Screened

% database screened	LigandFit −PMF	LigandFit Jain	LigandFit LUDI	LigandFit −PLP	LigandFit Ligscore	GOLD GoldScore	GOLD Chemscore	ROCS weighted C_{score}
5	2.56	2.56	2.56	1.53	2.05	3.07	3.07	3.07
10	2.30	2.56	2.56	1.79	2.56	2.82	2.05	3.07
15	2.05	2.22	2.22	2.05	2.39	2.73	1.53	2.90
20	1.79	2.30	2.30	2.18	2.05	2.30	1.92	2.69
25	1.84	2.15	2.15	2.05	1.84	1.94	2.05	2.46

of conjugate gradient method. This was performed to remove bad contacts and to improve the hydrogen bond geometries because the locations of hydrogen are not specified in X-ray structures.

ROCS-Based GOLD Docking. To maximize the benefits of the structural and chemical information in hand, we decided to use ROCS, a 3D shape superposition program, to guide our docking protocol.[31] ROCS require an input of pregenerated conformers. OpenEyes native conformer generation program OMEGA was not considered because ligand-based conformation generation programs work on a narrow energy window range and are more likely to overlook bioactive conformations, which in many instances does not correspond even to a local minimum.[32] Consequently, we decided to generate conformational ensembles on-the-fly using docking tools which seek to identify a restricted number of conformations in the context of the active site. To ensure adequate conformational coverage, two stochastic search algorithms, namely, Monte Carlo (MC) and Genetic algorithm (GA), implemented in Ligandfit[17] and GOLD,[13] respectively, were employed as conformation generation tools. The top 20 conformations generated by the respective software's in turn provided a pool of 40 conformers for each compound. This pregenerated pool of conformations was processed using ROCS to approximate the binding mode of the reference compound, generally termed as "active analog approach". Although this approach sounds to be prohibitive, employing it in the second phase for post processing the limited number of hits generally retrieved by structure-based pharmacophores would be probable.

A shape-based overlay process of the pregenerated conformers, optimized with respect to chemical features as defined by Mills and Dean force field,[33] was carried out on the reference compound. The conformations provided by ROCS were ranked on the basis of the combo score function, a sum of scaled color score and shape Tanimoto score. The best conformation thus obtained was rigidly docked using GOLD.[12] The binding pocket was defined using the crystallographic coordinates of the bound ligand (residues within 8 Å from the ligand). GOLD was run in the "Rigid-docking mode" comprising 100 000 genetic operations on an initial population of 100 members divided into five subpopulations (number of islands = 5). The annealing parameters of fitness function were set at 4.0 for van der Waals and 2.5 for hydrogen bonding. A niche size of 2 and a selection pressure of 1.1 were used. The early termination option was turned off, and rescoring was performed using the ChemScore function. The docked poses thus obtained were also rescored using the scoring functions implemented in LigandFit.

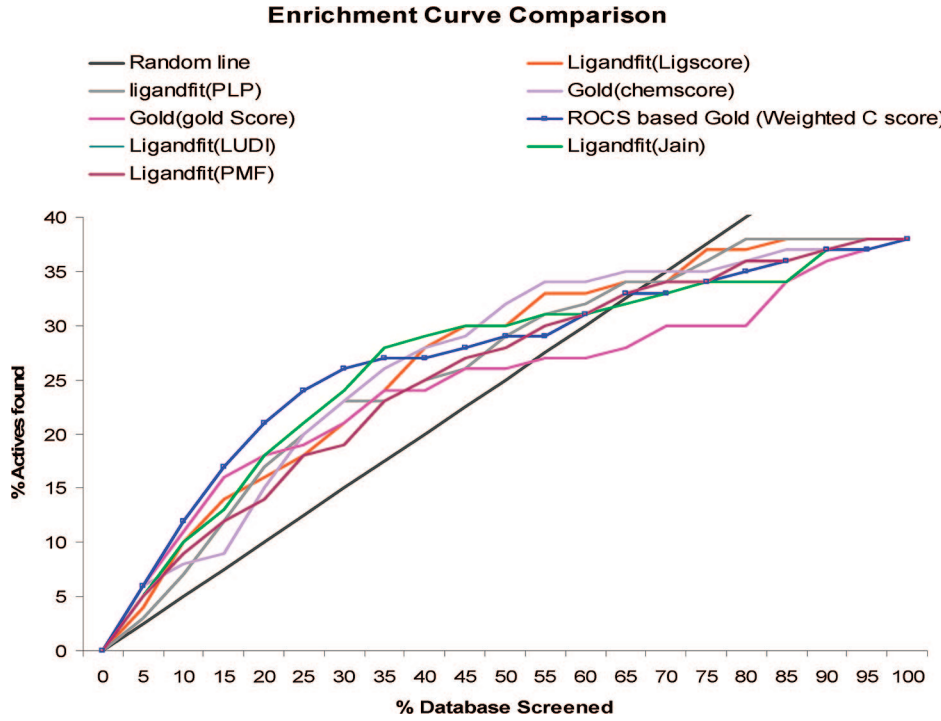


Figure 5. Comparison of enrichment curves obtained from different docking methods plotted at various levels of database screened in relation to the number of actives obtained.

Since each scoring function evaluates the protein ligand complex from its own perspective and uses various assumptions and simplifications for scoring, an idealized approach would be to use a consensus scoring scheme.

A modified consensus-scoring scheme that combines information from different scoring schemes to balance errors and a normalization factor to handle molecular weight biasness was pursued to improve the probability of identifying "true" hits. Large molecules tend to form many hypothetical interactions in binding sites and are prone to generate better scores than smaller compounds; hence a standardization procedure to normalize the bias resulting from molecular weight[34] was carried out using the formula

$$\text{standardized score} = \frac{\text{raw score}}{N^{1/2}} \quad (1)$$

where N represents the number of heavy atoms in the molecule. The standardized scores were further scaled to unit variance using average of auto scaled scores (AASS) method[33] to enable different scores to be used simultaneously regardless of the difference between them. Scaling was performed using the following formula:

$$\text{scaled}_{\text{score}} = (X - X_{\text{min}})/(X_{\text{max}} - X_{\text{min}}) \quad (2)$$

Scoring functions representing diverse classes, namely, knowledge-based (PMF[35]), force-field-based (Ligscore,[17] GoldScore[13]), and empirical (PLP,[36] Jain,[37] LUDI,[38] ChemScore[39]), were incorporated to maintain a subtle balance.

Therefore the consensus score for each pose was calculated using the formula

$$\text{consensus score } (C_{\text{score}}) = \sum \text{scaled}_{\text{scores}}/N \quad (3)$$

where, N represents the numbers of scoring functions used for the generation of C_{score}.

Consensus score calculations were performed in an automated fashion using an in-house program developed during the course of the work. Further, a similarity weighed function was incorporated that implicitly penalizes conformations that deviate from the ideal binding mode of the reference ligand.[40]

$$\text{weighted } C_{\text{score}} = C_{\text{score}} \times \text{Tanimoto score} \quad (4)$$

Where Tanimoto score (T_{score}) represents the Gaussian shape overlap (O) between two molecules A and B calculated using a similarity metric function.[31]

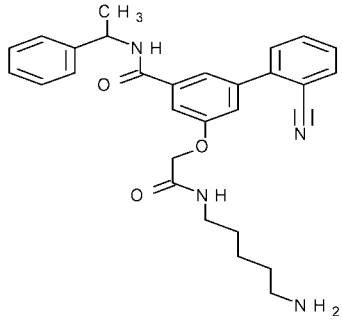
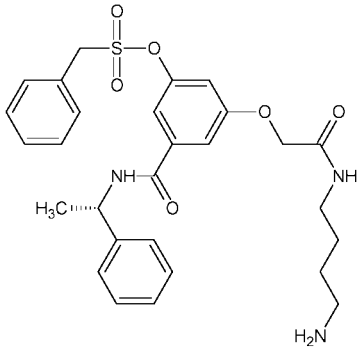
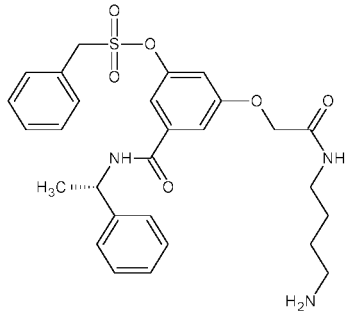
$$T_{\text{A,B}} = O_{\text{A,B}}/(O_{\text{A,A}} + O_{\text{B,B}} - O_{\text{A,B}}) \quad (5)$$

The weighted C_{score} will remain unaltered if T_{score} equals one, which is possible only when the reference ligand and the target ligand perfectly superimpose. When T_{score} decreases from the ideal value of 1, the weighted C_{score} will be penalized proportionally to the deviation observed from the ideal binding mode as defined by the reference ligand.

RESULTS AND DISCUSSION

Pharmacophore Model. The automated pharmacophore model generated by LigandScout was too restrictive in retrieving hits because it was over defined. Information gleaned from structure activity relationship (SAR) and the crystal bound structure of oxaacetamides was used to condense the pharmacophore model. The structure-based pharmacophore model developed in this study significantly differs from the Vertex Pharmacophore model made available through a patent. Our pharmacophore model has two hydrogen bond donor (HBD) features, two hyrodrophobic and one hydrogen bond acceptor (HBA) feature. The excluded volume spheres presented in our model provide an insight regarding the disallowed regions in the binding site. In an attempt to account protein flexibility and reorganization effects at the pharmacophore level, the size of the excluded volume was decreased by 0.5 Å to increase the

Table 3. Tanimoto Shape Overlap (T_{Shape}) Values Obtained for Oxacetamide Series with Respect to the Crystal Structure of 1TQF

Compound	Stereochemistry	ROCS based Gold docking	Gold	Ligand Fit
	R	0.521	0.445	0.228
	R	0.614	0.60	0.261
	S	0.62	0.454	0.176

effective size of the binding cavity. Pharmacophoric feature HBD (hydrogen bond donor) signifies the water (HOH35) mediated hydrogen bonding interactions with catalytic residues Asp 32 and Asp 228. "Hydrophobic I" and "Hydrophobic II" features reflecting the hydrophobic interaction with the side chain of Ile118 and Lys321, respectively.

The pharmacophoric features and the corresponding amino acids involved in interactions are shown in Figure 1, and their interfeature distance relationship is given in Table 1.

Validation of the screening protocol. In silico validation of our screening protocol was performed, beforehand, to ensure its ability to discern true positives. Enrichment calculations were carried out on a screening library consisting of known actives and decoys. Known actives with experimentally reported K_i and IC_{50} were pooled from BindingDB.[41] A decoy set (presumed nonbinders) was prepared in a fashion similar to the DUD decoy set using the Library Analysis module of Discovery Studio[38] because compound compilation used in a screening library can have profound effects on enrichment studies. Compounds having comparable physicochemical properties to the known inhibitor were retrieved from the NCI compound library[42] using Euclidean distance as similarity metric. This process ensures that enrichment calculations carried out from such data sets to be free of bias and accurate.[43] For validation purposes, our virtual screening strategy was compared with the standard flexible docking protocol employed in LigandFit and GOLD.

There are a number of approaches to gauge the success of a virtual screening tool. One commonly used measure that assesses the performance of a virtual screening strategy is

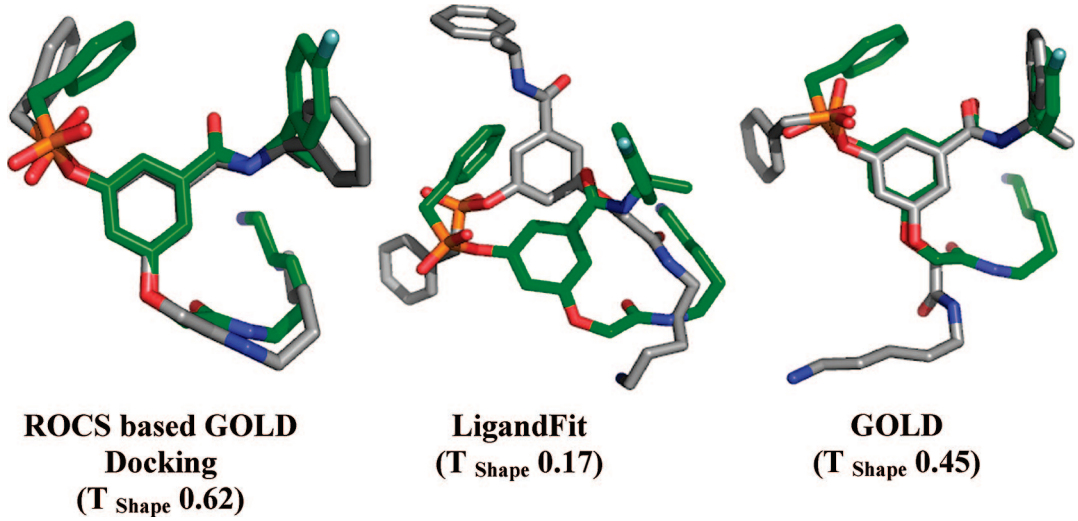


Figure 6. The docked conformation of an oxacetamide member overlaid to the reference structure shown in green. Image rendered using Pymol.

Table 4. Conformational Energy Landscape Analysis

	Omega conformational energy (Kcal/mol)		Catalyst/best conformational energy (Kcal/mol)		ROCS conformational energy (Kcal/mol)	allowed energy window range (Kcal/mol)
compound ID	lowest energy	highest energy	lowest energy	highest energy		
KM01595	168.36	182.23	148.73	203.34	192.68	148−203
KM01602	167.32	176.15	149.61	190.17	174.96	149−190
KM03141	127.13	134.97	133.89	137.61	130.39	127−137
HTS02812	111.97	149.70	88.9	107.10	120.47	88−149
RH01508	108.49	114.50	95.67	97.12	96.52	95−114

the receiver operating characteristic (ROC) curve.[44] The ROC curves illustrate the performance of a given method across a benchmark data set seeded with both actives (38) and decoys (79). ROC curves are quite insensitive to the number of actives present in the screening library. A theoretically perfect performance would correspond to an area under the curve (AUC) value of 1.0, while a random performance gives an AUC value of 0.5. AUC values less than 0.5 represents an unfavorable case with a systematic ranking of decoys higher than the rankings of known actives. AUC values signify the discriminative capacity of the protocol, when 1-specifity (false positives) is plotted as a function against selectivity (true positives).

The performance of LigandFit, together with its scoring functions, was evaluated against our ROCS-based GOLD docking approach method. Flexible docking in LigandFit was carried out using the options (N_{max} trials = 15 000, N_{save} = 10, distance dependent dielectric constant = 3r soft potential with $\alpha = 1$ and $\beta = 0.05$; improved interpolation with m = n = 2; grid resolution of 0.5 Å, CFF force field) and was scored using its inbuilt scoring functions. The top ranking conformer as stated by Dock Score was considered in the study. The ROC curve obtained from the study is shown in Figure 2.

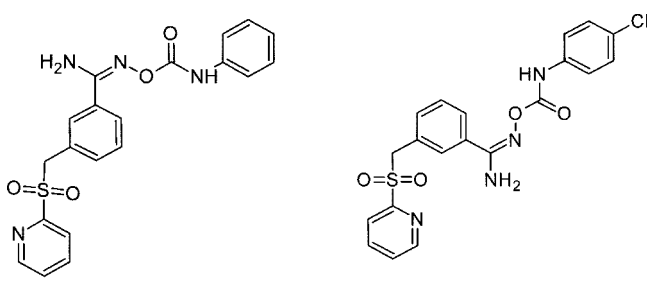


Figure 7. 2D Structure of the probable hits obtained from our screening program.

The obtained ROC curve indicates that our ROCS-based GOLD docking protocol, together with the weighted C_{score} function, performed better (AUC value of 0.843) when compared to LigandFit and its individual scoring components. Among the scoring functions implemented in LigandFit, it was found that the empirical scoring function, Jain score (AUC value of 0.788) performed better followed by Ligscore 0.778, −PLP 0.760, LUDI 0.750, and −PMF with a value of 0.736. Validation studies against GOLD flexible docking were carried out using the settings as stated earlier. Initial scoring was based on GoldScore, followed by rescoring using ChemScore. The ROC plot obtained as shown in Figure 3 reveals that the ROCS-based GOLD docking strategy, together with the weighted C_{score}, performs superior with AUC value of 0.845, followed closely behind by ChemScore fitness function (AUC 0.801), whereas the GOLD fitness score performed fairly with an AUC value of 0.637.

The observation that ChemScore performs better than GoldScore for druglike molecules is consistent with literature reports. The ROC method only makes a binary distinction between the predicted actives and inactives based on a threshold that generally performs well in many cases. Therefore, it is imperative to employ an enrichment measure that addresses the "early recognition problem" of virtual screening by quantifying the number of actives retrieved at the beginning of a rank-ordered list. Hence, the enrichment

Table 5. Examples of Hits with Their Compound ID and Their Respective Score As Obtained from Our Virtual Screening Program[a]

S.NO	compound ID	Goldscore	Chemscore	PLP	Ligscore	Jain	PMF	LUDI	weighted C_{Score}
1	AW00880	53.159	9.398	75.095	2.255	2.86	95.22	459.33	0.075
2	BTB01218	59.315	14.191	95.83	3.82	4.26	79.46	454.33	0.255
3	DP00785	62.381	15.578	79.355	3.205	3.76	40.09	296.66	0.070
4	DSHS00780	51.445	12.374	69.035	1.37	2.94	65.14	391	0.013
5	HTS02812	66.099	25.002	91.455	3.68	4.36	80.35	404	0.270
6	HTS05289	53.005	9.214	53.68	0.19	1.86	47.59	421	0.064
7	HTS06652	53.002	10.668	55.595	2.025	1.62	48.39	332.66	0.081
8	HTS10500	69.333	16.084	65.555	2.595	5.18	95.6	598.33	0.199
9	**KM01595**	**76.608**	**20.303**	**110.285**	**5.03**	**5.23**	**82.94**	**566.33**	**0.412**
10	KM03141	72.535	11.158	106.135	3.315	4.84	68.99	561.33	0.278
11	**KM01602**	**79.32**	**19.791**	**104.62**	**4.21**	**5.43**	**68.92**	**609.66**	**0.402**
12	RH 01508	57.033	20.853	91.66	3.865	2.66	78.8	568	0.258
13	SEW03010	57.582	17.884	93.895	1.455	5.38	90.82	538	0.120

[a] The most probable hits are shown in **bold**.

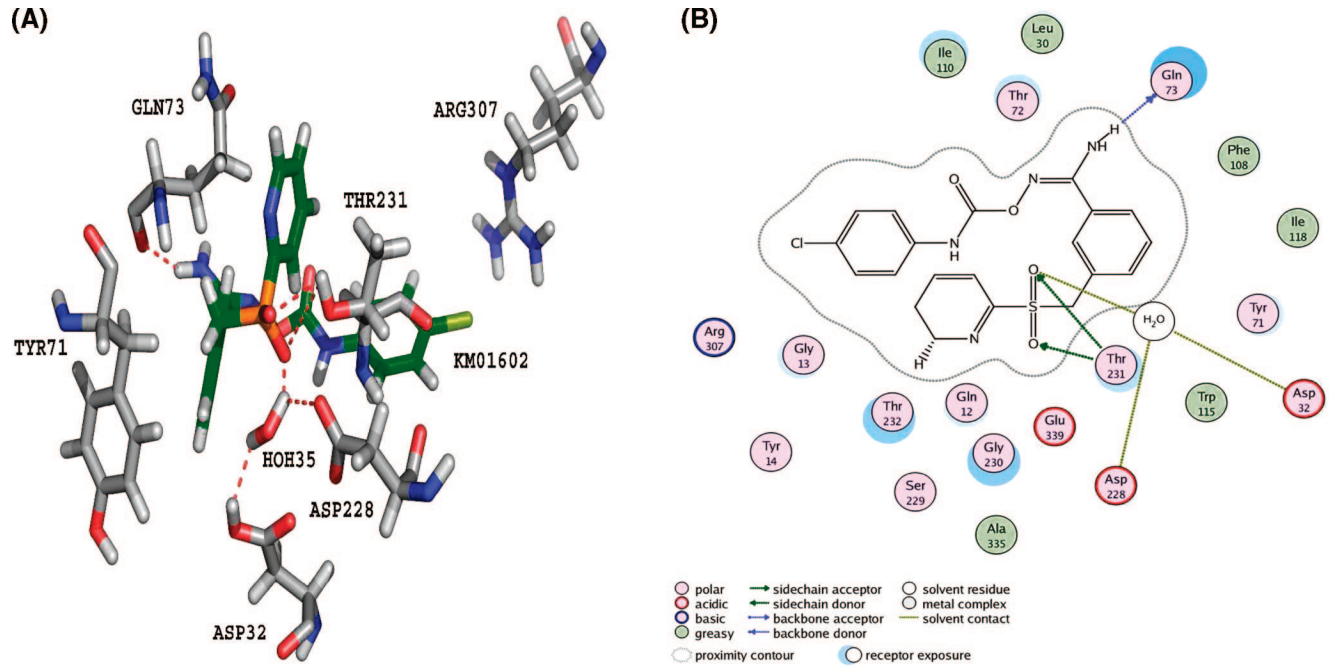


Figure 8. (A) Docked conformation revealing the binding mode of KM01602. Amino acids involved in nonbonded interaction are labeled. (B) 2D plot of the ligand interactions generated using MOE is shown together with the graphical key.

factor (EF), which reveals the success at a fixed percentage of database screened, was calculated using the formula

$$\text{EF} = \frac{\text{hits}_{\text{sampled}}^{x\%}}{N_{\text{sampled}}^{x\%}} \times \frac{N_{\text{total}}}{\text{hits}_{\text{total}}} \qquad (6)$$

where EF denotes the enrichment factor, $\text{hits}_{\text{sampled}}^{x\%}$ is the number of hits found at $x\%$ of the database screened, $N_{\text{sampled}}^{x\%}$ is the number of compounds sampled at $x\%$ of the database, $\text{hits}_{\text{total}}$ is the number of known active compounds in the entire database, and N_{total} is the number of compounds in the entire database. This method is sensitive to the number of known actives present in the database. The enrichment factors obtained are summarized in Table 2 and presented graphically in Figure 4, and the enrichment curve is shown in Figure 5.

GOLD, in conjunction with its scoring function Goldscore and Chemscore, yields comparable enrichment factors at 5% of the database screened, but it appears to result in degradation of performance when enrichment factors were assessed at 10%, 15%, 20%, and 25% of the database screened. Our docking methodology, together with weighted C_{score}, proves to yield enrichments that are consistently superior to other alternative methods. The enrichment factor obtained from Ligandfit, together with its scoring function (−PLP, −PMF, Jain, Ludi, and Ligscore), yielded modest EF values in the current study.

The enrichment curve obtained from our study was also compared with that of the GOLD and Ligandfit docking method. The weighed C_{score} used in our ROCS-based GOLD docking strategy is closest to the upper left corner (shown in blue in Figure 5) in the beginning of an ordered list. This implies that our method addresses the "early recognition" problem of virtual screening better than GOLD and Ligandfit. Therefore, it becomes more obvious that the hits obtained by our docking strategy are more probable to be true positives based on the current line of investigations.

The ability of our docking methodology in producing reliable binding modes was also validated. Self-docking experiments, the defacto standard used for validating the accuracy of a docking method in reproducing the bound conformation of a ligand by docking it back into the cocrystallized protein, cannot be used as measure of success

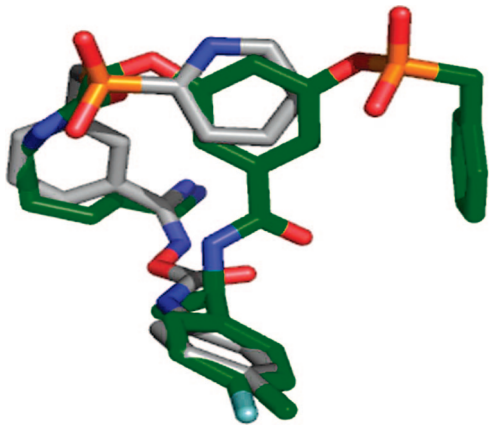


Figure 9. Probable hit KM01602 obtained in our study overlapped to the X-ray bound reference ligand ($T_{shape} = 0.562$). The reference ligand is shown in green.

in the current context. Because the shape-based overlay method was carried out using the native ligand structure, rmsd metrics obtained by redocking the same native ligand could obfuscate the validation process. To overcome this issue, a set of cognate ligands from the oxacetamide series, whose binding modes were not evident from crystal structures,[7] was used for validation (structures shown in Table 3). This presupposes the fact that similar ligands binds in a similar fashion; hence the T_{Shape} overlap with reference to the crystal bound structure was used as a metric for validating binding modes. The T_{Shape} overlap was calculated in ROCS using the shapeonly and scoreonly flags to ensure that the coordinates of the input structures remain unaltered during calculation. The Tanimoto shape overlap values, as evident from Table 3, endorse the ability of our ROCS-based docking method in predicting reliable binding modes.

The results obtained from the validation study clearly indicated that the search space of the ligand when restrained using ROCS presents an easier problem for the sampling algorithms incorporated in docking tools, and hence, they could overcome soft errors encountered in docking. It should be noted that GOLD was also able to reproduce near and comparable binding modes to our method. Though it was able to position the conformationally constrained regions of the ligands at the appropriate places, the highly flexible amino pentyl part of the ligand was not positioned properly. The performance of LigandFit was not satisfactory in producing reliable binding modes. It is imperative to note that LigandFit and GOLD were run using default routines as stated earlier to reflect everyday usage, hence exploration of different parameters to obtain better poses was not considered. The docked conformations thus obtained could be biased toward our assigned parameters. The binding mode obtained from ROCS-based GOLD docking, LigandFit, and GOLD for the cognate ligand superimposed on to the crystal bound inhibitor is shown in Figure 6.

This clearly indicates that the ROCS-based docking approach that constrains the search space by incorporating a value -added restrain, followed by rigid docking in GOLD performs superior when compared to LigandFit and GOLD when used in their default routines. ROCS, in principle, neglects the protein environment when it tries to predict a bioactive conformation, hence docking the obtained conformation rigidly in the context of the active site enhances the pose sampling stage of the docking engine. Mutually compensatory roles of these two independent tools account for its ability to attain better enrichment values. One possible shortcoming from such a type of screening could be the decreased diversity of the hits obtained. This issue could be overcome in cases where multiple bound ligands are available. The hits obtained by initial screening could be clustered in to families based on multiple reference structures and a "family-based docking" using the respective reference structure could help to identify high ranking member from each family.[45]

Binding Mode Analysis of Hits. In the context of drug design, the conformational energy, which a small molecule adopts when bound to the target, is of fundamental importance because the internal energy of the ligand contributes to the total binding free energy. Hence, the energies of the conformations, obtained after superimposition with ROCS, were analyzed to check for the presence of biologically unrealistic conformers with increased strain energy. The conformational energy of the ROCS superimposed conformer was compared with those of coverage-based low energy conformation generation tools, namely, Omega[21] and Catalyst. These tools are proven to have a broad coverage of the bioaccessible conformational space within their specified energy window. Conformational energy comparisons were carried out using a uniform force field because each of the stated tools uses different force field for calculating energies. Atom typing was done using the Dreiding 2.21[47] force field, and Gasteiger charges[48] were assigned in Cerius[2], followed by single-point energy calculation in gas phase. The conformational energy of the ROCS-processed conformer was well within the allowed energy window as defined by Omega and Catalyst. The values obtained for some of the key molecules are reported in Table 4. The obtained energy values ensure that the conformations considered in the study have energies within the allowed energy threshold commonly employed to identify bioactive conformations.

The most probable hits identified by our screening strategy were KM01595 and KM01602. 2D structures of the hits are shown in Figure 7, and their respective scores are given in Table 5. The predicted binding mode of KM01602 mimics the critical interaction pattern required for inhibition and is shown in Figure 8. Panel A of Figure 8 depicts the 3D binding mode, and panel B of Figure 8 depicts the 2D plot of protein−ligand interactions.

A visual inspection of the docked complexes shows that the hits obtained (KM01602 and KM01595) occupy the same general space in the active site as exhibited by oxacetamides and isopthalamides.[46]

The phenyl carbamate group present in the hit replaces the α-methylbenzyl group attached to the nitrogen of the benzamide moiety at the $S3^{sp}$ pocket occupied by residues Gly11, Gln12, and Thr232. The electron-withdrawing monovalent halogen atom chlorine, present at the p-position of phenylcarbamate moiety of the hit mimics the interaction pattern exerted by the fluorine atom present at the p-position of the α-methylbenzyl group of oxaacetamides. This indicates the importance of an electronegative group at the $S3^{sp}$ pocket. The sulfonyl oxygen present in KM01602 forms stabilizing hydrogen bonding interaction with the side chain hydroxyl group of THR231. The pyridine moiety present in KM01595 and KM01602 takes the place of benzoid core present in

the oxacetamides. The amino group attached to 3-(pyridin-2-ylsulfonylmethyl)phenyl methylidene moiety of KM01595 and KM01602 occupies the S1 pocket that is occupied by the amino group present in the amino pentyl chain of the crystal bound inhibitor of oxaacetamide. The only difference observed in the binding mode of the hits is the unoccupancy of the S2 pocket. A conserved water molecule, also present in the active site of other apocrystal structures of aspartyl proteases, was treated as a part of the protein system during pose evaluation because its crystallographic coordinates were known from the crystal structure. We noticed that in the presence of the conserved water molecule, HOH35, had the ability to mediate a primary complex, which was not evident when HOH35 was knocked out. On the basis of an informed judgment, evident from the holocrystal structure of 1TQF, we are led to believe that HOH35 is more likely to remain buried in the active site of the protein, thereby stabilizing the protein−ligand complex.

The sulfonyl group present in our hits forms a strong hydrogen bond (2.1 Å) with the conserved water molecule, which in turn forms strong hydrogen bonding with the key catalytic residues ASP32 and ASP228, mimicking the primary mediated H bonding, observed in the crystal structure of oxyacetamide.

The aromatic face of the phenyl ring attached to the methyledine group is involved in $\pi-\pi$ stacking interaction resulting from a parallel-displaced geometry with the aromatic ring of TYR71. These interactions are akin to the holo crystal structure of 1TQF. A thorough and in-depth structural analysis of the docked conformation of KM01602 reveals it to be a promising candidate for further studies. Work is currently underway to validate the biological activity of the obtained hits in cell-based assays.

CONCLUSIONS

Because in silico screening of chemical compounds has become a standard practice in drug discovery, we emphasize the need to tailor structure-based drug design workflow to maximize the benefits for individual targets. ROCS, in conjunction with GOLD, was implemented for identification of novel BACE1 inhibitors. Such types of guided docking approaches are more ideal, especially for a system like BACE1, where all known inhibitors are highly flexible and the accuracy level in producing the right binding mode decreases. Hence, the shape-based overlay approach that helps in orienting the ligand during sampling, ensures reliable binding mode predictions. Integrating ligand-based virtual screening tools with structure-based virtual screening tools is currently uncommon in SBVS. This study exemplifies the importance of integrating ligand-based and structure-based drug discovery tools in real time screening by its ability to impose a real value added constraint in docking. Constraining the ligand pose using ROCS could help in circumventing soft errors and the modified weighted C_{score} functionality could minimize the occurrence of hard errors, which are the two common elements that dampen the prospects of docking as a virtual screening tool.

Our prime focus in the study centers on the identification of novel inhibitors and the enrichment results reported here are related to a single test system. As a consequence, we emphasize that the validation studies carried out by us should not be extrapolated as a benchmark measure because results are known to vary for every individual system and programs like GOLD and LigandFit also contain many tunable parameters which can significantly influence the results. The comparative performance stated in the study was carried out only to ensure the efficiency of our method in real time virtual screening. No attempt was made to compare the performance of either the scores or the softwares cited in the study because the literature is replete with such interesting articles.[49−53]

ACKNOWLEDGMENT

The authors thank Council for Scientific and Industrial Research (CSIR), New Delhi, for providing financial grant under Mission Mode Program CMM 0017. Software support granted by OpenEye and Accelrys is greatly acknowledged. Support service rendered by Inte:Ligand is greatly acknowledged. We also acknowledge Mr. Ram Kumar for helping us in the development of a tool for calculating consensus scores. R.S.K.V and N.M.M. thank CSIR, New Delhi for a Senior Research Fellowship.

REFERENCES AND NOTES

(1) Alzheimer, About a peculiar diseases of the cerebral cortex. By Alois Alzhimer, 1907 (Translated by L. Jarvik and H. Greenson). *Alzheimer Dis. Assoc. Disord.* **1987**, *1*, 3−8.
(2) Glenner, G. G.; Wong, C. W. Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **1984**, *120*, 885–90.
(3) Vassar, R.; Bennett, B. D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E. A.; Denis, P.; Teplow, D. B.; Ross, S.; Amarante, P.; Loeloff, R.; Luo, Y.; Fisher, S.; Fuller, J.; Edenson, S.; Lile, J.; Jarosinski, M. A.; Biere, A. L.; Curran, E.; Burgess, T.; Louis, J.-C.; Collins, F.; Treanor, J.; Rogers, G.; Citron, M. β-Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **1999**, *286*, 735–41.
(4) Yan, R.; Bienkowski, M. J.; Shuck, M. E.; Miao, H.; Tory, M. C.; Pauley, A. M.; Brashier, J. R.; Stratman, N. C.; Mathews, W. R.; Buhl, A. E.; Carter, D. B.; Tomasselli, A. G.; Parodi, L. A.; Heinrikson, R. L.; Gurney, M. E. Membrane-anchored aspartyl protease with Alzheimer's disease β-secretase activity. *Nature* **1999**, *402*, 533–37.
(5) Esler, W. P.; Wolfe, M. S. A portrait of Alzheimer secretases−New features and familiar faces. *Science.* **2001**, *293*, 1449–54.
(6) Potter, H.; Dressler, D. The potential of BACE inhibitors for Alzheimer's therapy. *Nat. Biotechnol.* **2000**, *18*, 125–26.
(7) Coburn, C. A.; Stachel, S. J.; Li, Y. M.; Rush, D. M.; Steele, T. G.; Chen-Dodson, E.; Holloway, M. K.; Xu, M.; Huang, Q.; Lai, M. T.; DiMuzio, J.; Crouthamel, M. C.; Shi, X. P.; Sardana, V.; Chen, Z.; Munshi, S.; Kuo, L.; Makara, G. M.; Annis, D. A.; Tadikonda, P. K.; Nash, H. M.; Vacca, J. P.; Wang, T. *J. Med. Chem.* **2004**, *47*, 6117–19.
(8) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–69.
(9) Schuster, D.; Nashev, L. G.; Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T.; Odermatt, A. Discovery of nonsteroidal 17 β-hydroxysteroid dehydrogenase 1 inhibitors by pharmacophore-based screening of virtual compound libraries. *J. Med. Chem.* **2008**, *51*, 4188–99.
(10) Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. Docking: Success and challenges. *Curr. Pharm. Des.* **2005**, *11*, 323–33.
(11) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.
(12) Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein−protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–95.
(13) (a) 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*, 727–48. (b) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609–23.
(14) (a) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from

docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–09. (b) Feher, M. Consensus scoring for protein -ligand interactions. *Drug Discovery Today* **2006**, *11*, 421–28. (c) Schulz-Gasch, T.; Stahl, M. Scoring functions for protein ligand interaction a critical perspective. *Drug Discovery Today* **2004**, *1*, 231–39.

(15) (a) Pan, Y.; Huang, N.; Cho, S.; MacKerell, A. D., Jr. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267–72.

(16) Böhm, H.; Stahl, M. *Reviews in Computational Chemistry*; Lipkowitz, K. V., Boyd, B. D., Ed.; John Wiley & Sons, Inc: New York, 2003; Vol. 18, Chapter 2.

(17) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **2003**, *4*, 289–07.

(18) Gruninger-Leitch, F.; Schlatter, D.; Kung, E.; Nelbock, P.; Dobeli, H. Substrate and inhibitor profile of BACE (β-Secretase) and comparison with other mammalian aspartic proteases. *J. Biol. Chem.* **2002**, *277*, 4687–93.

(19) Accelrys. *Discovery Studio*, version 2.0; Accelrys Inc.: San Diego, CA, 2007.

(20) The Cambridge Crystallographic Data Centre. *GOLD*, version 3.2; CCDC: Cambridge, U.K., 2006.

(21) OpenEye Scientific Software, 9 Bisbee Court, Suite D, Santa Fe, NM 87508, 2006.

(22) Inte:Ligand. *LigandScout*, version 1.03; Inte:Ligand Software-Entwicklungs und Consulting GmbH Clemens Maria Hofbauer-G: Maria Enzersdorf, Austria, 2007.

(23) Accelrys. *TSAR*, version 3.0; Accelrys Inc.: San Diego, CA, 2007.

(24) Accelrys. *Catalyst*, version 4.1; Accelrys Inc.: San Diego, CA, 2006.

(25) Accelrys. *Cerius*^{2}, version 4.10; Accelrys Inc.: San Diego, CA, 2006.

(26) Accelrys. *InsightII*, version 4.1; Accelrys Inc.: San Diego, CA, 2007.

(27) Greenidge, P. A.; Carlsson, B.; Bladh, L. G.; Gillner, M. Pharmacophores incorporating numerous excluded volumes defined by X-ray crystallographic structure in three-dimensional database searching: Application to the thyroid hormone receptor. *J. Med. Chem.* **1998**, *41*, 2503–12.

(28) 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 Delivery Rev.* **2001**, *46*, 3–26.

(29) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–47.

(30) Polgár, T.; Keserü, G. M. Virtual screening for β-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.* **2005**, *48*, 3749–55.

(31) Rush, T. S.; Grant, J. A.; Mosyak, L,; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–95.

(32) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499–10.

(33) Mills, J. E.; Dean, P. M. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput.-Aided Mol. Des.* **1996**, *607*, 607–22.

(34) Jacobsson, M.; Karlén, A. Ligand bias of scoring functions in structure-based virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 1334–43.

(35) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–04.

(36) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317–24.

(37) Jain, A. N. Scoring noncovalent protein−ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J. Comput.-Aided. Mol. Des.* **1996**, *10*, 427–40.

(38) Böhm, H. J. LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided. Mol. Des.* **1992**, *6*, 593–06.

(39) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Computer.-Aided Mol. Des.* **1997**, *11*, 425–45.

(40) Fradera, X.; Knegtel, R. M.; Mestres, J. Similarity-driven flexible ligand docking. *Proteins* **2000**, *40*, 623–36.

(41) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic. Acids. Res.* **2007**, *35*, 198–201; http://www.bindingdb.org (accessed Dec 20, 2007).

(42) http://cactus.nci.nih.gov/ncidb2/download.html (accessed Dec 7, 2007).

(43) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–01.

(44) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection−What can we learn from earlier mistakes? *J. Comput.-Aided. Mol. Des.* **2008**, *22*, 213–28.

(45) Su, A. I.; Lorber, D. M.; Weston, G. S.; Baase, W. A.; Matthews, B. W.; Shoichet, B. K. Docking molecules by families to increase the diversity of hits in database screens: Computational strategy and experimental evaluation. *Proteins* **2001**, *42*, 279–93.

(46) Stachel, S. J.; Coburn, C. A.; Steele, T. G.; Jones, K. G.; Loutzenhiser, E. F.; Gregro, A. R.; Rajapakse, H. A.; Lai, M.T.; Crouthamel, M. C.; Xu, M.; Tugusheva, K.; Lineberger, J. E.; Pietrak, B. L.; Espeseth, A. S.; Shi, X. P.; Chen-Dodson, E.; Holloway, M. K.; Munshi, S.; Simon, A. J.; Kuo, L.; Vacca, J. P. Structure-based design of potent and selective cell-permeable inhibitors of human β-secretase (BACE-I). *J. Med. Chem.* **2004**, *16*, 6447–50.

(47) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. "DREIDING: A generic force field". *J. Phys. Chem.* **1990**, *94*, 8897–09.

(48) Gasteiger, J.; Marsili, M. A new model for calculating atomic charges in molecules. *Tetrahedron. Lett.* **1978**, *34*, 3181–84.

(49) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–03.

(50) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **2005**, *48*, 962–71.

(51) Cole, J. C.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Comparing protein-ligand docking programs is difficult. *Proteins* **2005**, *60*, 325–31.

(52) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035–42.

(53) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759–67.

CI800386V