Computational fragment-based screening using RosettaLigand: the SAMPL3 challenge

Ashutosh Kumar · Kam Y. J. Zhang

Abstract SAMPL3 fragment based virtual screening challenge provides a valuable opportunity for researchers to test their programs, methods and screening protocols in a blind testing environment. We participated in SAMPL3 challenge and evaluated our virtual fragment screening protocol, which involves RosettaLigand as the core component by screening a 500 fragments Maybridge library against bovine pancreatic trypsin. Our study reaffirmed that the real test for any virtual screening approach would be in a blind testing environment. The analyses presented in this paper also showed that virtual screening performance can be improved, if a set of known active compounds is available and parameters and methods that yield better enrichment are selected. Our study also highlighted that to achieve accurate orientation and conformation of ligands within a binding site, selecting an appropriate method to calculate partial charges is important. Another finding is that using multiple receptor ensembles in docking does not always yield better enrichment than individual receptors. On the basis of our results and retrospective analyses from SAMPL3 fragment screening challenge we anticipate that chances of success in a fragment screening process could be increased significantly with careful selection of receptor structures, protein flexibility, sufficient conformational sampling within binding pocket and accurate assignment of ligand and protein partial charges.

Keywords Fragment based drug design · Molecular docking · Receptor flexibility · Partial charges · Scoring function

A. Kumar · K. Y. J. Zhang (✉)
Zhang Initiative Research Unit, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
e-mail: kamzhang@riken.jp

Introduction

The fragment based drug design (FBDD) approach represents a rapid, resource efficient and productive route to the identification of novel ligand hits in the early phase of drug discovery process. This method was proposed by Fesik et al. in 1996 and has gained prominence in recent years and is now recognized as a successful method of lead identification in a drug discovery program [1]. FBDD approach focuses on the identification of low molecular-weight compounds that target sub-pockets within the overall active site. These fragment hits are expected to be more suitable starting points for hit to lead optimization due to their reduced complexity, which leaves more freedom for multidimensional property optimization of the fragment hits usually by adding new functional groups or by linking of two fragment hits binding in adjacent pockets. Furthermore, with fragments it is possible to improve physicochemical and pharmacokinetic properties while maintaining ligand efficiency [2].

Recently, FBDD has led to the discovery of new scaffolds that were later optimized into high affinity inhibitors [3–10]. However, the small size of fragments and low binding affinity makes it particularly difficult to detect in standard biochemical assays. Instead, biophysical methods such as NMR [11], X-ray crystallography [12], and Surface Plasmon Resonance (SPR) [13] were used to identify these low molecular weight compounds. Although biophysical methods have several advantages but despite their utility only up to several hundreds or thousands compounds can be tested. Also, biophysical screening involves significant time, labor and materials costs. The limitations associated with experimental biophysical screening and broad application of FBDD presents the need for alternative screening methodologies.

Molecular docking based virtual screening of drug-like libraries has already been proved to be an efficient technology in hit discovery. Furthermore, docking based virtual screening was integrated with experimental screening providing drug-like hits for large variety of targets [14]. Molecular docking has also been used in FBDD for variety of targets. There are studies reported in literature that identified fragment molecules for further optimization utilizing structure based docking for screening fragment-like libraries [15–18]. However, docking fragment libraries is still considered not very reliable with present methods and protocols [17] because of promiscuous binding modes and inability of scoring function to discriminate near native from irrelevant binding poses [19]. Most of these docking methods and scoring functions were developed for drug-like ligands with molecular weight and properties significantly different from fragments. This may be one of the reasons for the poor performance of the docking algorithm and scoring functions in predicting the binding mode and affinity of small molecule fragments. The adaptability of these docking methods and scoring functions against fragment based virtual screening deserves further exploration.

A key concept in predicting the binding mode and affinity of fragments is the incorporation of receptor flexibility along with ligand flexibility. Accounting for receptor flexibility is very important in fragment based discovery and it could prevent the finding of correct pose if receptor flexibility is not considered in docking [20]. Recently, several approaches have been proposed to address the problem of receptor flexibility. These approaches simultaneously sample both receptor and ligand flexibility during docking [21–23]. One of the most commonly used approach is the use of protein sidechain rotamer libraries and it has been shown that use of sidechain rotamers significantly improve docking program's ability to find accurate binding poses [23, 24]. Another most common approach is the use of multiple receptor conformations which is a straightforward way to mimic receptor flexibility [25]. These multiple receptor conformations are carefully selected either from molecular dynamics trajectories or from multiple X-ray crystal structures. Moreover, a small number of docking algorithm can handle full ligand and receptor flexibility. RosettaLigand [22, 26, 27] is one of such programs that uses Monte-Carlo sampling and Rosetta full-atom energy function to explicitly model full sidechain, backbone, and ligand flexibility with ligand and receptor degrees of freedom explored simultaneously. RosettaLigand extensively samples receptor sidechain conformations near the binding pocket and it has been shown that incorporation of receptor flexibility increases the probability of finding near-native poses with low energies [22, 26]. RosettaLigand was evaluated against benchmark protein ligand docking datasets and performed

well in retrieving the native poses in both self and cross docking experiments [22, 26, 27]. Furthermore, RosettaLigand generated binding energies correlate well with the biological activities of 229 inhibitors against diverse targets with a correlation coefficient around 0.6 for most of the targets [22, 26, 27].

RosettaLigand performs fairly well in docking ligands with drug-like properties, however, docking of fragments or fragment like molecules using RosettaLigand is still unexplored. The most common approach for evaluating a program for fragment docking is to test against benchmark datasets containing crystal structures with bound fragment like molecules. The accuracy of Glide has been recently evaluated for docking fragments using these datasets [28, 29]. However, the performance of a docking program on benchmark datasets in a retrospective study tends to be higher than the virtual screening performances in real life cases. The real test for any docking based screening approach would be in a blind testing environment; only in this prospective study the true predictive ability can be accessed. Blind assessment would avoid the bias associated with ligand pose sampling and scoring functions and provides more strict assessment of the strength and weakness of a docking program.

SAMPL3 fragment based virtual screening challenge is a blind assessment platform provided by OpenEye [30, 31] where researchers working in FBDD can test their methods, protocols and programs, share their experiences and learn from their experiences for the development of novel and accurate methods for screening of fragment like molecules. The test data for SAMPL3 fragment based virtual screening challenge was provided by Newman et al. [32, 33]. It was a blind fragment screening study where 500 fragments from the Maybridge fragment library were soaked into crystals of bovine pancreatic trypsin and their structures determined by X-ray crystallography. Binding affinity data were obtained from SPR.

In order to assess our ability to use RosettaLigand for fragment based virtual screening, we participated in SAMPL3 fragment based virtual screening challenge. Here we first describe our fragment-based virtual screen protocol using RosettaLigand in a non-blind environment where a test database consisting of trypsin fragment like inhibitors and decoys were screened. We next report the results of our fragment screening protocol in SAMPL3 blind testing environment and provide a retrospective analysis of factors that affected the performance our protocol. Our study indicates that with careful selection of receptor structures, proper handling of receptor flexibility, enough conformational sampling within binding pocket and accurate ligand and protein partial charges, it is possible to identify low molecular weight inhibitors for protein targets utilizing docking based fragment screening approach.

Computational methods

Our protocol of virtual screening of a fragment library consists of a number of steps: First, multiple receptor structures for docking are selected from a set of trypsin crystal structures from Protein Data Bank (PDB). Ligand conformations were then generated for each of the fragment in SAMPL3 challenge fragment library in order to take care of ligand flexibility. After assigning the partial charges for fragments and protein, all the fragments were docked using multiple receptor docking approach to each of the representative receptor conformation using RosettaLigand followed by selection of top binding poses for each receptor and each fragment. Finally, consistency in the binding mode is evaluated to prepare final rankings.

### Selection and preparation of receptor structures for docking

The crystal structures of trypsin were retrieved from PDB [34] and were analyzed in order to select suitable receptors for virtual screening. In a preliminary analysis of PDB structures, more than 100 crystal structures for trypsin in complex with ligands determined by X-ray crystallography at different resolutions were found. As the goal of our study was to screen fragment library only those crystal structures containing bound ligands with fragment like properties were selected. Fragment likeness for crystal structure ligands was defined by the rule of three [35], i.e., molecular weight ≤300 dalton, number of H bond acceptors (HBA),

H bond donors (HBD) ≤3, ClogP ≤3 and number of rotatable bonds ≤3. The crystal structure with the highest resolution was selected in case many structures with the same ligand were found. Finally our selection resulted in 14 trypsin complexes bound with unique fragment like trypsin inhibitors. These ligands and their corresponding complex PDB codes are depicted in Fig. 1. Analysis and filtering was carried out using MOE2010.10 [36]. To prepare receptor structures for molecular docking, hydrogens were added, bond orders were assigned and all the water molecules and atoms of the inhibitor were removed. Protonation states of charged residues were determined and implemented using Protonate3D [37] in MOE2010.10.

Molecular docking and scoring

All molecular dockings were performed using RosettaLigand [22, 26, 27], which is a fully flexible receptor and ligand docking program. RosettaLigand employs a stochastic conformational search inside a user defined cube centered on the binding site to identify low-energy protein ligand complexes. At the start of docking procedure, each conformation from an ensemble was placed into receptor binding site for docking calculations. RosettaLigand then simultaneously places probable side-chain amino acid rotamers around the ligand and optimizes the randomly sampled flexible ligand pose using a Metropolis Monte Carlo simulated annealing algorithm. The docked poses were ranked using an energy function dominated by van der Waals attractive and repulsive forces, electrostatic

Fig. 1 Fragment like inhibitors bound to trypsin in the X-ray complex structures used in the present study
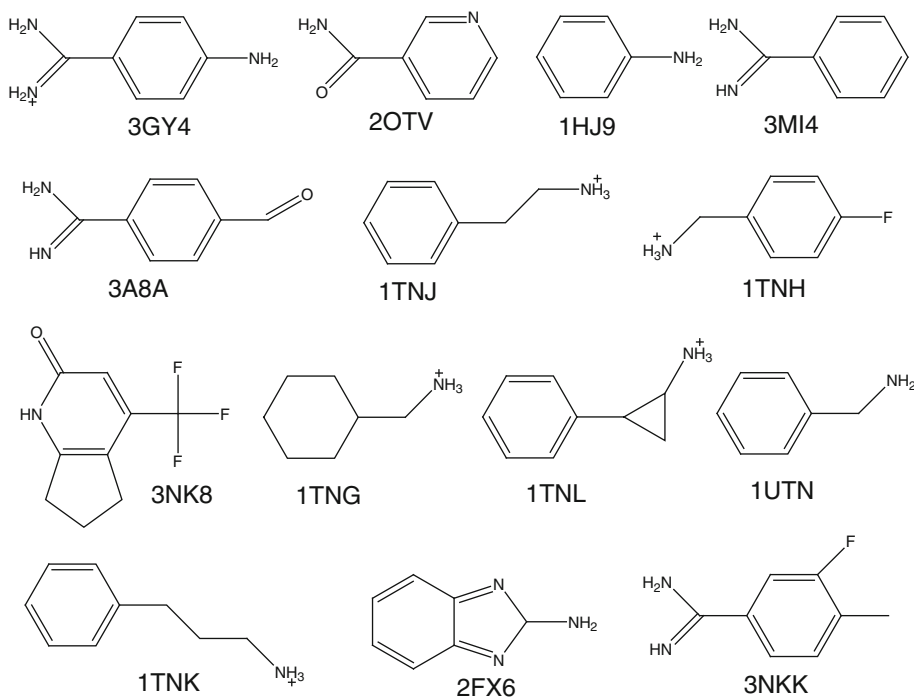
interactions between pairs of amino acids, and solvation assessing the effects of both side-chain side-chain interactions and side-chain ligand interactions, statistical energy derived from the probability of observing a side-chain conformation in the PDB and an orientation dependent hydrogen bonding potential. For each fragment and each receptor in consideration, approximately 5,000 docked poses were generated, and the top 5% of best scoring structures were re-ranked by ligand–protein interface scores (InterfaceDelta term in RosettaLigand scores), which is the difference between the energies of protein in ligand bound and unbound states. The pose with the lowest ligand protein interface score was then chosen as the best docked pose.

Preparation of ligands and fragment library for docking

As RosettaLigand handles the flexibility of a ligand by using a set of diverse ligand conformations, therefore, a conformational ensemble for each fragment in SAMPL3 test fragment library was generated using LowModeMD search algorithm [38] in MOE2010.10 that uses implicit vibrational analysis to focus a MD trajectory along the low-mode vibrations. For each fragment, a maximum of 200 conformations were requested and the resulting geometries were minimized in MOE2010.10 using MMFF94s forcefield [39]. All other parameters were set to their default values.

Receptor interaction fingerprint scoring

Docked poses generated by RosettaLigand were used to generate receptor interaction fingerprints. Receptor interaction fingerprints were generated using FingerPrintLib program [40] which uses OEChem C++ library from OpenEye, Inc. [31]. On the basis of a list of atom flags inferred from OEChem [31], positions of a bit vector are switched either on or off depending on whether or not predefined intermolecular interactions agree with user-defined rules. In the present work, only the first seven bits (hydrophobic interactions, aromatic face to face, aromatic face to edge, hydrogen bond acceptor, hydrogen bond donor, positively charged and negatively charged) which correspond to the most frequent protein–ligand interactions are calculated. The distance between two interaction fingerprints was calculated using a Tanimoto similarity coefficient (T_c) [41].

Hardware

The present work was carried out using Dell Precision T5400 workstation with 2.0 GHz Intel Xeon CPU and 97.4TFLOPS Intel Xeon 5570 based Massively Parallel PC Cluster of RIKEN Integrated cluster of clusters (RICC).

Results and discussion

Performance of RosettaLigand-based virtual screening protocol

Molecular docking of SAMPL3 challenge fragment library with RosettaLigand was the core part of our virtual screening protocol; therefore the reliability of RosettaLigand in docking fragment like compounds was assessed by three procedures. First, a self-docking analysis was carried out in which each ligand was docked back into its native crystal structure. After that, cross-docking was carried out where each crystal structure ligand was docked into all the 14 crystal structures of trypsin. Finally, RosettaLigand docking performance for fragment like compounds was evaluated using an external test set compiled from PDB-bind database [42, 43].

Self docking

A principle aim in docking fragments is to find energetically favorable binding modes of these fragments. To access the ability of RosettaLigand to accurately dock fragments, docking of fragments to their corresponding crystal structures was performed. The goal of this experiment is to evaluate how well RosettaLigand recapitulates experimentally determined binding modes. All ligands were extracted from the selected 14 trypsin X-ray crystal structures. They were then re-docked into their corresponding proteins. The docking results were evaluated through a comparison of the top ranked docked pose of the ligand with the one in the crystal structure. The pose with lowest ligand protein interface scores (Interface Delta in RosettaLigand energy function) was regarded as the top ranked pose. Interface Delta is the difference between the energies of protein in ligand bound and unbound state. For the purpose of comparison of top ranked pose with crystal structure, the root-mean-square deviation (RMSD) between the positions of the heavy atoms of the ligand in the calculated and experimental structures was calculated. The plot of RMSD value between the top ranked docking solution and native ligand binding pose in all the 14 crystal structures is given in Fig. 2. In most of the 14 trypsin complexes, RosettaLigand was able to reproduce the native binding mode of fragment with a RMSD value of <2 Å. RosettaLigand failed to obtain native binding pose of the corresponding crystal structure ligand in only two cases with RMSD of 4.44 and 3.63 Å respectively. As seen from self docking results, RosettaLigand demonstrated acceptable accuracy in docking fragments to their crystal structures without any special fragment specific optimization in the sampling and scoring function.
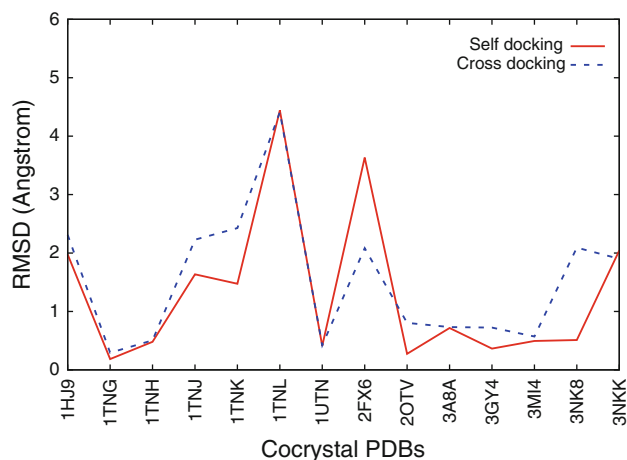
Fig. 2 RMSD distribution of docked poses generated by RosettaLigand in self and cross docking

Fig. 3 Distribution of RosettaLigand docking scores in cross docking

Fig. 4 RMSD distribution of docked poses produced by RosettaLigand for fragment like dataset compiled from PDBbind database

Cross docking

Cross docking studies were then performed where each trypsin fragment like inhibitor was docked into a trypsin protein structure bound with a different ligand. As in self docking, the performance of cross docking was assessed by comparing the RMSD between top ranked docking pose in non-native receptor with that of native trypsin ligand complex. The average RMSD between the best docked pose and native pose is plotted in Fig. 2. Comparative analysis of cross docking with self docking revealed that the success rate in reproducing native binding modes was reduced for cross docking. However, RosettaLigand was still able to find the native binding pose in majority of the cases with an average RMSD of <2 Å. The cross docking performance of RosettaLigand was further evaluated by checking the consistency in the docking scores produced for each ligand docked to all 14 target receptors. Figure 3 shows the distribution of RosettaLigand Interface delta scores produced by RosettaLigand for each ligand docked to 14 trypsin cocrystal structures. As seen from the Fig. 3, docking scores produced by RosettaLigand which reflect the binding free energy are quite consistent for each co-crystal ligand with respect to each target used for docking. This also shows the docking poses produced by RosettaLigand are reliable.

Docking of PDBbind set

To further evaluate the efficiency of RosettaLigand to dock fragment like ligands, a fragment test set was compiled from the PDBbind database [42, 43]. All the PDB entries in PDBbind dataset, which are not in complex with fragment like ligands were removed. Rule of three [35] for fragment likeness was used to define crystal structure ligands as fragment like. All the trypsin crystal structures were also
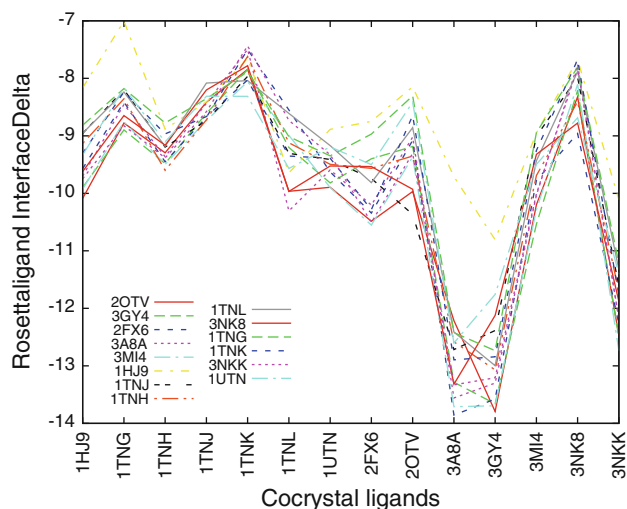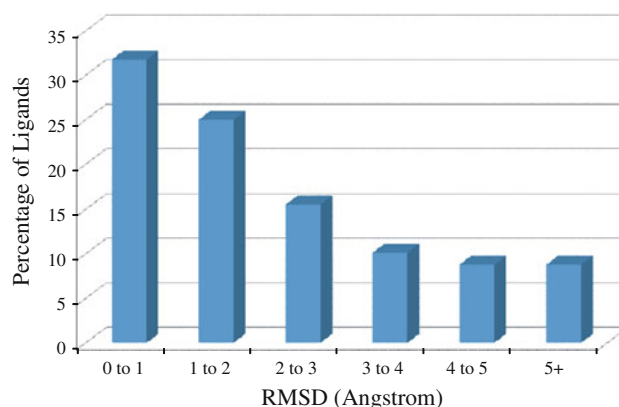
removed from the dataset. Finally, a dataset of 148 PDB cocrystal structures bound with fragment like ligands were used to test RosettaLigand. The test set used here consists of diverse protein targets and provides broader coverage of different types of proteins bound with fragment like ligands. The distribution of RMSD values between the top ranked docking solutions and native ligand binding poses produced by RosettaLigand is given in Fig. 4. RosettaLigand demonstrated acceptable accuracy in docking fragments compiled from PDBbind database with 56.75% of fragment like ligands docked to their respective proteins with a RMSD <2 Å. Another 15% of fragment like ligands show RMSD between 2 and 3 Å.

Docking based virtual screening performance

In virtual screening, rank ordering of compound libraries is a major goal for any computational method. The performance of our virtual screening protocol was evaluated by

screening a dataset comprising of both native and non-native ligands. Fourteen trypsin fragment inhibitors were mixed with 148 fragment like ligands compiled from PDBbind dataset [42, 43] as described previously. The virtual screening performance was quantified by the area under curve (AUC) of its receiver operating characteristic (ROC) plot. The AUC is a measure of how high a randomly selected active compound is ranked compared to a randomly chosen decoy. The AUC will be around 0.5 if the fragments are ranked randomly, while a perfect ranking of the fragments will result in an AUC of 1. The ROC plot is constructed by using ranked list of dataset compounds (both native and non-native) arranged in order of increasing RosettaLigand Interface Delta. The ROC plot is a plot of sensitivity (how many true positives are retrieved) versus specificity (how many false positives are retrieved) calculated at each position by assuming that all compounds ranked higher as active. The AUC and ROC plots were calculated using pROC package [44] of R program [45]. For evaluating the screening performance of our fragment screening protocol standard screening run were carried out independently on each of the 14 crystallographically derived receptor conformations of trypsin. Initially, we focused on single receptor runs for the assessment of virtual screening performance by plotting the relative rank of trypsin fragment like inhibitors against set of decoys. Table 1 presents the AUC values for 14 trypsin crystal structures. The maximum AUC value of 0.722 is found for 3MI4, which is also one of the highest resolution structure of trypsin. As seen from this table, the AUC value ranges from 0.438 to 0.722 for 14 crystal structures. In some of the

receptor conformation RosettaLigand outperforms random selection, however, some performed worse than the random selection like PDB 1TNH and 1TNJ.

To evaluate the performance of our fragment screening protocol in multiple receptors docking, the docking of test dataset containing the native ligands and decoys was performed on all 14 trypsin crystal structures. The aim of using multiple crystal structures in docking is to use additional protein plasticity represented by several ligand bound receptor conformations. First, a ranked list for multiple receptors docking was created by averaging the top ranking docking score for ligand docked to each of the 14 trypsin receptor. Another ranked list was created where the ligands are ranked according to their lowest energy score from 14 multiple receptor docking runs. The outcome of this study was the comparison of single receptor docking runs with the multiple receptor docking runs. The plot between sensitivity and specificity values obtained after ranking the test dataset by docking scores is shown in Fig. 5. The single receptor docking here outperformed multiple receptors docking in both cases. The single receptor docking displayed best AUC value of 0.722 as compared to AUC value of 0.620 when average docking scores are used and AUC value of 0.559 when lowest energy scores from all 14 receptors are used to prepare ranked list. Although single receptor docking performed fairly well as compared to multiple receptors

Table 1 Area under receiver operating characteristic curve (AUC) values for trypsin crystal structures used for single receptor docking of test library consisting of fragment like trypsin inhibitors and decoy molecules

PDB	AUC	CI. AUC (95%)
1HJ9	0.653	0.758–0.549
1TNG	0.504	0.366–0.642
1TNH	0.449	0.317–0.581
1TNJ	0.438	0.306–0.571
1TNK	0.603	0.731–0.476
1TNL	0.579	0.708–0.449
1UTN	0.677	0.786–0.568
2FX6	0.592	0.722–0.462
2OTV	0.708	0.822–0.593
3A8A	0.613	0.746–0.481
3GY4	0.706	0.815–0.597
3MI4	0.722	0.832–0.611
3NK8	0.578	0.701–0.456
3NKK	0.612	0.737–0.488

CI.AUC values represent the AUC values at 95% confidence interval
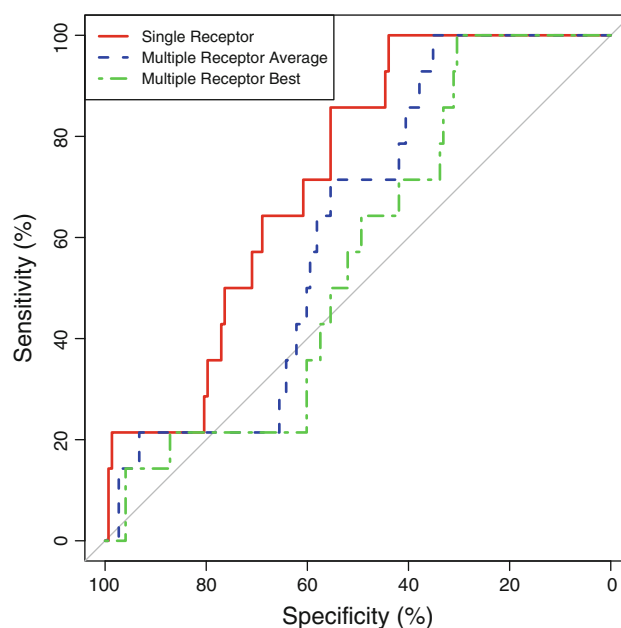

Fig. 5 The receiver operating characteristic (ROC) curve describing tradeoff between sensitivity and specificity for three different scoring procedures; single receptor docking, multiple receptor average (when average docking scores from all receptors are used to prepare rankings) and multiple receptor best (when best docking scores among all receptors are used to prepare rankings) for test library consisting of fragment like trypsin inhibitors and decoy molecules

docking here, it is contrary to previous results where multiple receptors docking is shown to improve the virtual screening performance [46–49]. Single receptor docking is a good measure to evaluate a docking program's performance when information about the native ligand is available and enrichment of native ligand could be used to identify best receptor for virtual screening. However, in real virtual screening runs where information about native and decoys is not available, the selection of best performing conformation is really problematic as it is hard to establish which conformer is going to perform best.

Re-ranking with receptor interaction fingerprint weighted scores

Receptor interaction fingerprints are used to incorporate knowledge from protein ligand interaction in a docking based scoring scheme and are known to improve the enrichment of active compounds against a set of decoys in a virtual screening campaign [40]. Here, we have tried to evaluate whether receptor interaction fingerprints based weighting on docking scores generated by RosettaLigand improve the virtual screening performance. The results show that receptor based weighting improves the virtual screening performance as reflected by AUC of 0.843 for single receptor docking and AUC value of 0.797 for multiple receptor docking. This improvement in the virtual screening performance might be because of the fact that receptor interaction fingerprints tend to favor ligands with similar chemotypes.

Molecular docking of SAMPL3 challenge fragment library

The SAMPL3 challenge fragment library obtained from OpenEye was prepared as described previously in the manuscript. The docking protocol as used in the evaluation of RosettaLigand was used with no fragment specific settings. To generate near native poses, RosettaLigand requires sufficient conformational sampling of the fragment within trypsin binding site. Therefore, 5,000 fragment poses for each fragment of SAMPL3 challenge fragment library were generated and docked onto each receptor. For each receptor structure, the pose with the best RosettaLigand protein–ligand interface score for each fragment of the SAMPL3 test fragment library was retained and a ranked list was created in order of increasing RosettaLigand protein–ligand interface scores for each of the 14 trypsin structures. The fragment poses were then evaluated for consistency in the binding mode and the interaction formed with active site residues of 14 trypsin structures. This evaluation was based on the assumption that a real inhibitor would bind to multiple conformations of the same

protein target in the same manner highlighting similar interactions with active site residues. To elucidate consistency in the binding mode automatically, average root mean square deviation (aRMSD) was calculated by comparing all the fragment poses for each receptor with the most representative pose in the cluster center. The most representative pose in the cluster center was selected as the one with the smallest aRMSD to the geometric center. The geometric centre was calculated by averaging the Cartesian coordinates of 14 poses for each fragment. aRMSD deviation and average RosettaLigand protein–ligand interface scores were then used to rank the fragments in SAMPL3 test fragment library. Finally, the ranking of the fragment library was manually curated by visually inspecting all poses for critical interactions displayed by trypsin crystal structure ligands. The final ranked list of fragments was then submitted to SAMPL3 for evaluation.

Retrospective analysis of SAMPL3 challenge results

Computationally predicted rank lists were evaluated against experimental results where fragments with $K_D < 1,000 \mu M$ against bovine pancreatic trypsin as measured by SPR are considered as actives [32]. Our ranked list submitted for evaluations resulted in ROC curve AUC value of 0.505 with AUC at 95% confidence interval ranges in between 0.39 and 0.62. This prediction was far worse than what we would have expected and our prediction was like random selection of fragments from the library. Table 2 presents the AUC values for SAMPL3 fragment library docked to 14 trypsin crystal structures.

Table 2 Area under receiver operating characteristic curve (AUC) values for trypsin crystal structures used for single receptor docking of SAMPL3 challenge fragment library

PDB	AUC	CI. AUC (95%)
1HJ9	0.616	0.727–0.505
1TNG	0.493	0.588–0.398
1TNH	0.479	0.582–0.375
1TNJ	0.543	0.441–0.645
1TNK	0.495	0.388–0.602
1TNL	0.557	0.468–0.645
1UTN	0.601	0.707–0.495
2FX6	0.559	0.448–0.670
2OTV	0.632	0.739–0.524
3A8A	0.518	0.628–0.409
3GY4	0.651	0.753–0.550
3MI4	0.619	0.725–0.513
3NK8	0.574	0.458–0.690
3NKK	0.516	0.618–0.414

CI.AUC values represent the AUC values at 95% confidence interval

The AUC values obtained after ranking fragment library with RosettaLigand Interface Delta scores obtained after docking to each of the 14 trypsin crystal structures ranges from 0.479 to 0.651. The best performing receptors were 3GY4, 2OTV and 3MI4 with AUC value of 0.651, 0.632 and 0.619 respectively. The worst performing receptors were 1TNH, 1TNG and 1TNK with AUC value of 0.479, 0.493 and 0.495 respectively which is worse than random picking.

Generally, the top three receptors according to AUC value in Table 2 have significantly higher AUC value than the bottom three of the receptors. We checked the structure of these best and worst performing receptor PDBs and observed that there is no big difference in the position and orientation of the backbone and sidechain atoms. The difference in the AUC values may be related with the absence of a sulphate ion in the inhibitor binding site of worst performing receptors. The sulphate ion sits at a prime position in the inhibitor binding site and could affect the placement of fragment within the binding pocket. Overall, ranking using single receptor could outperform random picking if receptor used for docking studies were selected carefully; however even with the best performing receptor it was unable to guarantee the enrichment of true binders among the top few hits. The performance of multiple receptors docking in prioritizing the fragments in the library was then analyzed. The performance of using multiple receptors in ranking the SAMPL3 fragment library was even worse than using single receptors. The use of average of all docking scores obtained after docking to 14 receptor PDBs resulted in an AUC of 0.468, whereas, AUC value of 0.472 was obtained when lowest energy scores from all 14 receptors are used to prepare ranked list. The virtual screening performance of multiple receptors docking versus single receptor docking is shown in Fig. 6. The ROC plot shows virtual screening enrichment achieved using best performing single receptor 3GY4 outperformed enrichments when multiple receptors are used in docking. Here, ranking is prepared using all the 14 receptor PDBs without any careful selection of receptors. This finding pointed out the fact that careful selection of receptors is prerequisite for multiple receptors docking. The performance of some of the trypsin PDBs like 1TNH, 1TNG and 1TNK was below random selection in single receptor docking run and use of these PDBs brought the AUC down.

It has already been reported previously that use of multiple receptor conformation improves the virtual screening performance if receptor ensemble is chosen carefully among the available crystal structures or molecular dynamics generated conformers [50, 51]. We observed that if top five receptors that performed best in single receptor docking are chosen, the AUC value goes up. Figure 7 presents the ROC
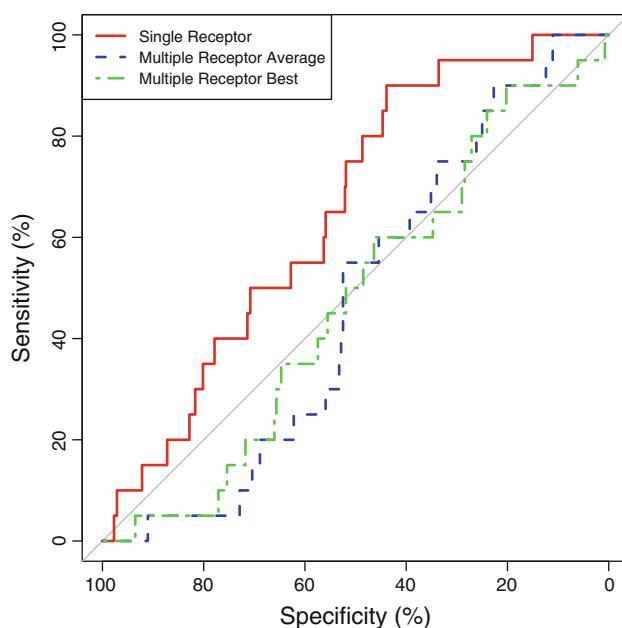

Fig. 6 The receiver operating characteristic (ROC) curve describing tradeoff between sensitivity and specificity for three different scoring procedures; single receptor docking, multiple receptor average (when average docking scores from all receptors are used to prepare rankings) and multiple receptor best (when best docking scores among all receptors are used to prepare rankings) for SAMPL3 challenge fragment library

plot when the five best performing receptors were used in preparing the ranking. The use of average docking scores obtained while docking to 3GY4, 2OTV, 3MI4, 1UTN and 1HJ9 in ranking the fragment library resulted in AUC value of 0.6317, whereas, AUC value of 0.624 was obtained when lowest energy scores were used. Our results suggest that the choice of most appropriate receptor conformations is a key for successful virtual screening. However, incorporation of too many receptor conformations can lead to reduced performance.

It has been shown earlier in the manuscript that receptor interaction fingerprint based weighting improves virtual screening performance of RosettaLigand. Therefore, to test whether using receptor interaction fingerprint based weighting improves the virtual screening performance of single and multiple receptors docking using RosettaLigand to screen the SAMPL3 fragment library. The receptor interaction fingerprints of top scoring docked poses generated by RosettaLigand were calculated for all the receptor using the protocol described in the method section. The receptor interaction fingerprints of SAMPL3 fragment library were then compared with that of reference ligands using Tanimoto correlation coefficient [41]. This Tanimoto correlation coefficient was then used as a weighting factor for RosettaLigand Interface Delta scores which were then
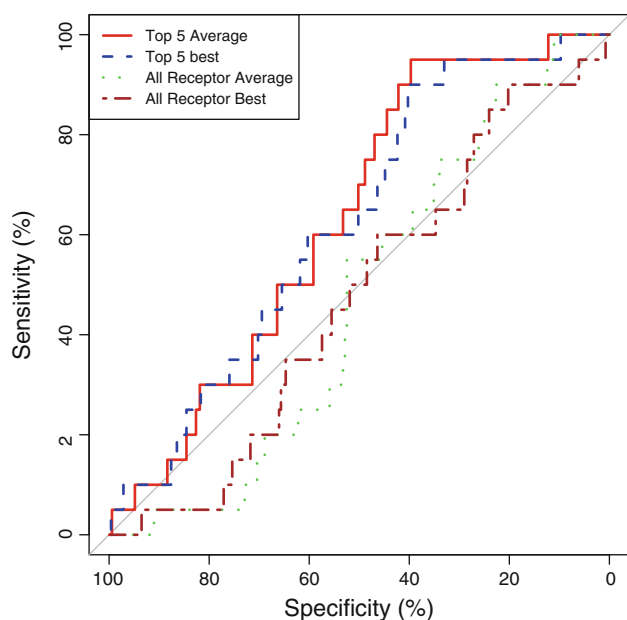
Fig. 7 The comparison of receiver operating characteristic (ROC) plots obtained after preparing the ranked list using top performing five receptors PDBs with all the receptor PDBs

used to rank the fragment library. The AUC values were calculated after ranking the fragment library with the receptor interaction fingerprint weighted docking scores. The use of receptor interaction fingerprint weighted scores did not improve the virtual screening performance. Instead it deteriorated the virtual screening performance and brought down the AUC value from 0.652 to 0.565 for single receptor docking. The AUC value for multiple

receptors docking remained almost the same and there was no improvement in the virtual screening performance. This result was expected as the reference ligands used for comparison of receptor interaction fingerprints do not belong to diverse chemical classes. The interaction fingerprint generated from these very similar reference ligands tends to favour fragments belonging to the same structural class and this may be the reason for low AUC values from ranking based on receptor interaction fingerprints.

Factors that affect the virtual screening performance

We have observed that careful selection of multiple receptors improves the virtual screening performance. However, this is only marginal improvement and the virtual screening performance is far from ideal. We then tried to explore the reasons for the meagre performance of our protocol in docking SAMPL3 fragment library. The crystal structures of seven of the active fragments in the SAMPL3 challenge fragment library (frag. vs. 115, frag. vs. 129, frag. vs. 188, frag. vs. 198, frag. vs. 236, frag. vs. 339 and frag. vs. 366) were obtained from Peat and Newman [32] and the top ranking poses generated by RosettaLigand were compared with that of crystal structures. Here the results from the best performing receptor 3GY4 are shown. As seen from Fig. 8, in five out of seven cases, the top scoring pose generated by RosettaLigand was different from the native fragment pose with RMSD of 3.90, 5.04, 4.21, 4.90 and 5.04 Å for frag. versus 115, frag. versus 129, frag. versus 198, frag. versus 236 and frag. versus 339 fragment

Fig. 8 The comparison of RosettaLigand generated docked poses with crystal structure conformation for active fragments in SAMPL3 challenge fragment library. Docked poses are shown in *green* while the crystal structure conformations are in *magenta*. The direction of *arrows* point to the improvement in the pose prediction after using Gasteiger ligand partial charges instead of MMFF94s forcefield based charges
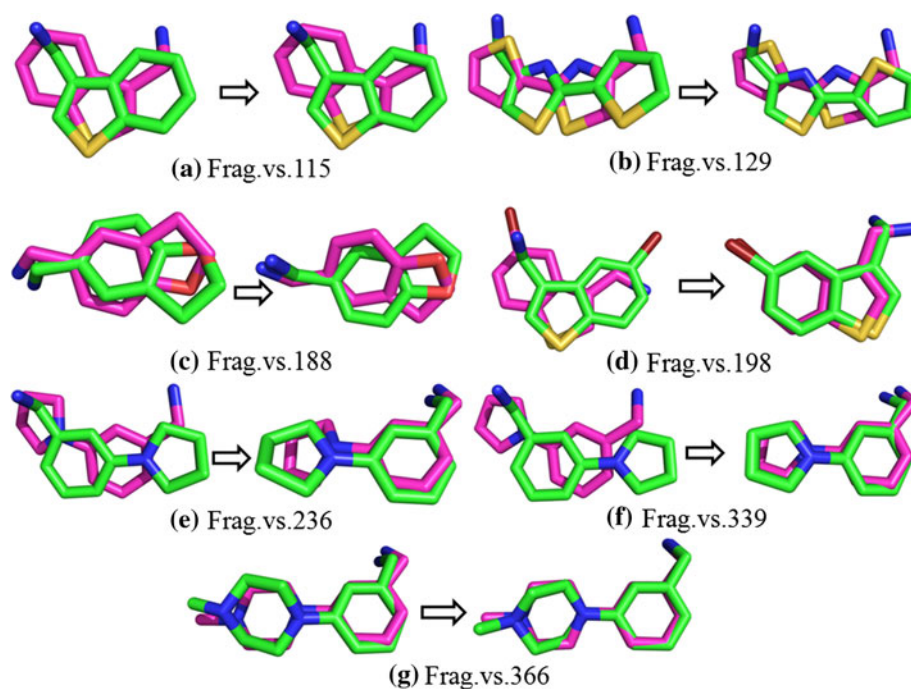

(a) Frag.vs.115

(b) Frag.vs.129

(c) Frag.vs.188

(d) Frag.vs.198

(e) Frag.vs.236

(f) Frag.vs.339

(g) Frag.vs.366

Table 3 Effect of ligand conformation, ligand pose sampling, ligand partial charges and scoring function on virtual screening performance

Method	AUC	CI. AUC (95%)
Ligand conformation		
LowModeMD	0.651	0.753–0.55
Stochastic	0.628	0.741–0.515
Systematic	0.621	0.723–0.519
Ligand pose sampling		
5,000 poses	0.651	0.753–0.55
10,000 poses	0.617	0.726–0.508
Ligand charges		
MMFF94s	0.651	0.753–0.55
AM1-BCC	0.637	0.745–0.528
Gasteiger	0.706	0.793–0.619
Scoring function		
RosettaLigand	0.706	0.793–0.619
Autodock	0.731	0.839–0.622
DrugScore	0.673	0.792–0.554

AUC represents area under receiver operating characteristic curve and CI.AUC values represent the AUC values at 95% confidence interval. The AUC values shown here are calculated after docking SAMPL3 challenge fragment library to best performing receptor PDB 3GY4

respectively. In only two cases lower RMSD of 1.89 and 1.30 Å was found for frag. versus 188 and frag. versus 366. In most cases, the top scoring pose was flipped by 180°. We then checked whether the near native pose was present in the top 10 scoring poses generated by RosettaLigand. To our surprise, in four cases, near native poses were not present in the top 10 scoring poses generated by RosettaLigand. Overall, more than 50% of the time native poses were not present among the top scoring 10 poses. As no correct poses could be picked up in the first place, therefore RosettaLigand's scoring function cannot predict the binding affinity accurately. We inferred that there was some problem in our docking protocol and we need to rethink where we might have made a mistake. In our virtual screening protocol, there were four putative sources of prediction errors: (1) ligand starting conformations, (2) sampling of the docking poses, (3) protein and ligand charges and (4) scoring function. All these sources of error were addressed in a retrospective analysis to highlight problems in our docking protocol and the results are summarized in Table 3 and will be described in detail below.

Influence of starting ligand conformations

RosettaLigand requires ligand conformations to be pre-generated to consider the ligand flexibility in docking. Ensemble of carefully generated ligand conformations generally produces better results than using single lowest

energy conformation [52]. In this study, fragment ligand conformations were generated using LowModeMD search algorithm [38] in MOE2010.10. In case of fragment-size ligands there are few rotatable bonds and conformational diversity is not as vast as drug-like ligands and the conformational search algorithm can sample the ligand conformational space well for fragment like ligands. Although inaccurate ligand conformation might not be the source of error, we tested this by generating ligand conformations using a systematic search algorithm and a stochastic search algorithm in MOE2010.10. We repeated the docking calculation using 3GY4 as receptor and there was no improvement in the virtual screening performance.

Ligand pose sampling

RosettaLigand is a Monte Carlo simulated annealing algorithm that employs a stochastic conformational search to identify low-energy protein ligand complexes [22, 26, 27]. Therefore, sufficient sampling is an ultimate requirement to discover near native poses. In our docking protocol, 5,000 protein ligand binding poses were generated and ranked by our scoring scheme. Considering that these 5,000 poses may not sample the binding site sufficiently to generate the near native pose, we raised the number of output poses from docking run to 10,000. However no significant improvement in the virtual screening performance was observed as AUC value only increased from 0.652 to 0.660. This marginal improvement in virtual screening performance is not worth the doubling of computational time.

Ligand partial charges

RosettaLigand's energy function involves the calculation of contributions from various energy components like electrostatic, van der Waals, hydrogen bond, solvent effects [22, 26, 27]. Therefore, the correct docking pose and docking score of the pose heavily depends upon the accurate partial charges on the ligand. The partial charges are even more important when the protein system is bovine pancreatic trypsin, as the inhibitor binding site is composed of negative charged residue Asp189 which plays an important role in cleaving substrate peptide. The presence of positively charge functionalities is hence desired for compounds able to inhibit trypsin. In this study, MMFF94s partial charges [39] were assigned to the SAMPL3 challenge fragment library. The MMFF94s partial charges were used because of the positive experiences derived from other studies in our group using these set of charges. To retrospectively study the influence of ligand partial charges on the performance of docking based virtual screening we have applied AM1-BCC [53] and Gasteiger [54] partial
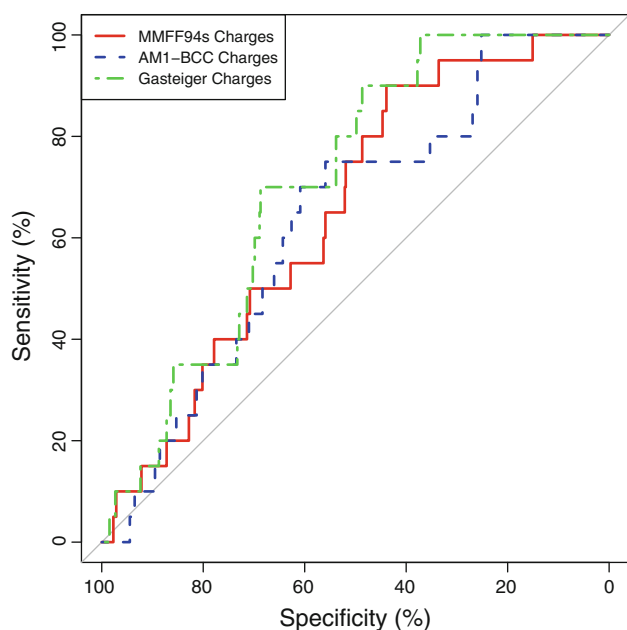
Fig. 9 The effect of three different partial charge calculation methods on molecular docking of SAMPL3 fragment library

charges to SAMPL3 challenge fragment library using MOE2010.10 and docked the library to trypsin best performing receptor 3GY4. As seen in Fig. 9 and Table 3, application of Gasteiger partial charges to the fragment library improved the performance of virtual screening with AUC values increases from 0.652 to 0.706. Using AM1-BCC charges, there was no improvement in the virtual screening performance as the AUC value was 0.637 which is even worse than using MMFF94s charges. This improvement in the virtual screening performance of RosettaLigand using Gasteiger charges may be attributed to the fact that empirical charge calculation method gives higher absolute value for partial charge to a selected atom than semi-empirical AM1-BCC method and MMFF94s forcefield based charges. This results in accurate handling of electrostatic interactions between the atomic pair and thus results in better and near to native binding pose as compared to other charge calculation methods.

The top ranked docking poses produced by RosettaLigand using Gasteiger charges were then compared with the native binding poses for seven fragments (frag. vs. 115, frag. vs. 129, frag. vs. 188, frag. vs. 198, frag. vs. 236, frag. vs. 339 and frag. vs. 366) in the SAMPL3 challenge fragment library. As seen from Fig. 8, the docking poses for five out of seven fragments were close to the binding mode displayed in the crystal structures with the RMSD value of 3.72, 4.95, 1.91, 0.96, 1.50, 1.58 and 0.84 Å for frag. versus 115, frag. versus 129, frag. versus 188, frag. versus 198, frag. versus 236, frag. versus 339 and frag. versus 366 respectively. The improvement in the identification of near native poses for these fragments by

RosettaLigand may be attributed to the accurate description of partial charges that results in more pronounced effect of electrostatic interactions between each atomic pair. This results into well defined protein ligand geometry as compared to MMFF94s charges used earlier.

Scoring function

The scoring function inside RosettaLigand was used to perform the prospective virtual screening of the SAMPL3 fragment library. It was desirable to access whether the virtual screening performance of our protocol can be further improved by rescoring the RosettaLigand generated poses with some other scoring function. We have chosen Autodock scoring function [55–57] as representative for empirical scoring function and DrugScore [58, 59] for knowledge based scoring function. The docking scores of RosettaLigand generated binding poses were re-calculated using Autodock and DrugScore scoring function and the ranked list for SAMPL3 challenge fragment library was produced. The ROC curve representing the performance of Autodock and DrugScore scoring function is presented in Fig. 10 and the AUC values are shown in Table 3. The use of Autodock scoring function on RosettaLigand generated poses slightly improved the virtual screening performance as the AUC increased from 0.706 to 0.731. However, there was no improvement while using knowledge based scoring function DrugScore as the AUC deteriorated from 0.706 to 0.674. The results showed that although there is no big
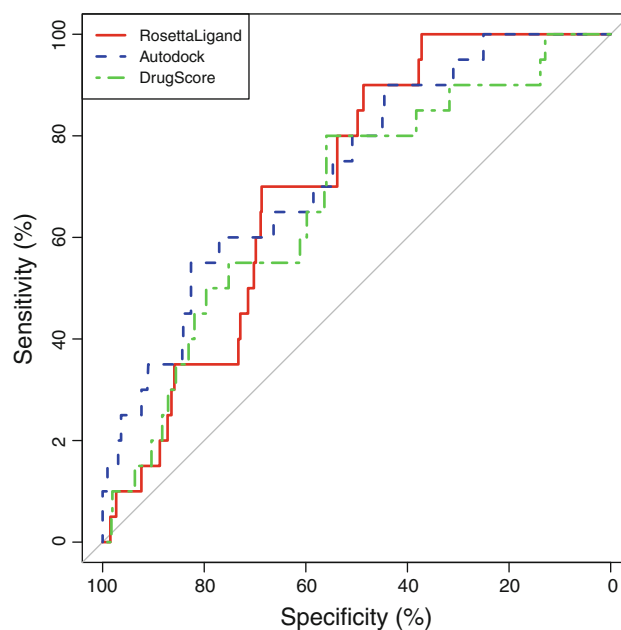

Fig. 10 The receiver operating characteristic (ROC) curve describing the comparison of three different scoring functions on ranking the active fragments ahead of inactives from SAMPL3 challenge fragment library

difference in the performance of RosettaLigand and Autodock scoring function, there is some margin for improvement in RosettaLigand's scoring function. The results pointed out the need for improvement in the scoring function of RosettaLigand in order to increase its fragment docking performance.

Lessons learnt from SAMPL3 challenge

SAMPL3 fragment based virtual screening challenge was a great platform for researchers working in the field of utilizing computational approaches for fragment based drug discovery. SAMPL3 challenge gives scientific community an opportunity to evaluate their programs, methods and protocols in a blind testing environment. As the test data for SAMPL3 challenge were not previously published anywhere, the blind testing environment would remove any bias associated with performing fragment screening and selecting hits. SAMPL3 challenge also presents an opportunity for researchers to learn from their experiences, share the results and experiences from their approach with those working with different approaches. We participated in SAMPL3 fragment screening challenge to evaluate our fragment based virtual screening protocol that we are planning to use in our drug discovery projects. Although the results of our SAMPL3 fragment screening challenge were poor, there are lots of things we have learned from the SAMPL3 challenge. One of the specific things we learned is the importance of accurate assignment of partial charges to the atoms of ligands. There are several charge methods available and because of the fundamental differences in their algorithm, significant differences may occur in electrostatic assignment on atoms. The charge models not only could affect the docking scores but also the docked poses and hence the docking accuracy. This is even more important in the protein system under study i.e. trypsin in which electrostatic interactions is the dominant force of inhibitor binding and the presence of charged group is essential [60]. We have assigned MMFF94s forcefield based partial charges [39] to SAMPL3 fragment library but the best virtual screening performance could be drawn using Gasteiger charges [54]. The better performance may be due to the accurate description charges on each ligand atoms. Another important thing we learned is using multiple receptor PDBs for docking does not always help. It has been reported in literature that using multiple crystallographic or molecular dynamics derived receptor ensemble in docking improves the virtual screening performance [46–48]. However in our case, even using the best performing receptor could not improve virtual screening performance displayed by using single receptor. The selection of appropriate receptor conformation for docking is very important in virtual screening. We have found that virtual screening performance varied from moderate to worse for different trypsin PDBs used in our virtual screening protocol. In this case, careful selection of crystallographically generated receptor PDBs is very important and non-performing receptor PDBs could be eliminated by performing enrichment analyses if some actives against target protein are known. In our case too many receptor conformations were considered which led to the reduced performance of our virtual screening protocol. The only problem here is that when no inhibitor information is available it is very difficult to find out a priori which receptor PDB is going to perform better than the other. The big take away message from SAMPL3 fragment screening challenge is that the success of a particular charge calculation method/scoring function/receptor conformation/ docking program highly depends upon the target protein system, and performing preliminary evaluation test runs would help to pick up the best combination. SAMPL3 fragment based virtual screening challenge also highlighted one direction for our future research as there is still wide margin of improvement in the fragment based virtual screening protocol involving full flexible docking simulation from RosettaLigand. As seen from the retrospective analysis of our SAMPL3 results, moderate success can be obtained without using any fragment specific settings. The performance may be further enhanced if RosettaLigand scoring function were improved with some fragment specific weighting to energetic terms. Rosetta sampling methodology which simultaneously optimizes protein sidechain, protein backbone and ligand degree of freedom is a major strength of RosettaLigand. However, weights to energetic terms to its scoring function are derived using linear fitting to experimental data composed of mostly drug-size ligands. Therefore, fragment specific weighting may improve docking performance.

Conclusion

Fragment based virtual screening remains a challenging area for computational tools and protocols. Nevertheless, our SAMPL3 fragment screening challenge study suggests that current tools and protocols can be used to identify initial fragment hits for further optimization in a fragment based drug discovery program. Our study provided important points that need to be considered carefully to improve success rate with structure based docking:

(1) Availability of information about inhibitors and protein ligand interaction information always helps. Program parameters and screening protocols that provide better enrichment of actives should be used.

(2) Receptor PDBs for either single receptor docking or multiple receptors docking should be selected after careful analysis of all available ones.

(3) Protein flexibility, along with the ligand flexibility should be considered in docking, either by using full flexible programs like RosettaLigand or by docking to receptor ensembles.

(4) Methods for calculating ligand and protein partial charges should be selected depending upon the protein active site environment and protein ligand interaction information.

(5) Sufficient conformational sampling of fragments within the binding pocket is required to find out the near native binding pose.

References

1. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534

2. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9(10):430–431. doi:10.1016/S1359-6446(04)03069-7

3. Artis DR, Lin JJ, Zhang C, Wang W, Mehra U, Perreault M, Erbe D, Krupka HI, England BP, Arnold J, Plotnikov AN, Marimuthu A, Nguyen H, Will S, Signaevsky M, Kral J, Cantwell J, Settachatgull C, Yan DS, Fong D, Oh A, Shi S, Womack P, Powell B, Habets G, West BL, Zhang KY, Milburn MV, Vlasuk GP, Hirth KP, Nolop K, Bollag G, Ibrahim PN, Tobin JF (2009) Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent. Proc Natl Acad Sci USA 106(1):262–267. doi:10.1073/pnas.0811325106

4. Card GL, Blasdel L, England BP, Zhang C, Suzuki Y, Gillette S, Fong D, Ibrahim PN, Artis DR, Bollag G, Milburn MV, Kim SH, Schlessinger J, Zhang KY (2005) A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. Nat Biotechnol 23(2):201–207. doi:10.1038/nbt1059

5. Tsai J, Lee JT, Wang W, Zhang J, Cho H, Mamo S, Bremer R, Gillette S, Kong J, Haass NK, Sproesser K, Li L, Smalley KS, Fong D, Zhu YL, Marimuthu A, Nguyen H, Lam B, Liu J, Cheung I, Rice J, Suzuki Y, Luu C, Settachatgull C, Shellooe R, Cantwell J, Kim SH, Schlessinger J, Zhang KY, West BL, Powell B, Habets G, Zhang C, Ibrahim PN, Hirth P, Artis DR, Herlyn M, Bollag G (2008) Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. Proc Natl Acad Sci USA 105(8):3041–3046. doi:10.1073/pnas.0711741105

6. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. J Med Chem 51(13):3661–3680. doi:10.1021/jm8000373

7. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6(3):211–219. doi:10.1038/nrd2220

8. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. Nat Chem 1(3):187–192. doi:10.1038/nchem.217

9. de Kloe GE, Bailey D, Leurs R, de Esch IJ (2009) Transforming fragments into candidates: small becomes big in medicinal chemistry. Drug Discov Today 14(13–14):630–646. doi:10.1016/j.drudis.2009.03.009

10. Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H, Spevak W, Zhang C, Zhang Y, Habets G, Burton EA, Wong B, Tsang G, West BL, Powell B, Shellooe R, Marimuthu A, Nguyen H, Zhang KY, Artis DR, Schlessinger J, Su F, Higgins B, Iyer R, D'Andrea K, Koehler A, Stumm M, Lin PS, Lee RJ, Grippo J, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, Chapman PB, Flaherty KT, Xu X, Nathanson KL, Nolop K (2010) Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature 467(7315):596–599. doi:10.1038/nature09454

11. Zartler ER, Mo H (2007) Practical aspects of NMR-based fragment discovery. Curr Top Med Chem 7(16):1592–1599

12. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005) Fragment-based lead discovery using X-ray crystallography. J Med Chem 48(2):403–413. doi:10.1021/jm0495778

13. Neumann T, Junker HD, Schmidt K, Sekul R (2007) SPR-based fragment screening: advantages and applications. Curr Top Med Chem 7(16):1630–1642

14. Kolb P, Ferreira RS, Irwin JJ, Shoichet BK (2009) Docking and chemoinformatic screens for new ligands and targets. Curr Opin Biotechnol 20(4):429–436. doi:10.1016/j.copbio.2009.08.003

15. Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenamin R, Patel S, Wallis N (2007) Application of fragment screening by X-ray crystallography to beta-secretase. J Med Chem 50(6):1116–1123. doi:10.1021/jm0611962

16. Huang JW, Zhang Z, Wu B, Cellitti JF, Zhang X, Dahl R, Shiau CW, Welsh K, Emdadi A, Stebbins JL, Reed JC, Pellecchia M (2008) Fragment-based design of small molecule X-linked inhibitor of apoptosis protein inhibitors. J Med Chem 51(22):7111–7118. doi:10.1021/jm8006992

17. Chen Y, Shoichet BK (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. Nat Chem Biol 5(5):358–364. doi:10.1038/nchembio.155

18. Teotico DG, Babaoglu K, Rocklin GJ, Ferreira RS, Giannetti AM, Shoichet BK (2009) Docking for fragment inhibitors of AmpC beta-lactamase. Proc Natl Acad Sci USA 106(18):7455–7460. doi:10.1073/pnas.0813029106

19. Hubbard RE, Chen I, Davis B (2007) Informatics and modeling challenges in fragment-based drug discovery. Curr Opin Drug Discov Dev 10(3):289–297

20. Babaoglu K, Shoichet BK (2006) Deconstructing fragment-based inhibitor discovery. Nat Chem Biol 2(12):720–723. doi:10.1038/nchembio831

21. May A, Zacharias M (2008) Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking. J Med Chem 51(12):3499–3506. doi:10.1021/jm800071v

22. Meiler J, Baker D (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins 65(3):538–548. doi:10.1002/prot.21086

23. Nabuurs SB, Wagener M, de Vlieg J (2007) A flexible approach to induced fit docking. J Med Chem 50(26):6507–6518. doi:10.1021/jm070593p

24. Anderson AC, O'Neil RH, Surti TS, Stroud RM (2001) Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. Chem Biol 8(5):445–457

25. Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 18(2):178–184. doi:10.1016/j.sbi.2008.01.004

26. Davis IW, Baker D (2009) RosettaLigand docking with full ligand and receptor flexibility. J Mol Biol 385(2):381–392. doi:10.1016/j.jmb.2008.11.010

27. Davis IW, Raha K, Head MS, Baker D (2009) Blind docking of pharmaceutically relevant compounds using RosettaLigand. Protein Sci 18(9):1998–2002. doi:10.1002/pro.192

28. Kawatkar S, Wang H, Czerminski R, Joseph-McCarthy D (2009) Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. J Comput Aided Mol Des 23:527–539. doi:10.1007/s10822-009-9281-4

29. Sandor M, Kiss R, Keseru GM (2010) Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes. J Chem Inf Model 50(6):1165–1172. doi:10.1021/ci1000407

30. Skillman G, Geballe M, Nicholls A (2012) The SAMPL3 challenge. J Comput Aided Mol Des, this issue

31. OpenEye Scientific Software, Inc., Santa Fe, NM, USA. www.eyesopen.com

32. Peat TS, Newman J (2012) High throughput fragment screening-Trypsin. J Comput Aided Mol Des, this issue

33. Newman J, Fazio VJ, Caradoc-Davies TT, Branson K, Peat TS (2009) Practical aspects of the SAMPL challenge: providing an extensive experimental data set for the modeling community. J Biomol Screen 14(10):1245–1250. doi:10.1177/1087057109348220

34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242

35. Congreve M, Carr R, Murray C, Jhoti H (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov Today 8(19):876–877

36. Molecular Operating Environment (MOE), version 2010.10; Chemical Computing Group Inc.: Montreal, QC, Canada, 2010

37. Labute P (2009) Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. Proteins 75(1):187–205. doi:10.1002/prot.22234

38. Labute P (2010) LowModeMD—implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. J Chem Inf Model 50(5):792–800. doi:10.1021/ci900508k

39. Halgren TA (1999) MMFF VI. MMFF94s option for energy minimization studies. J Comput Chem 20(7):720–729. doi:10.1002/(sici)1096-987x(199905)20:7<720::aid-jcc7>3.0.co;2-x

40. Marcou G, Rognan D (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 47(1):195–207. doi:10.1021/ci600342e

41. Rogers DJ, Tanimoto TT (1960) A computer program for classifying plants. Science 132(3434):1115–1118. doi:10.1126/science.132.3434.1115

42. Wang R, Fang X, Lu Y, Wang S (2004) The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 47(12):2977–2980. doi:10.1021/jm030580l

43. Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) The PDBbind database: methodologies and updates. J Med Chem 48(12):4111–4119. doi:10.1021/jm048957q

44. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinforma 12:77. doi:10.1186/1471-2105-12-77

45. R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

46. Bottegoni G, Rocchia W, Rueda M, Abagyan R, Cavalli A (2011) Systematic exploitation of multiple receptor conformations for virtual ligand screening. PLoS One 6(5):e18845. doi:10.1371/journal.pone.0018845

47. Craig IR, Essex JW, Spiegel K (2010) Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. J Chem Inf Model 50(4):511–524. doi:10.1021/ci900407c

48. Huang SY, Zou X (2007) Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins 66(2):399–421. doi:10.1002/prot.21214

49. Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W, Farid R (2008) Improving database enrichment through ensemble docking. J Comput Aided Mol Des 22(9):621–627. doi:10.1007/s10822-008-9182-y

50. Rueda M, Bottegoni G, Abagyan R (2010) Recipes for the selection of experimental protein conformations for virtual screening. J Chem Inf Model 50(1):186–193. doi:10.1021/ci9003943

51. Bolstad ES, Anderson AC (2008) In pursuit of virtual lead optimization: the role of the receptor structure and ensembles in accurate docking. Proteins 73(3):566–580. doi:10.1002/prot.22081

52. Plewczynski D, Lazniewski M, Augustyniak R, Ginalski K (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem 32(4):742–755. doi:10.1002/jcc.21643

53. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23(16):1623–1641. doi:10.1002/jcc.10128

54. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36(22):3219–3228

55. Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. Proteins 8(3):195–202. doi:10.1002/prot.340080302

56. Morris GM, Goodsell DS, Huey R, Olson AJ (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J Comput Aided Mol Des 10(4):293–304

57. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19(14):1639–1662. doi:10.1002/(sici)1096-987x(19981115)19:14<1639:aid-jcc10>3.0.co;2-b

58. Neudert G, Klebe G (2011) DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. J Chem Inf Model. doi:10.1021/ci200274q

59. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 295(2):337–356

60. Shi Y, Jiao D, Schnieders MJ, Ren P (2009) Trypsin-ligand binding free energy calculation with AMOEBA. Conf Proc IEEE Eng Med Biol Soc 2009:2328–2331. doi:10.1109/IEMBS.2009.5335108