J Mol Model (2012) 18:4263–4274
DOI 10.1007/s00894-012-1416-1

ORIGINAL PAPER

FDA approved drugs complexed to their targets: evaluating pose prediction accuracy of docking protocols

Mohammed H. Bohari · G. Narahari Sastry

Received: 21 November 2011 / Accepted: 26 March 2012 / Published online: 8 May 2012

Abstract Efficient drug discovery programs can be designed by utilizing existing pools of knowledge from the already approved drugs. This can be achieved in one way by repositioning of drugs approved for some indications to newer indications. Complex of drug to its target gives fundamental insight into molecular recognition and a clear understanding of putative binding site. Five popular docking protocols, Glide, Gold, FlexX, Cdocker and LigandFit have been evaluated on a dataset of 199 FDA approved drug-target complexes for their accuracy in predicting the experimental pose. Performance for all the protocols is assessed at default settings, with root mean square deviation (RMSD) between the experimental ligand pose and the docked pose of less than 2.0 Å as the success criteria in predicting the pose. Glide (38.7 %) is found to be the most accurate in top ranked pose and Cdocker (58.8 %) in top RMSD pose. Ligand flexibility is a major bottleneck in failure of docking protocols to correctly predict the pose. Resolution of the crystal structure shows an inverse relationship with the performance of docking protocol. All the protocols perform optimally when a balanced type of hydrophilic and hydrophobic interaction or dominant hydrophilic interaction exists. Overall in 16 different target classes, hydrophobic interactions dominate in the binding site and maximum success is achieved for all the docking protocols in nuclear hormone receptor class while performance for the rest of the classes varied based on individual protocol.

M. H. Bohari · G. N. Sastry (✉)
Molecular Modeling Group, Indian Institute of Chemical Technology,
Hyderabad, 500 607 Andhra Pradesh, India
e-mail: gnsastry@gmail.com

Keywords Docking · Drug repositioning · Drug-target complex · FDA approved drugs · Root mean square deviation (RMSD)

Introduction

Drugs designed to interact specifically with certain targets frequently also interact with other proteins, therefore showing several off-target interactions leading to harmful side effects. These interactions can be explored by the pharmaceutical industry to not only derive insight into side effects but also to discover novel therapeutic applications for old drugs [1–3]. These off-target interactions are of great interest. Getting FDA approval for a new chemical entity is a major objective of any drug discovery process. Depending on the therapeutic area, this process can extend over years and the current estimate of developing a novel agent is up to $800 million [4]. Approved drugs have acceptable pharmacokinetics, safety profiles and are accepted by regulatory agencies for human use, and it is important to address whether it is preferable to use them as a whole or as template for further design of new molecules. In addition, the complex of the FDA approved drug to its target gives a) valuable insight into various types of interactions, which may be potentially responsible to attain its required pharmacological action, b) a clear understanding of putative binding site of drug and c) the bioactive conformation of the drug. Therefore, it is desirable to utilize the existing pool of knowledge from approved drugs efficiently before directly developing novel molecules. With the integrative knowledge from the pharmacokinetic profile, bioactive conformation of the drug and binding site characteristics of the target, more potent and promising drug like molecules can be designed.

Computational methodologies have emerged as an indispensible tool in any drug discovery program, playing a key role from hit identification to lead optimization. Amongst these methodologies, molecular docking is the most commonly used method for lead identification and optimization. Its role can be appreciated by looking at the drugs developed in part by computer aided structure based drug design methods, which are in late stages of clinical trials or have reached market [5–7]. In molecular docking, a theoretically modeled complex of protein and ligand is employed for understanding the phenomenon of molecular recognition. This understanding gives fundamental insights into some of the important aspects of drug designing like, mechanism of drug action, subtype selectivity within a target class, which assists in designing more efficient and target selective molecules. Molecular docking has also been acknowledged with significant attention for its role in lead optimization where the known active structure can be tested in a computer model before actually proceeding to synthesis [8–12].

With the increase in the use of computational methodologies in drug discovery, growth in the number of molecular modeling tools with docking capability is increased. Equally, publications evaluating docking algorithms based on their ability to predict correct pose and to score active molecules preferentially over inactive molecules, have been increased and thus studies based on enrichment have become very important [13–29]. The success of a protocol in predicting a ligands' binding pose is usually measured by the RMSD between the experimentally observed ligand and the top ranked solution predicted by the docking protocol. Kontoyianni et al. [13] carried a study of 69 diverse protein-ligand complexes with five docking programs, the overall results ranged from 38 % to 69 % using 2 Å RMSD as a threshold. Warren et al. [14] showed that results varied from 0 % of poses satisfactorily to other combinations where >90 % of poses were within 2 Å RMSD cutoff. Comparison of new program with other ultimately requires a large and diverse test set in order to establish that differences between program success rates are statistically significant. Friesner et al. [15] in evaluation of Glide program have used 202 protein-ligand complexes from Protein Data Bank (PDB) which used most of the proteins of well-known Gold and FlexX test sets. In a number of articles docking programs were evaluated with a small number of targets (e.g., Jones et al. [16] five complexes, Bursulaya et al. [17] 37 complexes, Muryashev et al. [18] 19 complexes, Wang et al. [19] 12 complexes). Perola et al. [20] compared three docking protocols on a dataset of high pharmaceutical relevance and their dataset consisted of 150 proteins-ligand complexes which were initially selected from protein data bank and vertex structure collection according to different criteria. Chen et al. [21] tested four well known commercially available docking programs ICM, Gold, Glide, and FlexX on 164 high resolution protein-ligand complexes with success rate of 83, 79, 55, and 45 percent respectively. Li et al. [22] conducted extensive evaluation on 195 protein-ligand complexes using Glide, Gold, LigandFit, and Surflex, with the former two performing accurately over the latter two programs. Plewczynski et al. [23] evaluated Surflex, LigandFit, Glide, Gold, FlexX, eHiTS, and AutoDock, on the extensive dataset composed of 1300 protein–ligand complexes from PDBbind database, with reported experimental binding affinity values. The results showed that the ligand binding conformation could be identified in most cases, but the lack of universal scoring function for all types of molecules and protein families is still observed. Several studies also evaluate scoring efficiency of docking protocols along with evaluating pose prediction ability [24–29]. There are reviews specially discussing the pros and cons associated with the evaluation of docking protocols [30–34]. In order to assess performance of docking protocols, the diversity of the test set is of primary importance since performance varies as the target is varied. There are various test sets available including that of vendor sets for benchmarking the docking protocols like, Astex-85 [35] having 85 high-quality crystal structures of therapeutically relevant targets and "drug-like" ligands; GOLD benchmarking set [36, 37] derived from the most widely used docking benchmarks; DUD [38] set 38 protein targets for which sets of annotated actives and corresponding property-matched decoys are available for each target.

Docking is one of the promising methods when looking at *in silico* repositioning of drugs as it gives information about the possible mode of binding and to a certain extent the strength of binding [39, 40]. Recently, several studies have been reported where in the database of FDA approved marketed drugs were successfully repositioned using structure based approaches [41–46]. Several databases and servers have been created to identifying possible targets for drugs to be considered for repositioning and to predict drugs polypharmacology for identification of possible side effects [47–50]. Since the performance of docking protocols is highly dependent on the target and the ligand, it becomes mandatory to benchmark the docking protocols for their ability to predict experimental pose on focused dataset of drugs and their targets, before employing docking and scoring for repositioning, virtual screening and lead optimization. Our group's research is focused on practicing different structure-based and analogue-based approaches to design novel molecules with varied application [51–62]. There are several studies carried out in our group using Glide, Gold, Cdocker, LigandFit and FlexX, either singly or in combination and it was observed that the performance varied with the targets [54–60]. These protocols works on an independent principle to produce a ligand docked into the protein, hence comparing the performance on a large and diverse dataset will help us understand the efficiency of these algorithms. The evaluation of these protocols is also essential,

when employing them on larger scale for drug repositioning and for other lead design methods. In this study, we report the results of an extensive evaluation of five popular docking protocols, Glide, Gold, FlexX, Cdocker and LigandFit on a dataset comprised of all the FDA approved drugs in complex to their targets found in PDB. Availability of such complexes in protein data bank in significant number ensured us that our dataset would be one of its kind and highly relevant to drug discovery. To the best of our knowledge there is no study yet specifically addressing evaluation of protocols on dataset of drug target complexes. Hence, we have been motivated to carry out the evaluation of pose prediction ability of docking protocols. Present study provides important information which will help in designing improvised protocols not only for drug repositioning purpose but also for drug-like compound docking and subsequent virtual screening campaign. Performance of docking protocols is also assessed based on different factors like ligand flexibility, resolution of crystal structure, target class and nature of binding site.

Materials and methods

Dataset preparation

The dataset in the study was designed to retain high pharmaceutical relevance and for this purpose none other than drugs themselves would be of choice. Therefore, all the FDA approved drugs present in PDB [63] bound to their targets are considered. There are about 337 FDA approved drugs and nutraceuticals found in 6531 crystal structures in PBD, since focus of the study is on drug-target complexes, 51 nutraceuticals are excluded [64]. From the remaining 286 drugs, 187 are bound to proteins which are not specified as its target in DrugBank [65] and hence they are excluded. Therefore a total of 97 drugs finally remain after removing two molecules viz. DMSO and guanidine, which are not actually drugs. These 97 drugs are present in complex with 207 crystal structures in PDB of which three solution NMR, one theoretical model and four obsolete structures are removed thereby retaining 199 crystal structures in the dataset. The scheme for dataset construction is shown in Fig. 1. We have considered drug-target complex as a different entry if it had different resolution and different source of organism although same target. This was useful in studying the effect of resolution of proteins on the performance of docking protocol though target structure is the same in all other aspects. The distribution of dataset proteins in various resolution ranges in comparison to overall distribution of structures in PDB is shown in Table 1.

Protein preparation

Coordinates for each structure were taken from the RCSB protein data bank and were prepared using the protein preparation wizard incorporated in the Schrodinger software package. Bond orders and formal charges were added for hetero groups and hydrogens were added to all atoms in the system. To optimize the hydrogen bond network, His tautomers and ionization states were predicted, flip of side chains of Asn, Gln, and His residues were assigned and hydroxyl and thiol hydrogens were sampled. Water molecules in all structures were removed. For structures with missing side chain atoms in the active site, the refinement module in Prime was used to correct structure and predict their conformations. For each structure, a brief relaxation was performed using an all atom-constrained minimization carried out with the impact refinement module (Impref) using OPLS 2001 force field to alleviate steric clashes that may exist in the original PDB structures. The minimization terminates when the energy converged or RMSD reaches the maximum cutoff of 0.3 Å.

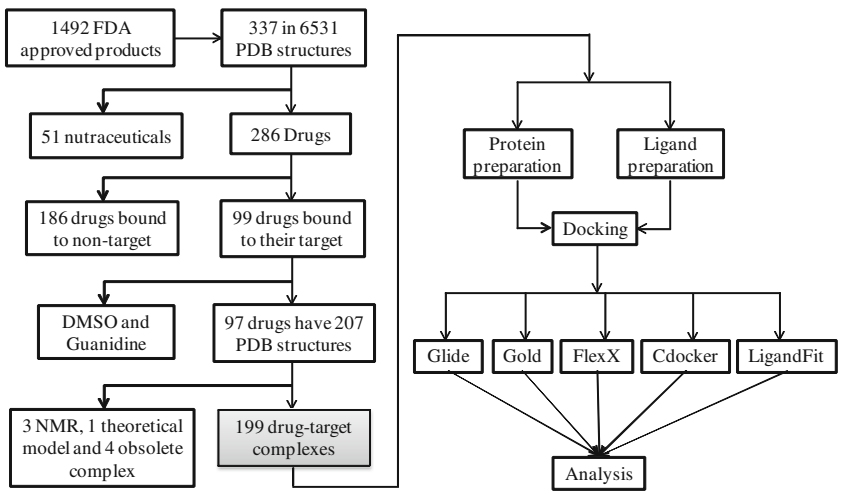


Fig. 1 Overall methodology followed in the study along with the criteria to reach final number of drug-target complexes. Blocks in the leftmost corner shown by thick arrow, are the ones that were excluded from dataset

Table 1 Distribution of dataset proteins in different resolution ranges

Sr. no.	Range (Å)	# of PDB in in dataset	# of PDB protein data bank
1	0.50 – 1.00	1	396
2	1.01-1.50	11	4896
3	1.51 – 2.00	60	22878
4	2.01 – 2.50	74	18001
5	2.51 – 3.00	45	8918
6	3.01 – 3.50	8	2541

Ligand preparation

Co-crystallized drug molecule was extracted from the protein after assigning correct bond order, ionization state and heteroaromatic state. Ligands were extracted from the protein prior to protein preparation and prepared by Ligprep wizard of Schrodinger Inc. Ionization state of ligand was retained without generating tautomers and stereochemistry was determined from 3D structure. Randomization of the starting geometry and the position of the ligands is important and it has been seen that the use of experimentally observed ligand geometries as a starting point has led to increased success rate, an observation rationalized as indicating a lack of coverage of conformational space in docking protocol's internal conformer generator.

Docking protocols

Docking in all the protocols was performed at default settings and without any prior optimization of its protocols except definition of the binding site, so that we can compare how well these protocols perform with vendor optimized default settings on dataset of our interest. Twenty poses were saved for each docking. The comparison of result was done by estimating the RMSD of the ligand heavy atoms between the docked pose and the reference (unoptimized) coordinates from a crystal structure of the complex, using the superposition tool incorporated in maestro. The cumulative fraction employed for analyzing distribution of docked pose in various RMSD ranges, was calculated by counting the number of poses within all the RMSD ranges then taking the cumulative sum followed by fractionating the sum from the total of 199 complexes.

Glide 4.5 [15, 66]

It is grid-based ligand docking with energetic, an algorithm that approximates a systematic search of positions, orientations, and conformations of ligands in the receptor-binding site using a series of hierarchical filters. Flexible ligand docking was performed with standard precision mode and refined poses were scored using GlideScore function [67]. The choice of best pose is made by using Emodel energy score that combines energy grid score, Glide score and internal strain of the ligand, which is better at predicting correct pose than docking score [15]. Default settings were used for both the grid generation and docking. This included scaling of ligand van der Waal radii by 0.8 Å to partially account the suboptimal fits that could be accommodated by minor receptor movements.

Gold 3.2 [16, 68]

Genetic optimization for ligand docking utilizes a genetic algorithm (GA) which mimics the process of evolution by representing ligand descriptors on a chromosome. Default GA settings having population size of 100, selection pressure 1.1, number of operations 100000, niche size 2 and number of island 5 were used for all calculations.

Cdocker 2.1 [69]

It is molecular dynamics (MD) simulated-annealing based algorithm. MD simulated annealing process is performed using rigid protein and flexible ligands. The final minimization of the ligand poses was performed using full potential. The final ranking of the docked poses are based upon total docking energy, including the intramolecular energy for ligand and the ligand-protein interactions. Heating steps of 2000 with target heating temperature of 700ºC and cooling target temperature of 300º in 5000 steps were used for simulated annealing.

LigandFit 2.1 [70]

It employs a shape comparison filter, in combination with a Monte Carlo conformational search for generating ligand poses consistent with the active site shape. Candidate poses were minimized in the context of active site using a grid based method for evaluating protein-ligand interaction energy. In Monte Carlo conformational search, 15,000 trials were employed with energy grid using Dreiding force field and smart minimizer algorithm did the final minimization of the docked poses.

FlexX 3.14 [71, 72]

It is a fast flexible docking method that uses an incremental construction algorithm to place the ligand into active site. The base fragment is automatically selected and is placed into the active site using a pattern recognition algorithm called pose clustering. The remainder of the ligand is built up incrementally from other fragments. Placement of ligand

is scored based on protein ligand interaction. Finally, the binding energy is estimated and placements are ranked. Ligands were docked starting with a base fragment and the best solution is selected according to ChemScore, which takes into account aromatic, hydrogen bonding, ionic, lipophilic interactions, and entropy.

Binding site

In Glide active site was defined by 1000 $Å^3$ box centered on center of mass of native ligand to confine the center of mass of docked ligand; in Gold as all the protein atoms within 5.0 Å of native ligand; in FlexX as residues falling within 5.0 Å radius from the cognate ligand coordinates; in Cdocker as sphere of 10.0 Å radius around native ligand. LigandFit defines binding site based on site finding algorithm, all free grid points (i.e., grid points not occupied by the protein) that lie within the radius of any ligand atoms are determined. The radius of ligand heavy atom is set at 2.5 Å while for the ligand hydrogens the radius is set to 2.0 Å. Thus, site definition is a collection of all grid points occupied by the ligand and un-occupied by the protein.

Results and discussion

Of the numerous poses generated during a docking run, the top 20 are saved, among them the top energy pose (hereafter referred as TE) and top RMSD (hereafter referred as TR) pose are used for analysis, for all the docking protocols. Top energy (TE) pose is the pose with lowest energy, i.e., ranked number one (# 1) by the corresponding scoring function and top RMSD (TR) pose is the pose with lowest RMSD i.e., closest to the experimental pose. The best value obtained from Emodel energy, Gold score Fitness, Cdocker energy and Dockscore of Glide, Gold, Cdocker and LigandFit respectively are considered for TE analysis. General performance of all the docking protocols is assessed by comparing the docked pose with the experimental pose. Docking is considered accurate if the RMSD between docked pose and the experimental ligand pose is less than or equal to 2.0 Å.

Among TE and TR, number and percentage of complexes qualifying the arbitrary accuracy criteria of 2.0 Å and stringent success criteria of 1 Å is shown in Table 2. At 2.0 Å cutoff, performance of Glide and Cdocker is found to be equivalent in TE pose 38.7 % and 34.7 % and TR pose 53.3 % and 58.8 % respectively. While at stringent RMSD cutoff of 1.0 Å, Glide performs comparatively (22.1 %) better than other docking protocols. Cdocker produced lowest average RMSD of 3.93 followed by almost similar results for Glide, LigandFit and Gold, 4.21, 4.32 and 4.33 respectively while FlexX with highest avg. RMSD of 4.93, among the top ranked poses. Figure 2 shows, cumulative fraction of TE and TR pose representing the distribution of docked poses at various RMSD cutoffs (see methods section for method of calculating cumulative fraction). The thin lines represent the TR pose while the thick ones represent TE pose. The magnitude of difference between the respective lines of TE and TR pose for a particular protocol indicates the inefficiency of scoring function to rank the sampled poses. Highest magnitude of difference (Table 2) is observed for Cdocker (48) and least difference is observed for Glide (28). FlexX performance is poor throughout, indicating inefficiency of both the search strategy to sample the correct conformation and the scoring function to rank the sampled poses. There are 36 complexes where metal-drug coordination is found and for the majority of them, docking protocols are unable to produce correct binding pose. This poor performance of protocols indicates the need to include specific parameters to acknowledge metal ligand interaction in the scoring function.

Ideally, the TE pose and TR pose should be the same but it is not true, as in the case of Glide only 40 instances where TE and TR pose are the same and are in success, followed by 33 instances in LigandFit, as compared to fewer cases with other protocols are observed. Top ranked (TE) pose is one that is filtered from the compound library during docking-based virtual screening funnel and therefore is of prime importance over TR pose when looking at

Table 2 General performance of docking protocols with the number (#) and the percentage (%) of complexes within the success criteria at 1.0 Å and 2.0 Å for top ranked pose (TE) and top RMSD (TR) pose

Docking protocol	Top ranked pose (TE)				Top RMSD pose (TR)				Average RMSD	
	#		%		#		%			
	1 Å	2 Å	1 Å	2 Å	1 Å	2 Å	1 Å	2 Å	TE	TR
Glide	44	77	22.1	38.7	68	105	34.2	52.8	4.21	2.74
Gold	20	53	10.1	26.6	37	89	18.6	44.7	4.33	2.96
FlexX	15	44	7.5	22.1	28	77	14.3	38.7	4.92	3.65
Cdocker	25	69	12.6	34.7	40	117	20.1	58.8	3.63	2.34
LigandFit	31	60	15.6	30.1	47	98	23.6	49.2	4.32	2.82

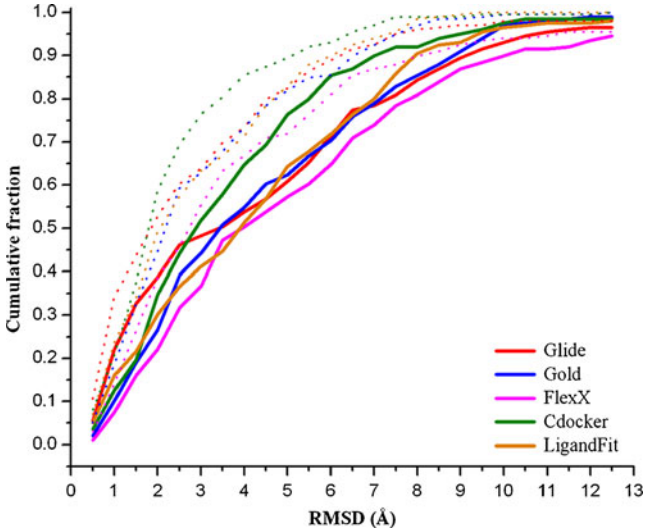


Fig. 2 Cumulative fraction of docked poses for docking protocols at various RMSD cutoffs, when top ranked [in bold line (TE)] pose and top RMSD [in dotted line (TR)] pose are considered

performance of docking protocols. The performance of docking protocols is analyzed as functions of various subsets like resolution of the proteins, ligand flexibility, different target class in dataset and nature of interaction in binding site, and they are discussed in the following sections.

Ligand flexibility

Inadequacy to sample the conformational space of ligand is a well-known problem in docking, which increases with increase in ligand flexibility. The thoroughness of sampling is partially sacrificed by the docking method to maintain reasonable computation time. However, low ligand flexibility can also be difficult to handle although not much conformation sampling ligand is required but precise positioning of ligand in cavity is required. In order to study the effect of ligand flexibility on accuracy of docking protocols, we have classified the docking results based on number of rotatable bonds of co-crystals, into complexes with low (<5), medium (5–9) and high (>9) ligand flexibility. From Table 3, it is clear that the docking protocols perform best when ligands have low or medium flexibility. In low flexibility class LigandFit, Cdocker and Glide perform almost equivalently, with LigandFit best among them. Performance of hierarchical systematic search of Glide is consistent in producing correct pose in low and medium flexibility class while for the rest of the protocols success in medium class varied with significant margin from low flexibility class. Overall efficiency of conformational search strategy of all the docking protocols decreased as ligand flexibility increased, with highest sensitivity for incremental construction of FlexX. Monte Carlo sampling method of Ligandfit is found to be most efficient in handling high ligand flexibility, as compared to search strategy of other protocols.

Resolution of crystal structure

In our dataset, lesser quality structures are purposely retained, which most of the researchers do not include in their dataset as it may badly affect the docking results. This have been done to study the effect of resolution on efficiency of docking protocols in pose prediction. It is worthy to note that for many proteins in drug discovery projects, the 3D structures are solved at low resolutions (>2.0 Å) due to difficulties during crystallization procedures. Nonetheless, they still provide excellent starting point for drug hunting. For the purpose of analyzing the docking results based on resolution of crystal structure, the data set complexes are divided into three class as high (≤2.0; 71), medium (>2.0 - ≤2.50; 75) and low (> 2.50; 53) resolution. From Fig. 3, performance of protocols in different resolution ranges among TE and TR indicates an inverse relationship of resolution with the performance of protocols, since the trend of accurate results for high resolution structure is not observed. Although LigandFit is found to be sensitive to resolution as it followed the trend, producing correct results for higher resolution structures. For Glide, Gold, FlexX and Cdocker the order of accuracy is reversed with resolution, medium or low resolution complexes are in success more than high resolution complexes. One possibility could be the complexity of ligands which to a certain extent is reflected from average number of rotatable bonds in each class. For

Table 3 Percentage of accurately docked drugs, classified based on ligand complexity in terms of number of rotatable bonds and nature of interaction between drug-target

Docking protocol	Ligand complexity			Binding site		
	Low	Medium	High	Hydrophobic	Medium	Hydrophilic
Glide	46	42	19	44.0	62.5	57.5
Gold	41	22	14	33.5	57.0	47.5
FlexX	35	19	7	24.5	53.0	46.5
Cdocker	49	27	21	55.0	57.0	62.5
LigandFit	48	17	28	37.5	50.0	52.5

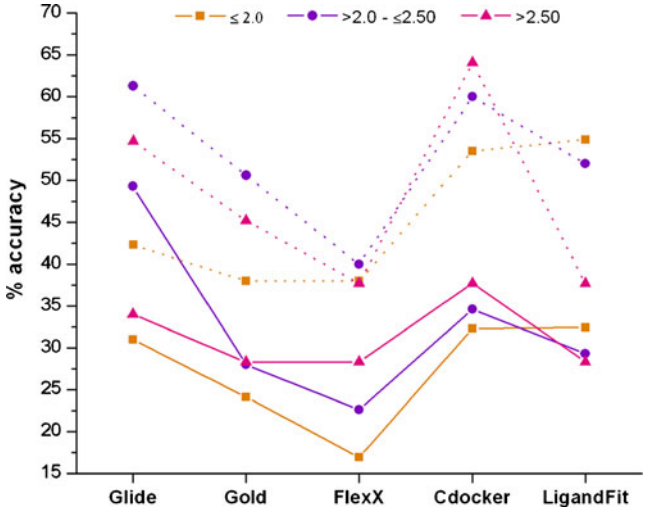


Fig. 3 Percentage accuracy of docking as a function of resolution of crystal structure of protein at 2.0 Å RMSD cutoff among top ranked pose (normal lines) and top RMSD pose (dotted lines)

high it is 9.4, for medium it is 8.4 and for low it is 7.4 although the number of complexes in each class varies. Overall structural variations due to resolution of the protein structures in the given resolution range (0.50-3.50 Å) are not very significant in lowering the accuracy of docking protocols on our dataset which contradicts the earlier reported results [30].

Different target classes

For 97 drugs (199 crystal structures) in the dataset, a total of 85 unique targets (39 redundant) exist. To assess the performance of docking protocols on different target families, grouping of these 85 targets is done based on their biological functions. A group of target family is created if more than two crystal structures are available for that particular target in the dataset. This grouping criteria reduced the dataset complexes to 130 which are then grouped into 16 different target classes and the remaining complexes are included in the miscellaneous class. Targets with different isoforms are included in the same class like in the case of carbonic anhydrase where CA-1, CA-2, CA-14, CA-13 are present and phospholipase A2 where PLA2 isoform2 are present. All the grouped target classes along with the drugs bound, number and PDBID of crystal structure are presented in Table 4.

The distribution of number of rotatable bonds (RB), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA) of drugs (in column bar) and hydrophilic and hydrophobic interactions between drug-target (in stack bar), for the dataset complexes classified into different target classes are presented in Fig. 4. The column bar properties are computed using QikProp module of Schrodinger package, while the stack bar properties are obtained by calculating the proportion of different types of interactions viz. hydrogen bonding, metal interaction and hydrophobic interaction, in the binding site. Each stack column represents interaction involved in a drug-target complex. The X-axis, for column bars shows drug id while, that for stack bar shows the corresponding PDBIDs for each drug-target complex. In graph, D, E, F, G and H the PDBIDs are grouped into different boxes representing a target class and each sub-box in a box represents a particular drug. While, in graph A, B and C each box represents a particular drug. From Fig. 4, it is found that in the majority of complexes hydrophobic type of interaction is dominate while there are only four classes where hydrophilic interactions are dominate.

Among the four target classes viz. carbonic anhydrase (Zn), farnesyl pyrophosphate synthase (Zn), hydroxylase (Fe), three of them have metal drug co-ordination type of interaction. Analysis of results revealed that Cdocker gave maximum success in TE pose in CA and FS class. For hydroxylase class all the protocols perform equivalently for the same two complexes (1PBC, 1PBF) and the remaining three complexes (3PAH, 4PAH, 6PAH) out of five which are in failure are phenylalanine hydroxylase containing iron in the active site. Table 5 gives the details about the number of accurately docked complexes in all the protocols, among TE and TR. Best performance by all the protocols except Gold is observed for nuclear hormone receptor class. In HMG CoA reductase class Glide and Gold perform better. Cdocker and LigandFit outperform all the other protocols in HIV protease class by reproducing correct pose five out of six times. For the complexes of phospholiapseA2 (PL), none of the protocols are able to sample the experimental pose. We analyze the results obtained for PLA2 class; drugs involved are (Table 4) aspirin, atropine, amino salicylic acid and niflumic acid involved in the class with rotatable bonds 3, 8, 4 and 4 respectively and corresponding six crystal structures with resolution <2.50. These two factors, i.e., resolution and ligand *flexibility* are within acceptable range, i.e., protocol should result in success, therefore, failure in producing correct pose remains unclear. Active site analysis is done for PLA2 class and it is found that active site cavity is shallow and dominated by hydrophobic residues. For Glide and Gold, the docked pose is obtained in such a way that the major part of the ligand is buried in the cavity as compared to the experimental ligand pose that partially covers the cavity. For atropine with two crystal structures in the study, only single geometry was used for docking, but for analysis of failure due to insufficient sampling of ligands conformational space all the possible conformers of atropine are generated and used for docking. For both the complexes, correct docked pose is obtained after conformation generation, indicating lack of ability of protocols to cover the conformational space of ligand.

Table 4 Target class name with the drug/drugs involved with number of crystal structure for that particular drug in the dataset in brackets and their respective PDBIDs

Sr. no.	Class	Drugs involved	PDBID
1	Carbonic anhydrase (CA)	Acetazolamide (1); Topiramate (2)	3 dc3, 1rj6, 1jdo, 1azm, 1ydb, 2h4n, 1zsb, 1ydd, 3hku
2	Phospholipase A2 (PA)	Aspirin (1); Atropine (2); Aminosalicylic acid (3); Niflumic acid (4)	1oxr, 1tgm, 1th6, 2arm, 1sxk, 1td7
3	HMG CoA reductase (HR)	Atrovastatin (1); Fluvastatin (2); Simvastatin (4); Rosuvastatin (3)	1hwk, 1hwi, 1hw9, 1hwl
4	Betalactamase (BL)	Cefalotin (1); Cefoxitin (2); Cloxacillin (3)	1kvl, 1ymx, 1i2w, 1fcm
5	Nuclear hormone receptor (NH)	Dexamethasone (2); Diethylstilbesterol (3); Mefepristone (4); Norethindrone (5); Norgestrel (6); Progesterone (7); Raloxifene (8); Spironolactone (9); Conjugated estrogen (1); Rosiglitazone (10)	1p93, 1m2z, 1s9p, 3erd, 2w8y, 1a52, 1qku, 1qkt, 1gwr, 1 g50, 1nhz, 1sqn, 3d90, 1a28, 1err, 2qxs, 2jfa, 2ab2, 2oax, 1 fm6, 3cs8, 3dzy, 1zgy, 2prg
6	Kinase (KN)	Erlotinib (1); Ganciclovir (2); Gefitinib (3); Imatinib (4); Sorafenib (7); Sunitinib (8); Amiloride (5); Dasatinib (6)	1 m17, 1ki2, 2ito, 2ity, 2itz, 1 t46, 2hyy, 1uwh,1uwj, 3g0e, 3gof, 1f5l, 2gqg
7	Phosphodiestrase 5A (PD)	Sildenafil(1); Tadalafil(2); Vardenafil (3)	3jwq, 2 h42, 1udt, 1tbf, 1udu, 1x0z, 1xp0, 1uho, 3b2r
8	Farnesyl pyrophosphate synthase (FS)	Alendronate (1); Ibandronate (2); Pamidronate (3); Zoledronate (4)	2f92, 1yhm, 2f94, 2f89, 1zw5, 2f8z, 2f8c, 2f9k
9	Transthyretin (TT)	Diethylstilbesterol (1); Diflunisal (2); Levothyroxine (3)	1tt6, 1tz8, 3d2t, 2rox, 1eta, 1etb, 1ie4, 1sn0, 1ict
10	Plasma protein (PP)	Diclofenac (1); Diflunisal (2); Ibuprofen (3); Iodipamide (4); Levothyroxine (5); Propofol (6); Salicylic acid (7)	3cfq, 2bxe, 2bxg, 2bxn, 1hk1, 1hk2, 1hk3, 1hk4, 1hk5, 1e7a, 2i2z, 2i30, 3b9m
11	Acetyltransferase (AT)	Chloramphenicol (1); Fusidic acid (2); Isonazide (3); Carnitine (4)	3cla, 4cla, 2xat, 1cla, 1qca, 1w6f, 1s5o
12	HIV protease (HP)	Darunavir (1); Lopinavir (2)	2hs1, 2idw, 2hs2, 2ien, 2ieo, 1rv7
13	Alpha thrombin (TR)	Proflavin (1); Suramin (2)	3bf6, 2h9t, 1bcu
14	Thymidylate synthase (TS)	Raltitrexed (1); Pamitrexed (2); Pyrimethamine (3)	1hvy, 1i0o, 1ju6, 1juj, 1j3j
15	Hydroxylase (HY)	Epinephrine (1); Levodopa (2); Norepinephrine (3); Aminosalicylic0 acid (4)	3pah, 6pah, 4pah, 1pbc, 1pbf
16	Dehydrogenase (DG)	Flurouracil (1); Mannitol (2); Mycophenolic acid (3); Ursedeoxycholic acid(4)	1h7x, 1m2w, 1me7, 1meh, 1mei, 1jr1, 1ihi

Type of interaction in binding site

In order to divide the dataset based on type of interaction, degree of hydrogen bonding (DHB) for each complex is calculated. DHB is ratio of number of hydrogen bonds between protein and ligand to the number of heavy atoms in ligand [20]. DHB indicates whether the interaction between drug and its target is hydrogen bond driven or hydrophobic driven. It is an indirect measure of nature of binding site, since hydrophilic binding site favors more hydrogen bonding and there by a high degree of hydrogen bonding while hydrophobic binding site favors the hydrophobic interactions. Table 3 shows the distribution of success based on the interaction driven criteria between drug and its target. Complexes with degree of hydrogen bonding, 0.1 or less are classified as hydrophobic driven, i.e., low DHB and ones with 0.2 or more are hydrogen bond driven, including ligand-metal interaction, i.e., high DHB while the remaining complexes have medium DHB. Performance of docking protocols remains almost similar for binding site with balance of hydrophilic and hydrophobic interaction (see Table 3). There is a decrease in performance of all the docking protocols as hydrophobicity of the binding site increases, with highest sensitivity in the case of FlexX. The variation in performance of Cdocker is less, indicating its independent performance on all types of binding site.

Analyzing pose prediction failures

Of the 199 crystal structures, for 76 complexes correct pose are not predicted by any of the docking protocols employed, while for 40 of them the correct conformation is not even sampled. Here we analyze the possible reasons of the failure in prediction of correct binding pose for these 40 complexes. As ligand flexibility is found to be one of the significant factor in affecting the performance (i.e., number of rotatable bonds >10), these 40 complexes are analyzed for this factor and 18 of them come under this category. For the remaining

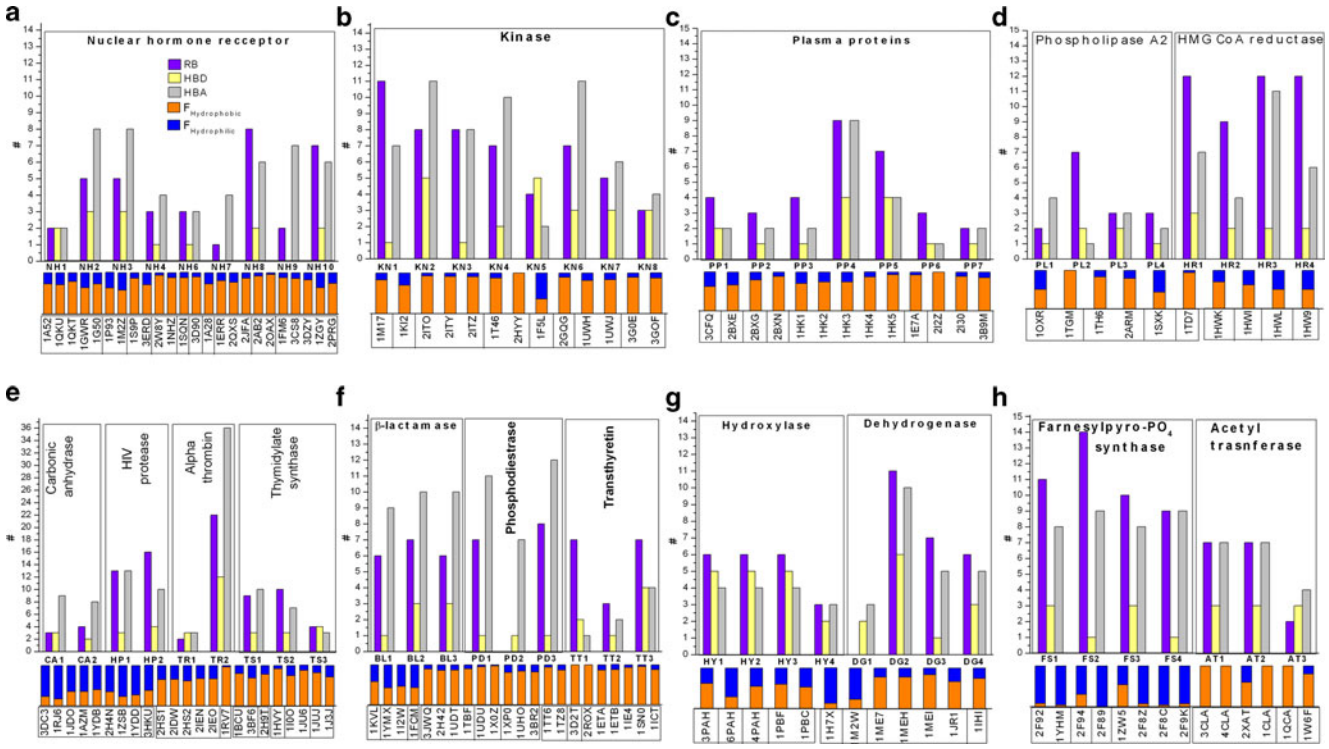


Fig. 4 Distribution of number of rotatable bonds (RB), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA) of drugs (in column bar) and hydrophilic (blue) and hydrophobic (orange) interactions between drug-target (in stack bar), for the dataset complexes in grouped target classes

Table 5 Effect of different target class on performance of docking protocols among top rank pose (TE) and top RMSD (TR) pose. Table shows number of accurately docked complexes at 2.0 Å RMSD cutoff, number of drugs in the class and their corresponding number of crystal structures

Sr. no.	Target class	# of complexes										# of drugs	# of PDB
		Glide		Gold		FlexX		Cdocker		LigandFit			
		TE	TR	TE	TR	TE	TR	TE	TR	TE	TR		
1	Carbonic Anhydrase (CA)	2	7	1	4	1	4	3	6	2	5	2	9
2	Phospholipase A_2 (PL)	0	0	0	0	0	0	0	0	0	0	4	6
3	HMG CoA reductase (HR)	3	3	3	3	1	1	2	3	0	2	4	4
4	Betalactamase (BL)	1	2	1	1	0	2	0	2	1	3	3	4
5	Nuclear hormone receptor (NH)	18	22	16	23	10	15	17	21	18	19	9	24
6	Kinase (KN)	8	8	4	4	3	5	6	10	2	5	8	13
7	Phosphodiestrase 5A (PD)	2	3	2	6	0	1	5	8	1	4	3	9
8	Farnesylpyrophosphate synthase (FS)	0	0	1	3	1	2	3	6	0	0	4	8
9	Transthyretin (TT)	4	6	0	2	1	2	1	3	1	2	3	9
10	Plasma protein (PP)	4	8	1	6	1	1	3	6	2	3	7	13
11	Acetyltransferase (AT)	0	1	0	2	1	2	0	3	1	2	4	7
12	HIV protease (HP)	0	0	0	1	0	4	5	5	5	5	2	6
13	Alpha thrombin (TR)	0	1	0	0	0	1	1	1	1	1	2	3
14	Thymidylate synthase (TS)	1	2	0	1	1	1	0	1	0	0	3	5
15	Hydroxylase (HY)	2	2	2	2	2	2	2	2	2	2	4	5
16	Dehydrogenase (DG)	2	4	2	4	3	5	2	5	0	6	4	7
17	Miscellaneous	31	38	20	27	19	29	21	37	25	40	38	67

complexes ligand flexibility does not seem to be the issue. The majority of the remaining complexes have their co-crystals partially buried in protein or the active site is shallow and solvent exposed. For predicting correct pose for a complex with small ligands, docking protocols are required to place the ligand correctly more than sampling the conformation. For a partially buried ligand docking algorithms have more options to place ligand than for a completely buried ligand and therefore this becomes one of the reasons for failure in pose prediction. In a couple of complexes (2ARM, 1TH6) the co-crystal is conformationally strained as compared to docked ligand and hence becomes one of the reasons for failure. In a couple of complexes (1TBF, 2ITO) the ligand is bigger in size of which the core fragment of ligand is placed correctly by protocols, but presence of flexible side chain leads to overall failure (see Fig. 5). In one particular example (1P7R) heme is present in active site and the co-crystallized ligand did not show any interaction with heme iron which is very common for any heme protein, but TE pose from all the docking protocols show the presence of electrostatic interaction between iron of heme and nitrogen atom of ligand (see Fig. 6). Kellenberger et al. [27] reported similar types of observations for the failure in pose prediction.

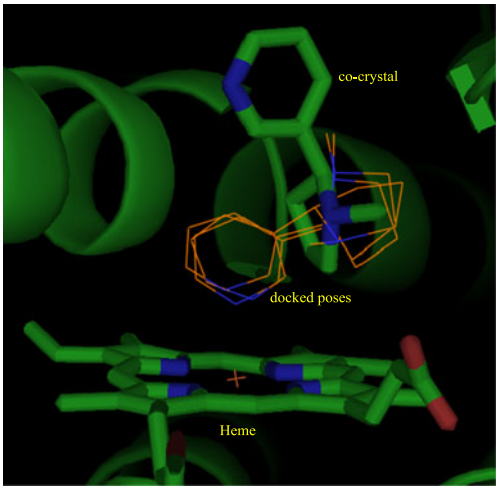
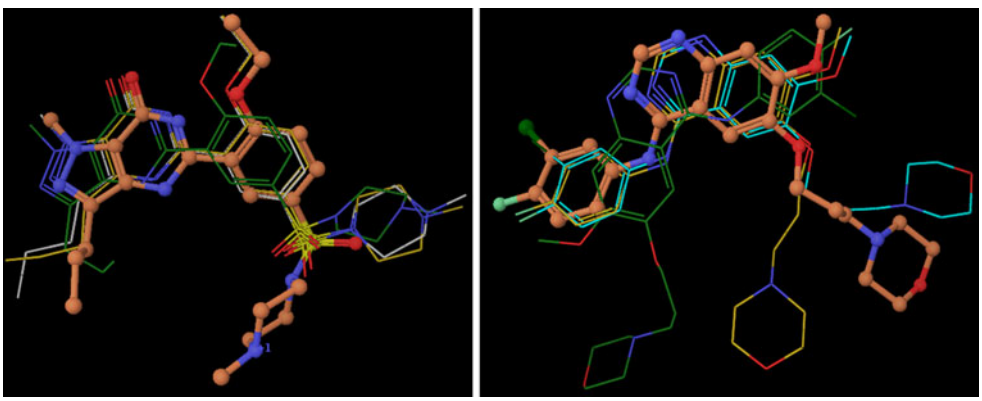


Fig. 6 Co-crystallized ligand along with docked pose from different docking protocols for 1P7R complex

Conclusions

FDA approved drugs form a potential starting point for structure-based drug design campaign. 3D structures of these drugs in complex to their targets will provide fundamental insights into phenomenon of molecular recognition. The goal of the present study is to evaluate Glide, Gold, FlexX, Cdocker and LigandFit for their ability to accurately predict the experimental pose on a dataset of FDA approved drug target complexes (97 drugs with 199 corresponding crystal structures). Also, to characterize docking outcomes for the test set as a whole and subsets based on resolution of proteins, ligand flexibility, nature of interaction between target and drug and target family. Performance of all the protocols are compared at vendor optimized default settings, providing optimal protocol for a particular target. Glide appeared to perform best (37.8 %) in pose prediction as compared to other protocols, but pose sampling efficiency of Cdocker (58.8 %) is higher than that of Glide (53.7 %). LigandFit, Cdocker and Glide perform well at low ligand flexibility; Glide performs equivalently with both low and medium ligand flexibility, while highly flexible ligands are sampled well with Monte Carlo simulation of LigandFit. It is also found that number of rotatable bonds and different target class affects accuracy of docking protocol very significantly while effect of resolution range is not very significant. Performance of all the docking protocols decreases as the binding of ligand to its target is dominated by hydrophobic interactions except Cdocker which is found to be independent of it. During analysis of binding site interaction in different target classes, it was found that hydrophobic interactions are the dominating forces governing the binding of ligand. In terms of performance of protocols on different



1TBF 2ITO

Fig. 5 Docked poses (shown in lines) from different docking protocols superimposed on experimental conformation (shown in ball and sticks)

targets class, all the protocols perform well on nuclear hormone receptor class with the rest of the classes performance varied based on individual protocol. A lot more effort is needed for understanding complex phenomenon of drug-target interaction and improving the way in which docking protocols can explain it to a better extent, though in the present scenario with our findings we propose that Glide is the most efficient protocol in predicting experimental pose of a drug bound to its target.

Acknowledgments GNS acknowledges financial support from Department of Biotechnology and Department of Science and Technology for Swarnajayanthi fellowship. MB gratefully acknowledges National Institute of Pharmaceutical Education and Research (NIPER) Hyderabad for financial support.

References

1. Chong CR, Sullivan DJ Jr (2007) New uses for old drugs. Nature 448:645–646
2. Nielsch U, Schafer S, Wild H, Busch A (2007) One target - multiple indications: a call for an integrated common mechanisms strategy. Drug Discov Today 12:1025–1031
3. Tobinick EL (2009) The value of drug repositioning in the current pharmaceutical market. Drug News Perspect 22:119–125
4. O'Connor KA, Roth BL (2005) Finding new tricks for old drugs: an efficient route for public-sector drug discovery. Nat Rev Drug Discov 4:1005–1014
5. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3:935–949
6. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein small molecule docking methods. J Comput Aided Mol Des 16:151–166
7. Borman S (2005) Drugs by design. Chem Eng News 83:28–30
8. Srivani P, Usharani D, Jemmis ED, Sastry GN (2008) Subtype selectivity in phosphodiesterase 4 (PDE4): a bottleneck in rational drug design. Curr Pharma Des 14:3854–3872
9. Badrinarayan P, Sastry GN (2011) Virtual high-throughput screening in new lead identification. Comb Chem High Thr Scr 14:840–860
10. Murthy JN, Nagaraju M, Sastry GM, Rao AR, Sastry GN (2006) Active site acidic residues and structural analysis of human aromatase: molecular modeling study based on mammalian CYP2C. J Comput Aided Mol Des 19:857–870
11. Reddy AS, Pati PS, Kumar PP, Pradeep HN, Sastry GN (2007) Virtual screening in drug discovery- A computational perspective. Curr Prot Pep Sci 8:331–353
12. Srivastava HK, Bohari M, Sastry GN (2012) Modeling anti-HIV compounds: The role of analogue based approaches. Curr Comput Aided Drug Des (in press)
13. Kontoyianni M, McClennan LM, Sokol GS (2004) Evaluation of docking performance: comparative data on docking algorithm. J Med Chem 47:558–565
14. Warren GL, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindrall M, Nevins N, Sesmus SF, Senger S, Tedesco G, Wall ID, Wolren JM, Peishoff CG, Head MS (2006) Critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931
15. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelly M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749
16. Jones G, Willett P, Glen C, Leach AR, Taylor R (1997) Development and validation of genetic algorithm for flexible ligands docking. J Mol Biol 267:727–748
17. Bursulaya BD, Totrov M, Abagyan R, Brooks CL (2003) Comparative study of several algorithms for flexible ligands docking. J Comput Aided Mol Des 17:755–763
18. Muryashev AE, Tarasov DN, Butygin AV, Butigina O, Aleksandrov AB, Nikitin SM (2003) A novel scoring function for molecular docking. J Comput Aided Mol Des 17:597–605
19. Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 46:2287–2303
20. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on system of pharmaceutical relevance. Proteins Struct Funct Bioinf 56:235–249
21. Chen H, Lyne PD, Giordanetto F, Lovell T, Li J (2006) On evaluating molecular docking methods for pose prediction and enrichment factors. J Chem Inf Model 46:401–415
22. Li X, Li Y, Cheng T, Liu Z, Wang R (2010) Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. J Comput Chem 31:2109–2125
23. Plewczynski D, Łaźniewski M, Augustyniak R, Ginalski K (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem 32:742–755
24. Bissantz C, Folkers G, Rognan D (2000) Protein based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 43:4759–4767
25. Schulz-Gasch T, Stahl M (2003) Binding site characteristics in structure based virtual screening: Evaluation of current docking tools. J Mol Model 9:47–57
26. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL III (2004) Assessing scoring functions for protein-ligand interactions. J Med Chem 47:3032–3047
27. Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins Struct Funct Bioinf 57:225–242
28. Englebienne P, Fiaux H, Kuntz DA, Corbeil CR, Gerber-Lemaire S, Rose DR, Moitessier N (2007) Evaluation of docking programs for predicting binding of golgimannosidase II inhibitors: a comparison with crystallography. Proteins Struct Funct Bioinf 69:60–176
29. Cross JB, Thompson DC, Rai BK (2009) Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. J Chem Inf Model 49:1455–1474
30. Verkhiver GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, Colson AB, Freer SII, Larson V, Luty BA, Marrone T, Rose PW (2000) Deciphering common failures in molecular docking of ligand protein complexes. J Comput Aided Mol Des 14:731–751
31. Cole JC, Murray CW, Nissink WMJ, Taylor RD, Taylor R (2005) Comparing protein-ligands docking programs is difficult. Proteins Struct Funct Bioinf 60:325–332
32. Leach A, Shoichet B, Peishoff C (2006) Prediction of protein-ligand interactions. Docking and scoring: success and Gaps. J Med Chem 49:5851–5855
33. Jain A, Nicholls A (2008) Recommendations for evaluation of computational methods. J Comput Aided Mol Des 22:133–139
34. Jain AN (2008) Bias, reporting and sharing: Computational evaluations of docking methods. J Comput Aided Mol Des 22:201–212
35. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WT, Mortenson PN, Murray CW (2007) Diverse, high quality test set for the validation of protein-ligand docking performance. J Med Chem 50:726–741
36. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, Rizzo RC (2006) Development and validation of a modular,

extensible docking program: DOCK 5. J Comput Aided Mol Des 20:601–619
37. Nissink JW, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R (2002) A new test set for validating predictions of protein-ligand interaction. Proteins Struct Funct Bioinf 49:457–471
38. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49:6789–6801
39. Li YY, An J, Jones SJM (2006) A large scale computational approach to drug repositioning. Genome Inform 17:239–247
40. Ekins S, Williams AJ, Krasowski MD, Freundlich JS (2011) In silico repositioning of approved drugs for rare and neglected diseases. Drug Discov Today 16:298–310
41. Bisson WH, Cheltsov AV, Bruey-Sedano N, Lin B, Chen J, Goldberger N, May LT, Christopoulos A, Dalton JT, Sexton PM, Zhang X-K, Abagyan R (2007) Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. Proc Natl Acad Sci 17:11927–11932
42. Clouser CL, Patterson SE, Mansky LM (2010) Exploiting drug repositioning for discovery of a novel HIV combination therapy. J Virol 84:9301–9309
43. Kinnings SL, Liu TPJ, Jackson RM, Xie L, Bourne PE (2011) A machine learning-based method to improve docking scoring functions and its application to drug repurposing. J Chem Inf Model 51:408–419
44. Bernard P, Dufresne-Favetta C, Favetta P, Do QT, Himbert F, Zubrzycki S, Scior T, Lugnier C (2008) Application of drug repositioning strategy to TOFISOPAM. Curr Med Chem 15:3196–3203
45. Leung CH, Chan DSH, Kwan MHT, Wong ZCCY, Zhu GY, Fong WF, Ma DL (2011) Structure-based repurposing of FDA-approved drugs as TNF-a inhibitors. Chem Med Chem 6:765–768
46. Baures PW, Oza VB, Peterson SA, Kelly JW (1999) Synthesis and evaluation of inhibitors of transthyretin amyloid formation based on the non-steroidal anti-inflammatory drug, flufenamic acid. Bioorg Med Chem 7:1339–1347
47. Xu K, Cote TR (2011) Database identifies FDA-approved drugs with potential to be repurposed for treatment of orphan diseases. Brief Bioinform 12:341–345
48. Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 7:1–9
49. Luo H, Chen J, Shi L, Mikailov M, Zhu H, Wang K, He L, Yang L (2011) DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical–protein interactome. Nucleic Acids Res 39:W492–W498
50. von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissne R (2011) PROMISCUOUS: A database for network-based drug-repositioning. Nucleic Acids Res 39:D1060–D1066
51. Bohari M, Srivastava HK, Sastry GN (2011) Analogue based approaches in anticancer compound modelling: The relevance of QSAR models. Org Med Chem Lett 1:3–15
52. Janardhan S, Srivani P, Sastry GN (2006) 2D and 3D quantitative structure-activity relationship studies on a series of bis-pyridinium compounds as choline kinase inhibitors. QSAR Comb Sci 25:860–872
53. Srivani P, Sastry GN (2009) Potential choline kinase inhibitors: a molecular modeling study of bis-quinolinium compounds. J Mol Graph Mod 27:676–688
54. Badrainarayan P, Sastry GN (2011) Sequence, analysis of p38 MAP kinase: exploiting DFG-out conformation as a strategy to design new type II leads. J Chem Inf Model 51:115–129
55. Badrinarayan P, Sastry GN (2012) Virtual screening filters for the design of type II p38 MAP kinase inhibitors: a fragment based library generation approach. J Mol Graph Model 34:89–100
56. Ravindra GK, Achaiah G, Sastry GN (2008) Molecular modeling studies of phenoxypyrimidinyl imidazoles as p38 kinase inhibitors using QSAR and docking. Eur J Med Chem 43:830–838
57. Ravindra GK, Srivani P, Achaiah G, Sastry GN (2007) Strategies to design pyrazolyl urea derivatives for p38 kinase inhibition: A molecular modeling study. J Comput Aided Mol Des 25:155–166
58. Srivastava HK, Chourasia M, Kumar D, Sastry GN (2011) Comparison of computational methods to model DNA minor groove binders. J Chem Inf Model 51:558–571
59. Srivani P, Srinivas E, Raghu R, Sastry GN (2007) Molecular modeling studies of pyridopurinone derivatives - Potential phosphodiesterase 5 inhibitors. J Mol Graph Model 26:378–390
60. Srivani P, Kiran K, Sastry GN (2006) Understanding the structural requirements of triarylethane analogues towards PDE-IV inhibitors: A molecular modeling study. Ind J Chem A 45A:68–76
61. Chourasia M, Sastry GM, Sastry GN (2005) Proton binding sites and conformational analysis of H+K+−ATPase. Biochem Biophysic Res Commun 336:961–966
62. Bindu PH, Sastry GM, Murty US, Sastry GN (2004) Structural and conformational changes concomitant with the E1-E2 transition in H+K+−ATPase: a comparative protein modeling study. Biochem Biophysic Res Commun 319:312–320
63. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Veissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
64. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM (1997) PDBsum: a web based database of summaries and analysis of all PDB structures. Trends Biochem Sci 22:488–490 [Data retrieved in November 2009]
65. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34:D668–D672
66. Halgen TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. Enrichment factor in database screening. J Med Chem 47:1750–1759
67. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring function 1. The development of fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 11:425–445
68. Jones G, Willett P, Glen C (1995) Molecular recognition of receptor site using genetic algorithm with a description of desolvation. J Mol Biol 245:43–53
69. Wu G, Robertson DH, Brooks CL III, Vieth M (2003) Detailed analysis of grid based molecular docking: a case study of C-DOCKER – A CHARMm based molecular docking algorithm. J Comput Chem 24:1549–1562
70. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph Model 22:289–307
71. Rarey M, Kramer B, Lengauer T (1996) A fast flexible docking method using incremental construction algorithm. J Mol Biol 261:470–489
72. Kramer B, Rarey M, Lengauer T (1999) Evaluation of FlexX incremental construction algorithm for protein ligands docking. Proteins Struct Funct Genet 37:228–241