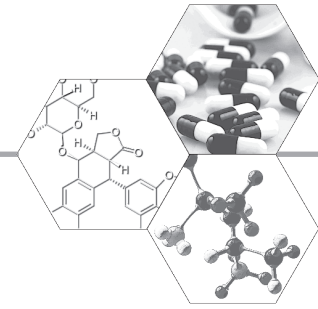

Reverse docking: a powerful tool for drug repositioning and drug rescue

Reverse or inverse docking is proving to be a powerful tool for drug repositioning and drug rescue. It involves docking a small-molecule drug/ligand in the potential binding cavities of a set of clinically relevant macromolecular targets. Detailed analyses of the binding characteristics lead to ranking of the targets according to the tightness of binding. This process can potentially identify novel molecular targets for the drug/ligand which may be relevant for its mechanism of action and/or side effect profile. Another potential application of reverse docking is during the lead discovery and optimization stages of the drug-discovery cycle. This review summarizes the state-of-the-art and future prospects of the reverse docking with particular emphasis on computational molecular design.

Prashant S Kharkar*, Sona Warrier & Ram S Gaud

SPP School of Pharmacy & Technology Management, SVKM's NMIMS, V. L. Mehta Road, Vile Parle (West), Mumbai-400 056, India
*Author for correspondence:
Tel.: +91 22 4233 2016
Fax: +91 22 2618 5422
prashant.kharkar@nmims.edu

Drug discovery has become extremely difficult lately. The cost of bringing a **new molecular entity** into the market is in excess of US$2 billion [1]. The increased expenditure coupled to the 'over-cautious' stand of the regulatory bodies have only made the situation worse in terms of 'productivity' as measured by the number of new molecular entities launched per year. The return-on-investment has become a serious concern. It has been said that the 'blockbuster era is over'. The future will tell us how the pharmaceutical industry worldwide survives unless the present approaches to drug discovery are modified critically.

Drug repositioning or repurposing and **drug rescue** are two promising approaches to address the 'productivity gap' that the global pharmaceutical giants are currently facing [2]. Such 'lateral' strategies will lead to identification of novel and better treatment options with substantial cost and time savings as demanded by the 'starting-from-scratch' version of the conventional drug discovery. This is particularly important for the discovery of innovative medicines needed for the treatment of neglected and/or orphan diseases. The leading pharmaceutical companies are mostly reluctant to invest in such therapeutic areas, fearing negligible return-on-investment. There are several success stories of drug repositioning and drug rescue, such as sildenafil and thalidomide [3]. NIH (MD, USA) have initiated a program named 'Discovering New Therapeutic Uses for Existing Molecules' in 2012 with the aim of discovering innovative treatments using abandoned and existing molecules/drugs [101,102]. In addition to the abovementioned initiatives based on experimentation for drug repositioning and drug rescue, computational (virtual or *in silico*) approaches can prove extremely valuable in identifying potential opportunities in these fields. Of the several techniques for generating computational repositioning hypotheses, systematic analysis of transcriptomics, side effects and genetics (genome-wide association study), in addition to the data obtained from electronic health records and phenotypic screening campaigns, are proving to be very useful [4].

Virtual screening using molecular docking represents a structure-based computational strategy wherein a large collection of synthesized and/or hypothetical or designed molecules is screened for a macromolecular binding-site complementarity (shape and electrostatics) [5]. The molecules are then ranked according to the docking score, binding energy or related parameters. Such a ranking helps in identifying potential binders, which are then prioritized for procurement or synthesis. In brief, this is 'one target-many ligands' scenario which is common in the conventional drug discovery. The critical parameters in molecular docking are the docking process itself and scoring of the docked poses. Several docking algorithms/programs exist which use various **scoring functions**. The scoring functions may be suitable for certain ligand chemotypes and the macromolecular targets. The performance of the docking programs and the scoring functions therein may, thus, be variable depending on the ligand chemical matter and the target. Hence, development of universal docking algorithms with consensus scoring functions to address the inherent difficulties involved in the molecular docking, is an intensive area of computational research [5,6].

FUTURE SCIENCE part of fsg

Key Terms

New molecular entity: Drug that contains an active moiety that has never been approved by US FDA or marketed in the USA.

Drug repositioning: Process of finding 'new' uses for the approved (old or existing) drugs.

Drug rescue: Research using small molecules/biologics previously used in studies but not progressed or developed further and submitted to the regulatory bodies for approval.

Scoring function: Fast, usually approximate mathematical models used for estimating binding affinity (the strength of the noncovalent interactions between two molecules, a small-molecule ligand and a macromolecular target) after they have been docked. Scoring functions have also been developed to predict the strength of other types of intermolecular interactions, such as protein–protein or protein–DNA.

Computational drug repositioning and drug rescue can be viewed as a complete turnaround to the conventional 'one target-many ligands' scenario. Here 'one ligand-many targets' view is considered (**Figure 1**). In contrast to docking, in the so-called 'reverse or inverse docking', a small-molecule drug/ligand is screened for its binding complementarity against an array of clinically relevant macromolecular targets. The end result is a list of targets ranked according to 'some score'. The top-ranking targets are likely to result in potential binding with the drug/ligand and may be relevant for drug repositioning and/or rescue. Since the basic philosophy behind docking and reverse docking is the same, the critical parameters (docking and scoring), though common, present problems of different nature. The current review article discusses the recent developments and future prospects of the reverse docking tool, with particular focus on drug repositioning and drug rescue.

Reverse docking: concept & methods

The PubMed search using keywords 'inverse docking', 'reverse docking', '*in silico* reverse screening' and 'reverse virtual screening' resulted in few hits relevant to the subject matter of this article. Similar to the conventional drug-design methods, both the ligand-based and structure-based approaches exist for 'reverse virtual screening'. The structure-based approach is the 'reverse or inverse' docking whereas the ligand-based strategies utilize 2D fingerprints (FPs; 2D FP-based reverse virtual screening) [7], 3D similarity search [8] and pharmacophore (reverse pharmacophore mapping) [7]. Many of these ligand-based approaches at times depend on the target structural information. For reverse pharmacophore mapping, for example, the 'pharmacophore model' of proteins in the target list is the starting point for comparing each of these models with every putative ligand in the flexible conformation [7]. Based on a similar concept, PharmMapper, a freely accessible web server, attempts to identify potential target candidates for small-molecule drugs, natural products or novel compounds whose targets are unknown [9,103]. The core database consists of pharmacophore database (PharmTargetDB [9]) containing over 7000 receptor-based pharmacophore models (covering >1500 drug targets). PharmMapper identifies 'best mapping poses of the query' against all pharmacophore models in the core database. The jobs are completed in high-throughput manner. The server represents a valuable resource for pharmacophore-based reverse screening. In all, the ligand-based methods may offer advantages such as faster computations and availability of a large variety of data. If one is interested in finding out potential targets for a set of ligands, direct or indirect use of target structural information becomes mandatory. Similar to traditional virtual screening, the combined use of structure- and ligand-based tools in reverse virtual screening may lead to higher enrichment of the hits.

■ Ligand & macromolecular target databases

The basic requirements for any docking study are a set of ligands and macromolecular targets (proteins, DNA, and so forth). Reverse docking, as applied to drug repositioning and drug rescue, would require a collection of approved and experimental/investigational drugs. DrugBank [10,104] is such a collection containing 1528 US FDA-approved small-molecule drugs, 87 nutraceuticals and 5080 experimental drugs. It represents a valuable resource for applications, such as reverse docking. Another such database is the NIH Chemical Genomics Center Pharmaceutical Collection, which contains 2750 small molecules approved by the leading regulatory bodies across the globe [105]. This comprehensive physical collection of approved and investigational drugs is publically available for high-throughput screening. Similarly, other databases, such as World Drug Index [106] and ChEMBL [107], can also be used as sources of drug/ligand structures.

A database of 3D structures of carefully selected macromolecular targets is crucial for the successful outcome of the reverse docking campaign aimed at drug repositioning. A protein structure database called Potential Drug Target Database (PDTD) [11,108] and a web server called Target Fishing Dock (TarFisDock) [12,109] have been established to facilitate identification of the potential binding targets *in silico*. The PDTD contains 1207 entries covering 841 known and potential drug targets with known 3D structures in PDB [110]. The targets in PDTD belong to several classes such as enzymes, receptors, transport proteins, ion channels, RNA, signaling proteins and many others with enzyme targets comprising 80% in all [108]. The targets, according to therapeutic area, are involved in bacterial/fungal/viral/parasitic infections, inflammation, cancer, renal, cardiovascular, gastrointestinal, blood and neuronal disorders. An interesting

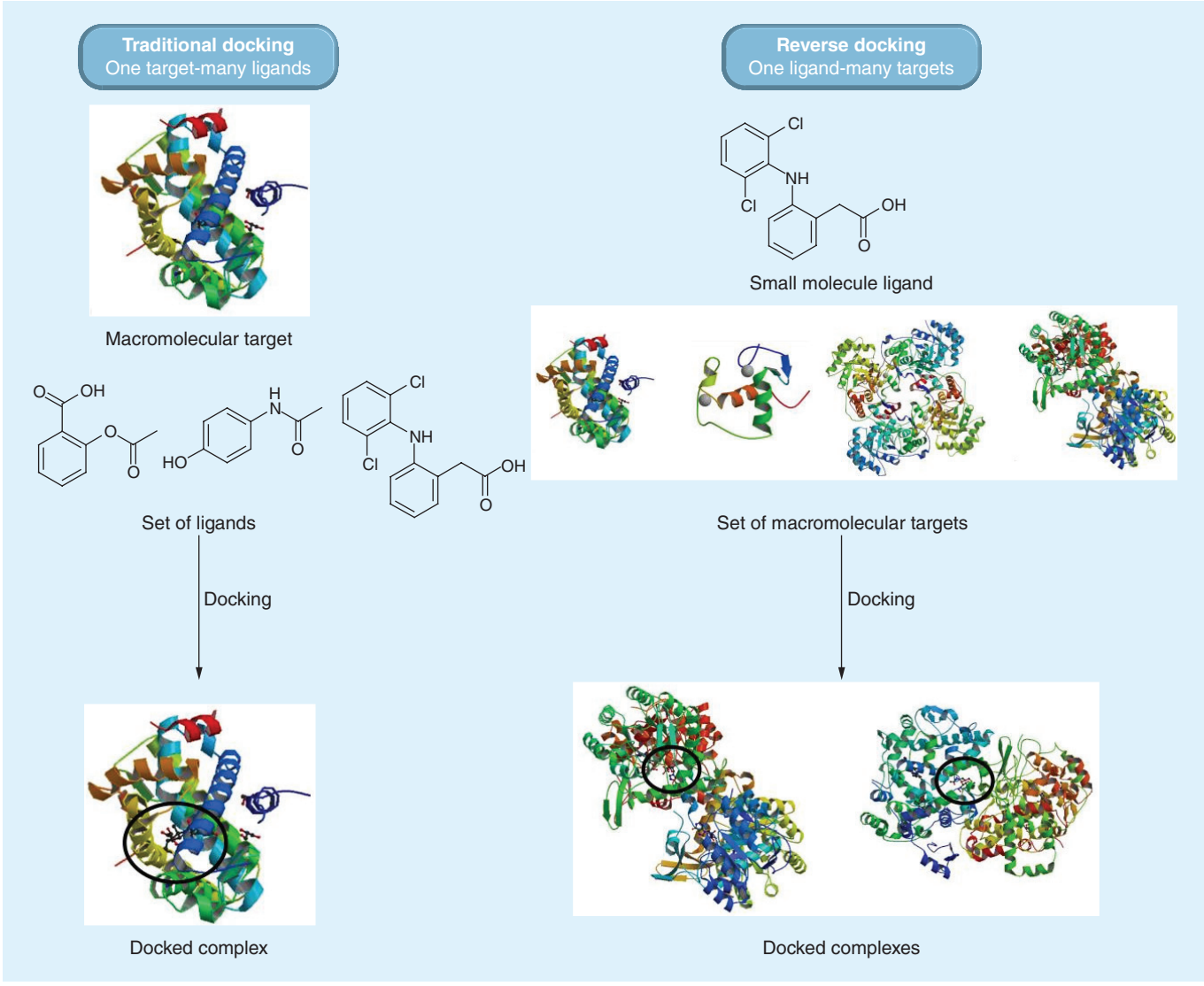

Figure 1. Schematic representations highlighting distinction between traditional docking and reverse docking.

work focused on 'Tropical Disease Research' [13] culminated in Tropical Disease Research Targets Database version 5 [111], which lists potential drug targets in seven tropical disease pathogens such as *Mycobacterium tuberculosis*, *Plasmodium falciparum*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, *Brugia malayi* and *Schistosoma mansoni*. This focused database is a valuable resource for many applications in drug-discovery research.

Another database of known and explored therapeutic protein and nucleic acid targets is Therapeutic targets database version 4.3.02 [14,112]. The database contains 2025 targets (364 successful, 286 clinical trials, 44 discontinued and 1331 research targets), 17,816 drugs (1540 approved, 1423 clinical trials and 14,853 experimental drugs), 3681 multitarget agents (14,170 small molecules and 652 antisense drugs). The targets belong to 61 biochemical classes and the drugs are distributed in 140 therapeutic classes. Therapeutic targets database is the most comprehensive of similar databases created, which provides information for both the approved and experimental drugs and the targets at one place. Interested readers may find several focused databases (disease or target class) located on various web servers. Most of these databases are likely to be linked to the available 3D structural information in PDB.

sc-PDB, an annotated database of druggable binding sites compiled from PDB, represents yet another valuable resource for reverse docking applications [15]. The binding sites were extracted

from high-resolution crystal structures and contain nucleotides, peptides, cofactors and organic compounds as ligands. The database was carefully annotated and contains critical information about the protein targets and the ligands. It is periodically updated and is accessible on the web [113].

■ Docking programs

Compared to a variety of programs and algorithms available for traditional ligand–protein docking, only a handful of such programs are accessible to the scientific community for reverse docking. The reasons for this scarcity could be many, including but not limited to, the non-availability of the 3D structural information for the macromolecular target class in PDB (e.g., G protein-coupled receptors [GPCRs], which are targets of >30% of the marketed drugs [16], ion-channels, nuclear receptors, and so forth), limited reliability of the homology models of protein targets, excessive reliance on 'one-target-one ligand' strategy, and so on. In last few years, substantial number of GPCRs has been crystallized and their 3D structures solved. This is a step forward in understanding this therapeutically important class of targets. The increasing amount of structural information will help in solving many mysteries in biological sciences; finding a target of action for a drug is just one example. Hence, target 3D structures may be a limiting factor for successful use of the reverse docking tool.

The process of reverse docking can be broken down into n steps per ligand where n represents the total number of targets to be used for docking. An algorithm capable of performing simple docking could be used for reverse docking, provided the process of selection of a different target with every successive step is automated. In that case, ideally, any docking program can be used with the similar settings for docking and scoring. Such a program should also be able to automate protein structure preparation (e.g., deleting unbound water and extra copies of the monomers present in the PDB structure, assigning correct ionization states and rotamers to the protein residues lining the active site), which needs to be done only once. In addition to automation, these docking programs need to be computationally efficient.

Among these, TarFisDock [12] is a front-runner. It uses PDTD as a protein target database for docking the input small-molecule ligand. The ligand–receptor interaction energy terms of the DOCK program (DOCK scoring function) are used for ranking the potential targets of the input molecule. It also uses a fragment-based incremental searching algorithm as implemented in DOCK 4.0. This approach has been tested by searching the potential binding proteins for vitamin E and 4H-tamoxifen [12]. The predictions by TarFisDock are in agreement with the experimental results (top 2 and 10% vitamin E targets identified by the tool, cover 30 and 50%, respectively, of the reported targets validated by experiments). These results unravel the potential utility of TarFisDock for drug repositioning and related applications.

A docking software package MDock, which can be used for reverse docking, uses a novel knowledge-based scoring function called ITScore [17,18]. It has been extensively evaluated and validated using diverse test sets in virtual screening applications. For each protein in the target database, molecular surface of the binding site is generated, along with the potential initial positions for ligand atom centers represented by associated sphere points. The ligand atoms are then matched to these sphere points and the orientations are sampled. This is followed by ranking the targets using ITScore. Using reverse docking performed in MDock (PDTD database), potential target of PRIMA-1 (p53 reactivation and induction of massive apoptosis), a small-molecule reactivator of tumor suppressor gene *p53*, was predicted to be OSC, a component of the cholesterol biosynthetic pathway. Ro 48–8071, a known potent inhibitor of OSC, was shown to increase binding of mutant p53 to DNA in BT-474 cells [18]. The experimental findings, guided by a well-planned reverse docking investigation, only emphasized the potential utility of MDock in identifying potential targets of small-molecule ligands.

INVDOCK, the first reverse docking program, developed back in 2001, uses a flexible docking algorithm in an attempt to dock a molecule to each cavity of every protein in the database [19]. It uses an automated method for the identification of binding cavities and derivation of the cavity models. The first stage involves a vector-based docking of a ligand to a cavity, followed by limited conformational optimization on the ligand and the side chains lining the cavity. In the next step, energy minimization is performed for all atoms in the binding cavity. Scoring of the docked molecules uses a ligand-protein interaction energy function DELP [19]. Interested readers are advised to refer literature

reports demonstrating satisfactory performance of INVDOCK in identifying therapeutic and toxicity targets of the approved drugs and medicinal natural products [20,21].

Recently, development of programs for enabling very large virtual drug screens using high-performance computing is reported [22]. The program VinaMPI has been specifically developed for decreasing the time required for completing the screen. It is based on a multithreaded virtual docking program, Autodock Vina. The tasks are distributed evenly while multithreading is used to speed-up individual docking tasks. VinaMPI is is able to handle efficiently multiple proteins in a virtual screen which makes it useful for reverse docking applications. The code is freely available and downloadable [114].

Reverse docking: applications

■ Target identification for small-molecule drugs/ligands

In an interesting application of reverse docking for the identification of putative molecular targets of four structurally related antineoplastic polyphenols from tea, PDTD was used as a source of target structures while Autodock [23] and TarFisDock were used for comparative reverse virtual screening [24]. The study was performed using two separate workflows to rank the targets in PDTD by the energy score. Furthermore, the identified potential targets and the polyphenols were subjected to exploration of the binding mode to validate the docking results. Various targets implicated in cancer or related diseases were correctly identified as potential targets for tea polyphenols (prediction accuracy ~55%). This is the first time when Autodock was used for reverse docking purposes. Analysis of the binding mode of one of the polyphenols into Leukotriene A4 hydrolase active site confirmed the possibility of the ligand being a competitive inhibitor. Overall, the study provided important information related to the putative mechanism of action of tea polyphenols and targeted molecular design efforts.

A similar application of reverse docking using TarFisDock featured the use of 16 known cytotoxic arylaminopyridine and arylaminoquinoline derivatives [25]. The aim of the study was to investigate and rationalize the observed cytotoxic effects of these compounds against a pool of antineoplastic targets. In the initial stage, TarFisDock reverse docking yielded 14 putative targets for every molecule (top 10%). This was used as a working hypothesis supplemented by extensive literature search. The final target pool consisted of 16 potential targets and 24 PDB entries. In view of the limitations of the scoring function used in TarFisDock, another round of virtual target screening using AutoDock 4.0 [23] against each of the 24 protein targets was performed. The study finally culminated in narrowing down the potential targets of the cytotoxic aminopyridines as tyrosine- and serine kinases and topoisomerase I.

Reverse docking was applied to identify potential target(s) for a compound isolated from folk medicine in yet another impressive study [26]. The lead compounds were identified during random screening of the herbal extracts (and subsequent analog syntheses) for anti-*H. pylori* agents. The database PDTD was searched for putative targets of the lead molecules using TarFisDock. DC and PDF were proposed to be the targets based on the interaction energy scores. Both the compounds were then evaluated in enzymatic assays for *H. pylori* DC and PDF. Enzyme inhibition data confirmed PDF as the molecular target of the lead molecules. Furthermore, x-ray crystal structure data showed that both the compounds bound well in the binding pocket of *H. pylori* PDF.

In a related application, inverse virtual screening was performed for a set of chemically diverse phenolic natural products against a panel of targets responsible for the genesis and progression of cancer [27]. Based on the predicted binding energy, few compounds were selected for *in vitro* inhibition of definite targets in the panel. The experimental results identified two compounds, xanthohumol and isoxanthohumol, as PDK1 and PKC inhibitors.

Overall, these systematic investigations resulted in identifying and validating the macromolecular targets of the natural product leads. It can be concluded that the computational reverse docking can be used complementarily with functional genomics and chemical biology in target identification.

A retrospective study featuring structure-based virtual screening of the sc-PDB target library for the identification of putative targets of small molecule ligands was performed for four compounds – biotin, methotrexate, 4-hydroxytamoxifen and 6-hydroxy-1,6-dihydropurine ribo-nucleoside [28]. The docking program GOLD (GOLD © 2013 CCDC Software Ltd., Cambridge, UK) was used along with several target ranking protocols based on GOLD fitness score and topological molecular interaction FP

Key Term

Off-target effect: Undesirable effect of a drug arising from its binding to unintended targets, in addition to its molecular target (responsible for the desired pharmacological effect).

comparison. Most of the targets of the ligands were recovered, simultaneously reducing number of false positives. This study conclusively demonstrates the potential utility of reverse docking in identification of putative targets for small-molecule ligands.

In addition to the small-molecule drugs/ligands, one interesting application featured target identification of a 1,3,5-triazepan-2,6-dione scaffold-focused library using high-throughput docking against a collection of 2150 druggable active sites from the PDB [29]. Of the five targets (among the top 2% scoring targets) prioritized for experimental evaluation, secreted PLA2 was confirmed to be the real target of the set of 1,3,5-triazepan-2,6-dione-containing library molecules with affinity in the micromolar range.

To summarize, small-molecule drug/ligand target identification using reverse docking has been successfully applied and validated using relevant experimental results. Interested readers are advised to refer the literature for related applications of this approach [30–32].

■ Prediction of off-target effects

Small-molecule drugs (and biologicals), in addition to their intended target, interact with a variety of other macromolecules, leading to 'off-target' effects. These interactions may have pharmacokinetic and/or toxicological consequences. Examples include cytochrome P450 isoforms (induction/inhibition, drug–drug interactions), plasma protein-binding (drug–drug interactions), hERG channel (cardiotoxicity), and many others. Using network pharmacology and reverse docking, the **off-target effects** (leading to increased mortality and adverse cardiovascular effects) of torcetrapib, a CETP were investigated [33]. The program CDOCKER, as implemented in Discovery Studio® (Accelrys Software Inc., Discovery Studio Modeling Environment, CA, USA), was used for docking torcetrapib into a set of protein targets based on enriched signaling pathways. The results indicated four potential off-targets for torcetrapib. The combined use of network pharmacology, systems biology and reverse docking led to the identification of potential targets responsible for off-target effects.

In a similar application addressing kinase inhibitor selectivity and off-target effects, indirubin derivatives were subjected to virtual 'reverse' screening to identify their potential targets [34]. In addition to known targets, PDK1 was found to be the target of one of the derivatives, 6BIO. The hypothesis was quickly tested in an *in vitro* kinase assay and related functional assays dependent on PDK1. The experimental results validated the 'reverse' screening hypothesis.

In conclusion, experimental validation is crucial to prove, beyond doubt, the usefulness of such *in silico* tools.

Macromolecular target druggability analysis

Successful completion of the human genome project led to the identification of unprecedented number of proteins which could be potential targets for novel drugs [16]. Systematic target validation studies of these newly identified targets may result in unique opportunities for disrupting a disease process. Assessing the 'druggability' of the target is of paramount importance with reference to its utility for drug-discovery efforts [35]. One possible solution is to look for homologous proteins whose druggability is already known. The low availability of the druggable targets limits the usefulness of this approach. Another option is to search for 'bindability' – an ability to bind to putative drugs [35]. In such a study involving large-scale reverse docking profiles using a non-redundant set of druggable and less-druggable targets [36] and a set of 35 drugs (used for generating reverse docking profiles), the average docking scores of druggable set were found to be greater than the corresponding scores of less-druggable set. The study, based on the docking fitness score, differentiated potential druggable binding sites from less druggable binding sites. Of the six enzyme classes, oxidoreductases were found frequently in the druggable class, whereas hydrolases, lyases and isomerases were part of less druggable class [36]. The presented study involved the largest number of target structures to date in a computational reverse docking campaign.

■ Prediction of protein functions based on similarity in docking profiles

In an extension of the largest reverse docking study discussed above [36], a large-scale protein function prediction of enzymes was undertaken. The aim of the study was to assess whether the docking profiles contain useful information which can be utilized for the prediction of protein function. For 3,874,883 pairs of 5989 and 647 human and yeast enzymes, respectively, BLAST e-value [37] was used as the sequence similarity measure and Euclidean distance between the two docking profiles as the similarity measure. It was found from the receiver operating characteristic

curve that the docking profile similarity contained useful information for the prediction of protein function which was mutually orthogonal to the sequence information. Such an outcome of the reverse docking combined with sequence, structure and binding-site similarities can be utilized to infer protein function.

Future perspective

For any reverse docking investigation, the macromolecular target database is, by far, the most crucial factor affecting the outcome. 'How many structures to include?' will always be a matter of discussion. For one particular protein, there could be several structures in the PDB bound to a variety of ligands, with alternative conformations of the binding site. Including many structures would obviously increase the computational cost. The structures in PDTD is a subset of the structures present in PDB for a protein, for example, for glucocorticoid receptor, two structures are present in PDTD (PDB ID 1NHZ – antagonist form, 1P93 – agonist form). Currently there are >10 structures of the same receptor in the PDB including those containing mutations in the binding cavity. Careful scrutiny of the structures to be included in the target database is needed and must be accompanied by regular updation.

In an interesting application of the reverse docking for studying the selectivity of protein kinase inhibitors, the authors constructed a database of 422 kinase structures taken from PDB or derived from homology modeling [38]. The reverse docking protocol and the scoring function were optimized initially followed by validation using a set of seven selective kinase inhibitors. These inhibitors showed higher values of the scoring function for their original targets. This study attempted to prove the utility of reverse docking for selectivity analysis of the target classes. Dedicated efforts in constructing a target class-specific or universal targets database would pay rich dividends in terms of successful outcomes of the reverse docking method.

Another concern is preparation of the protein structures before they are usable for large-scale reverse docking. It could range from detection of putative binding sites, treatment of cofactors or ligands, residues surrounding the binding pocket, the protonation states, rotamer states, and so forth. Reverse docking, in contrast to the focused docking, may involve exploration of the entire protein surface for possible ligand binding site(s). Even if the binding site for a ligand (e.g., inhibitor) is known from the target 3D structure, the drug/ligand to be used in reverse docking investigation may bind at other site. Hence, a powerful, fast, cavity/binding-site detection algorithm is an important requirement [39]. A program Fpocket [40,115] is a fast protein-cavity detection algorithm available publicly. It has been used for large-scale applications involving reverse docking previously [35] along with Open Babel [41] for protonating all the residues in the pockets. Since the place where docking is likely to take place must accommodate the entire protein and is therefore much larger in size than a conventional docking box, the running time is going to be significantly higher due to large number of energy evaluations. One possible approach could be the use of binding sites calculated directly from the docking grid. This strategy distributes the problem in multiple independent docking runs focused on smaller boxes as opposed to one larger blind docking run covering the entire protein surface. Such a focused docking approach has been shown to be advantageous over the blind docking approach [39].

The critical issues related to the docking part are the prediction of correct binding pose and the estimation of some measure of the binding affinity. Several available docking programs are successful in reproducing the crystal structure binding mode but, at times, struggle to predict the 'tightness' of the binding [42]. An interesting article on critical assessment of ten popularly used docking programs and 37 scoring functions concluded that none of the docking programs or scoring functions was able to predict the ligand binding affinity correctly, although these were able to identify active compounds from a set of decoy compounds [42]. It was also shown that none of the docking programs could do well for all the macromolecular targets belonging to seven different protein types such as kinases, proteases and isomerases, to name a few. Even, handling dissimilar ligands becomes an issue in few cases. The scoring function used in these algorithms is, obviously, critical for ranking the ligands; in reverse docking, the macromolecular targets. Scoring function, if nonspecific in nature, may be fast in evaluating the docked complexes, but may lead to significant number of false positives, such as, TarFisDock. It may be advisable to use consensus scoring (e.g., consensus of force field scoring functions, knowledge-based scoring functions and/or empirical scoring functions) to further make the target selection more stringent [43]. In addition, an ensemble of conformations to represent each small-molecule drug/ligand in

the unbound state as an input for the reverse docking run may lead to enrichment of the list of putative macromolecular targets.

In 2012, a freely available web server, idTarget, was constructed in an attempt to identify targets of small molecules utilizing robust scoring functions and a divide-and-conquer docking approach [44,116]. The scoring functions are based on robust regression analysis and quantum chemical charge models. Target-affinity profiles are provided for comparison with the predictions. The web server screens against nearly all target structures in PDB. It has been demonstrated to reproduce known off-target effects of drugs or drug-like structures.

Benchmarking is yet another important aspect of the reverse docking method. Unlike the benchmarks for traditional molecular docking [45], there is scarcity of such sets for reverse docking. These sets for docking mainly focused on re-docking the cognate ligand of a 3D protein–ligand complex to measure prediction accuracy of the binding pose. Also, the use of property-matched decoy sets for evaluating the performance of virtual screening campaigns is commonplace [46]. For reverse docking, PDTD and TarFisDock provide three examples as benchmarks [108,109]. Out of these, two are the computational results in comparison with the experimental results for the possible binding proteins of vitamin E and 4H-tamoxifen, an anticancer drug. The third benchmark is the computational results and the experimental validation for the discovery of the binding protein of an anti-*H. pylori* natural product [26]. Overall, several such benchmarking sets for reverse docking are needed.

Similar to benchmarking, evaluation of several docking strategies for reverse docking applications is another area which needs more attention of the scientific community since the effectiveness of these strategies in multiple target identification is unclear. In a related study, the authors tried to compare and evaluate five docking schemes in order to identify multiple targets from a dataset of 1714 entries (1594 known drug targets) [47]. TarSearch-X [47], an in-house target search scheme from Jian and Wei's group, was found to be the most effective method in identification and validation of multiple targets. In addition, the method was useful for the prediction of toxic effects.

To summarize, the newer reverse docking programs/algorithms should take into consideration the issues mentioned above to be of potential interest to the scientific community.

Acknowledgements

The authors would like to thank A Ramdhave, doctoral student at SPP School of Pharmacy and Technology Management (SPPSPTM), SVKM's NMIMS, Mumbai, for his help during preparation of this manuscript.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Reverse docking & drug repositioning

- Reverse docking represents a useful tool to explore the potential targets of small molecule ligands with potential applications in the areas of drug repositioning and drug rescue.

Current status

- The field, even with the potential utility to the scientific community, is still far from maturity for several reasons. The issues related to the macromolecular target database, docking program, scoring function and the computational power need to be addressed to enhance usefulness of the reverse docking tool.

Potential applications

- With several novel targets being identified and validated in the post-genomic era, reverse docking along with functional genomics, systems biology, network pharmacology and related bioinformatics tools, will only prove indispensable to invent tool molecules and small-molecule drugs against these newly discovered targets.

What next?

- The scientific community needs to take a greater interest in this approach to replenish the drying pharmaceutical pipelines of new molecular entities. Several experimental tools such as high-throughput protein crystallography, biophysical characterization to investigate possible ligand–receptor interactions, and so forth, can validate the hypotheses generated from reverse docking analysis.
- In addition, the ligand-based approaches for reverse virtual screening can be used in combination with the reverse docking tool leading to better outcome.

References

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1 Pammolli F, Magazzini M, Riccaboni M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discov.* 10(6), 428–438 (2011).

2 Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3(8), 673–683 (2004).

3 Liu Z, Fang H, Reagan K *et al. In silico* drug repositioning – what we need to know. *Drug Discov. Today* 18(3–4), 110–115 (2013).

▪▪ **Excellent overview of the *in silico* drug repositioning approaches.**

4 Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational drug repositioning: from data to therapeutics. *Clin. Pharmacol. Ther.* 93(4), 335–341 (2013).

5 Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J.* 14(1), 133–141 (2012).

6 Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3(11), 935–949 (2004).

7 Bhattacharjee B, Chatterjee J. Identification of proapoptopic, anti-inflammatory, anti-proliferative, anti-invasive and anti-angiogenic targets of essential oils in cardamom by dual reverse virtual screening and binding pose analysis. *Asian Pac. J. Cancer Prev.* 14(6), 3735–3742 (2013).

▪ **Demonstrates the utility of reverse virtual screening in identifying targets of known compounds.**

8 Kinnings SL, Jackson RM. ReverseScreen3D: a structure-based ligand matching method to identify protein targets. *J. Chem. Inf. Model.* 51(3), 624–634 (2011).

9 Liu X, Ouyang S, Yu B *et al.* PharmMapper server: A web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* 38, W609–W614 (2010).

10 Knox C, Law V, Jewison T *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* 39(D1), D1035–D1041 (2011).

11 Gao Z, Li H, Zhang H *et al.* PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics* 9, 104 (2008).

12 Li H, Gao Z, Kang L *et al.* TarFisDock: a web server for identifying drug targets with docking approach. *Nucl. Acids Res.* 34, W219–W224 (2006).

13 Crowther GJ, Shanmugam D, Carmona SJ *et al.* Identification of attractive drug targets in neglected-disease pathogens using an *in silico* approach. *PLoS Negl. Trop. Dis.* 4(8), e804 (2010).

14 Zhu F, Shi Z, Qin C *et al.* Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* 40(D1), D1128–D1136 (2012).

15 Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D. sc-PDB: an Annotated database of druggable binding sites from the protein data bank. *J. Chem. Inf. Model.* 46(2), 717–727 (2006).

▪▪ **Describes the development of sc-PDB database and its applications including reverse docking.**

16 Hopkins AL, Groom CR. The druggable genome. *Nat. Rev. Drug Discov.* 1(9), 727–730 (2002).

17 Huang SY, Zou X. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. derivation of the interaction potentials. *J. Comput. Chem.* 27(15), 1866–1875 (2006).

18 Grinter SZ, Liang Y, Huang S-Y, Hyder SM, Zou X. An inverse docking approach for identifying new potential anti-cancer targets. *J. Mol. Graph. Model.* 29(6), 795–799 (2011).

▪ **Interesting application of reverse docking for target identification.**

19 Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* 43(2), 217–226 (2001).

20 Chen YZ, Ung CY. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. *J. Mol. Graph. Mod.* 20(3), 199–218 (2001).

▪ **Applications of reverse docking for identifying side effects and toxicity.**

21 Chen YZ, Ung CY. Computer automated prediction of putative therapeutic and toxicity protein targets of bioactive compounds from chinese medicinal plants. *Am. J. Chin. Med.* 30(1), 139–154 (2002).

22 Ellingson SR, Smith JC, Baudry J. VinaMPI: Facilitating multiple receptor high-throughput virtual docking on high-performance computers. *J. Comput. Chem.* 34(25), 2212–2221 (2013).

▪▪ **Latest development in the faster docking algorithms suited for reverse docking applications.**

23 Morris GM, Huey R, Lindstrom W *et al.* Autodock4 and AutoDockTools4: Automated docking with selective receptor flexiblity. *J. Comp. Chem.* 30(16), 2785–2791 (2009).

24 Zheng R, Chen T, Lu T. A comparative reverse docking strategy to identify potential antineoplastic targets of tea functional components and binding mode. *Int. J. Mol. Sci.* 12(8), 5200–5212 (2011).

25 Erić S, Ke S, Barata T, *et al.* Target fishing and docking studies of the novel derivatives of arylaminopyridines with potential anticancer activity. *Bioorg. Med. Chem.* 20(17), 5220–5228 (2012).

26 Cai J, Han C, Hu T *et al.* Peptide deformylase is a potential target for anti-Helicobacter pylori drugs: reverse docking, enzymatic assay, and x-ray crystallography validation. *Protein Sci.* 15(9), 2071–2081 (2006).

27 Lauro G, Masullo M, Piacente S, Riccio R, Bifulco G. Inverse virtual screening allows the discovery of the biological activity of natural compounds. *Bioorg. Med. Chem.* 20(11), 3596–3602 (2012).

28 Kellenberger E, Foata N, Rognan D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. *J. Chem. Inf. Model.* 48(5), 1014–1025 (2008).

29 Muller P, Lena G, Boilard E *et al. In silico*-guided target identification of a scaffold-focused library: 1,3,5-triazepan-2,6-diones as novel phospholipase A2 inhibitors. *J. Med. Chem.* 49(23), 6768–6778 (2006).

30 Lamb ML, Burdick KW, Toba S *et al.* Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* 42(3), 296–318 (2001).

31 Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. *J. Med. Chem.* 46(14), 3045–3059 (2003).

32 Paul N, Kellenberger E, Bret G, Müller P, Rognan, D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* 54(4), 671–680 (2004).

33 Fan S, Geng Q, Pan Z *et al.* Clarifying off-target effects for torcetrapib using network pharmacology and reverse docking approach. *BMC Syst. Biol.* 6, 152 (2012).

34 Zahler S, Tietze S, Totzke F *et al.* Inverse *in silico* screening for identification of kinase inhibitor targets. *Chem. Biol.* 14(11), 1207–1214 (2007).

35 Lee M, Kim D. Large-scale reverse docking profiles and their applications. *BMC Bioinformatics* 13(Suppl. 17), S6 (2012).

36 Krasowski A, Muthas D, Sarkar A, Schmitt S, Brenk R. DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *J. Chem. Inf. Model.* 51(11), 2829–2842 (2011).

37 Karlin S, Altschul SF. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl Acad. Sci. USA* 87(6), 2264–2268 (1990).

38 Liu TL, Xie HZ, Wei YQ, Yang SY. The feasibility of application of reverse docking method to the selectivity studies of protein kinase inhibitors. *Yao. Xue. Xue. Bao.* 44(7), 758–763 (2009).

▪ **Reverse docking application focused on target selectivity.**

39 Ghersi D, Sanchez R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites. *Proteins* 74(2), 417–424 (2009).

40 Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* 10, 168 (2009).

41 O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J. Cheminform.* 3, 33 (2011).

42 Warren GL, Andrews CW, Capelli A-M *et al.* A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49(20), 5912–5931 (2006).

43 Huang SY, Grinter SZ, Zou X. Scoring functions for protein-ligand docking: Recent advances and future directions. *Phys. Chem. Chem. Phys.* 12(40), 12899–12908 (2010).

44 Wang JC, Chu PY, Chen CM, Lin JH. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Res.* 40, W393–W399 (2012).

45 Huang N, Stoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J. Med. Chem.* 49(23), 6789–6801 (2006).

46 Spitzer R, Jain AN. Surflex-Dock: Docking benchmarks and real-world application. *J. Comput. Aided Mol. Des.* 26(6), 687–699 (2012).

47 Hui-fang L, Qing S, Jian Z, Wei F. Evaluation of various inverse docking schemes in multiple target identification. *J. Mol. Graph. Model.* 29(3), 326–330 (2010).

■ Websites

101 Discovering new therapeutic uses for existing molecules.
www.ncats.nih.gov/research/reengineering/rescue-repurpose/therapeutic- uses/therapeutic-uses.html

102 National center for advancing translational sciences (NCATS) pharmaceutical Collection.
www.ncats.nih.gov/research/tools/preclinical/npc/pharmaceutical-collection.html

103 PharmMapper Server: an updated integrated pharmacophore matching platform with statistical method for potential target identification.
59.78.96.61/pharmmapper/index.php

104 DrugBank: Open Data Drug and Drug Target Database.
www.drugbank.ca

105 NCGC Pharmceutical Collection (NPC).
http://tripod.nih.gov/npc

106 World Drug Index (WDI).
www.daylight.com/products/wdi.html

107 ChEMBL.
www.ebi.ac.uk/chembl

108 Potential Drug Target Database (PDTD).
www.dddc.ac.cn/pdtd

109 Target fishing dock (TarFisDock).
www.dddc.ac.cn/tarfisdock

110 RCSB PDB.
www.rcsb.org/pdb/home/home.do

111 The tropical disease research (TDR) Targets database v5.
http://tdrtargets.org

112 Therapeutic Targets Database (TTD).
http://bidd.nus.edu.sg/group/ttd/ttd.asp

113 sc-PDB: an annotated database of druggable binding sites from PDB.
http://bioinfo-pharma.u-strasbg.fr/scPDB

114 VinaMPI: facilitating multiple receptor high-throughput virtual docking on high-performance computers.
http://cmb.ornl.gov/~sek

115 fpocket: program for scalable high-performance pocket detection.
http://fpocket.sourceforge.net

116 idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach.
http://idtarget.rcas.sinica.edu.tw