What is the potential of structure-based target prediction methods?

"...inverse structure-based virtual screening has to be further abstracted from just using established 'classic' docking tools in a reverse manner in order to account for the challenges of (off-)target prediction."

Protein target identification for drugs or drug candidates is essential for avoiding unwanted side-effects evoked by off targets. Additionally, it opens the route for drug repurposing and supports the rational design of multitarget drugs. Since wet lab experiments for target identification, involving many different proteins, are inefficient in terms of cost and effort (different activity assays for each protein, protein isolation difficulties, etc.), computational methods are frequently applied. These either use collected activity data for deriving predictions by ligand similarity following the paradigm 'similar molecules bind to the same targets' (ligand-based and network-based methods) or exploit the information contained in the 3D structures of protein–ligand complexes to identify potential target proteins for small molecules (structure-based methods). Next to finding – or 'fishing' – targets, structure-based methods elucidate the protein–ligand binding mode and thus provide valuable insights for rational drug selectivity enhancement. Target prediction methods should be part of any drug design project to avoid late attrition due to adverse effects resulting from binding to off targets.

While the choice of the computational method should be guided by the available data (ligand activity data vs protein structure data), the advantages and caveats of each respective method also have to be considered. Next to elucidating the binding mode and thus providing a starting point for lead optimization, structure-based methods are also better at detecting activity cliffs. In a descriptor-based representation as used in ligand-based methods (e.g., [1]), the chemical differences leading to an activity change might get lost due to the abstraction of the descriptor. Also, the information needed to detect the activity cliff might not be represented in the available data. In structure-based methods, however, the complementarity between protein and ligand is evaluated, which allows the assessment of even subtle molecular differences of ligands. Clearly, these methods are dependent on available 3D protein structures.

Structure-based target prediction methods cover molecular docking in an inverse screening context, pharmacophore-based predictions or protein-ligand interaction fingerprints (for an overview of structure-based target prediction methods see [2]). Molecular docking is the most promising approach for detecting activity cliffs, since the structural reason for a large activity difference lies in the (in-)complementarity between protein and ligand which might not be detected by the more abstract pharmacophore-based or interaction fingerprint-based approaches.

In protein–ligand docking, the complementarity between protein and ligand concerning shape and interactions is evaluated (for a detailed review on docking in the context of inverse screening see [3]). However, the methods used for classical screening, in other words, one target, numerous ligands, can in principle not be applied to the reverse case without adjustments. The following limitations have to be considered if a

Karen T Schomburg
Center for Bioinformatics, University of Hamburg, Hamburg, Germany

Matthias Rarey
Author for correspondence:
Center for Bioinformatics, University of Hamburg, Hamburg, Germany
rarey@zbh.uni-hamburg.de

method created for structure-based virtual screening is applied for target prediction:

Since docking programs were not designed for handling the computationally demanding structure data of large numbers of proteins, high computing times emerge [4], which are difficult to integrate into a drug design pipeline. Furthermore, many classic docking programs rely on the user to define a pocket that the program shall use as a binding site. However, if thousands of protein structures shall be considered, an automatic mode is needed to release the user of this task. A third limitation concerns the interpretation of the results: scoring functions as applied for classic virtual screening were not designed for being capable of intertarget differentiation and thus often contain a bias toward, for example, highly hydrophilic protein binding sites. While this bias does not impede ranking ligands for one target, the absolute score of different proteins (especially of proteins with diverse binding sites) is not comparable if no correction is applied to remove this bias [5–7].

Therefore, the use of classical docking and scoring approaches for target prediction is only feasible for a limited number of related proteins. However, some software applications have already been introduced that face the challenges of target prediction by protein–ligand docking and thus allow more complex target prediction tasks.

> "...we developed an inverse screening approach enabling the efficient processing of many protein structures as well as straightening out the pocket-specific scoring bias."

One of the very first inverse docking approaches, InvDock [8,9], already accounted for the high computing cost of screening proteins by aborting the full search for the best binding mode at a first pose in each pocket. However, this approach impairs one of the greatest gains of docking-based target prediction: the concurrent prediction of binding modes.

The protein ranking problems of classic protein–ligand scoring functions have already been observed more than 5 years ago [4], however, so far few methodic improvements have been introduced [5]. In order to address these issues, we developed an inverse screening approach enabling the efficient processing of many protein structures as well as straightening out the pocket-specific scoring bias [10]. About 8000 protein structures were screened in an enrichment evaluation which found the first true target among 0.28% of the ranked list.

Some recent approaches combine different computational methods into 'hybrid' prediction tools [11,12]

and thus overcome each single method's shortcomings. One approach successfully identifies 72% of the main targets of 189 clinical candidates by choosing automatically from four ligand-based and two structure-based approaches depending on ligand and target properties. Using docking only, the success rate dropped to 25% [11]. Nevertheless, the advantages from method combination are limited making further efforts into individual method improvements mandatory.

> "...structure-based methods will in future continue to gain importance in drug design."

Although inverse structure-based virtual screening is still subject to the mentioned limitations, they were already used successfully to support numerous target fishing application studies by adding measures to overcome these. One example is the work of Eric *et al.* [13], who rationalized the mechanism of action of the cytotoxicity of new aryl-aminopyridines. TarFisDock [14] was applied for a preselection of targets followed by computationally more demanding AutoDock calculations. The authors identify cyclin-dependent kinases as well as tyrosine kinases as protein targets and provide with binding mode studies valuable insights for follow-up studies. In another example, Zahler *et al.* studied kinase targets for indirubin derivatives with docking and confirmed PDK1 as a beforehand unknown target of one of the derivatives experimentally [15]. In order to achieve this, a modification of the scoring function was applied to circumvent a bias toward metal-containing proteins.

These examples show the high potential of target prediction by inverse screening – but also emphasize the need for improved methods which can be used without expert knowledge of scoring bias corrections as an example. With the exponentially rising number of available structures of different proteins and protein conformations, structure-based methods will in future continue to gain importance in drug design. Computational methods are challenged with handling the rising amount of structure data. New approaches for choosing a protein structure from an ensemble for docking, for example using ligand similarity [16], might become essential for performance in an acceptable time frame.

In summary, from the point of view of method developers, we believe that inverse structure-based virtual screening has to be further abstracted from just using established 'classic' docking tools in a reverse manner in order to account for the challenges of (off-) target prediction. The differences have to be apprehended and cared for in order to uncover the full potential of computational target prediction.

References

1   Keiser MJ, Setola V, Irwin JK *et al.* Predicting new molecular targets for known drugs. *Nature* 462, 175–181 (2009).

2   Rognan D. Structure-based approaches to target fishing and ligand profiling. *Mol. Inform.* 29(3), 176–187 (2010).

3   Kharkar PS, Warrier S, Gaud RS. Reverse docking: a powerful tool for drug repositioning and drug rescue. *Future Med. Chem.* 6(3), 333–342 (2014).

4   Meslamani J, Li J, Sutter J, Stevens A, Bertrand H-O, Rognan D. Protein–ligand-based pharmacophores: generation and utility assessment in computational ligand profiling. *J. Chem. Inf. Model.* 52(4), 943–955 (2012).

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

6   Wang W, Zhou X, He W, Fan Y, Chen Y, Chen X. The interprotein scoring noises in Glide docking scores. *Proteins* 80(1), 169–183 (2012).

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

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

9   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).

10   Schomburg KT, Bietz S, Briem H, Henzler AM, Urbaczek S, Rarey M. Facing the challenges of structure-based target prediction by inverse virtual screening. *J. Chem. Inf. Model.* 54(6), 1676–1686 (2014).

11   Meslamani J, Bhajun R, Martz F, Rognan D. Computational profiling of bioactive compounds using a target-dependent composite workflow. *J. Chem. Inf. Model.* 53(9), 2322–2333 (2013).

12   Li GB, Yang L-L, Xu Y, Wang W-J, Li L-L, Yang S-Y. A combined molecular docking-based and pharmacophore-based target prediction strategy with a probabilistic fusion method for target ranking. *J. Mol. Graph. Model.* 44, 278–285 (2013).

13   Eric S, Ke S, Barata T, Solmajer T *et al.* Target fishing and docking studies of the novel derivatives of aryl-aminopyridines with potential anticancer activity. *Bioorg. Med. Chem.* 20, 5220–5228 (2012).

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

15   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).

16   Broccatelli F, Brown N. Best of both worlds: On the complementarity of ligand-based and structure-based virtual screening. *J. Chem. Inf. Model.* 54(6), 1634–1641 (2014).