

This is the peer reviewed version of the following article:

Selection of protein conformations for structure-based polypharmacology studies / Pinzi, Luca; Caporuscio, Fabiana; Rastelli, Giulio. - In: DRUG DISCOVERY TODAY. - ISSN 1359-6446. - 23:11(2018), pp. 1889-1896. [10.1016/j.drudis.2018.08.007]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

13/12/2025 22:11

(Article begins on next page)

Short Review

Informatics

Selection of protein conformations for structure-based polypharmacology studies

Luca Pinzi¹, Fabiana Caporuscio¹, Giulio Rastelli¹

¹Department of Life Sciences, University of Modena and Reggio Emilia, Via Giuseppe Campi 103, 41125, Modena, Italy.

Corresponding author: Giulio Rastelli (giulio.rastelli@unimore.it)

Phone number: +39 059 2058564 (Office) / +39 059 2058609 (Laboratory)

Keywords: polypharmacology, multi-target ligands, protein conformation, virtual screening, ligand-based methods, structure-based methods

Teaser

This review will focus on currently available methods for the selection of the most suitable protein conformations for multi-target structure-based drug design.

Highlights

- Structure-based approaches play a central role in polypharmacology drug design.
- Similar protein conformations are able to accommodate similar ligands.
- The use of selected protein conformations affects structure-based drug design results.
- *In silico* approaches for protein conformation selection are presented.
- Examples of protein conformation selection for structure-based drug design are reported.

ABSTRACT

Several drugs exert their therapeutic effect through the modulation of multiple targets. Structure-based approaches hold great promise for identifying compounds with the desired polypharmacological profiles. These methods employ the knowledge of the protein binding sites to identify stereoelectronically complementary ligands. The selection of the most suitable protein conformations to be used in the design process is of paramount importance, especially for multi-target drug design in which the same ligand has to be accommodated in multiple binding pockets. Herein, we will focus on currently available techniques for the selection of the most suitable protein conformations for multi-target drug design, compare potential advantages and limitations of each method, and comment on how their combination could help in polypharmacology drug design.

Introduction

The simultaneous modulation of multiple targets by means of a single drug, *i.e.*, polypharmacology, is increasingly recognized as a valuable opportunity in drug discovery [1]. Several drugs are known to simultaneously bind to multiple biological targets, thus eliciting either unwanted side effects or beneficial synergistic activities [2]. However, designing drugs with a polypharmacological profile represents a challenging task [1,3,4] and the few examples of intended polypharmacology are usually within related protein families. However, recent efforts in predicting polypharmacology for distantly related targets have started to yield promising results [5–7]. Indeed, computational approaches have certainly proved to play a key role in exploiting the available structural information, and to perform *de novo* multi-target drug design and *in silico* profiling [8]. Moreover, a multitude of molecular modeling methods, which can

be broadly classified into structure- and ligand-based approaches, are currently available to aid in polypharmacology drug design [9,10].

In general, structure-based methods allow the discovery of potentially active compounds according to their structural complementarity to a receptor, being molecular docking one of the most frequently used. In this context, the selection of the most suitable protein conformations to be used in a structure-based virtual screening (VS) campaign undoubtedly represents a challenging task. In fact, proteins are known to undergo significant conformational changes depending on the activation/inactivation state, catalytic activity, protein-protein interactions and binding of substrates or different modulators. Therefore, selected protein conformations and peculiar amino acid rearrangements induced by co-crystallization might bias the outcome of a structure-based drug design project [11,12]. Moreover, it has already been reported that: (i) better enrichments can be obtained in docking-based VS when the screened compounds are structurally similar to the co-crystallized ligands [13–15]; and (ii) the ability of structure-based methods (*e.g.*, molecular docking) in discriminating active from inactive compounds is highly dependent on the used protein structure [14,16,17]. If this is not negligible in standard *in silico* structure-based drug discovery (*i.e.*, the one drug – one target paradigm), this concept is even truer when designing molecules that are expected to simultaneously bind to more than one biological target.

Importantly, the degree of similarity of the binding pockets of the studied targets is a critical aspect for the selection of protein conformations in a multi-target structure-based drug design project. In fact, it is well known that proteins with a high degree of binding site similarity are likely to bind similar ligands, as reported in drug repurposing or ligand promiscuity investigations [18,19]. Moreover, the more two proteins display similar binding sites and common interaction *hot spots*, the higher are the chances of identifying low molecular weight compounds that are stereoelectronically complementary to the targets under study. In the light of this, the selection of the most suitable protein conformations has a dramatic effect on *in silico* structure-based polypharmacology approaches, such as ligand profiling and *de novo* design. However, it has to be pointed out that the design of compounds with multi-target activity is also possible by developing ambivalent, chimeric ligands that bind *head-to-tail* in their respective protein targets. Nevertheless, these chimeric ligands usually have high molecular weight and unfavorable drug-like properties.

In this article, we will focus on the utility of different *in silico* approaches to select the most suitable protein conformations for structure-based multi-target drug design. This is still an emerging and largely unexplored topic. The following sections will provide the reader with an *excursus* of the currently available *in silico* approaches for protein conformation selection. A few illustrative examples will also be provided, demonstrating potential shortcomings and complementarities of the different approaches.

***In silico* approaches for the selection of the most suitable protein conformations**

The selection of the most suitable protein conformations in multi-target drug design projects represents a challenging task, especially when the studied targets are structurally distant and belong to different protein families. However, medicinal chemistry data available from public repositories, which in the last decades have entered the so called “big data era” [20], could offer valuable opportunities for inferring new relationships between known modulators and different proteins. Indeed, several computational approaches have been successfully applied to identify potential activities of molecules on other targets,

e.g., on the basis of their structural similarity with already reported modulators [21,22]. At present, a variety of structure and ligand-based methods are available to analyze the wealth of information provided by public databases [23,24], and to guide the selection of the most suitable protein conformations for structure-based drug design [13,25].

Ligand-based approaches in protein conformation selection

Ligand-based methods enable the identification of new small-molecules through the knowledge of the structural details of known active and inactive compounds. Several approaches with different levels of abstraction (*e.g.*, molecular descriptors, sets of bit-strings and spatial arrangements of pharmacophoric features) and algorithms allow encoding ligand properties and estimating the degree of similarity with respect to a reference compound [26,27]. Such methods could be profitably applied in protein conformation selection. In particular, conformations of different targets can be selected on the basis of the cross similarities of their ligands (Figure 1, panel *a*). For example, a recent study by Broccatelli F. *et al.* reported how ligand-based methods can be used to select the most suitable receptor conformations for a docking-based VS [13]. In this study, the authors compared the performance of different docking protocols (*i.e.*, Glide HTVS, SP, and XP protocols) [28] in retrieving compounds active on the CDK2 protein kinase (PK) within the ChEMBL [29,30] and Directory of Useful Decoys [31] datasets. In particular, they performed docking calculations by selecting CDK2 conformations on the basis of the similarity of the screened compounds with the co-crystallized ligands. The ligand-based analyses were performed by using the Pipeline Pilot Extended Connectivity Fingerprints (ECFP4) and the ROCS software that allowed taking into account the different aspects of the ligand similarity (*i.e.*, chemical structure or shape and chemical pattern) [32,33]. The authors also suggested the use of multiple protein conformations in a VS to achieve higher hit rates. In fact, by combining docking to different CDK2 conformations and ligand-based similarity approaches, Broccatelli F. *et al.* demonstrated that the more the screened compounds are structurally similar to the co-crystallized ligand in a given protein conformation, the higher are the hit rates obtained with docking [13]. Similar considerations to those argued by Broccatelli F. *et al.* can also be drawn by a more recent study proposed by Xu X. *et al.*, in which the authors were able to predict binding modes and affinities of a series of Hsp90 inhibitors [34]. Results were in good agreement with the experimental data and were obtained by docking the ligands to Hsp90 conformations selected on the basis of the similarity with the co-crystallized compounds. Moreover, this study also demonstrated that if a number of ligand-protein complexes are available for the target under study, significant improvements in binding mode prediction can be achieved by performing docking in conformations carefully selected on the basis of ligand similarity rather than through ensemble-based strategies, which use multiple protein conformations [34]. Similar improvements in binding mode prediction were also achieved by Duan R. *et al.* [35] that successfully implemented a workflow combining 3D ligand similarity evaluations for protein conformation selection and MD simulations of the obtained complexes to prospectively predict binding modes of a series of FXR ligands [35]. Despite the aforementioned studies were not aimed at identifying multi-target ligands, similar results, in principle, could be obtained when selecting the most suitable conformations of different proteins in multi-target projects, provided that the ligands co-crystallized with the two targets present overlapping pharmacophoric and/or structural features. In general, protein conformations in complex with the same inhibitor, or with similar ligands, should be preferred in polypharmacology studies. This rationale has been applied in the selection of protein conformations to get some insights into the mechanism of dual inhibition of a PI3K/mTOR dual inhibitor [36] and for prospective multi-target drug

design of dual p53-MDM2/MDMX interaction inhibitors [37]. However, the ability of ligands to bind to different biological targets by using separate chemical moieties (*head-to-tail* ligands) and/or adopting different conformations cannot be neglected [38,39].

Ligand information derived by protein-ligand complexes has also been exploited for structure-based drug repositioning. In particular, Kalinina O.V. *et al.* [40] were able to predict thousands of known and unknown protein-drug interactions by first superimposing 3D structures of distant proteins in complex with a common active ligand and then, evaluating the structural complementarity of the newly identified cross protein-ligand complexes [40]. Even if this study did not focus on protein conformation selection, the obtained results highlighted the importance of selecting the most suitable protein conformations for drug repurposing. In particular, the identification of ligand-target associations with such a method was possible considering that: (i) proteins sharing a common ligand will likely accommodate other binders of the targets [40]; and (ii) cross protein-ligand compatibility can be appropriately evaluated using receptor conformations of the targets whose binding pockets present similarities and/or accommodate similar ligands. Interestingly, these considerations are, to some extent, in line with those reported in recent papers by Tyzack J.D. *et al.* and Das S. *et al.*, in which the authors compared ligand-enzyme complexes from the PDB on the basis of the similarity of the bound ligands and discussed on how such results could be used to select the most relevant complexes for structure-based drug design [41,42]. Nevertheless, the selection of the most suitable protein conformations on the basis of the similarity of the co-crystallized ligands obviously presents some limitations. Firstly, it is possible only for those targets for which an adequate number of protein-ligand complexes are available [13]. Moreover, intrinsic limitations of ligand-based approaches should also be considered [43]. In addition, the identification of multi-target ligands structurally similar to that already co-crystallized might be unfavorable when searching for novel chemotypes, *e.g.*, to avoid patent issues. On the other hand, several advantages can also be envisioned. In particular, if the structure of the identified hits is similar to those of the co-crystallized ligands, the newly identified multi-target hits are less likely to be false positives [39]. In addition, structure-activity relationships of already reported crystallographic congeners might also help to identify which chemical moieties are important for the activity. However, one should be aware that the application of different ligand-based approaches in the similarity assessments may result in different protein conformation selections. In fact, different degree of similarity might be evaluated for the same pair of compounds on the basis of the chosen method (*e.g.*, 2D vs. 3D ligand-based approaches). In this regard, a recent study by Chen *et al.*, which reported an exhaustive analysis of a subset of the PDB structures by using different protein-centric and ligand-centric approaches, demonstrated that different performances in retrieving protein pairs sharing common ligands could be obtained according to the employed method [44]. Interestingly, the reported results allow to derive some guidelines for multi-target drug-design as well. In particular, they showed that 3D-based approaches outperformed 2D methods in retrieving structurally unrelated ligands and sequence unrelated proteins (but with similar binding sites), whereas 2D methods outperformed 3D ones in recognizing “obvious” structural relationships [44]. For the purpose of this review, these results could be reworded as follows: if the targets under study share a number of significantly similar co-crystallized ligands, 2D methods should be preferred for identifying the most suitable protein conformations for polypharmacology design. On the contrary, if available crystal structures of the targets are in complex with structurally unrelated ligands, 3D methods might help in identifying the most appropriate conformations.

Structure-based approaches in protein conformation selection

A variety of protein-centered approaches might represent efficient tools to identify suitable protein conformations for polypharmacology studies [9,18,25,45]. In fact, the shape and properties of a binding site are mostly dependent on amino acid rearrangements, switches between active and inactive protein states, and the presence and chemical nature of the bound ligand. For example, the conformations of active and inactive states of PKs are known to be significantly different, leading to distant binding pockets. Selecting an active or an inactive PK conformation for polypharmacology ligand design would most likely lead to different sets of retrieved hits [46]. Moreover, PK structures may show additional allosteric pockets or sub-pockets that are not accessible in crystal structures with non-allosteric ligands.

Binding site similarity evaluations allow the comparison of protein binding cavities [18]. Methods enabling the evaluation of the 3D similarity (on the basis of shape or geometrical similarities) between binding sites of structurally unrelated proteins have gained a considerable attention in medicinal chemistry, especially for the prediction of off-target activities [18,19]. Such methods allow the identification of binding pockets with similar anchoring points (or similar shapes) among different biological targets [18]. Moreover, they might also help identifying common “hot spots” across different targets derived by the presence of specific residues or water molecules involved into relevant H-bond networks that could explain ligand promiscuity [47]. Methods for 3D binding site similarity prediction have been successfully applied to assist polypharmacology-related tasks [18,48,49]. For example, BioGPS [50], which performs a binding site comparison by first detecting cavities in a given crystal structure, and, then, by evaluating the overlap in the molecular interaction fields (MIFs) between two different targets, allowed Siragusa L. *et al.* to unveil new potential target combinations for multi-target drug design [50]. In particular, the authors reported similarities between the binding site of the Sarcoplasmic Reticulum Ca^{2+} ion channel ATPase and the Estrogen Receptor alpha, even if the two proteins present a distant structural folding. Moreover, the authors also identified the structural determinants of sub-family selectivity in both the p38a/ERK2 and PPAR α /PPAR γ proteins [50]. Similarly, Niu M. *et al.* [51] identified Gefitinib as an inhibitor of β -secretase (BACE-1) by performing: (i) an extensive similarity analysis between BACE-1 and PK structures with the TM-align tool [52], which suggested the EGFR protein as one of the most structurally similar PKs to BACE-1; (ii) local binding site comparisons between BACE-1 and EGFR by using the SMAP software [53] further demonstrating that these proteins present a high local binding site similarity; and, finally, (iii) *in vitro* testing 13 FDA-approved drugs targeting EGFR on BACE-1 [51]. Combinations of different approaches for binding site comparisons have also successfully explained the promiscuity of ligands from a receptor-based point of view, outperforming the use of a single method [47]. However, to the best of our knowledge, no systematic study has been performed so far to explore the possibility of identifying the most suitable protein conformations by means of binding site similarity in a multi-target drug design context. Given the potentialities of currently available binding site comparison methods, we believe that such approaches will play an important role for the selection of the most suitable protein conformations in future structure-based polypharmacology drug design. Sequence similarity-based methods have also been shown to be particularly efficient in identifying closely related targets, being often more efficient than 3D methods [44]. However, as the 3D coordinates of protein residues cannot be taken into account with such methods, they cannot be used for the protein conformation selection.

Although recent progresses in structural biology and crystal structure determination have increased the number of accessible protein conformations [54,55], the applicability of structure-based approaches to identify suitable receptor conformations accommodating similar ligands remains limited by the availability of 3D structures. Moreover, chosen criteria for the binding site definition (*i.e.*, the residues selected for the binding site definition) may affect the final results. Interestingly, these methods could be used in combination with molecular dynamics (MD) to further explore the conformational space of the investigated binding cavity. MD is mainly used to: (i) describe the dynamic behavior of proteins at different timescales; (ii) understand the events occurring during the ligand-protein interaction process; and (iii) interpret the results of biophysical experiments and modeling studies [56,57]. However, MD approaches have also been applied in combination with the analyses of molecular descriptors, as molecular shape and volumes (POVME) [58], and statistical procedures (TRAPP) [59] to identify transient receptor conformations and help accounting for protein flexibility. Indeed, protein flexibility could strongly bias VS results [60,61] so that a variety of computational approaches has been developed to tackle this issue [62–66]. Combinations of MD and appropriate analysis tools could help exploring binding site flexibility, and thus obtaining representative structures, including alternative protein conformations such as different protein functional states, metastable intermediate conformations, and structures with transient pockets, to be employed for drug design [67]. For example, Osguthorpe D.J. *et al.* successfully performed MD simulations and binding site clustering to identify a subset of conformations of the Androgen Receptor for VS, thus improving database enrichment and the chemical diversity of the retrieved actives with respect to screening to a single conformation [64]. Moreover, the authors also applied the same binding site clustering approach on sets of HIV protease and CDK2 crystal structures, demonstrating that enrichments and ligand diversity can be also improved when docking is performed on conformational ensembles of crystallographic structures [64]. Structure-based approaches combining rigid receptor docking with protein structure modeling have also successfully overcome receptor flexibility issues [68] and have been applied to rationalize the selectivity of multi-target ligands. In particular, in a recent study of Selvam B. and coworkers, an application of MD simulations with Desmond [69] and probe mapping with the “*multifragment*” search module of the MOE2011.10 program [70] was used to address ligand selectivity [45]. In this study, the authors generated the conformational space of a set of similar bioaminergic G-protein-coupled receptors through MD, allowing the identification of common and distinct interaction “*hot-spots*” that were used to investigate selectivity and side effects. These features could be useful to guide the design of molecules able to circumvent metabolic side effects and pulmonary hypertension, which are commonly associated with antipsychotic drugs targeting bioaminergic receptors. Then, they performed retrospective docking screenings (with Glide) on a set of conformations of the different proteins that allowed to discriminate known ligands from decoys [28,45]. Despite its utility in single target drug design [67], MD has not been thoroughly explored for selecting protein conformations for polypharmacology studies. We envision that a combination of MD simulations (to generate an ensemble of protein conformations) and binding site comparisons (to assess binding site similarity of the generated conformations) could significantly increase the possibility of identifying overlapping conformations between two or more targets (Figure 2).

The combination of structure refinement, MD, binding energy predictions, and docking has also been proposed to account for protein flexibility and improve the docking accuracy [71–73]. In fact, it is well known that currently available docking algorithms might provide inaccurate binding energy estimations

[74] as they might not properly account for entropic and solvation terms [75]. Therefore, the use of more accurate scoring methods has been explored. In this context Greenidge P.A. *et al.* demonstrated that advanced workflows, including combinations of MM-GBSA binding free energy rescoring and docking, improved pose prediction and binding affinity estimations in a VS [76]. However, it should be noted that such methods might present intrinsic limitations, one of them being the computational cost. More recently, scoring functions (SC) based on machine-learning approaches have also been developed. For example, Wójcikowski M. *et al.* have recently reported a scoring function (RF-Score-VS) based on the machine-learning random forest approach [77]. They also demonstrated that machine-learning based SC could exceed classical SC in correctly scoring training set compounds across different biological targets [77]. Similar results were also achieved by Kinnings S.L. *et al.* in the drug repurposing field, where they successfully applied a classification support vector machine approach to improve SC performance in docking [78]. In particular, the authors were able to identify Phosphodiesterase (PDE) inhibitors with potential activity on InhA, a recognized target for treating *Mycobacterium Tuberculosis* infections. Moreover, the authors were also able to prioritize a set of compounds known to bind to estrogen-like receptors, further confirming the ability of machine-learning based SC in predicting binding affinity [78]. Despite the potentialities of machine-learning approaches in drug discovery, applications in protein conformation selection have not yet been reported.

Molecular docking can be considered as well a useful tool for the selection of the most suitable protein conformations for structure-based VSs. In this regard, a recent study of Wang *et al.* reported how integration of molecular docking and statistical analysis could help in identifying the most suitable protein conformations for structure-based drug design [25]. In particular, the authors investigated 249 protein structures of 14 autophagy-related targets by using Surflex-dock [79]. Then, they prioritized protein conformations on the basis of their ability to distinguish active ligands from inactive ones evaluated in terms of docking score distributions. However, as also stated by the authors, the applicability of this methodology strongly depends on the availability of active and inactive ligands for the target under study [25]. Therefore, the applicability of docking-based protein conformation selection in polypharmacology design requires an adequate number of known inhibitors for all the targets under study. More recently, Jaiteh M. *et al.* have reported the identification of novel dual-target ligands of A_{2A}AR and MAO-B by means of docking screenings [80]. In particular, the authors performed first parallel validations of a series of A_{2A}AR and MAO-B docking protocols, without focusing on structures of already reported dual inhibitors to avoid potential biases in the prospective screening. This allowed the identification of structurally distant protein conformations of the two targets, providing the best logAUC on a training set of ligands. Then, they carried out structure-based VS, prioritizing screening compounds on the basis of a consensus rank on the two targets. Thanks to this approach, Jaiteh M. *et al.* identified two nanomolar dual inhibitors of A_{2A}AR and MAO-B. However, subsequent structure-activity relationship studies did not provide improvements of the dual activity profiles, as even slight modifications of the identified scaffolds negatively affected the activity toward one target or the other. Therefore, these results provided some insights into the use of docking methods for selecting the suitable protein conformations for multi-target drug design.

Concluding Remarks and Future Perspectives

Structure-based methods represent essential tools for designing multi-target ligands. Despite a variety of methods and algorithms are currently available to estimate the complementarity of ligands with the binding site of the targets under study, the selection of the proper protein conformations is still of paramount concern. In fact, it is well known that structure-based VS results are strongly dependent on the used protein conformation and that structural details or peculiarities in the selected binding sites may allow potential biases toward the identification of certain chemical scaffolds instead of others [13,14]. These issues become even more relevant in polypharmacology applications, even if they have received very little attention so far. However, the information provided by available protein-ligand complexes, which allows encoding key receptor-ligand interactions and anchoring points of a protein binding site, should be always taken into account when performing polypharmacology structure-based drug design. Several approaches have been proposed to partially overcome potential pitfalls related to the use of a single receptor conformation in standard VS. In particular, methods based on the use of an ensemble of protein conformations and docking algorithms incorporating receptor flexibility in the calculations have demonstrated to provide higher hit rates with respect to the use of a single protein conformation [65,66,81–84]. In multi-target drug design, considering an ensemble of protein conformations may increase the chances of retrieving compounds with dual activity, as a consequence of the higher probability of identifying conformations of the two targets sharing common anchor points. Moreover, the selection of appropriate protein conformations becomes even more relevant when the two targets display low similarities because of the presence of water molecules in the binding site. In such cases, the importance of specific water molecules for ligand binding should be carefully assessed *a priori* in order to: (i) avoid including in the binding site thermodynamically unfavorable waters (potentially replaceable by the inhibitors) and; (ii) retain those involved in highly conserved H-bond networks. In this regard, the reported protein-ligand complexes may allow the identification of conserved water-mediated protein-ligand interactions that cannot be neglected in the ligand design process [85]. However, even though several drug design applications have been reported [86,87], this issue has not yet been thoroughly considered in multi-target drug design.

In general, binding site conformations of the different targets should be selected by trying to: (i) limit biases on compound selection derived by the use of peculiar protein conformations, which may represent low populated protein states; and (ii) identify, either by ligand similarity or binding site comparison techniques, protein conformations able to accommodate compounds with similar scaffolds. Therefore, selecting multiple conformations across different targets to enable structure-based drug design represents a challenging task. However, a number of ligand- and structure-based approaches are available to analyze the wealth of information provided by publicly available databases of either protein-ligand complexes or apo structures [9]. Moreover, the combination of different computational approaches could allow to thoroughly exploit information derived by the structural details of both co-crystallized ligands and residues lining the binding site. In particular, when an adequate number of structures of protein-ligand complexes are available, structure- and ligand-based analyses can be performed in parallel to guide the selection of protein conformations. Such approaches may allow the identification of several pairs of protein conformations with sufficient similarity to be used in docking. For example, ligand-based similarity approaches on the co-crystallized ligands could help in identifying pairs of protein conformations able to accommodate similar compounds, whereas structure-centered approaches might allow the identification of common anchoring points or similar structural features shared by different

protein cavities. These latter approaches are especially useful when dealing with apo structures. In the same line, molecular dynamics techniques combined with binding site comparison tools, machine-learning approaches, analyses of molecular descriptors, *etc.* may also allow to sample the conformational landscape of proteins, thus enabling the use of alternative protein conformations for multi-target ligand design [67]. Overall, even if it received very little attention in polypharmacology so far, the selection of protein conformations operated upon these considerations will likely provide structure-based models with improved forecasting performances, thus creating novel opportunities for future multi-target drug design.

Acknowledgements

The project leading to this article has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 676434, "Big Data in Chemistry" ("BIGCHEM", <http://bigchem.eu>). The article reflects only the authors' view and neither the European Commission nor the Research Executive Agency are responsible for any use that may be made of the information it contains.

References

- [1] Anighoro A, Bajorath J, Rastelli G. Polypharmacology: Challenges and opportunities in drug discovery. *J Med Chem* 2014; 57:7874–87.
- [2] Frantz S. Drug discovery: Playing dirty. *Nature* 2005; 437:942–3.
- [3] Talevi A. Multi-target pharmacology: possibilities and limitations of the "skeleton key approach" from a medicinal chemist perspective. *Front Pharmacol* 2015; 6:205.
- [4] Reddy AS, Zhang S. Polypharmacology: drug discovery for the future. *Expert Rev Clin Pharmacol* 2013; 6:10.1586/ecp.12.74. DOI:10.1586/ecp.12.74. (<https://www.tandfonline.com/toc/ierj20/current>).
- [5] Anighoro A, Pinzi L, Marverti G, Bajorath J, Rastelli G. Heat shock protein 90 and serine/threonine kinase B-Raf inhibitors have overlapping chemical space. *RSC Adv* 2017; 7:31069-74.
- [6] Peng F-W, Xuan J, Wu T-T, Xue J-Y, Ren Z-W, Liu D-K, et al. Design, synthesis and biological evaluation of N-phenylquinazolin-4-amine hybrids as dual inhibitors of VEGFR-2 and HDAC. *Eur J Med Chem* 2016; 109:1–12.
- [7] Kalash L, Val C, Azuaje J, Loza MI, Svensson F, Zoufir A, et al. Computer-aided design of multi-target ligands at A1R, A2AR and PDE10A, key proteins in neurodegenerative diseases. *J Cheminform* 2017; 9:67.
- [8] Zhang W, Pei J, Lai L. Computational Multitarget Drug Design. *J Chem Inf Model* 2017; 57:403–12.
- [9] Rastelli G, Pinzi L. Computational polypharmacology comes of age. *Front Pharmacol* 2015;6:157.
- [10] Ma XH, Shi Z, Tan C, Jiang Y, Go ML, Low BC, et al. In-silico approaches to multi-target drug discovery : computer aided multi-target drug design, multi-target virtual screening. *Pharm Res* 2010; 27:739–49.
- [11] Śledź P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol* 2018; 48:93–102.

- [12] Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J Med Chem* 2004; 47:45–55.
- [13] Broccatelli F, Brown N. Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. *J Chem Inf Model* 2014; 54: 1634–41.
- [14] Jain AN. Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. *J Comput Aided Mol Des* 2009; 23:355–74.
- [15] Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW. Protein–Ligand Docking against Non-Native Protein Conformers. *J Chem Inf Model* 2008; 48:2214–25.
- [16] Sheridan RP, McGaughey GB, Cornell WD. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J Comput Aided Mol Des* 2008; 22:257–65.
- [17] Zhou S, Li Y, Hou T. Feasibility of using molecular docking-based virtual screening for searching dual target kinase inhibitors. *J Chem Inf Model* 2013; 53:982–96.
- [18] Ehrt C, Brinkjost T, Koch O. Impact of Binding Site Comparisons on Medicinal Chemistry and Rational Molecular Design. *J Med Chem* 2016; 59:4121–51.
- [19] Sturm N, Desaphy J, Quinn RJ, Rognan D, Kellenberger E. Structural Insights into the Molecular Basis of the Ligand Promiscuity. *J Chem Inf Model* 2012; 52:2410–21.
- [20] Costa FF. Big data in biomedicine. *Drug Discov Today* 2014; 19:433–40.
- [21] Liu X, Xu Y, Li S, Wang Y, Peng J, Luo C, et al. In Silico target fishing: addressing a “Big Data” problem by ligand-based similarity rankings with data fusion. *J Cheminform* 2014; 6:33.
- [22] Hu Y, Bajorath J. Entering the “big data” era in medicinal chemistry: molecular promiscuity analysis revisited. *Futur Sci OA* 2017; 3:FSO179. DOI: 10.4155/fsoa-2017-0001 (<https://www.future-science.com/loi/fso>).
- [23] Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN. Virtual screening in drug discovery -- a computational perspective. *Curr Protein Pept Sci* 2007; 8:329–51.
- [24] Xie L, Xie L, Kinnings SL, Bourne PE. Novel computational approaches to polypharmacology as a means to define responses to individual drugs. *Annu Rev Pharmacol Toxicol* 2012; 52:361–79.
- [25] Wang N, Wang L, Xie X-Q. ProSelection: A Novel Algorithm to Select Proper Protein Structure Subsets for in Silico Target Identification and Drug Discovery Research. *J Chem Inf Model* 2017; 57:2686–98.
- [26] Stahura FL, Bajorath J. New methodologies for ligand-based virtual screening. *Curr Pharm Des* 2005; 11:1189–202.
- [27] Ripphausen P, Nisius B, Bajorath J. State-of-the-art in ligand-based virtual screening. *Drug Discov Today* 2011; 16:372–6.
- [28] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* 2004; 47:1739–49.
- [29] Voigt JH, Elkin C, Madison VS, Duca JS. Cross-Docking of Inhibitors into CDK2 Structures. 2. *J Chem Inf Model* 2008; 48:669–78.
- [30] Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res* 2014; 42:D1083–90.
- [31] Huang N, Shoichet BK, Irwin JJ. Benchmarking Sets for Molecular Docking. *J Med Chem* 2006; 49:6789–

801.

- [32] Pipeline Pilot v8.0.1.500. (<http://accelrys.com/products/pipeline-pilot/>).
- [33] Hawkins PCD, Skillman AG, Nicholls A. Comparison of shape-matching and docking as virtual screening tools, *J Med Chem* 2007; 50:74-82.
- [34] Xu X, Yan C, Zou X. Improving binding mode and binding affinity predictions of docking by ligand-based search of protein conformations: evaluation in D3R grand challenge 2015. *J Comput Aided Mol Des* 2017; 31:689-99.
- [35] Duan R, Xu X, Zou X. Lessons learned from participating in D3R 2016 Grand Challenge 2: compounds targeting the farnesoid X receptor. *J Comput Aided Mol Des* 2018; 32:103-11.
- [36] Takeda T, Wang Y, Bryant SH. Structural insights of a PI3K/mTOR dual inhibitor with the morpholino-triazine scaffold. *J Comput Aided Mol Des* 2016; 30:323-30.
- [37] Chen S, Li X, Yuan W, Zou Y, Guo Z, Chai Y, et al. Rapid identification of dual p53-MDM2/MDMX interaction inhibitors through virtual screening and hit-based substructure search. *RSC Adv* 2017; 7:9989-97.
- [38] Rizeq N, Georgiades SN. Investigation of 'Head-to-Tail'-Connected Oligoaryl N,O-Ligands as Recognition Motifs for Cancer-Relevant G-Quadruplexes. *Molecules* 2017; 22:2160. DOI: <https://doi.org/10.3390/molecules22122160>. (<http://www.mdpi.com/journal/molecules>) .
- [39] Gilberg E, Stumpfe D, Bajorath J. X-ray-Structure-Based Identification of Compounds with Activity against Targets from Different Families and Generation of Templates for Multitarget Ligand Design. *ACS Omega* 2018; 3:106-11.
- [40] Kalinina O V, Wichmann O, Apic G, Russell RB. Combinations of protein-chemical complex structures reveal new targets for established drugs. *PLoS Comput Biol* 2011; 7:e1002043. DOI:10.1371/journal.pcbi.1002043. (<http://journals.plos.org/ploscompbiol/>)
- [41] Tyzack JD, Fernando L, Ribeiro AJM, Borkakoti N, Thornton JM. Ranking Enzyme Structures in the PDB by Bound Ligand Similarity to Biological Substrates. *Structure* 2018; 26:565-571.e3.
- [42] Das S, Orengo C. Choosing the Best Enzyme Complex Structure Made Easy. *Structure* 2018; 26:528-30.
- [43] Venkatraman V, Perez-Nueno VI, Mavridis L, Ritchie DW. Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. *J Chem Inf Model* 2010; 50:2079-93.
- [44] Chen Y-C, Tolbert R, Aronov AM, McGaughey G, Walters WP, Meireles L. Prediction of Protein Pairs Sharing Common Active Ligands Using Protein Sequence, Structure, and Ligand Similarity. *J Chem Inf Model* 2016; 56:1734-45.
- [45] Selvam B, Porter SL, Tikhonova IG. Addressing selective polypharmacology of antipsychotic drugs targeting the bioaminergic receptors through receptor dynamic conformational ensembles. *J Chem Inf Model* 2013; 53:1761-74.
- [46] Kontopidis G, McInnes C, Pandalaneni SR, McNae I, Gibson D, Mezna M, et al. Differential binding of inhibitors to active and inactive CDK2 provides insights for drug design. *Chem Biol* 2006; 13:201-11.
- [47] Sturm N, Quinn RJ, Kellenberger E. Similarity between Flavonoid Biosynthetic Enzymes and Flavonoid Protein Targets Captured by Three-Dimensional Computing Approach. *Planta Med* 2015; 81:467-73. DOI:10.1055/s-0035-1545697. (<https://www.thieme-connect.de/products/ejournals/journal/10.1055/s-00000058>).

- [48] Jalencas X, Mestres J. Identification of Similar Binding Sites to Detect Distant Polypharmacology. *Mol Inform* 2013; 32:976–90.
- [49] Duran-Frigola M, Siragusa L, Ruppin E, Barril X, Cruciani G, Aloy P. Detecting similar binding pockets to enable systems polypharmacology. *PLOS Comput Biol* 2017; 13:e1005522. DOI:10.1371/journal.pcbi.1005522. (<http://journals.plos.org/ploscompbiol/>).
- [50] Siragusa L, Cross S, Baroni M, Goracci L, Cruciani G. BioGPS: navigating biological space to predict polypharmacology, off-targeting, and selectivity. *Proteins* 2015; 83:517–32. DOI:10.1002/prot.24753. (<https://onlinelibrary.wiley.com/journal/10970134>).
- [51] Niu M, Hu J, Wu S, Xiaoe Z, Xu H, Zhang Y, et al. Structural bioinformatics-based identification of EGFR inhibitor gefitinib as a putative lead compound for BACE. *Chem Biol Drug Des* 2014; 83:81–8. DOI: <https://doi.org/10.1111/cbdd.12200>. ([http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1747-0285](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1747-0285)).
- [52] Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005; 33:2302–9.
- [53] Xie L, Xie L, Bourne PE. A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery. *Bioinformatics* 2009; 25:i305–12.
- [54] Bard J, Ercolani K, Svenson K, Olland A, Somers W. Automated systems for protein crystallization. *Methods* 2004; 34:329–47.
- [55] Gavira JA. Current trends in protein crystallization. *Arch Biochem Biophys* 2016; 602:3–11.
- [56] Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol* 2011; 9:71. DOI:10.1186/1741-7007-9-71 (<https://bmcbiol.biomedcentral.com/>).
- [57] De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* 2016; 59:4035–61.
- [58] Durrant JD, de Oliveira CAF, McCammon JA. POVME: An Algorithm for Measuring Binding-Pocket Volumes. *J Mol Graph Model* 2011; 29:773–6.
- [59] Kokh DB, Richter S, Henrich S, Czodrowski P, Rippmann F, Wade RC. TRAPP: A Tool for Analysis of Transient Binding Pockets in Proteins. *J Chem Inf Model* 2013; 53:1235–52.
- [60] Brown SP, Hajduk PJ. Effects of Conformational Dynamics on Predicted Protein Druggability. *ChemMedChem* 2006; 1:70–2.
- [61] Sgobba M, Caporuscio F, Anighoro A, Portioli C, Rastelli G. Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. *Eur J Med Chem* 2012; 58:431–40.
- [62] Buonfiglio R, Recanatini M, Masetti M. Protein Flexibility in Drug Discovery: From Theory to Computation. *ChemMedChem* 2015; 10:1141–8.
- [63] Nichols SE, Baron R, Ivetac A, McCammon JA. Predictive Power of Molecular Dynamics Receptor Structures in Virtual Screening. *J Chem Inf Model* 2011; 51:1439–46.
- [64] Osguthorpe DJ, Sherman W, Hagler AT. Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chem Biol Drug Des* 2012; 80:182–93. DOI:10.1111/j.1747-0285.2012.01396.x. ([http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1747-0285](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1747-0285)).
- [65] Rueda M, Bottegoni G, Abagyan R. Recipes for the Selection of Experimental Protein Conformations for

Virtual Screening. *J Chem Inf Model* 2010; 50:186–93.

- [66] Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W, Farid R. Improving database enrichment through ensemble docking. *J Comput Aided Mol Des* 2008; 22:621–7.
- [67] Caporuscio F, Rastelli G. Exploiting computationally derived out-of-the-box protein conformations for drug design. *Future Med Chem* 2016; 8:1887–97.
- [68] Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 2006; 49:534–53.
- [69] Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, November 11-17.
- [70] Molecular Operating Environment (MOE); Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018.
- [71] Gioia D, Bertazzo M, Recanatini M, Masetti M, Cavalli A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. *Molecules* 2017; 22(11). DOI:10.3390/molecules22112029. (<http://www.mdpi.com/journal/molecules>).
- [72] Degliesposti, G.; Portioli, C.; Parenti, M. D.; Rastelli G. BEAR, a novel virtual screening methodology for drug discovery. *J Biomol Screening* 2011; 16(1):129–33.
- [73] Parenti MD, Rastelli G. Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnol Adv* 2012; 30:244–50.
- [74] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004; 3:935–49.
- [75] Gilson MK, Zhou H-X. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* 2007; 36:21–42.
- [76] Greenidge PA, Kramer C, Mozziconacci J-C, Sherman W. Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. *J Chem Inf Model* 2014; 54:2697–717.
- [77] Wojcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci Rep* 2017; 7:46710. DOI:10.1038/srep46710. (<http://www.nature.com/srep/index.html>)
- [78] Kinnings SL, Liu N, Tonge PJ, Jackson RM, Xie L, Bourne PE. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J Chem Inf Model* 2011; 51:408–19.
- [79] Jain AN. Surfex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* 2003; 46:499–511.
- [80] Jaiteh M, Zeifman A, Saarinen M, Svenningsson P, Brea J, Loza MI, et al. Docking Screens for Dual Inhibitors of Disparate Drug Targets for Parkinson’s Disease. *J Med Chem* 2018; 61:5269–78.
- [81] Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* 2008; 18:178–84.
- [82] Zhao H, Huang D, Caflisch A. Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics. *ChemMedChem* 2012; 7:1983–90.

- [83] Bottegoni G, Rocchia W, Rueda M, Abagyan R, Cavalli A. Systematic Exploitation of Multiple Receptor Conformations for Virtual Ligand Screening. PLoS One 2011; 6:e18845. DOI:10.1371/journal.pone.0018845. (<http://www.plosone.org/>)
- [84] Sherman W, Beard HS, Farid R. Use of an induced fit receptor structure in virtual screening. Chem Biol Drug Des 2006; 67:83–4. DOI:10.1111/j.1747-0285.2005.00327.x. ([http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1747-0285](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1747-0285)).
- [85] Sgobba M, Rastelli G. Structure-based and in silico design of Hsp90 inhibitors. ChemMedChem 2009;4:1399–409.
- [86] Lenselink EB, Beuming T, Sherman W, van Vlijmen HWT, IJzerman AP. Selecting an optimal number of binding site waters to improve virtual screening enrichments against the adenosine A2A receptor. J Chem Inf Model 2014; 54:1737–46.
- [87] Knegtel RMA, Robinson DD. A Role for Hydration in Interleukin-2 Inducible T Cell Kinase (Itk) Selectivity. Mol Inform 2011; 30:950–9.

Figure Legends

Figure 1. Ligand-based methods can be used to select the most suitable protein conformations to be used in structure-based multi-target drug design. Panel *a*) shows that ligands extracted from the complexes in their crystallographic conformations can be used to identify pairs of ligands (and, therefore, pairs of protein conformations) for the subsequent VS. Different approaches can be used for the assessment of the ligand similarity. Moreover, the identified ligand cross-similarities could also be used, in principle, to prescreen a database of ligands in a VS campaign - Panel *b*).

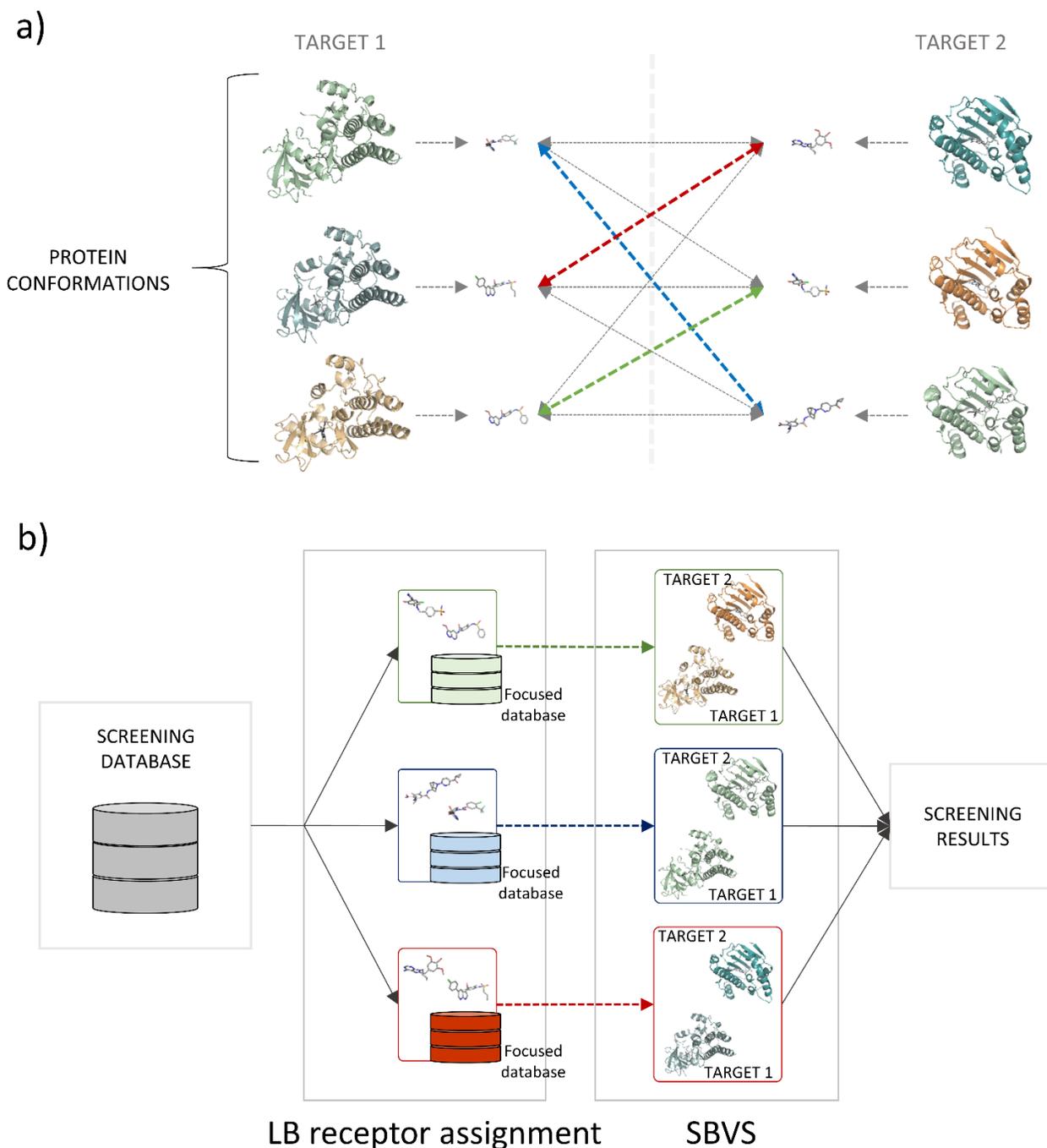


Figure 2. Structure-based approaches can be used to compare the binding site of different protein conformations or validate screening protocols, when an appropriate number of crystal structures conformation is available for the targets under study. These analyses could allow the identification of protein conformations across different targets that are able to accommodate the same ligands. Moreover, molecular dynamics simulations can be used to further sample the protein conformational space. The identified pairs of protein conformations can then be used in a parallel structure-based VS workflow to

increase the chances of identifying multi-target inhibitors, either as a single conformation *per* target or as ensembles.

