Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches

Javier Vázquez [1,2,*], **Manel López** [3], **Enric Gibert** [1], **Enric Herrero** [1,*] and **F. Javier Luque** [2,*]

1   Pharmacelera, Plaça Pau Vila, 1, Sector C 2a, Edificio Palau de Mar, 08039 Barcelona, Spain; enric.gibert@pharmacelera.com

2   Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, Institute of Biomedicine (IBUB), and Institute of Theoretical and Computational Chemistry (IQTC-UB), University of Barcelona, Av. Prat de la Riba 171, E-08921 Santa Coloma de Gramanet, Spain

3   AB Science, Parc Scientifique de Luminy, Zone Luminy Enterprise, Case 922, 163 Av. de Luminy, 13288 Marseille, France; manel.lopez@ab-science.com

*   Correspondence: javier.vazquez@pharmacelera.com (J.V.); enric.herrero@pharmacelera.com (E.H.); fjluque@ub.edu (F.J.L.)

Abstract: Virtual screening (VS) is an outstanding cornerstone in the drug discovery pipeline. A variety of computational approaches, which are generally classified as ligand-based (LB) and structure-based (SB) techniques, exploit key structural and physicochemical properties of ligands and targets to enable the screening of virtual libraries in the search of active compounds. Though LB and SB methods have found widespread application in the discovery of novel drug-like candidates, their complementary natures have stimulated continued efforts toward the development of hybrid strategies that combine LB and SB techniques, integrating them in a holistic computational framework that exploits the available information of both ligand and target to enhance the success of drug discovery projects. In this review, we analyze the main strategies and concepts that have emerged in the last years for defining hybrid LB + SB computational schemes in VS studies. Particularly, attention is focused on the combination of molecular similarity and docking, illustrating them with selected applications taken from the literature.

1. Introduction

Predicting with chemical accuracy the biological activity that a small drug-like compound can attain against its target is a major challenge in drug discovery. In the late stages of lead optimization, this task can be accomplished by resorting to enhanced sampling techniques, such as free energy calculations [1–3], which can estimate the binding affinity between a ligand and its macromolecular target. Remarkably, the effort spent in developing robust algorithms in conjunction with efficient configurational sampling methods permit to estimate the binding affinity with a chemical accuracy close to the 1 kcal/mol limit [4–7], though at the expense of a significant computational cost that prevents their application in large datasets. Nevertheless, attempts have been made to alleviate this limitation through the development of automated workflows for the in silico prediction of binding affinities [8,9], which will facilitate the usage of these sophisticated techniques to nonexpert researchers in computational chemistry.

A different scenario occurs in the early stages of drug discovery, where attention is focused on the identification of potential hit compounds endowed with a promising activity against a druggable target or, alternatively, the search of novel chemotypes that may lead to innovative strategies in the treatment of diseases and pathological disorders. In these stages, the suitability of several computationally demanding methods, such as steered molecular dynamics (MD) and quantum mechanical-based approaches, has been explored, generally when the interest is focused on a reduced set of compounds [10–12]. Nevertheless, when one keeps in mind the diversity of the chemical universe that can a priori be explored [13–15], the ability to discriminate between actives and inactives still represents a formidable challenge that makes it necessary to resort to simplified computational approaches. At this point, it is worth noting that the number of different compounds that could be synthesized has been estimated to be around 10^{20}–10^{24} molecules [16]. The vast amount of drug-like chemical libraries can be explored using in silico virtual screening (VS) techniques, which encompass a variety of computational algorithms and formalisms in the search of novel bioactive molecules. They have proven their applicability in numerous studies, leading to hit rates competitive with the results derived from experimental high-throughput screening and at a much lower cost [17–19].

VS techniques can be grouped into two major categories, depending on the available structural information. The term structure-based virtual screening (SBVS), often denoted as target-based VS, encompasses methods that exploit the three-dimensional (3D) structure of the target. The most widely used SBVS technique is molecular docking, which uses the structural and chemical complementarity resulting from the interaction between a fragment-like or drug-like compound and its target receptor, predicting the preferred pose of ligands in the binding site through the use of scoring functions, often supplemented with pharmacophoric constraints [20–23]. On the other hand, ligand-based virtual screening (LBVS) relies on the structural information and physicochemical properties of the chemical scaffold of known active and inactive molecules, which are examined under the molecular similarity principle [24]. Accordingly, the relationships between compounds in a given library and one or more known actives are examined by similarity measurements using suitable molecular descriptors. These measurements can be performed based on 1D and 2D descriptors, generally encoding information about the chemical nature of compounds and their topological features [25–27], and 3D descriptors associated to molecular fields [28–31], shape and volume [32,33], and pharmacophores [28,34].

The combined integration of SBVS and LBVS techniques may be a promising strategy when data about both the structure of ligand-target complexes and similarity relationships to active compounds are available, leading to a holistic framework suitable to enhance the success of drug discovery projects [35,36]. As an example of the potential impact of combining SBVS and LBVS, we limit ourselves to cite a couple of representative studies. The first is the work by Spadaro et al. [37], who used a pharmacophoric model derived from the analysis of X-ray crystallographic data in conjunction with LBVS techniques for disclosing novel inhibitors of the 17β-hydroxysteroid dehydrogenase type 1 (17β-HSD1) enzyme, leading to the identification of a keto-derivative compound with an inhibitory potency in the nanomolar range (Figure 1A). In the second example, Debnath et al. [38] used a combined VS strategy to identify selective non-hydroxamate histone deacetylase 8 (HDAC8) inhibitors (Figure 1B). To this end, a database of 4.3×10^6 molecules was explored using a pharmacophore model, and the top 500 hits retrieved were filtered using ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) criteria. The selected compounds were subsequently assessed by molecular docking. Among the final hits selected for in vitro biological evaluation, compounds SD-01 and SD-02 inhibited the HDAC8 enzyme with IC_{50} (i.e., the concentration of inhibitor that gives half-maximal response) values of 9.0 and 2.7 nM, respectively. These two examples suffice to demonstrate that a judicious choice of LB and SB techniques, adapted to the available information about the ligands and target, may be powerful in disclosing drug-like compounds.
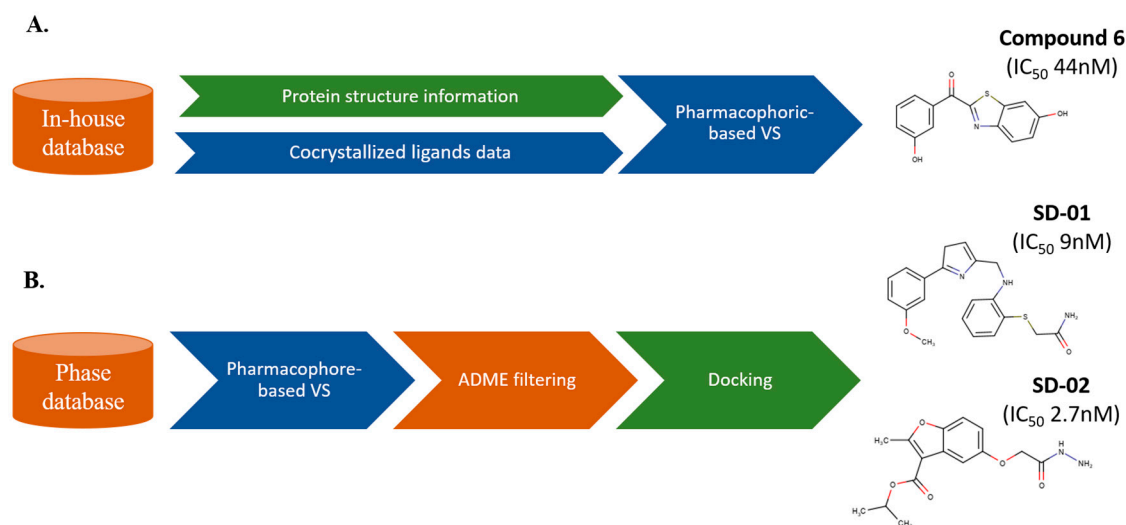
Figure 1. Representative cases of two combined ligand-based (LB) and structure-based (SB) strategies leading to the discovery of potent inhibitors of (**A**) 17β-hydroxysteroid dehydrogenase type 1 (17β-HSD1) and (**B**) histone deacetylase 8 (HDAC8) enzymes. LB and SB methods are highlighted in blue and green, respectively. VS: virtual screening.

Several strategies have been proposed to combine LBVS and SBVS in order to reinforce the mutual complementarity of these approaches and palliate their individual weaknesses [39–41]. The major shortcoming in LBVS is the bias toward the reference template, which may result in overfitting to the input structures. When a pharmacophore is used to guide the screening of the compounds, the chemical features of the ligands present in the training set may affect the optimal choice of the pharmacophoric restraints. Moreover, the available activity data may turn out to be inadequate for selecting a structural and functional pool of compounds, often limited by the absence of data relative to poorly active or inactive compounds, which may be valuable to calibrate the merits of the pharmacophore model in distinguishing between actives and inactives. On the other hand, accounting for protein flexibility is a major drawback for docking methods. The binding site of a protein is flexible and can adopt diverse conformational states, generally at the level of side chain residues but often also involving structural changes in loops and the remodeling of secondary structural elements induced upon ligand binding [42–45]. Furthermore, the outcome of docking studies may be largely affected by the identification of water molecules that mediate the interactions of the ligand in the binding pocket, making it necessary to explore the potential role of bridging waters or networks of ordered waters in docking calculations [46–50]. On the other hand, providing an accurate score and even estimating the binding affinity at a reasonable cost compatible with the screening of large chemical libraries is still challenging for docking methods [51–54]. Finally, the outcome of LBVS and SBVS also appear to exhibit a strong target dependency [55,56]. For the sake of brevity, a detailed discussion of these weaknesses is omitted here, and the reader is addressed to previous studies in the literature [57–60].

In this context, searching for computational strategies that can mitigate the limitations of LB and SB methods has been actively pursued in the last years. One alternative is to focus the screening effort on targeted chemical libraries that would facilitate the task of hit identification [61–65]. This can be achieved via automated algorithms of molecular generation or de novo design, often assisted by artificial intelligence techniques, which aim to create sets of compounds endowed with properties similar to the structural and chemical features found in real cases, including a bias toward specific ranges of physicochemical properties or toward compounds active against a given target. Alternatively, a balanced combination of LB and SB methods may be devised to exploit synergistically the merits of these VS techniques, while counterbalancing their limitations, in order to increase the success rate in the screening of large chemical libraries.

Here, our attention is focused on the methodologies and computational approaches undertaken to enrich the outcome of VS by combining LB and SB techniques. In particular, we review the main strategies that have been proposed combining molecular similarity and docking. The strengths and weaknesses of the combined approaches are illustrated by selecting representative studies reported in the literature, primarily dealing with the efforts reported in the last five years. Overall, the aim of this review is to provide useful guidelines for the application of combined LB and SB methods in drug discovery.

2. LB and SB Strategies in VS

Different schemes can be adopted to combine LB and SB methods. The classification proposed by Drwal and Griffith will be adopted in this review [40]. Accordingly, the discussion of the combined LB and SB strategies can be completed following three main categories: sequential, parallel, and hybrid, which are summarized in Figure 2.
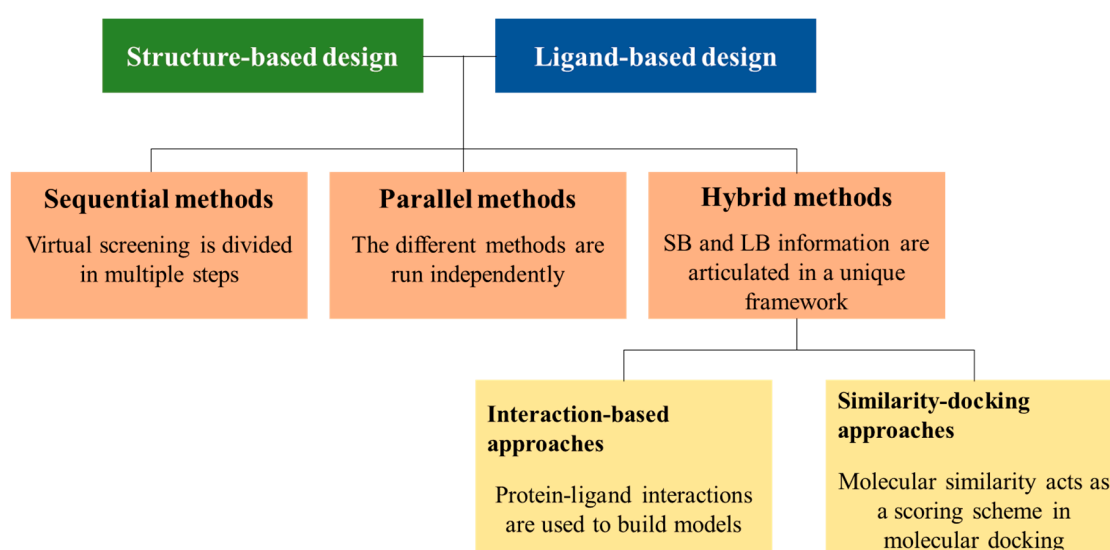

Figure 2. Schematic representation of the three main strategies adopted for combining LB and SB methods.

(i) Sequential approaches divide the VS pipeline in consecutive steps with the aim to perform a progressive filtering in the library of chemical compounds toward the most promising candidates, which will be selected for biological testing at the end of this multi-step process. Generally, prefiltering is performed at the beginning of the VS process using LB techniques due to their reduced computational cost, whereas the most computationally demanding SB methods are exploited in the final stages of the selection process. Thus, this strategy attempts to optimize the tradeoff between the computational expensiveness and the complexity of the formalism that underlies the filtering technique along the VS process. However, they do not exploit all the available information at once and maintain the limitations of the individual methods.

(ii) In the parallel approach, both LB and SB methods are run independently, and the best candidates identified from each separate method are selected for biological testing. Swann et al. reported a prospective application of this approach in 2011 [66], and subsequent studies have examined distinct functional forms for combining the ranks obtained from LB and SB methods (see below). In particular, the compounds obtained in the final rank order lead to meaningful increases in both performance and robustness over the single-modality approaches, but the results also demonstrate the sensitivity of the performance to the target structural details (i.e., the nature of the template ligand in measurements of molecular similarity and the reference protein pocket in docking studies) [67,68].

(iii) Finally, the hybrid strategies comprise approaches that represent a true combination of LB and SB techniques into a standalone method. Two main combinations have been followed to achieve this goal: (i) interaction-based methods and (ii) similarity-docking methods. The former translates the observed protein-ligand interactions into pharmacophoric features and quantitative structure-activity relationship (QSAR) models [39,69,70], which have been used for several applications, such as VS, the profiling of ligands, the analysis of pseudo-receptors, and de novo designs [71–74]. On the other hand, the combination of molecular similarity and docking techniques has been examined in the last years as an alternative procedure to assess the reliability of predicted poses of ligands by measuring the overlay against suitable templates [75–80].

3. Sequential LB and SB Methods

High-throughput VS may be computationally demanding when large sets of compounds have to be evaluated. In this scenario, decomposing the VS pipeline into a multi-step process can be valuable to reduce progressively the number of compounds and enrich the chemical library toward the most promising scaffolds before screening with more expensive methods.

LB techniques are generally used in the prefiltering step, as illustrated in different works that have exploited 2D fingerprints [81,82], 3D molecular similarity [83–85] and pharmacophore models [86–88]. To enhance the drug-likeness of the compounds, knowledge-based in silico ADMET or pan-assay interference compounds (PAINS; [89]) filters can also be applied. For cases with a reduced number of compounds, the results obtained from the SBVS can be further refined, resorting to the structural stability observed in MD simulations [90–93].

As an example that illustrates the sequential application of LB and SB techniques, Khan et al. [84] performed multi-step LBVS and SBVS to identify G protein-coupled estrogen receptor-1 (GPER-1) modulators (Figure 3A). LBVS was performed based on a GPER-1 selective agonist (1-((3a*R*,4*S*,9b*S*)-4-(6-bromobenzo[*d*][1,3]dioxol-5-yl)-3a,4,5,9b-tetrahydro-3*H*-cyclopenta(c)quinolin-8-yl)ethan-1-one) as a query model for screening of the eMolecules library (about 7.2 million compounds used in [84]) using Rapid Overlay of Chemical Structures (ROCS; [32]) and electrostatic potential screening (EON; [94]). Then, after generation of a GPER-1 homology model, FRED [95,96] was used to screen the top-scored hits from LBVS. Next, the top-ranked hits retrieved by molecular docking were clustered based on the similarity between their scaffolds. Finally, the prospective validation in SK-BR-3 and MCF-7 cell lines resulted in two compounds with an EC_{50} (i.e., effective drug concentration that gives half-maximal response) antiproliferative activity in the micromolar range.

Alternative LB and SB sequential protocols have also been adopted, as in the study by Dawood et al. [97], where the SBVS was followed by a LBVS (Figure 3B). An in-house database of 1720 phytochemicals used in traditional Egyptian medicine was screened to search for inhibitors of the human aromatase enzyme. The initial size of the library allowed the direct use of molecular docking using Glide [98–100]. Subsequently, a LB pharmacophore was used to filter the ranked compounds with PHASE [101,102]. In vitro testing revealed that the methylene chloride extract of *Artemisia annua* showed the most significant aromatase inhibitory activity with an IC_{50} of 2.2 µg/mL, thus opening a path for the use of secondary metabolites in the search for new therapeutic leads.
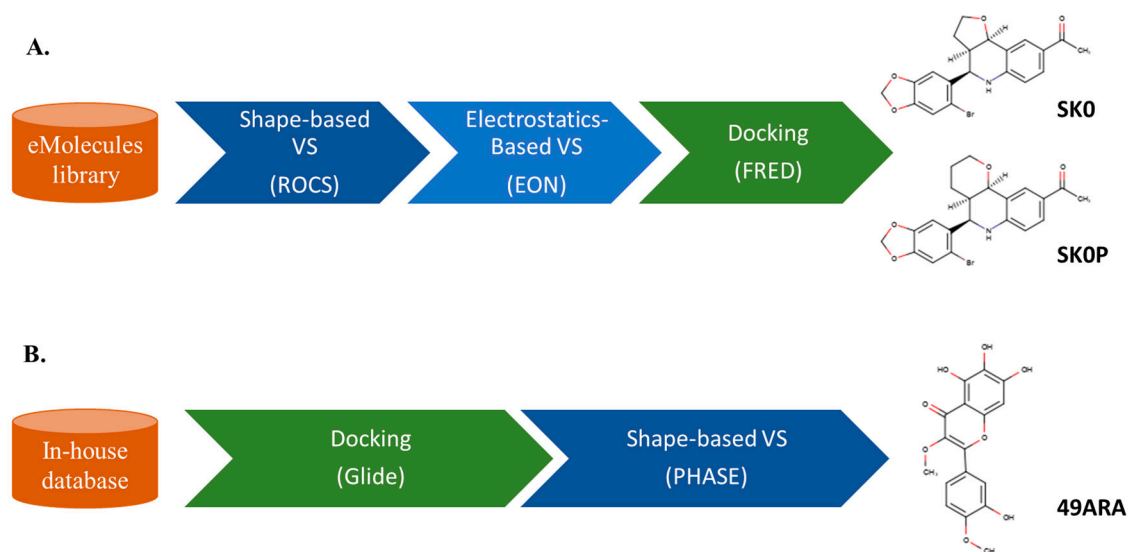
Figure 3. Schematic representation of two sequential VS processes (LB and SB methods are highlighted in blue and green, respectively). (**A**) Sequential application of LB (shape-based and electrostatic-based similarity search) followed by SB (docking). The two most active compounds found, SK0 and SK0P, exhibited antiproliferative activities against SK-BR-3 and MCF7 cell lines in the micromolar range. (**B**) Sequential VS of SB followed by LB over an in-house database. The compound 49ARA was detected among the top-ranked compounds included in the methylene chloride extract of *Artemisia annua* (IC_{50} of 2.2 µg/mL). ROCS: Rapid Overlay of Chemical Structures.

4. Parallel LB and SB Approaches

The application of different LB and SB methods generates distinct sets of ranked compounds for the same target. Given that there is no single method that consistently ranks a database of compounds in the best decreasing order, a combination of the ranks obtained from multiple LB and SB searches into a single ranking could lead to a better overall enrichment and a wider diversity of hit structures [103].

In this context, LBVS approaches have been combined under the framework of data fusion [104,105], targeting the search for new entities, drug repurposing, polypharmacology, and safety profile analysis [106–109]. With regard to SBVS, distinct methods have been examined to yield a "consensus scoring" [110–113], relying on three main strategies: (i) the same docked poses have been evaluated with different scoring functions to build the final ranking, (ii) the results obtained for an ensemble of different protein structures of the same target have been combined to obtain a final score, and (iii) multiple docking methods have been used against a single-protein structure [54,114,115].

Efforts have also addressed the development of parallel protocols for combining LB and SB methods [66,67,103]. Table 1 summarizes different fusion strategies that have been adopted to combine the results of LB and SB techniques in benchmarking studies. The parallel selection seems to perform better than other rank fusion metrics, although the quality of the structural information and the specific physicochemical features of the target system may influence the overall performance. For instance, Tan et al. [116] evaluated the performance of a parallel protocol that combined docking and similarity search calculations using 2D fingerprints on nine target enzymes. The results were combined through rank fusion, where the ranks from docking and similarity searching were added to generate the final ranking, or the parallel selection method, where compounds are alternately selected according to the ranks obtained separately for LB and SB screenings. These combinations yielded an overall improvement in compound recall in 25% of the calculations. Furthermore, parallel selection was found to be more effective than rank fusion.

Table 1. Description of selected fusion algorithms implemented in parallel ligand-based (LB) and structure-based (SB) strategies. VS: virtual screening.

Algorithm	Description	Case Studies
ADDITION algorithms	Adds together the ranks from the different VS methods rank lists. Standard statistical measures, weighted or not, are used (i.e., sum, average, and median or max. value) to combine rank positions.	[66,67,103,116]
PARETO ranking	Ranks a compound based on how many other compounds are better in all screening methods. Ties could be broken using the sum rank, as example.	[103]
PARALLEL selection	Compounds are alternatively selected among the top-ranked compounds obtained from each screening method until the desired number of compounds is reached.	[81,103]

Swann et al. examined the combination of LBVS (2D graph-based extended connectivity fingerprint (ECFP6) [117] and ROCS) and SBVS (chemical Gaussian overlay, CGO [95]) within a probabilistic framework that returns a quantitative likelihood (or probability) of observing bioactivity for the selected compounds [66]. The analysis of the results obtained for a set of 18 targets showed that the retrieval rates for the cumulative probability (obtained from the fusion of the individual LB and SB values) are equal to or better than the highest retrieval rate achieved with any single method. Similar trends were observed for an additional external validation set of six targets, and, importantly, the method was successful in the identification of novel hit compounds in a prospective study performed against four targets not included in the training and validation sets.

A number of studies dealing with the application of the parallel strategy in the search of novel hits have been reported in the last years [118–121]. An illustrative example is the work by Vucicevic et al. [119], who reported the identification of compounds with anticancer potential effects through a parallel LB and SB screening protocol (Figure 4A). Starting from a large virtual library with more than 9×10^6 compounds, those molecules that showed good ranking in both approaches were selected for biological testing. The most active compound exhibited a cytotoxic profile similar to the positive control and enhanced the apoptotic response to doxorubicin, thus representing an adjuvant chemotherapeutic strategy for doxorubicin-insensitive cancers.
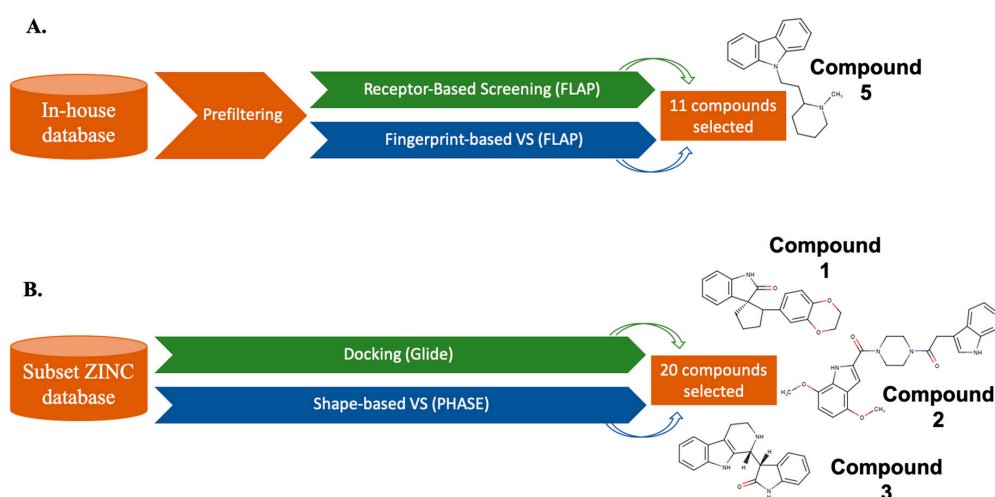

Figure 4. Schematic representation of two parallel VS processes (LB and SB methods are highlighted in blue and green, respectively). (**A**) Parallel application of LB (fingerprint similarity search) followed by SB (pharmacophoric receptor-based screening). FLAP, which allows ligand-ligand and ligand-protein similarity assays, was used for both approaches. The most active compound exhibited an IC_{50} in the micromolar range. (**B**) Parallel VS over a ZINC database subset. Glide was run for SB and PHASE for LB. Among the 20 hits selected, three compounds showed activity in the micromolar range.

Finally, a more recent example is the work by Costa et al. [121], where they performed a parallel VS application followed by MD simulations in the search of a novel compound able to inhibit human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) RNA-dependent DNA polymerase activity (Figure 4B). More than 143,000 natural compounds commercially available in the ZINC database were screened. As a result, 20 hit molecules were chosen and tested in biochemical assays. However, instead of merging the output rankings, compounds shared in both LB and SB VSs were selected. Among them, three compounds were identified as novel non-nucleoside RT inhibitors in the low micromolar range.

As a final remark, let us note that, compared to sequential methods, parallel LB and SB screenings imply a larger computational cost, as several VS techniques have to be simultaneously run in order to derive their respective rankings, which will subsequently be used to generate the final selection.

5. Hybrid Approaches

As noted above, hybrid LB and SB strategies can be grouped into two major categories, which are denoted as (i) interaction-based approaches and (ii) similarity-docking methods.

5.1. Interaction-Based Methods

These methods rely on the identification of patterns of protein–ligand interactions, which are subsequently used in the screening of compounds through the usage of pseudo-receptor and pseudoquery methods. Since SB information is not effectively incorporated in pseudo-receptor models, we limit ourselves to giving a brief description for the sake of completeness but omit a detailed discussion, which can be found elsewhere [73].

Pseudo-receptor methods rely on the mapping of the potential interactions that may be formed by a set of reference ligands suitably aligned in their bioactive conformation to mimic their overlaid arrangement in the binding pocket [122–124]. This process leads to a rough definition of the overall shape and key anchoring points of the binding pocket, which can be exploited for the screening of chemical libraries. The performance of these models is strongly affected by the chemical space of the ligand dataset and the overlay of the ligands. The model can only account for those features present in the starting set of ligands, and the superposition of ligands is sensitive to minor modifications in the chemical scaffold, especially for highly flexible ligands.

In contrast with the preceding approaches, pseudoquery methods exploit the experimental structures of protein–ligand complexes in order to extract a profile of the interaction pattern established by the ligands bound to the protein target. This pattern is generally translated into fingerprints that encode ligand–target interactions or, alternatively, into pharmacophoric features and then used in similarity searches to find ligands that match the interaction pattern [123–131] (see Table 2 for a brief description of several formalisms). In addition, the search of novel hits can be performed, imposing constraints related to the shape and volume of the binding site.

Table 2. Selection of pseudoquery methods categorized by the underlying protein-interaction model.

Model	Methods	Description
Interaction fingerprint-based	SIFt [132]	Fingerprint encoding seven predefined types of target–ligand interactions.
	PLIP [133]	A web service for the detection and visualization of seven protein–ligand interaction types considering a 3D space.
	FLIP [134]	For each residue, 7 different interactions are represented in 10 bits.
	PADIF [135]	Fingerprints with the inclusion of information relative to the strength of interactions and unfavorable ones.
Pharmacophore-based	LigandScout [125]	Pharmacophores derived from six types of nonbonded protein-ligand interactions and volume constraints.
	FLAP [126]	Four-point pharmacophore fingerprints with a shape component.
	IChem [136]	Converts the protein–ligand interaction pattern in fingerprints and graphs.
	TIFP [137]	Encodes a string of unique triplets (two interacting atoms and an interaction pseudo-atom).

The pioneering methods included key elements of the protein–ligand complex, such as the formation of hydrogen bonds, hydrophobic or aromatic interactions, or contacts with acidic and basic groups, often supplemented by isocontours of the binding site. As an example, Salentin et al. [138] resorted to PLIP to perform a pharmacophoric search of over more than 170,000 complexes using protein-ligand interaction profiles, leading to the disclosure of the FDA-approved malaria drug amodiaquine as the top-ranking hit, which was subsequently validated as a potential anticancer agent showing inhibitory activity on the target protein Hsp27. This demonstrates the potential of pseudoquery methods for drug repurposing.

Recent methods have evolved to include solvation and entropy effects. For instance, Tran-Nguyen et al. [131] included in their pseudoquery pharmacophoric tool the desolvation component of the protein–ligand interaction energy using a Poisson−Boltzmann treatment. Furthermore, the analysis was decomposed in three consecutive steps: (i) the detection of druggable cavities at the surface of the protein target and the identification of pharmacophoric features, (ii) the generation of cavity-based pharmacophore queries in the 3D space, and (iii) molecular alignment exploiting the cavity-based feature. The proposed pharmacophoric model was benchmarked using DUD-E [139]. Unique chemotypes were retrieved from high-throughput VS, being as efficient as state-of-the-art docking [140] and shape-matching [32] methods in both pose prediction and ranking power.

As noted above, pseudoquery methods have also exploited interaction fingerprint patterns (IFP) containing information about the contacts of the ligand with the protein, thus condensing the 3D structural binding information into a 1D binary string, leading to a drastic reduction in the cost of VS. This is exemplified by the Structural Interaction Fingerprint (SIFt; [132]) method, where crystallographic ligands are divided into two groups of fragments: (i) atoms involved in protein–ligand interactions (interaction fragments, IFs) and (ii) fragments generated by the random deletion of ligand atoms used as a control. Then, for each ligand and the corresponding fragments, MACCS (Molecular ACCess System) structural keys [141] were calculated and used as a fingerprint for similarity searching. The results of their validation work suggested that IFs used as templates can increase the similarity

search performance of conventional structural fingerprints. SIFt is included in Arpeggio [142], a web server for the analysis of protein interactions with small-molecule ligands, proteins, and DNA.

The concept of IF has been adopted in alternatives models to outperform conventional scoring functions in predicting the correct poses for drug-like compounds [143], similarity-based screening [137,144–148], binding/unbinding kinetics [146], and drug resistance [147].

5.2. Similarity-Docking Strategies

An important challenge in VS is to create accurate scoring and ranking functions to identify hit compounds active against specific targets. In this context, similarity measurements between compounds can be used to assist molecular docking to score sampled poses [148–154] and to discriminate between active and inactive molecules [155,156]. Exploiting the synergy between molecular similarity and docking has received increasing interest in the last years, leading to hybridized tools, such as HomDock [157] and Hybrid [96].

5.2.1. Predicting the Pose of Ligands

A direct application of merging molecular similarity and docking is related to improving the prediction of the ligand pose in the binding pocket. At this point, exploiting the experimental information on the binding mode of active compounds has been shown to enhance the performance of predicting the pose of drug-like compounds [158–160]. For instance, the participants of the Drug Design Data Resource (D3R) Grand Challenge 3 were challenged in predicting the binding poses of 24 cathepsin S ligands. Kumar and Zhang [161] tested the performance of three methods (PoPSS [150], CDVS [162], and PoPSS-Lite) based on the concept of ligand 3D shape similarity. PoPSS evaluates the shape similarity with existing crystallographic compounds bound to the target protein for predicting the poses of query ligands with unknown binding modes. The ligand with the highest shape similarity score was selected and placed into the binding pocket. After ligand placement, side-chain residues of the binding pocket were repacked based on the query ligand conformation, followed by Monte Carlo energy minimization of the protein-ligand complex. Finally, ligand-bound structures were scored using the Rosetta energy function [163]. For CDVS and PoPSS-Lite, 3D shape similarity calculations were also used to identify the ligand pose. However, once the suitable ligand–receptor pair was identified, CDVS performed a standard docking using Glide, and PoPSS-Lite refined the pose with an energy minimization. PoPSS-Lite exhibited an excellent performance in this challenge, leading to the lowest mean root mean square deviation (RMSD) values between the native and predicted poses. Moreover, CDVS and PoPSS were located among the best 15 methods tested for both metrics.

Shape similarity between the ligand conformation and the crystallographic ligand is the most common scheme adopted for guiding the pose prediction. However, other similarity measurements and methodological refinements have been explored. As an example Jacquemard et al. defined a benchmark constituted by 2376 high-quality structures representing 64 proteins and compared the performance of three rescoring schemes applying the similarity of IFP, graph matching of interaction patterns (GRIM [137]), and ROCS [78] (Figure 5). GRIM and ROCS were more efficient than IFP rescoring based on 2D fingerprints, even when the comparison involved structurally dissimilar molecules. In addition, the speed of calculation for all the methods was improved, facilitating the processing of a large number of poses.

The search for methodological innovations is also exemplified by Kumar and Zhang [151], as they modified PoPSS to account for water-mediated protein–ligand interactions using a continuum Poisson-Boltzmann (PB) solvation model, leading to the PoPSS-PB model. PoPSS-PB demonstrated an excellent performance in D3R GC4, with mean and median RMSDs of 1.20 (ranked 10th out of 74) and 1.13 (ranked 9th out of 74) Å, improving the performance obtained for PoPSS and PoPSS-Lite.
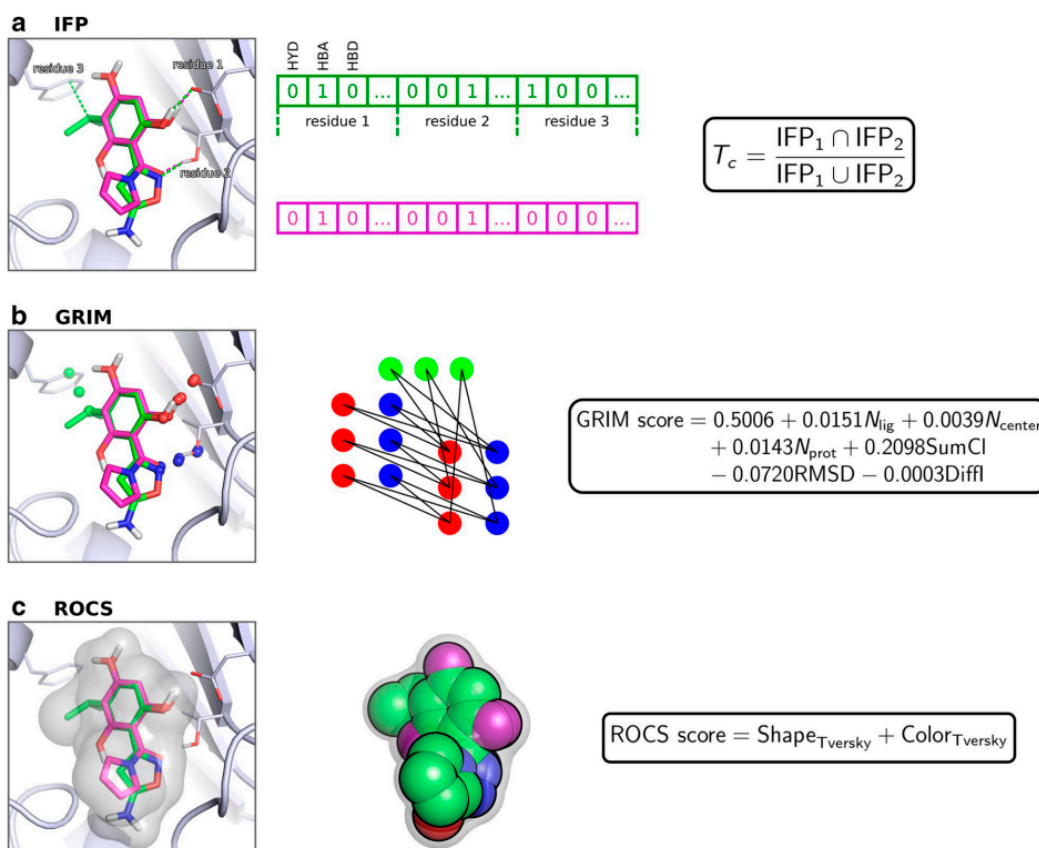
Figure 5. Overview of rescoring methods coupled to interaction fingerprints (IFP), graph matching (GRIM), and ROCS. (**a**) In the IFP score Tc denotes the Tanimoto coefficient. (**b**) In the GRIM score, N_{lig} is the number of aligned ligand points, N_{center} stands for the number of aligned centered points, N_{prot} is the number of aligned protein points, SumCl is the sum of clique weights over all weights, RMSD is the root mean square deviation of the matched cliques, and DiffI stands for the difference between the number of interaction points in the query and the template compound. (**c**) The ROCS score is based on the Tversky coefficient. Reprinted with permission from Springer Nature [79].

Finally, Varela-Rial et al. [164] also evaluated in the D3R Grand Challenge 4 an algorithm named SkeleDock to define the binding mode based on the structure of a protein−ligand complex. The algorithm defines graphs for the query and the template molecules, and then, these graphs are compared to extract a common subgraph, which describes a continuous set of atoms whose element (node) and bonds (edges) are equivalent in the two molecules (Figure 6). Thus, a mapping that links atoms in query and template compounds can be identified, facilitating the conformational adjustment of the atoms in the query ligand onto those in the template molecule, whereas atoms in the query molecule with no equivalent counterpart in the template are positioned by using a tethered template docking protocol. The algorithm was ranked 15th out of 74 according to the mean RMSD (1.33 Å) and 9th according to the median RMSD (1.02 Å).
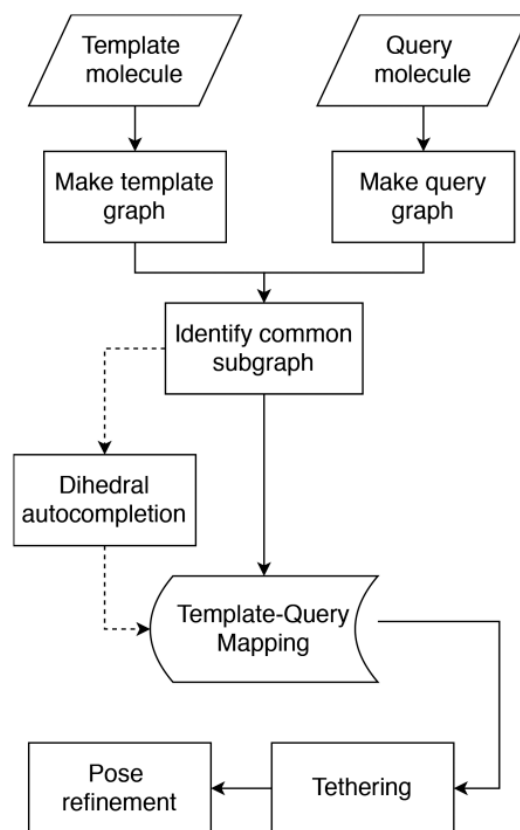
Figure 6. Schematic representation of the flowchart implemented in the SkeleDock algorithm. Reprinted with permission from the American Chemical Society [164].

5.2.2. Similarity-Guided Score Scheme

In addition, to assist the prediction of the ligand pose in the binding cavity, similarity measurements can also be used as a weighting factor in the reranking of the docked compounds. In fact, considering that LB 3D-shape matching algorithms often produced better enrichments than docking, assessing the overlay of docked poses relative to known crystallographic ligands could be valuable to retrieve active compounds in screening studies. To this end, the scoring function in docking calculations could be supplemented with 3D molecular similarity measurements to build the final ranking in the VS process, taking into account that positive candidates accommodated in the active site are expected to share similar structural and physicochemical features that resemble those of known actives in cocrystal structures and improve the ranking of the screened ligands.

The first implementations of this LB and SB scheme in docking programs were performed by Marialke et al. [157] and McGann [96] in the development of HomDock and Hybrid, respectively. HomDock is a combination of optimization methods with graph-based molecular alignment (GMA; [165]) that superposes a query molecule on a rigid template. GMA places candidate ligands over the template and optimizes their placement in the field of the protein. Then, the ligands are ranked according to their interaction with the protein and/or their structural similarity with the ligand. On the other hand, Hybrid uses an exhaustive search algorithm, treating ligand and protein structures as rigid bodies. Both the protein and ligand flexibility are addressed through multiple conformers. Subsequently, the CGO ligand-based scoring function is applied. CGO scores based on how well the docked molecule matches the shape and 3D arrangement of the chemical features of the crystallographic ligand bound to the active site.

In another vein, Anighoro and Bajorath published a series of comparative studies where the best poses of commercial docking software are directly scored using 3D similarity methods, such as the whole-ligand 3D shape similarity and protein–ligand IFP similarity [76,155,156]. The protocol was

validated by performing retrospective VS calculations for different targets, including dihydrofolate reductase, glucocorticoid receptor, HIV-1 protease, vascular endothelial growth factor receptor-2, adenosine A2A receptor, and β2 adrenergic receptor. As noted in Figure 7, the hybridized approach yielded better performance in retrieving active compounds against the targets included in the validation set. Thus, the results showed that ranking by whole-ligand 3D similarity calculations outperformed the force field-based ranking tested for both global performance and early enrichments. It was also shown that a ligand was less suitable as a reference for 3D similarity calculations if it contained large solvent-exposed groups not directly interacting with the target. Finally, they highlighted the importance that reference ligands should be engaged in interactions within the binding site as much as possible.
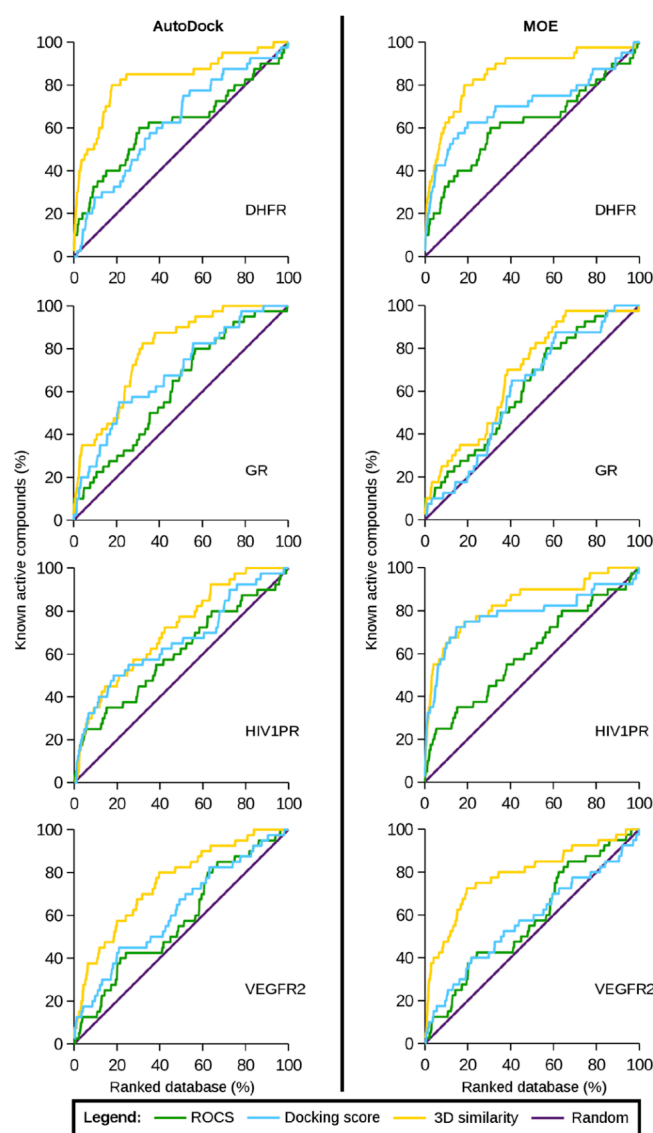

Figure 7. ROC plots obtained for four validation targets (DHFR: dihydrofolate reductase, GR: glucocorticoid receptor, HIV1PR: HIV-1 protease, and VEGFR2: vascular endothelial growth factor receptor-2) using shape-based similarity (ROCS; green), docking (AutoDock [166] and MOE [167]; cyan), and the hybridized method (yellow). Reprinted with permission from the American Chemical Society [76].

Under the same framework, our group has recently presented a 3D similarity scheme to enrich the docking performance based on the usage of lipophilic descriptors [168] determined from quantum

mechanical-based continuum solvation models [80]. The 3D similarity was determined by comparing the 3D distribution of atomic lipophilicity computed by PharmScreen [31,169] (Figure 8), and the similarity measurements were exploited in conjunction with the poses obtained by using three docking programs: Glide, rDock [170], and GOLD [171]. Two hybrid algorithms were examined over 44 sets: (i) rescoring ranking (RR), where the final ranking was determined according to the score obtained from the 3D lipophilic similarity of the best pose generated by the docking method and the co-crystallized ligand, and (ii) consensus ranking (CR), where the final score of the compounds was obtained by merging the rankings provided directly from the docking method and from the RR. The results obtained support the synergy of the hybrid LB and SB approaches, as the CR consistently showed better performance than using only either the LB or SB methods. In addition, the results suggest that CR may overcome the existence of multiple binding modes differing from the experimental pose of the co-crystallized ligand.
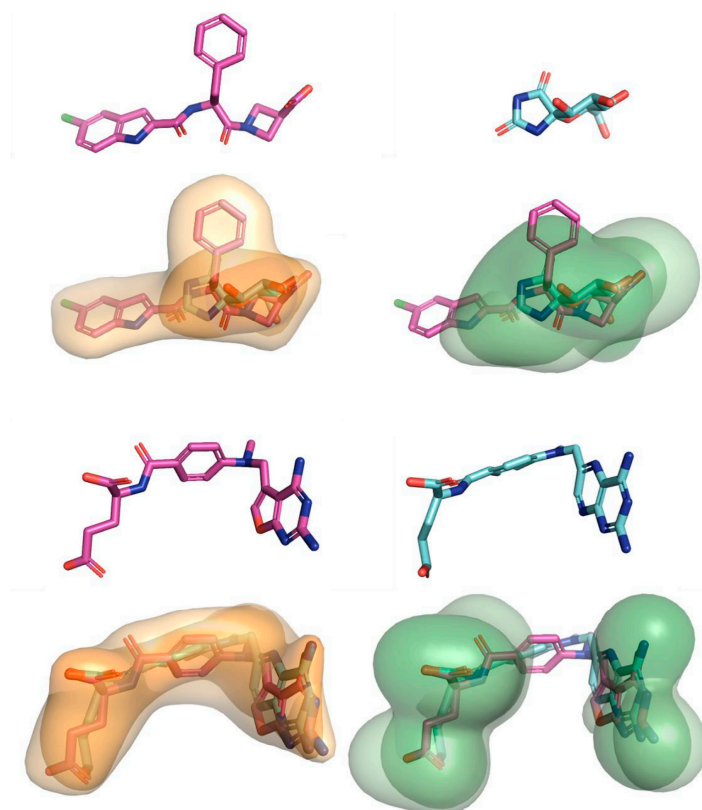

Figure 8. Representation of the hydrophobic molecular overlay exploited by PharmScreen. (Top) Overlay of ZINC02046793 (template) and ZINC1489956 (active) pertaining to glycogen phosphorylase®. (Bottom) Molecular alignment of ZINC0384989 (template) and ZINC1529323 (active) pertaining to the dihydrofolate reductase. Orange and green contours denote the fields originated from the cavitation and electrostatic components of the molecular lipophilicity. Reprinted with permission from the American Chemical Society [80].

6. Exploiting Chemical Libraries and Biological Data

While selecting different LB and SB strategies may provide alternative approaches to enrich the results of VS, the identification of a novel lead compound may be conditioned by the structural diversity of the chemical space encoded in compound libraries. Therefore, the choice of a representative dataset well-suited to the specific structural, physicochemical, and biological features of the macromolecular target may be crucial for the successful outcome of a VS campaign, especially keeping in mind that the available chemical libraries comprise only a small portion of the synthesizable chemical universe of compounds.

To facilitate this task, there has been continued progress in the availability of experimental data in the public domain during the last two decades [172], which is exemplified in the consolidation of curated databases of bioactive molecules with drug-like properties, as exemplified with ChEMBL [173], PubChem [174], and DrugBank [175,176], where the user may find a comprehensive compilation of diverse information, such as the chemical structure of the compounds, physicochemical properties, biological assays data, related targets, pharmacokinetics and pharmacodynamics properties, metabolic and signaling pathways, and patents. Currently, selection of the compounds can be performed through a variety of chemical databases, which can be categorized into three main groups: public, commercial (provided by vendors), and proprietary (Table 3; see references 177 and 178 for a detailed discussion).

Table 3. Examples of public and commercial databases (data taken from [177,178]).

Database	Type	No. Cpds
AstraZeneca with Enamine BBs	Proprietary	10^{17}
Boehr.-Ing. BICLAIM	Proprietary	5×10^{11}
CH/PMUNK	Public	$>95 \times 10^{6}$
eMolecules Plus	Commercial	5.9×10^{8}
Enamine Real	Commercial	$>300 \times 10^{6}$
EVOspace	Proprietary	1.6×10^{16}
GDB-17	Public	$\sim166 \times 10^{9}$
Lilly LPC	Proprietary	2×10^{11}
MASSIV	Proprietary	10^{20}
SAVI	Public	$\sim283 \times 10^{6}$
PGVL	Proprietary	3×10^{12}
PubChem	Public	9.6×10^{6}
SCUBIDOO	Public	$\sim21 \times 10^{6}$
Sigma Aldrich	Commercial	1.4×10^{7}
ZINC15	Commercial	2×10^{6}

In this context, rather than focusing the computational effort on the massive screening of larger databases, one might consider the possibility to enhance the success of LB and SB strategies in the search of novel hit compounds by resorting to the screening of targeted chemical libraries. An example is the work by Miyao et al., who reported an algorithm for the exhaustive generation of chemical structures based on inverse quantitative structure-property (QSPR)/activity (QSAR) relationships to build datasets of compounds endowed with a suitable range of desired properties [179]. The synthetic feasibility of the compounds may also be accounted from the availability of information about known chemical reactions [180–182]. More recently, the de novo design algorithm for exploring chemical space (DAECS) exploits the combination of a two-dimensional distribution of the chemical properties with the projection of the biological activity for a set of training compounds in order to generate structures in a specific target area of the chemical space [61,182]. A library of novel designed structures is constructed through an iterative process that involves the selection of seed structures characterized with selected chemical features and the generation of novel compounds by means of introducing slight structural changes from the seed dataset.

The design of target chemical libraries is actually an undertaking of increasing interest, as illustrated by a number of recent studies that have reported the implementation of artificial intelligence-based algorithms [63–65,183–186]. One of them is the development of ReLeaSE (Reinforcement Learning for Structural Evolution), which integrates a generative deep neural network with a predictive one into a joint framework for the design of novel compounds satisfying certain chemical requirements, as illustrated with the biased selection of compounds fulfilling a specific range of physical properties (i.e., melting temperature and lipophilicity) or inhibitory activity against the desired target protein (Janus protein kinase 2) [62]. Another example is the transfer-learning-based generation algorithm proposed by Amabilino et al. [186], where recurrent neural networks are used as SMILES (Simplified Molecular-Input Line-Entry System) generators and trained on a smaller set of molecules with the

biological activity of interest for the design of focused libraries. On the other hand, the REINVENT code proposed by Olivecrona et al. [187] also relies on a recurrent neural network model that operates on a SMILES representation of molecules for the automated creation of molecules with predicted biological activity. Very recently, this algorithm was adapted to enable the pair-based multi-objective optimization of several molecular features based on Pareto dominance [188] and applied to the de novo design of datasets of inhibitors targeting neuraminidase, acetylcholinesterase, and the main protease of the severe acute respiratory syndrome coronavirus 2.

Another issue that deserves a brief discussion concerns the discovery of ligands able to modulate protein–protein interactions (PPIs) in the early stages of drug discovery. Given the estimated 650,000 PPIs that comprise the human interactome, the stabilization and inhibition of PPIs may represent a valuable strategy to alter the oligomerization equilibria of supramolecular protein complexes, thus altering their physiological functions in the cell [189–191], which may thus be exploited in the search of novel therapeutic approaches [192,193]. Nevertheless, the success of these studies may be affected by the availability of chemical libraries with ligands suitable to interact with druggable pockets at the interface of protein–protein complexes. In this context, it is worth noting the efforts made toward the design of specific databases enriched in protein–protein modulators, such as PPI-HitProfiler, which was developed to provide for any drug-like compound collection a focused chemical library enriched in putative PPI inhibitors [194], 2P2I$_{HUNTER}$, which is a learning machine tool for filtering potential PPI modulators [195], and Fr-PPIChem, which reflects the collective effort of a French consortium to provide a unique chemical library for PPI inhibition [196]. A comparative analysis of different PPI-focused libraries was reported in the study by Zhang et al. [197], where they noticed that PPI inhibitors tend to be larger and more hydrophobic than standard drugs and that PPI-focused libraries, although designed using different strategies, tend to share common chemical subspaces. Efforts have also been conducted for the application of combined strategies in the identification of PPI modulators. As an example, Singh et al. [198] proposed a VS protocol using a hybrid SB and LB method, highlighting the benefits of 3D topological descriptors to assess the post-docking output. As a validation set, 11 PPI targets with known active and inactive compounds were considered. In a first step, the compounds were docked using Surflex, and the docked poses were post-processed to calculate the shape similarity and the structural interaction fingerprint similarity to the co-crystallized PPI inhibitor. Notably, the hybrid protocol showed an improved performance for numerous targets, supporting the application of these combined techniques for prospective studies of PPI modulators.

In a different context, it is also worth mentioning here the systemic chemogenomics/QSAR procedure introduced by Cruz-Monteagudo et al. [199], which aimed to generate a disease-relevant pool of ligands by combining phenotypic data with LB and SB information in a sequential process that ended up with a phenotypic VS performed with QSAR models (Figure 9). The first step is the selection of a representative set of ligands targeting a disease with a measurable therapeutic phenotype, such as compounds successfully evaluated in clinical trials. Then, information about the ligand–target interaction, key genes, or protein targets involved in the molecular interactions and reaction networks are compiled to build a disease-relevant chemogenomics space, which should encompass potential targets directly or indirectly (i.e., cascading effects) implicated in the physiological response. Gene ontology is subsequently used to encode the systemic effect of each ligand in fingerprints containing both chemical descriptors and biological information. The codified compounds are split into two classes: ligands that significantly interact with at least one target (phenotype-positive class) and compounds with no significant interaction with any of the targets associated with the desired phenotype (phenotype-negative class). Finally, a QSAR-based VS methodology is performed. This protocol was utilized in a retrospective study aimed at prioritizing ligands acting as neuroprotective agents in Parkinson's disease, and a significant fraction of the drug candidates used as starting points could be recovered at early fractions of the screened data.
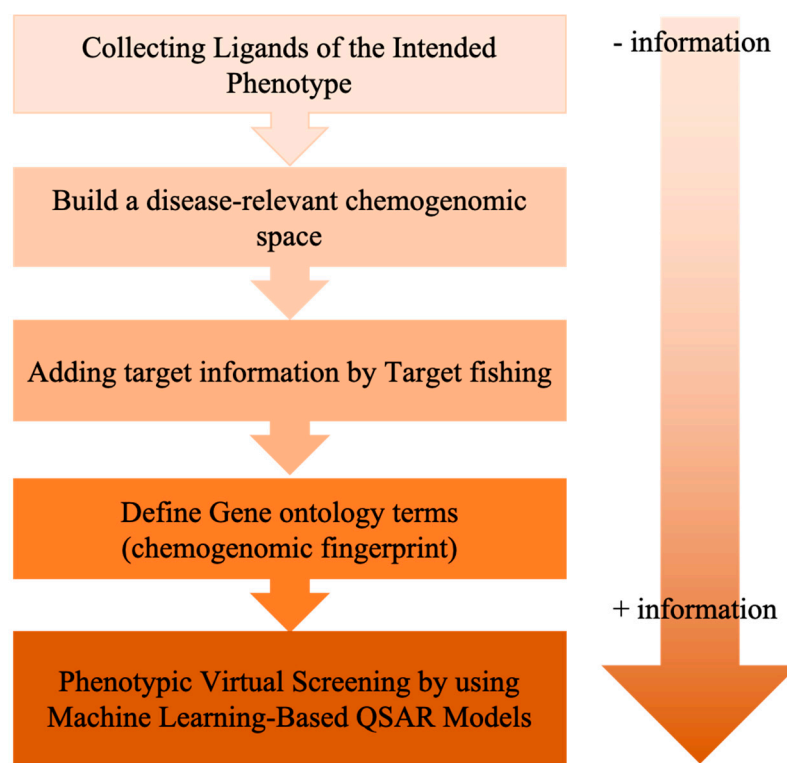
Figure 9. Schematic representation of systemic chemogenomics/quantitative structure-activity relationships (QSARs) for phenotypic VS.

As a final remark, it is worth emphasizing the relevance of curating the activity data in the analysis of the compounds and the preparation of targeted chemical libraries, minimizing the bias introduced by spurious hits, which can arise from a number of unexpected factors, such as covalent modification of specific protein residues or changes in the redox state of a range of ligands. In particular, a large number of cases have been attributed to self-aggregation of the ligand in an aqueous solution [200–202]. The tendency of small organic molecules to spontaneously form colloidal self-aggregates can lead to undesired artifacts in the screening of drug-like compounds, resulting in the identification of false positives [203,204]. The colloidal aggregates formed by these types of compounds exhibit common trends, such as the lack of robust structure-activity relationships, as well as the identification of time-dependent noncompetitive-like inhibition [200,205], likely reflecting the nonspecific inhibitory mechanisms related to adsorption of the target protein onto the aggregates or the induction of conformational alterations that affect the protein's activity [206]. In these cases, a critical parameter to be considered is the critical aggregation concentration of the compound, which turns out to be in the micromolar range for a significant number of aggregating drug-like compounds [202,207]. Overall, this discussion suffices to emphasize the need to perform a detailed curation of the biological data and of the conditions used in experimental assays, as this information may have an unexpected influence on the efficacy of the bioactive compounds.

7. Conclusions

In the past decades, VS has been a powerful alternative to high-throughput screening assays due to the reduced expensiveness, the continued progress in computer resources, and the refinement in LB and SB techniques, often leading to hit rate enrichments that outperform the results obtained with experimental screenings. Being the most widely used strategy for disclosing novel hit compounds in the early stages of drug discovery, the success of VS campaigns is also severely affected by the intrinsic shortcomings of both LB and SB methods, which makes it necessary to search for novel computational strategies that exploit the merits of the individual techniques synergistically.

In the last years, we have witnessed a flourishment of different combined LB and SB approaches, ranging from the hierarchical application of techniques in multi-step filtering process to novel methods that integrate LB and SB techniques into a standalone framework. The progress is encouraging, but it can be anticipated that the adoption of these integrated strategies will depend on two main factors. First, an extensive benchmarking of the distinct combination strategies, including a diverse sets of targets covering distinctive structural and physicochemical features, the calibration of different descriptors for similarity measurements, and docking algorithms in retrospective studies should be necessary, eventually complemented with the prospective application to drug discovery projects. These studies should be valuable to judge not only the improvement obtained with the usage of integrated methods relative to either pure LB or SB techniques but, also, to identify the optimal combination strategy in light of the druggability characteristics of the target protein. Second, the ability to implement the combined LB and SB strategies in automated modeling platforms should provide user-friendly access to the screening of targeted-oriented chemical libraries, guidelines for an appropriate design of the combination strategy, and graphical display facilities to analyze the results. The implementation in modern software will be necessary to facilitate the adoption of the combined strategies by the drug discovery community.

References

1. Free Energy Calculations. *Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2007.

2. Abel, R.; Wang, L.; Harder, E.D.; Berne, B.J.; Friesner, R.A. Advancing drug discovery through enhanced free energy calculations. *Acc. Chem. Res.* **2017**, *50*, 1625–1632. [CrossRef] [PubMed]

3. Williams-Noonan, B.J.; Yuriev, E.; Chalmers, D.K. Free energy methods in drug design: Prospects of "alchemical perturbation" in medicinal chemistry. *J. Med. Chem.* **2018**, *61*, 638–649. [CrossRef] [PubMed]

4. Christ, C.D.; Fox, T. Accuracy assessment and automation of free energy calculations for drug design. *J. Chem. Inf. Model.* **2013**, *54*, 108–120. [CrossRef] [PubMed]

5. Mondal, D.; Florian, J.; Warshel, A. Exploring the effectiveness of binding free energy calculations. *J. Phys. Chem. B* **2019**, *123*, 8910–8915. [CrossRef] [PubMed]

6. Cournia, Z.; Allen, B.; Sherman, W. Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937. [CrossRef]

7. Zhang, H.; Gattuso, H.; Dumont, E.; Cai, W.; Monari, A.; Chipot, C.; Dehez, F. Accurate estimation of the standard binding free energy of netropsin with DNA. *Molecules* **2018**, *23*, 228. [CrossRef]

8. Fu, H.; Gumbart, J.C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFFE: A user-friendly graphical interface facilitating absolute binding free-energy calculations. *J. Chem. Inf. Model.* **2018**, *58*, 556–560. [CrossRef]

9. Jespers, W.; Esguerra, M.; Aqvist, J.; Gutiérrez-de-Terán, H.J. QligFEP: An automated workflow for small molecule free energy calculations in Q. *J. Cheminform.* **2019**, *11*, 26. [CrossRef]

10. Gioia, D.; Bertazzo, M.; Recanatini, M.; Masetti, M.; Cavalli, A. Dynamic docking: A paradigm shift in computational drug discovery. *Molecules* **2017**, *22*, 2029. [CrossRef]

11. Ruiz-Carmona, S.; Schmidtke, P.; Luque, F.J.; Baker, L.; Matassova, N.; Davis, B.; Roughley, S.; Murray, J.; Hubbard, R.; Barril, X. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat. Chem.* **2017**, *9*, 201–206. [CrossRef]

12. Cavasotto, C.N.; Aucar, M.G. High-throughput docking using quantum mechanical scoring. *Front. Chem.* **2020**, *8*, 246. [CrossRef] [PubMed]

13. Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J. Am. Chem. Soc.* **2010**, *132*, 7361–7371. [CrossRef] [PubMed]

14. Gimeno, A.; Ojeda-Montes, M.J.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; García-Vallvé, S. The light and dark sides of virtual screening: What is there to know? *Int. J. Mol. Sci.* **2019**, *20*, 1375. [CrossRef] [PubMed]

15. Yasuo, N.; Sekijima, M. Improved method of structure-based virtual screening via interaction-energy-based learning. *J. Chem. Inf. Model.* **2019**, *59*, 1050–1061. [CrossRef]

16. Ertl, P. Cheminformatics analysis of organic substituents: Identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374–380. [CrossRef]

17. Shoichet, B.K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865. [CrossRef]

18. Schneider, G. Virtual screening: An endless staircase? *Nat. Rev. Drug Discov.* **2010**, *9*, 273–276. [CrossRef]

19. Westermaier, Y.; Barril, X.; Scapozza, L. Virtual screening: An in silico tool for interlacing the chemical universe with the proteome. *Methods* **2015**, *71*, 44–57. [CrossRef]

20. Taylor, R.D.; Jewsbury, P.J.; Essex, J.W. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151–166. [CrossRef]

21. Li, J.; Fu, A.; Zhang, L. An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdiscip. Sci. Comput. Life Sci.* **2019**, *11*, 320–328. [CrossRef]

22. Torres, P.H.M.; Sodero, A.C.R.; Jofily, P.; Silva, F.P., Jr. Key topics in molecular docking for drug design. *Int. J. Mol. Sci.* **2019**, *20*, 4574. [CrossRef] [PubMed]

23. Pagadala, N.S.; Syed, K.; Tuszynski, J. Software for molecular docking: A review. *Biophys. Rev.* **2017**, *9*, 91–102. [CrossRef] [PubMed]

24. *Concepts and Applications of Molecular Similarity*; Johnson, M.A.; Maggiora, G.M. (Eds.) John Wiley & Sons: New York, NY, USA, 1990.

25. Jørgensen, A.M.M.; Pedersen, J.T. Structural diversity of small molecule libraries. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 338–345. [CrossRef]

26. Ivanciuc, O.; Taraviras, S.L.; Cabrol-Bass, D. Quasi-orthogonal basis sets of molecular graph descriptors as a chemical diversity measure. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 126–134. [CrossRef] [PubMed]

27. Duan, J.; Dixon, S.L.; Lowrie, J.F.; Sherman, W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* **2010**, *29*, 157–170. [CrossRef]

28. Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. FLAP: GRID Molecular interaction fields in virtual screening. Validation using the DUD data set. *J. Chem. Inf. Model.* **2010**, *50*, 1442–1450. [CrossRef]

29. Mestres, J.; Rohrer, D.C.; Maggiora, G.M. MIMIC: A molecular-field matching program. Exploiting applicability of molecular similarity approaches. *J. Comput. Chem.* **1997**, *18*, 934–954. [CrossRef]

30. Cheeseright, T.J.; Mackey, M.D.; Melville, J.L.; Vinter, J.G. FieldScreen: Virtual screening using molecular fields. Application to the DUD data set. *J. Chem. Inf. Model.* **2008**, *48*, 2108–2117. [CrossRef] [PubMed]

31. Vázquez, J.; Deplano, A.; Herrero, A.; Ginex, T.; Gibert, E.; Rabal, O.; Oyarzabal, J.; Herrero, E.; Luque, F.J. Development and validation of molecular overlays derived from three-dimensional hydrophobic similarity with PharmScreen. *J. Chem. Inf. Model.* **2018**, *58*, 1596–1609. [CrossRef]

32. Hawkins, P.C.D.; Skillman, A.G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82. [CrossRef]

33. Sastry, G.M.; Dixon, S.L.; Sherman, W. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *J. Chem. Inf. Model.* **2011**, *51*, 2455–2466. [CrossRef] [PubMed]

34. Abrahamian, E.; Fox, P.C.; Nærum, L.; Thøger Christensen, I.; Thøgersen, H.; Clark, R.D. Efficient generation, storage, and manipulation of fully flexible pharmacophore multiplets and their use in 3-D similarity searching. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 458–468. [CrossRef] [PubMed]

35. Sperandio, O.; Miteva, M.; Villoutreix, B. Combining ligand- and structure-based methods in drug design projects. *Curr. Comput. Aided Drug Des.* **2008**, *4*, 250–258. [CrossRef]

36. Talevi, A.; Gavernet, L.; Bruno-Blanch, L. Combined virtual screening strategies. *Curr. Comput. Aided Drug Des.* **2009**, *5*, 23–37. [CrossRef]

37. Spadaro, A.; Negri, M.; Marchais-Oberwinkler, S.; Bey, E.; Frotscher, M. Hydroxybenzothiazoles as new nonsteroidal inhibitors of 17®-hydroxysteroid dehydrogenase type 1 (17®-HSD1). *PLoS ONE* **2012**, *7*, 29252. [CrossRef] [PubMed]

38. Debnath, S.; Debnath, T.; Bhaumik, S.; Majumdar, S.; Kalle, A.M.; Aparna, V. Discovery of novel potential selective HDAC8 inhibitors by combine ligand-based, structure-based virtual screening and in-vitro biological evaluation. *Sci. Rep.* **2019**, *9*, 17174. [CrossRef]

39. Wilson, G.L.; Lill, M.A. Integrating structure-based and ligand-based approaches for computational drug design. *Future Med. Chem.* **2011**, *3*, 735–750. [CrossRef]

40. Drwal, M.N.; Griffith, R. Combination of ligand- and structure-based methods in virtual screening. *Drug Discov. Today Technol.* **2013**, *10*, e395–e401. [CrossRef] [PubMed]

41. Wang, Z.; Sun, H.; Shen, C.; Hu, X.; Gao, J.; Li, D.; Cao, D.; Hou, T. Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* **2020**, *22*, 3149–3159. [CrossRef] [PubMed]

42. Spyrakis, F.; Bidon-Chanal, A.; Barril, X.; Luque, F.J. Protein flexibility and ligand recognition: Challenges for molecular modeling. *Curr. Top. Med. Chem.* **2011**, *11*, 192–210. [CrossRef]

43. Lexa, K.W.; Carlson, H.A. Protein flexibility in docking and surface mapping. *Q. Rev. Biophys.* **2012**, *45*, 301–343. [CrossRef] [PubMed]

44. Salmaso, V.; Moro, S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Front. Pharmacol.* **2018**, *9*, 923. [CrossRef] [PubMed]

45. Chen, Y.C. Beware of docking! *Trends Pharmacol. Sci.* **2015**, *36*, 78–95. [CrossRef] [PubMed]

46. Sridhar, A.; Ross, G.A.; Biggin, P.C. Waterdock 2.0: Water placement prediction for *Holo*-structures with a Pymol plugin. *PLoS ONE* **2017**, *12*, e0172743. [CrossRef] [PubMed]

47. Rudling, A.; Orro, A.; Carlsson, J. Prediction of ordered water molecules in protein binding sites from molecular dynamics simulations: The impact of ligand binding on hydration networks. *J. Chem. Inf. Model.* **2018**, *58*, 350–361. [CrossRef]

48. Sciebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T.E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing role of water in protein-ligand binding studies by neutro crystallography on trypsin complexes. *Nat. Commun.* **2018**, *9*, 3559. [CrossRef]

49. Maurer, M.; Oostenbrink, C. Water in protein hydration and ligand recognition. *J. Mol. Recog.* **2019**, *32*, e2810. [CrossRef] [PubMed]

50. Geschwindner, S.; Ulander, J. The current impact of water thermodynamics for small-molecule drug discovery. *Expert Opin. Drug Discov.* **2019**, *14*, 1221–1225. [CrossRef]

51. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular docking and structure-based drug design strategies. *Molecules* **2015**, *20*, 13384–13421. [CrossRef]

52. Liu, J.; Wang, R. Classification of current scoring functions. *J. Chem. Inf. Model.* **2015**, *55*, 475–482. [CrossRef]

53. Guedes, I.A.; Pereira, F.S.S.; Dardenne, L.E. Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Front. Pharmacol.* **2018**, *9*, 1089. [CrossRef] [PubMed]

54. Palacio-Rodríguez, K.; Lans, I.; Cavasotto, C.N.; Cossio, P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci. Rep.* **2019**, *9*, 5142. [CrossRef] [PubMed]

55. Hein, M.; Zilian, D.; Sotriffer, C.A. Docking compared to 3D-pharmacophores: The scoring function challenge. *Drug Discov. Today Technol.* **2010**, *4*, e229–e236. [CrossRef]

56. Eckert, H.; Bajorath, J. Molecular similarity analysis in virtual screening: Foundations, limitations and novel approaches. *Drug Discov. Today* **2007**, *12*, 225–233. [CrossRef]

57. Grinter, S.Z.; Zou, X. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules* **2014**, *19*, 10150–10176. [CrossRef]

58. Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. *J. Chem. Inf. Model.* **2014**, *54*, 1717–1736. [CrossRef]

59. Antunes, D.A.; Devaurs, D.; Kavraki, L.E. Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discov.* **2015**, *10*, 1301–1313. [CrossRef] [PubMed]

60. Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964–12975. [CrossRef]

61. Takeda, S.; Kaneko, H.; Funatsu, K. Chemical-space-based de novo design method to generate drug-like molecules. *J. Chem. Inf. Model.* **2016**, *56*, 1885–1893. [CrossRef]

62. Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4*, eaap7885. [CrossRef]

63. Fischer, T.; Gazzola, S.; Riedl, R. Approaching target selectivity by de novo drug design. *Expert. Opin. Drug Discov.* **2019**, *14*, 791–803. [CrossRef]

64. Méndez-Lucio, O.; Baillif, B.; Clevert, D.-A.; Rouquié, D.; Wichard, J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat. Commun.* **2020**, *11*, 10. [CrossRef]

65. Yuan, Y.; Pei, J.; Lai, L. LigBuilder V3: A multi-target de novo drug design approach. *Front. Chem.* **2020**, *8*, 142. [CrossRef] [PubMed]

66. Swann, S.L.; Brown, S.P.; Muchmore, S.W.; Patel, H.; Merta, P.; Locklear, J.; Hajduk, P.J. A Unified, probabilistic framework for structure- and ligand-based virtual screening. *J. Med. Chem.* **2011**, *54*, 1223–1232. [CrossRef] [PubMed]

67. Cleves, A.E.; Jain, A.N. Structure- and ligand-based virtual screening on DUD-E⁺: Performance dependence on approximations to the binding pocket. *J. Chem. Inf. Model.* **2020**, *60*, 4296–4310. [CrossRef]

68. Kooistra, A.J.; Vischer, H.F.; McNaught-Flores, D.; Leurs, R.; De Esch, I.J.P.; De Graaf, C. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Sci. Rep.* **2016**, *6*, 28288. [CrossRef]

69. Tan, L.; Lounkine, E.; Bajorath, J. Similarity searching using fingerprints of molecular fragments involved in protein-ligand interactions. *J. Chem. Inf. Model.* **2008**, *48*, 2308–2312. [CrossRef]

70. Tan, L.; Bajorath, J. Utilizing target-ligand interaction information in fingerprint searching for ligands of related targets. *Chem. Biol. Drug Des.* **2009**, *74*, 25–32. [CrossRef]

71. 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.* **2012**, *52*, 943–955. [CrossRef] [PubMed]

72. Larsson, M.; Fraccalvieri, D.; Andersson, C.D.; Bonati, L.; Linusson, A.; Andersson, P.L. Identification of potential aryl hydrocarbon receptor ligands by virtual screening of industrial chemicals. *Environ. Sci. Pollut. Res.* **2018**, *25*, 2436–2449. [CrossRef]

73. Tanrikulu, Y.; Schneider, G. Pseudoreceptor models in drug design: Bridging ligand- and receptor-based virtual screening. *Nat. Rev. Drug Discov.* **2008**, *7*, 667–677. [CrossRef] [PubMed]

74. Lloyd, D.G.; Buenemann, C.L.; Todorov, N.P.; Manallack, D.T.; Dean, P.M. Scaffold hopping in de novo design. Ligand generation in the absence of receptor information. *J. Med. Chem.* **2004**, *47*, 493–496. [CrossRef]

75. Lorenzo, V.P.; Barbosa Filho, J.M.; Scotti, L.; Scotti, M.T. Combined structure- and ligand-based virtual screening to evaluate caulerpin analogs with potential inhibitory activity against monoamine oxidase B. *Rev. Bras. Farmacogn.* **2015**, *25*, 690–697. [CrossRef]

76. Anighoro, A.; Bajorath, J. Three-dimensional similarity in molecular docking: Prioritizing ligand poses on the basis of experimental binding modes. *J. Chem. Inf. Model.* **2016**, *56*, 580–587. [CrossRef]

77. Anighoro, A.; Bajorath, J. A hybrid virtual screening protocol based on binding mode similarity. *Methods Mol. Biol.* **2018**, *1824*, 165–176. [PubMed]

78. Jacquemard, C.; Drwal, M.N.; Desaphy, J.; Kellenberger, E. Binding mode information improves fragment docking. *J. Cheminform.* **2019**, *11*, 24. [CrossRef]

79. Jacquemard, C.; Tran-Nguyen, V.-K.; Drwal, M.N.; Rognan, D.; Kellenberger, E. Local interaction density (LID), a fast and efficient tool to prioritize docking poses. *Molecules* **2019**, *24*, 2610. [CrossRef] [PubMed]

80. Vázquez, J.; Deplano, A.; Herrero, A.; Gibert, E.; Herrero, E.; Luque, F.J. Assessing the performance of mixed strategies to combine lipophilic molecular similarity and docking in virtual screening. *J. Chem. Inf. Model.* **2020**, *60*, 4231–4245. [CrossRef]

81. Ai, G.; Tian, C.; Deng, D.; Fida, G.; Chen, H.; Ding, L.; Ma, Y.; Gu, Y. A Combination of 2D similarity search, pharmacophore, and molecular docking techniques for the identification of vascular endothelial growth factor receptor-2 inhibitors. *Anticancer. Drugs* **2015**, *26*, 399–409. [CrossRef] [PubMed]

82. Staroń, J.; Kurczab, R.; Warszycki, D.; Satała, G.; Krawczyk, M.; Bugno, R.; Lenda, T.; Popik, P.; Hogendorf, A.S.; Hogendorf, A.; et al. Virtual screening-driven discovery of dual 5-HT6/5-HT2A receptor ligands with pro-cognitive properties. *Eur. J. Med. Chem.* **2020**, *185*, 111857. [CrossRef]

83. Oum, Y.H.; Kell, S.A.; Yoon, Y.; Liang, Z.; Burger, P.; Shim, H. Discovery of novel aminopiperidinyl amide CXCR4 modulators through virtual screening and rational drug design. *Eur. J. Med. Chem.* **2020**, *201*, 112479. [CrossRef] [PubMed]

84. Khan, S.U.; Ahemad, N.; Chuah, L.H.; Naidu, R.; Htar, T.T. Sequential ligand- and structure-based virtual screening approach for the identification of potential g protein-coupled estrogen receptor-1 (GPER-1) modulators. *RSC Adv.* **2019**, *9*, 2525–2538. [CrossRef]

85. Xu, X.; Ren, J.; Ma, Y.; Liu, H.; Rong, Q.; Feng, Y.; Wang, Y.; Cheng, Y.; Ge, R.; Li, Z.; et al. Discovery of cyanopyridine scaffold as novel indoleamine-2,3-dioxygenase 1 (IDO1) inhibitors through virtual screening and preliminary hit optimisation. *J. Enzyme Inhib. Med. Chem.* **2019**, *34*, 250–263. [CrossRef] [PubMed]

86. Lu, F.; Luo, G.; Qiao, L.; Jiang, L.; Li, G.; Zhang, Y. Virtual screening for potential allosteric inhibitors of cyclin-dependent kinase 2 from traditional chinese medicine. *Molecules* **2016**, *2*, 1259. [CrossRef] [PubMed]

87. Liang, J.W.; Wang, M.Y.; Wang, S.; Li, S.L.; Li, W.Q.; Meng, F.H. Identification of novel CDK2 inhibitors by a multistage virtual screening method based on SVM, pharmacophore and docking model. *J. Enzyme Inhib. Med. Chem.* **2020**, *35*, 235–244. [CrossRef] [PubMed]

88. Kaur, M.; Silakari, O. Ligand-based and e-pharmacophore modeling, 3D-QSAR and hierarchical virtual screening to identify dual inhibitors of spleen tyrosine kinase (Syk) and janus kinase 3 (JAK3). *J. Biomol. Struct. Dyn.* **2017**, *35*, 3043–3060. [CrossRef]

89. Baell, J.B.; Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740. [CrossRef]

90. Liu, K.; Watanabe, E.; Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *J. Comput. Aided Mol. Des.* **2017**, *31*, 201–211. [CrossRef]

91. Liu, K.; Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: A cross-docking study. *J. Chem. Inf. Model.* **2017**, *57*, 2514–2522. [CrossRef]

92. Mollica, L.; Theret, I.; Antoine, M.; Perron-Siera, F.; Charton, Y.; Fourquez, J.-M.; Wierzbicki, M.; Boutin, J.A.; Ferry, G.; Decherchi, S.; et al. Molecular dynamics simulations and kinetic measurements to estimate and predict protein-ligand residence times. *J. Med. Chem.* **2016**, *59*, 7167–7176. [CrossRef]

93. Majewski, M.; Barril, X. Structural stability predicts the binding mode of protein-ligand complexes. *J. Chem. Inf. Mod.* **2020**, *60*, 1644–1651. [CrossRef] [PubMed]

94. OpenEye Scientic Software. EON. 2.2.0.5. *Santa Fe, NM, USA.* Available online: https://www.eyesopen.com/ (accessed on 15 October 2020).

95. McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578–596. [CrossRef]

96. McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comput. Aided Mol. Des.* **2012**, *26*, 897–906. [CrossRef] [PubMed]

97. Dawood, H.M.; Ibrahim, R.S.; Shawky, E.; Hammoda, H.M.; Metwally, A.M. Integrated in silico-in vitro strategy for screening of some traditional egyptian plants for human artomatase inhibitors. *J. Ethnopharmacol.* **2018**, *224*, 359–372. [CrossRef]

98. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [CrossRef]

99. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759. [CrossRef] [PubMed]

100. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. [CrossRef] [PubMed]

101. Dixon, S.L.; Smondyrev, A.M.; Knoll, E.H.; Rao, S.N.; Shaw, D.E.; Friesner, R.A. PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3d database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.* **2006**, *20*, 647–671. [CrossRef]

102. Dixon, S.L.; Smondyrev, A.M.; Rao, S.N. PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug Des.* **2006**, *67*, 370–372. [CrossRef]

103. Svensson, F.; Karlén, A.; Sköld, C. Virtual screening data fusion using both structure-and ligand-based methods. *J. Chem. Inf. Model.* **2012**, *52*, 225–232. [CrossRef]

104. Willett, P. Enhancing the effectiveness of ligand-based virtual screening using data fusion. *QSAR Comb. Sci.* **2006**, *25*, 1143–1152. [CrossRef]

105. Willett, P. Combination of similarity rankings using data fusion. *J. Chem. Inf. Model.* **2013**, *53*, 1–10. [CrossRef] [PubMed]

106. Böttcher, J.; Dilworth, D.; Reiser, U.; Neumüller, R.A.; Schleicher, M.; Petronczki, M.; Zeeb, M.; Mischerikow, N.; Allali-Hassani, A.; Szewczyk, M.M.; et al. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nat. Chem. Biol.* **2019**, *15*, 822–829. [CrossRef] [PubMed]

107. Arany, A.; Bolgar, B.; Balogh, B.; Antal, P.; Matyus, P. Multi-aspect candidates for repositioning: Data fusion methods using heterogeneous information sources. *Curr. Med. Chem.* **2012**, *20*, 95–107. [CrossRef]

108. Huang, G.; Li, J.; Wang, P.; Li, W. A review of computational drug repositioning approaches. *Comb. Chem. High Throughput Screen.* **2017**, *20*, 831–838. [CrossRef]

109. Liu, X.; Xu, Y.; Li, S.; Wang, Y.; Peng, J.; Luo, C.; Luo, X.; Zheng, M.; Chen, K.; Jiang, H. In silico target fishing: Addressing a "big data" Problem by ligand-based similarity rankings with data fusion. *J. Cheminform.* **2014**, *6*, 1–14. [CrossRef]

110. Bajusz, D.; Rácz, A.; Héberger, K. Comparison of data fusion methods as consensus scores for ensemble docking. *Molecules* **2019**, *24*, 2690. [CrossRef]

111. Jaundoo, R.; Bohmann, J.; Gutierrez, G.; Klimas, N.; Broderick, G.; Craddock, T. Using a consensus docking approach to predict adverse drug reactions in combination drug therapies for gulf war illness. *Int. J. Mol. Sci.* **2018**, *19*, 3355. [CrossRef]

112. Charifson, P.S.; Corkery, J.J.; Murcko, M.A.; Walters, W.P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109. [CrossRef]

113. Feher, M. Consensus scoring for protein-ligand interactions. *Drug Discov. Today* **2006**, *11*, 421–428. [CrossRef]

114. Sun, H.; Pan, P.; Tian, S.; Xu, L.; Kong, X.; Li, Y.; Li, D.; Hou, T. Constructing and validating high-performance MIEC-SVM models in virtual screening for kinases: A better way for actives discovery. *Sci. Rep.* **2016**, *6*, 1–12. [CrossRef] [PubMed]

115. Shen, M.; Tian, S.; Pan, P.; Sun, H.; Li, D.; Li, Y.; Zhou, H.; Li, C.; Ming-Yuen Lee, S.; Hou, T. Discovery of novel ROCK1 inhibitors via integrated virtual screening strategy and bioassays. *Sci. Rep.* **2015**, *5*, 16749. [CrossRef] [PubMed]

116. Tan, L.; Geppert, H.; Sisay, M.T.; Gütschow, M.; Bajorath, J. Integrating structure- and ligand-based virtual screening: Comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets. *ChemMedChem* **2008**, *3*, 1566–1571. [CrossRef] [PubMed]

117. Rogers, D.; Hahn, M. Extended-connectivity fingerpirints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef]

118. Berry, M.; Fielding, B.C.; Gamieldien, J. Potential broad spectrum inhibitors of the coronavirus 3CLpro: A virtual screening and structure-based drug design study. *Viruses* **2015**, *7*, 6642–6660. [CrossRef]

119. Vucicevic, J.; Srdic-Rajic, T.; Pieroni, M.; Laurila, J.M.M.; Perovic, V.; Tassini, S.; Azzali, E.; Costantino, G.; Glisic, S.; Agbaba, D.; et al. A combined ligand- and structure-based approach for the identification of rilmenidine-derived compounds which synergize the antitumor effects of doxorubicin. *Bioorganic Med. Chem.* **2016**, *24*, 3174–3183. [CrossRef]

120. Jang, C.; Yadav, D.K.; Subedi, L.; Venkatesan, R.; Venkanna, A.; Afzal, S.; Lee, E.; Yoo, J.; Ji, E.; Kim, S.Y.; et al. Identification of novel acetylcholinesterase inhibitors designed by pharmacophore-based virtual screening, molecular docking and bioassay. *Sci. Rep.* **2018**, *8*, 14921. [CrossRef]

121. Costa, G.; Rocca, R.; Corona, A.; Grandi, N.; Moraca, F.; Romeo, I.; Talarico, C.; Gagliardi, M.G.; Ambrosio, F.A.; Ortuso, F.; et al. Novel natural non-nucleoside inhibitors of HIV-1 reverse transcriptase identified by shape- and structure-based virtual screening techniques. *Eur. J. Med. Chem.* **2019**, *161*, 1–10. [CrossRef]

122. Vedani, A.; Zbinden, P.; Snyder, J.P. Pseudo-receptor modeling: A new concept for the three-dimensional construction of receptor binding sites. *J. Recept. Signal Transduct.* **1993**, *13*, 163–177. [CrossRef]

123. Andrews, P.R.; Quint, G.; Winkler, D.A.; Richardson, D.; Sadek, M.; Spurling, T.H. Morpheus: A conformation-activity relationships and receptor modeling package. *J. Mol. Graph.* **1989**, *7*, 138–145. [CrossRef]

124. Pei, J.; Chen, H.; Liu, Z.; Han, X.; Wang, Q.; Shen, B.; Zhou, J.; Lai, L. Improving the quality of 3D-QSAR by using flexible-ligand receptor models. *J. Chem. Inf. Model.* **2005**, *45*, 1920–1933. [CrossRef] [PubMed]

125. Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–169. [CrossRef] [PubMed]

126. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J.S. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): Theory and application. *J. Chem. Inf. Model.* **2007**, *47*, 279–294. [CrossRef]

127. Jacob, L.; Vert, J.-P. Protein-ligand interaction prediction: An improved chemogenomics approach. *Bioinformatics* **2008**, *24*, 2149–2156. [CrossRef]

128. Salam, N.K.; Nuti, R.; Sherman, W. Novel method for generating structure-based pharmacophores using energetic analysis. *J. Chem. Inf. Model.* **2009**, *49*, 2356–2368. [CrossRef]

129. Sato, T.; Honma, T.; Yokoyama, S. Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening. *J. Chem. Inf. Model.* **2010**, *50*, 170–185. [CrossRef]

130. Koes, D.R.; Camacho, C.J. ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* **2012**, *40*, W409–W414. [CrossRef]

131. Tran-Nguyen, V.K.; Da Silva, F.; Bret, G.; Rognan, D. All in one: Cavity detection, druggability estimate, cavity-based pharmacophore perception, and virtual screening. *J. Chem. Inf. Model.* **2019**, *59*, 573–585. [CrossRef] [PubMed]

132. Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337–344. [CrossRef]

133. Salentin, S.; Schreiber, S.; Haupt, V.J.; Adasme, M.F.; Schroeder, M. PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **2015**, *43*, 443–447. [CrossRef]

134. Hajiebrahimi, A.; Ghasemi, Y.; Sakhteman, A. FLIP: An assisting software in structure based drug design using fingerprint of protein-ligand interaction profiles. *J. Mol. Graph. Model.* **2017**, *78*, 234–244. [CrossRef] [PubMed]

135. Jasper, J.B.; Humbeck, L.; Brinkjost, T.; Koch, O. A novel interaction fingerprint derived from per atom score contributions: Exhaustive evaluation of interaction fingerprint performance in docking based virtual screening. *J. Cheminform.* **2018**, *10*, 15. [CrossRef] [PubMed]

136. Da Silva, F.; Desaphy, J.; Rognan, D. IChem: A versatile toolkit for detecting, comparing, and predicting protein-ligand interactions. *ChemMedChem* **2018**, *13*, 507–510. [CrossRef] [PubMed]

137. Desaphy, J.; Raimbaud, E.; Ducrot, P.; Rognan, D. Encoding protein-ligand interaction patterns in fingerprints and graphs. *J. Chem. Inf. Model.* **2013**, *53*, 623–637. [CrossRef] [PubMed]

138. Salentin, S.; Adasme, M.F.; Heinrich, J.C.; Haupt, V.J.; Daminelli, S.; Zhang, Y.; Schroeder, M. From malaria to cancer: Computational drug repositioning of amodiaquine using PLIP interaction patterns. *Sci. Rep.* **2017**, *7*, 11401. [CrossRef] [PubMed]

139. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594. [CrossRef] [PubMed]

140. Jain, A.N. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput. Aided Mol. Des.* **2007**, *21*, 281–306. [CrossRef]

141. Symyx Software. *MACCS Structural Keys*; Symyx Technologies: San Ramon, CA, USA, 2002.

142. Jubb, H.C.; Higueruelo, A.P.; Ochoa-Montaño, B.; Pitt, W.R.; Ascher, D.B.; Blundell, T.L. Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* **2017**, *429*, 365–371. [CrossRef]

143. Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195–207. [CrossRef]

144. Pérez-Nueno, V.I.; Rabal, O.; Borrell, J.I.; Teixidó, J. APIF: A new interaction fingerprint based on atom pairs and its application to virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 1245–1260. [CrossRef]

145. Lenselink, E.B.; Jespers, W.; Van Vlijmen, H.W.T.; IJzerman, A.P.; Van Westen, G.J.P. Interacting with GPCRs: Using interaction fingerprints for virtual screening. *J. Chem. Inf. Model.* **2016**, *56*, 2053–2060. [CrossRef] [PubMed]

146. Zhang, D.; Huang, S.; Mei, H.; Kevin, M.; Shi, T.; Chen, L. Protein–ligand interaction fingerprints for accurate prediction of dissociation rates of P38 MAPK type II inhibitors. *Integr. Biol.* **2019**, *11*, 53–60. [CrossRef] [PubMed]

147. Zhao, Z.; Bourne, P.E. Revealing acquired resistance mechanisms of kinase-targeted drugs using an on-the-fly, function-site interaction fingerprint approach. *J. Chem. Theory Comput.* **2020**, *16*, 3152–3161. [CrossRef] [PubMed]

148. Kumar, A.; Zhang, K.Y.J. Application of shape similarity in pose selection and virtual screening in CSARdock2014 exercise. *J. Chem. Inf. Model.* **2016**, *56*, 965–973. [CrossRef]

149. Prathipati, P.; Mizuguchi, K. Integration of ligand and structure based approaches for CSAR-2014. *J. Chem. Inf. Model.* **2016**, *56*, 974–987. [CrossRef]

150. Kumar, A.; Zhang, K.Y.J. A pose prediction approach based on ligand 3D shape similarity. *J. Comput. Aided Mol. Des.* **2016**, *30*, 457–469. [CrossRef] [PubMed]

151. Kumar, A.; Zhang, K.Y.J. Improving ligand 3D shape similarity-based pose prediction with a continuum solvent model. *J. Comput. Aided Mol. Des.* **2019**, *33*, 1045–1055. [CrossRef]

152. Kumar, A.; Zhang, K.Y.J. Prospective evaluation of shape similarity based pose prediction method in D3R grand challenge 2015. *J. Comput. Aided Mol. Des.* **2016**, *30*, 685–693. [CrossRef]

153. Lee, H.S.; Choi, J.; Kufareva, I.; Abagyan, R.; Filikov, A.; Yang, Y.; Yoon, S. Optimization of high throughput virtual screening by combining shape-matching and docking methods. *J. Chem. Inf. Model.* **2008**, *48*, 489–497. [CrossRef]

154. Kelley, B.P.; Brown, S.P.; Warren, G.L.; Muchmore, S.W. POSIT: Flexible shape-guided docking for pose prediction. *J. Chem. Inf. Model.* **2015**, *55*, 1771–1780. [CrossRef]

155. Anighoro, A.; Bajorath, J. Binding mode similarity measures for ranking of docking poses: A case study on the adenosine A2Areceptor. *J. Comput. Aided Mol. Des.* **2016**, *30*, 447–456. [CrossRef] [PubMed]

156. Anighoro, A.; Bajorath, J. Compound ranking based on fuzzy three-dimensional similarity improves the performance of docking into homology models of g-protein-coupled receptors. *ACS Omega* **2017**, *2*, 2583–2592. [CrossRef] [PubMed]

157. Marialke, J.; Tietze, S.; Apostolakis, J. Similarity based docking. *J. Chem. Inf. Model.* **2008**, *48*, 186–196. [CrossRef]

158. Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J.A.; Kang, Y.N.; Delproposto, J.; Kubish, G.; Dunbar, J.B.; Carlson, H.A.; et al. D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. *J. Comput. Aided Mol. Des.* **2016**, *30*, 651–668. [CrossRef] [PubMed]

159. Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V.A.; Walters, W.P.; Kuhn, B.; Rudolph, M.G.; et al. D3R grand challenge 2: Blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **2018**, *32*, 1–20. [CrossRef] [PubMed]

160. Gaieb, Z.; Parks, C.D.; Chiu, M.; Yang, H.; Shao, C.; Walters, W.P.; Lambert, M.H.; Nevins, N.; Bembenek, S.D.; Ameriks, M.K.; et al. D3R grand challenge 3: Blind prediction of protein-ligand poses and affinity rankings. *J. Comput. Aided Mol. Des.* **2019**, *33*, 1–18. [CrossRef]

161. Kumar, A.; Zhang, K.Y.J. Shape similarity guided pose prediction: Lessons from D3R grand challenge 3. *J. Comput. Aided Mol. Des.* **2019**, *33*, 47–59. [CrossRef]

162. Kumar, A.; Zhang, K.Y.J. A cross docking pipeline for improving pose prediction and virtual screening performance. *J. Comput. Aided Mol. Des.* **2018**, *32*, 163–173. [CrossRef]

163. Alford, R.F.; Leaver-Fay, A.; Jeliazkov, J.R.; O'Meara, M.J.; DiMaio, F.P.; Park, H.; Shapovalov, M.V.; Renfrew, P.D.; Mulligan, V.K.; Kappel, K.; et al. The rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048. [CrossRef]

164. Varela-Rial, A.; Majewski, M.; Cuzzolin, A.; Martinez-Rosell, G.; De Fabritiis, G. SkeleDock: A web application for scaffold docking in PlayMolecule. *J. Chem. Inf. Model.* **2020**, *60*, 2673–2677. [CrossRef]

165. Marialke, J.; Körner, R.; Tietze, S.; Apostolakis, J. Graph-based molecular alignment (GMA). *J. Chem. Inf. Model.* **2007**, *47*, 591–601. [CrossRef]

166. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. Autodock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *16*, 2785–2791. [CrossRef]

167. *Molecular Operating Environment 2019.01*; Chemical Computing Group ULC: Montreal, QC, Canada, 2020.

168. Ginex, T.; Vázquez, J.; Gibert, E.; Herrero, E.; Luque, F.J. Lipophilicity in drug design: An overview of lipophilicity descriptors in 3D-QSAR. *Fut. Med.* **2019**, *11*, 1177–1193. [CrossRef] [PubMed]

169. PharmScreen. Pharmacelera, Barcelona. 2019. Available online: www.pharmacelera.com (accessed on 15 October 2020).

170. Ruiz-Carmona, S.; Alvarez-García, D.; Foloppe, N.; Garmendia-Doval, A.B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R.E.; Morley, S.D. rDock: A Dast, Versatile and open source porgram for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.* **2014**, *10*, e1003571. [CrossRef] [PubMed]

171. Jones, G.; Willet, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748. [CrossRef] [PubMed]

172. Chen, H.; Kogej, T.; Engkvist, O. Cheminformatics in drug discovery, an industrial perspective. *Mol. Inform.* **2018**, *37*, 1800041. [CrossRef] [PubMed]

173. Gaulton, A.; Hersey, A.; Nowotka, M.; Patrícia Bento, A.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrí an-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2016**, *45*, 945–954. [CrossRef]

174. Wang, Y.; Xiao, J.; Suzek, T.O.; Zhang, J.; Wang, J.; Bryant, S.H. PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, *37*, 623–633. [CrossRef]

175. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672. [CrossRef]

176. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [CrossRef]

177. Hoffmann, T.; Gastreich, M. The next level in chemical space navigation: Going far beyond enumerable compound libraries. *Drug Discov. Today* **2019**, *24*, 1148–1156. [CrossRef] [PubMed]

178. Van Hilten, N.; Chevillard, F.; Kolb, P. Virtual compound libraries in computer-assisted drug discovery. *J. Chem. Inf. Model.* **2019**, *59*, 644–651. [CrossRef] [PubMed]

179. Miyao, T.; Arakawa, M.; Funatsu, K. Exhaustive structure generation for inverse-QSPR/QSAR. *Mol. Inform.* **2010**, *29*, 111–125. [CrossRef] [PubMed]

180. Hartenfellar, M.; Proschak, E.; Schüller, G. DOGS: Reaction-driven de novo design of bioactive compounds. *PLoS Comput. Biol.* **2012**, *8*, 1–12. [CrossRef]

181. Pottel, J.; Moitessier, N. Customizable generation of synthetically accessible, local chemical subspaces. *J. Chem. Inf. Model.* **2017**, *57*, 454–467. [CrossRef]

182. Gao, W.; Coley, C.W. The synthesizability of molecules proposed by generative models. *J. Chem. Inf. Model.* **2020**. [CrossRef]

183. Mishima, K.; Kaneko, H.; Funatsu, K. Development of a new de novo design algorithm for exploring chemical space. *Mol. Inform.* **2014**, *33*, 779–789. [CrossRef]

184. Podlewska, S.; Czarnecki, W.M.; Kafel, R.; Bojarski, A.J. Creating the new from the old: Combinatorial libraries generation with machine-learning-based compound structure optimization. *J. Chem. Inf. Model.* **2017**, *57*, 133–147. [CrossRef] [PubMed]

185. Gao, K.; Nguyen, D.D.; Tu, M.; Wei, G.-W. Generative network complex for the automated generation of drug-like molecules. *J. Chem. Inf. Model.* **2020**. [CrossRef]

186. Amabilino, S.; Pogány, P.; Pickett, S.D.; Green, D.V.S. Guidelines for recurrent neural network transfer learning-based molecular generation of focused libraries. *J. Chem. Inf. Model.* **2020**. [CrossRef]

187. Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminform.* **2017**, *9*, 48. [CrossRef] [PubMed]

188. Domenico, A.; Nicola, G.; Daniela, T.; Fulvio, C.; Nicola, A.; Orazio, N. De novo drug design of targeted chemical libraries based on artificial intelligence and pair-based multiobjective optimization. *J. Chem. Inf. Model.* **2020**. [CrossRef] [PubMed]

189. Bosch, J. PPI inhibitor and stabilizer development in human diseases. *Drug Discov. Today Technol.* **2017**, *24*, 3–9. [CrossRef]

190. Sijbesma, E.; Hallenbeck, K.K.; Leysen, S.; De Vink, P.J.; Skóra, L.; Jahnke, W.; Brunsveld, L.; Arkin, M.R.; Ottmann, C. Site-directed fragment-based screening for the discovery of protein-protein interaction stabilizers. *J. Am. Chem. Soc.* **2019**, *141*, 3524–3531. [CrossRef]

191. Stevers, L.M.; Sijbesma, E.; Botta, M.; Mackintosh, C.; Obsil, T.; Landrieu, I.; Cau, Y.; Wilson, A.J.; Karawajczyk, A.; Eickhoff, J.; et al. Modulators of 14-3-3 protein-protein interactions. *J. Med. Chem.* **2018**, *61*, 3755–3778. [CrossRef] [PubMed]

192. Petta, I.; Lievens, S.; Libert, C.; Tavernier, J.; De Bosscher, K. Modulation of protein-protein interactions for the development of novel therapeutics. *Mol. Ther.* **2016**, *24*, 707–718. [CrossRef] [PubMed]

193. Zhong, M.; Lee, G.M.; Sijbesma, E.; Ottman, C.; Arkin, M.R. Modulating protein-protein interaction networks in protein homeostasis. *Curr. Opin. Chem. Biol.* **2019**, *50*, 55–65. [CrossRef]

194. Reynès, C.; Host, H.; Camproux, A.-C.; Laconde, G.; Leroux, F.; Mazars, A.; Deprez, B.; Fahraeus, R.; Villoutreix, B.O.; Sperandio, O. Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. *PLoS Comput. Biol.* **2010**, *6*, e1000695. [CrossRef]

195. Hamon, V.; Brunel, J.-M.; Combes, S.; Basse, M.J.; Roche, P.; Morelli, X. 2P2IChem: Focused chemical libraries dedicated to orthosteric modulation of protein-protein interactions. *MedChemComm* **2013**, *4*, 797–809. [CrossRef]

196. Bosc, N.; Muller, C.; Hoffer, L.; Lagorce, D.; Bourg, S.; Derviaux, C.; Gourdel, M.-E.; Rain, J.-C.; Miller, T.W.; Villoutreix, B.O.; et al. Fr-PPIChem: An academic compound library dedicated to protein-protein interactions. *ACS Chem. Biol.* **2020**, *15*, 1566–1574. [CrossRef]

197. Zhang, X.; Betzi, S.; Morelli, X.; Roche, P. Focused chemical libraries—Design and enrichment: An example of protein-protein interaction chemical space. *Future Med. Chem.* **2014**, *6*, 1291–1307. [CrossRef]

198. Singh, N.; Chaput, L.; Villoutreix, B.O. Fast rescoring protocols to improve the performance of structure-based virtual screening performed on protein-protein interfaces. *J. Chem. Inf. Model.* **2020**, *60*, 3910–3934. [CrossRef] [PubMed]

199. Cruz-Monteagudo, M.; Schürer, S.; Tejera, E.; Pérez-Castillo, Y.; Medina-Franco, J.L.; Sánchez-Rodríguez, A.; Borges, F. Systemic QSAR and phenotypic virtual screening: Chasing butterflies in drug discovery. *Drug Discov. Today* **2017**, *22*, 994–1007. [CrossRef] [PubMed]

200. Feng, B.Y.; Simeonov, A.; Jadhav, A.; Babaoglu, K.; Inglese, J.; Shoichet, B.K.; Austin, C.P. A High-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* **2007**, *50*, 2385–2390. [CrossRef] [PubMed]

201. Feng, B.Y.; Shoichet, B.K. A Detergent-based assay for the detection of promiscous inhibitors. *Nat. Protoc.* **2006**, *1*, 550–553. [CrossRef]

202. Duan, D.; Torosyan, H.; Elnatan, D.; McLaughlin, C.K.; Logie, J.; Shoichet, M.S.; Agard, D.A.; Shoichet, B.K. Internal structure and preferential protein binding of colloidal aggregates. *ACS Chem. Biol.* **2017**, *12*, 282–290. [CrossRef]

203. Owen, S.C.; Doak, A.K.; Wassam, P.; Schoichet, M.S.; Schoichet, B.K. Colloidal aggregation affects the efficacy of anticancer drugs in cell culture. *ACS Chem. Biol.* **2012**, *7*, 1249–1435. [CrossRef]

204. Dlim, M.M.; Shahout, F.S.; Khabir, M.K.; Labonté, P.P.; LaPlante, S.R. Revealing drug self-associations into nano-entities. *ACS Omega* **2019**, *4*, 8919–8925. [CrossRef]

205. Liu, Y.; Beresini, M.H.; Johnson, A.; Mintzer, R.; Shah, K.; Clark, K.; Schmidt, S.; Lewis, C.; Liimatta, M.; Elliott, L.O.; et al. Case studies of minimizing nonspecific inhibitors in HTS campaigns that use assay-ready plates. *J. Biomol. Screen.* **2012**, *17*, 225–236. [CrossRef]

206. Ghattas, M.A.; Al Rawashdeh, S.; Atatreh, N.; Bryce, R.A. How do small molecule aggregates inhibit enzyme activity? A molecular dynamics study. *J. Chem. Inf. Model.* **2020**, *60*, 3901–3909. [CrossRef]

207. Coan, K.E.D.; Schoichet, B.K. Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors. *J. Am. Chem. Soc.* **2008**, *130*, 9606–9612. [CrossRef] [PubMed]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.