

Review

# Protein–Ligand Docking in the Machine-Learning Era

Chao Yang <sup>1</sup> , Eric Anthony Chen <sup>1</sup> and Yingkai Zhang <sup>1,2,\*</sup> 

<sup>1</sup> Department of Chemistry, New York University, New York, NY 10003, USA; cy1063@nyu.edu (C.Y.); eac709@nyu.edu (E.A.C.)

<sup>2</sup> NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

\* Correspondence: yingkai.zhang@nyu.edu; Tel.: +1-212-998-7882

**Abstract:** Molecular docking plays a significant role in early-stage drug discovery, from structure-based virtual screening (VS) to hit-to-lead optimization, and its capability and predictive power is critically dependent on the protein–ligand scoring function. In this review, we give a broad overview of recent scoring function development, as well as the docking-based applications in drug discovery. We outline the strategies and resources available for structure-based VS and discuss the assessment and development of classical and machine learning protein–ligand scoring functions. In particular, we highlight the recent progress of machine learning scoring function ranging from descriptor-based models to deep learning approaches. We also discuss the general workflow and docking protocols of structure-based VS, such as structure preparation, binding site detection, docking strategies, and post-docking filter/re-scoring, as well as a case study on the large-scale docking-based VS test on the LIT-PCBA data set.

**Keywords:** molecular docking; virtual screening; protein–ligand scoring function; machine learning; deep learning; datasets



**Citation:** Yang, C.; Chen, E.A.; Zhang, Y. Protein–Ligand Docking in the Machine-Learning Era. *Molecules* **2022**, *27*, 4568. <https://doi.org/10.3390/molecules27144568>

Academic Editor: Chung F. Wong

Received: 3 July 2022

Accepted: 14 July 2022

Published: 18 July 2022

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



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Discovering bioactive compounds for a given target from a large compound library is one of the major tasks in drug development. It is laborious and costly to carry out binding affinity measurements on tens or hundreds of thousands of compounds. Hence, the overall cost of drug discovery will be greatly reduced if the binding affinities of compounds can be effectively predicted with computational methods before performing experiments. Employing computational methods to find active compounds is called Computer-Aided Drug Design (CADD) [1–6]. CADD has emerged as a powerful and promising technique in the development of new hit compounds and led to the discovery of several approved drugs, including the human immunodeficiency virus I (HIV-1) drugs (amprenavir and saquinavir), the fibrinogen antagonist (tirofiban), the carbonic anhydrase II inhibitor (dorzolamide), the angiotensin-converting enzyme (ACE) inhibitor (captopril) and the human rhinovirus 3C protease inhibitor (rupintrivir) [5,7–10].

CADD methods can be categorized into two general types, structure-based drug discovery (SBDD) and ligand-based drug discovery (LBDD) [11]. SBDD aims to find active compounds based on the physical interactions of 3-dimensional (3D) structures between the target protein and small molecule [2]. LBDD investigates existing activities using approaches of quantitative structure-activity relationship (QSAR) models, chemical similarity, pharmacophore and 3D shape matching to predict the property of a novel compound [12]. Both SBDD and LBDD are widely used in drug discovery processes and can be combined to use in virtual screening (VS). For example, scientists can first search for compounds similar to available and moderately active compounds from a large library using LBDD and then predict the protein–ligand interactions to find the favorable compounds using SBDD [8].

Understanding the binding mechanism of the protein and small molecule is crucial to discover and optimize drug molecules. The application of SBDD tools has gained significant interest in recent decades due to the recent explosion of high quality 3D macromolecule structures [2]. SBDD aims to identify binding sites and interactions that are important for the biological function of the protein. This structural information then suggests the design of therapeutic compounds that can compete with essential interactions involving the target protein and thus interrupt the abnormal biological pathways.

Structure-based inhibitor design approaches often use molecular docking, a computational procedure that efficiently predicts non-covalent interactions between macromolecules (receptor) and small molecules (ligand) [13–17]. This procedure mimics the lock-and-key model of drug action to predict the experimental binding pose and affinity of a small molecule within the binding site of the target protein [18]. Docking methods are commonly used in structure-based VS on large molecular libraries, since they are fast enough to scan over millions of compounds using a simplified scoring function [19]. Docking programs, such as DOCK, AutoDock, GOLD, Glide, FRED and Surflex-Dock, rely on scoring functions to evaluate protein–ligand binding [15,17,20–24]. Therefore, the critical component of molecular docking is a robust, fast and accurate scoring function.

In this review, we will first describe protein–ligand scoring functions including its classification, datasets, and evaluation metrics. Then, we discuss recent advances in ML-based scoring functions. Finally, molecular docking protocols and general workflow utilized in structure-based VS are examined.

## 2. Protein–Ligand Scoring Functions

The binding affinity between a protein and ligand is determined by their binding free energy. Rigorous prediction of binding free energy requires extensive sampling of complex conformations and explicit treatment of aqueous solution environment, such as free energy perturbation (FEP) [25,26] and thermodynamic integration (TI) [27,28], which are too computationally expensive to be suited for large scale VS. Alternatively, molecular docking typically employs a scoring function to estimate the protein–ligand binding free energy based on a single protein–ligand complex structure. This is much faster and facilitates its use in VS on large molecular libraries [3,19,29].

Scoring functions are a family of computational methods that have been widely applied in SBDD for fast evaluation of protein–ligand interactions [30,31]. They can be used in a molecular docking job to rank different putative ligand binding poses to select the most favorable one (best-scored pose). The score of the favorable pose is used to represent the binding affinity of the compound. This combined docking/scoring scheme has been widely applied to VS for hit identification as well as structure–activity relationship (SAR) analysis for hit-to-lead and lead optimization [6,29,32]. In this section, we will introduce protein–ligand scoring functions including its classification, datasets, and evaluation (as shown in Figure 1).

### 2.1. Classification

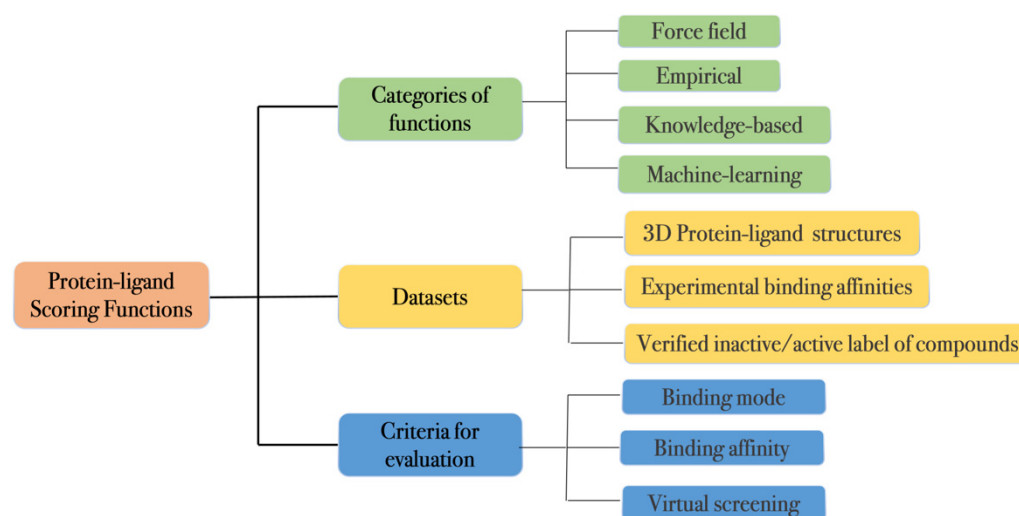
Scoring functions first emerged in the early 1990s and have inspired continuing research since then. Researchers have developed a variety of scoring functions formulated on different assumptions or algorithms [30,33,34]. These scoring functions can be roughly classified into four categories: (i) physics-based methods, (ii) knowledge-based statistical potentials, (iii) empirical scoring functions, and (iv) machine-learning scoring functions [35].

- (1) Physics-based scoring functions are centered on molecular mechanical calculations [20,21,36]. These scoring functions are often predicated on fundamental molecular physics terms such as Van der Waals interactions (Lennard-Jones potential), electrostatic interactions (Coulombic potential) and desolvation energies. These terms can be derived from both experimental data and *ab initio* quantum mechanical calculations. Due to the computational cost, solvation and entropy terms are usually oversimplified or ignored

in physics-based scoring functions. Programs such as GoldScore, DOCK and early versions of AutoDock use this type of scoring function [20,21,36].

- (2) Knowledge-based scoring functions consist of statistical potentials derived from experimentally determined protein–ligand structures. The frequency of specific interactions from many protein–ligand complexes are used to generate these potentials via the inverse Boltzmann distribution. This approach approximates complicated and difficult-to-characterize physical interactions using large numbers of the protein–ligand atom-pairwise terms. As a result, the scoring function lacks an immediate physical interpretation. DrugScore, ITScore and PMF are examples of knowledge-based scoring function [37–40].
- (3) Empirical scoring functions characterize the binding affinity of protein–ligand complexes based on a set of weighted scoring terms. These scoring terms may include descriptors for VDW, electrostatics, hydrogen bond, hydrophobic, desolvation, entropy, etc. The corresponding weights of the descriptors are determined by fitting experimental binding affinity data of protein–ligand complexes via linear regression. Empirical scoring functions draw from both physics-based and knowledge-based scoring functions. Empirical scoring functions use physically meaningful terms similarly to physics-based scoring functions. The contribution (weight) of each term is learned from the training data, similarly to knowledge-based scoring functions. Compared to knowledge-based scoring functions, empirical scoring functions are less prone to overfitting due to the constraints imposed by the physical terms. The scoring terms also provide insight into the individual contributions to the final binding affinity. Bohm pioneered the first empirical scoring function, LUDI, in 1994 [41,42]. Other famous empirical scoring functions, such as ChemScore, GlideScore, X-Score and Autodock Vina, were developed afterwards [21–23,43,44]. Autodock Vina is one of the widely used open-source docking programs, and its scoring function consists of five empirical interaction terms (two gaussian terms, a repulsion term, a hydrogen bond term, and a hydrophobic term) and a ligand torsion count term [43]. Recently, a linear empirical scoring function inspired from Vina scoring function, Lin\_F9, was developed to improve the scoring performance and overcome some of the limitations of Vina by introducing new empirical terms, such as the mid-range interactions and metal–ligand interactions. Trained on a small but high-quality protein–ligand dataset, Lin\_F9 achieved better scoring accuracy than Vina in binding affinity prediction [45].
- (4) Machine learning (ML) scoring functions are a group of methods that use ML techniques to learn the functional form of the binding affinity by associating patterns in the training data. Without employing a predetermined functional form, ML scoring functions can implicitly capture intermolecular interactions that are hard to model explicitly. ML scoring functions have shown marked improvements in binding affinity prediction in recent years [46,47]. In Section 3, we will discuss ML scoring functions in detail.

The first three types (1–3) can be grouped as classical scoring functions. These scoring functions usually adopt a linear form, which is a linear combination of several force-field or interaction descriptors. On the other hand, ML scoring functions can adapt much more complicated functional forms by utilizing ML methods, such as Support Vector Machine (SVM) [48], Random Forest (RF) [49], eXtreme Gradient Boosting (XGB) [50], Deep Neural Network (DNN), Convolutional Neural Network (CNN) and Graph Neural Network (GNN) [51,52].



**Figure 1.** Schematics of the categories and datasets and evaluations of the protein–ligand scoring functions.

## 2.2. Datasets

A representative dataset is the most substantial part of protein–ligand scoring function development and is crucial for the evaluation of scoring functions. Here, we introduce some widely used datasets:

- (1) Datasets that consist of 3D protein–ligand structures with experimentally measured binding affinities are typically used to evaluate methods in binding pose identification and binding affinity prediction [53–55]. One example, PDBbind, provides 3D protein–ligand structures with experimentally measured binding affinity data manually collected from their original references. PDBbind is currently one of the largest datasets of protein–ligand structures for the development and validation of docking methodologies and scoring functions. The current release (version 2020) of PDBbind general set contains 19,443 protein–ligand complexes with binding affinity data ( $K_d$ ,  $K_i$  or  $IC_{50}$ ) ranging from 1.2 pM to 10 mM, and is annually updated to keep up with the growth of Protein Data Bank (PDB) [56,57]. PDBbind also contains a refined subset of high-quality data according to several criteria concerning the quality of the structures and the affinity data. In addition, PDBbind provides a benchmarking “core set” used for the comparative assessment of scoring functions (CASF) [33,34], which will be discussed in Section 2.3 in detail. Similar datasets, such as the Community Structure-Activity Resource (CSAR) [58–63] exercises and the D3R Grand Challenge [64–67], are mainly curated to validate SBDD.
- (2) Datasets that label active/inactive compounds to protein structure or sequence targets are generally used to develop and evaluate methods in VS tasks, such as early hit enrichment and active/inactive classification [68]. The Database of Useful Decoys (DUD) [69] and Database of Useful Decoys-Enhanced (DUD-E) [70] have been widely used for benchmarking. DUD-E consists of 102 targets with 22,886 active compounds with binding affinities. For each active compound, DUD-E also includes 50 computer-generated decoy compounds, which have similar physiochemical properties but dissimilar two-dimensional topology to the active compound. Many decoys are presumed, without experimental verification, to be inactive compounds. This remains a major drawback for DUD-E dataset because false negative samples might exist in the dataset.

Maximum Unbiased Validation (MUV) database is constructed based on PubChem bioactivity data from 17 targets, each with 30 actives and 15,000 inactives, and is designed to avoid analog bias and artificial enrichments [71]. Unlike DUD and DUD-E, MUV provides experimentally verified inactive compounds mostly tested with cell-based assays. This raises questions into the suitability of using MUV as a structure-based VS benchmark

because many actives are not validated against their putative targets. Thus, MUV is more appropriate for benchmarking ligand-based VS approaches.

In 2020, Tran-Nguyen and co-workers have curated LIT-PCBA [68], a dataset derived from dose–response assays in the PubChem database [72–74]. LIT-PCBA consists of 15 targets, and for each target, all the actives and inactives were taken from the experimental data under homogeneous conditions. One main advantage of LIT-PCBA over prior efforts is the careful removal of potential false-positive results (dose–response curve of each active should have  $0.5 < \text{Hill slope} < 2.0$ ). However, the main limitation of LIT-PCBA dataset is that more than half of the primary assays (8 of 15 targets) are cell-based phenotypic assays. Thus, structure-based VS tests on this benchmark also have some limitations.

- (3) Other datasets contain a large variety of compounds with binding affinity data but contain few or lack annotated protein–ligand structures, such as the Binding Database (BindingDB) and ChEMBL. These are used in developing ligand-based or sequence-based approaches to predict binding affinities and can supplement protein–ligand scoring function development and validation [75–82]. As of 4 May 2022, BindingDB contains 2,513,948 binding data for 8839 protein targets and 1,077,922 small molecules. ChEMBL is a manually curated database of bioactive molecules with drug-like properties. The current release (version 30) contains 19,286,751 activities for 14,885 targets and 2,157,379 compounds.

### 2.3. Evaluation Metrics

Several metrics are commonly used to evaluate the performance of a scoring function in binding pose identification, binding affinity prediction, and VS tasks.

- (1) The goal of binding pose identification is to determine the native binding pose among computer-generated decoys. Given a set of decoys, a reliable scoring function should be able to rank the native binding pose to the top by their binding scores. The root-mean-square deviation (RMSD) between the top docking pose and the experimentally determined ligand pose is a commonly used evaluation metric. If the RMSD is  $\leq 2 \text{ \AA}$ , the binding pose prediction is considered successful. Due to its simplicity and ease of implementation, the RMSD metric for binding pose prediction has been widely used in the field [33,34,64–67,83]. It should be noted that the minimum symmetry-corrected RMSD should be calculated for small molecules with symmetric functional groups or whole-molecule symmetry [84–87].
- (2) Binding affinity prediction aims to predict the binding affinity for a given protein–ligand complex. Nevertheless, some scoring functions give a score that cannot be directly compared to experimental binding data [20,88]. Thus, a widely used criterion for affinity prediction is the Pearson correlation coefficient between the predicted scores and the experimental binding data on benchmark test sets [33,34]. Since the correlation between the predicted scores and experimental binding data does not have to be linear, an alternative criterion is the Spearman ranking correlation coefficient. This first ranks the predicted and experimental scores and then calculates the correlation between the two ranking sets [89].
- (3) VS aims to identify true actives in a compound library. The screening performance typically estimates whether a scoring function is able to rank the known binders above many inactive compounds in the library. There are several evaluation metrics, including enrichment factor (EF), area under the curve (AUC), and receiver operating characteristics (ROC) curve, to quantify the screening performance of a scoring function [90]. EF is defined as the accumulated rate of true binders found above a certain percentile of the ranked database that includes both the actives and inactives. A higher EF at a fixed percentage of ranked database indicates better early hit enrichment (a higher likelihood to select actives based on predicted scores). EF is computed as follows:

$$EF_{\alpha} = \frac{NTB_{\alpha}}{NTB_{total} \cdot \alpha}, \quad (1)$$

where  $NTB_{\alpha}$  is the number of true binders among top  $\alpha$  percentile of ranked candidates (e.g.,  $\alpha = 1\%$ ,  $5\%$ ,  $10\%$ ) based on predicted binding scores and  $NTB_{total}$  is the total number of true binders in the database. AUC-ROC is an evaluation method for classifiers assessing true binder identification. ROC plots the false positive rates (FPR, also called specificity) and true positive rates (TPR, also called recall or sensitivity) into a curve whose area under the curve (AUC) ranges from 0 to 1, where 0.5 reflects a random-level selection and 1 for perfect selection. This method is more appropriate when the number of inactive compounds is comparable to the number of active compounds.

The comparative assessment of scoring functions (CASF) benchmark is one of the most widely used retrospective benchmark for evaluation of scoring functions [33,34]. The current version (CASF-2016) is derived from PDBbind refined dataset, consisting of 57 targets with 5 crystal protein–ligand complexes for each target (5 different co-crystallized ligands from high affinity to low affinity in order) [54]. Scoring functions are evaluated on four different metrics (scoring, ranking, docking, and screening). The scoring metric calculates the Pearson correlation coefficient between predicted binding scores and experimentally measured binding affinities. The ranking metric measures the average Spearman's rank correlation coefficient after ranking all 57 targets by their predicted binding score. This quantifies the ability of a scoring function to correctly rank the known ligands of a certain target protein. The docking metric calculates the rate which predicted poses are found to be similar to the crystal pose ( $RMSD \leq 2 \text{ \AA}$ ). This establishes the ability of a scoring function to find the native pose among computer-generated decoys. The screening metric incorporates two indicators to measure the ability of a scoring function to find true binders within the top 1%, 5% and 10% ranked ligands for each target. The first indicator is the success rate in identifying the average highest-affinity binder over all the targets and the second indicator is the average EF over all the targets. The screening power assessment of CASF benchmark is limited due to the small size of the database (only 285 compounds in total) and the lack of verification of inactive compounds for targets. Therefore, a large dataset with many confirmed inactive compounds, such as LIT-PCBA benchmark [68], should be more suitable to evaluate the early hit enrichment performance of a scoring function.

Annually over the years 2015–2019, the Drug Design Data Resource (D3R) has provided blind and open competitions for pose prediction and affinity ranking to evaluate participants' computational methods [64–67]. In the last round (D3R Grand Challenge 4) [67], D3R has organized a multiple-stage competition for pose prediction (stage 1a and stage 1b) and affinity ranking (stage 2) for macrocyclic small molecule inhibitors targeting beta secretase 1 (BACE1), as well as a one-stage affinity ranking competition for Cathepsin S (CatS) inhibitors. Mean/median RMSD values are used to evaluate the participant's prediction performance. Spearman and Kendall ranking correlation coefficients are used to assess the affinity prediction.

In early 2022, Critical Assessment of Computational Hit-finding Experiments (CACHE) [91], a prospective benchmarking project to evaluate and improve VS methods in real screening, has been launched and is open to public. The CACHE aims to organize multiple rounds of challenges to provide opportunities for scientists to improve and test their VS methods in next several years. These prospective benchmarks would enable the improvement of computational methods to handle novel targets in the future.

### 3. Machine-Learning Scoring Function

ML is a branch of artificial intelligence that has gained attention in diverse research fields, including CADD. With the rapid progress of computational power and exponential increase of data, ML has been applied in many branches of CADD, such as chemical space exploration, molecular property prediction, protein structure prediction and VS [92–97]. ML algorithms have also been widely employed in SBDD, such as pose prediction, binder/nonbinder identification and binding affinity prediction [46,96]. This review will focus on the discussion of ML scoring functions, a supervised learning method that

learns from structure data labeled with experimental measured binding affinities. Early efforts used traditional ML methods, such as SVM, RF and GBT, to improve scoring performance on benchmark test sets. Their inputs were manually designed descriptors, such as molecular interaction fingerprints, ligand features, atom-pairwise terms and force field terms [46]. To date, many DL scoring functions have also been developed. However, they do not always significantly outperform traditional ML scoring functions [98]. In the following part, we will describe some ML scoring functions, as shown in Table 1.

**Table 1.** Machine learning scoring functions.

ML Algorithm	Name	Input Features	Dataset	Year
RF	RF-score [99]	Protein–ligand atom-type pair counts in predefined distance cutoff	PDBbind v2007	2010
	SFCscore <sup>RF</sup> [100]	Descriptors of ligand-dependent, specific interactions, surface area	PDBbind v2007	2013
	$\Delta_{\text{Vina}}\text{RF}_{20}$ [101]	Vina empirical terms, surface area terms	PDBbind v2014 CSAR dataset	2017
XGB	$\Delta_{\text{Vina}}\text{XGB}$ [102]	Vina empirical terms, surface area terms, ligand stability terms, bridge water terms A series of gauss terms characterizing protein–ligand interactions, surface area terms,	PDBbind v2016 CSAR dataset	2019
	$\Delta_{\text{LinF9}}\text{XGB}$ [103]	ligand descriptors, bridge water terms and pocket features	PDBbind CSAR dataset BindingDB	2022
ERT	ET-score [104]	Distance-weighted interatomic contacts between protein and ligand	PDBbind v2016	2021
GBT	AGL-Score [105]	Algebraic graph theory-based features of protein–ligand complex	PDBbind	2019
	ECIF-GBT [106]	Protein–ligand atom-type pair counts considering each atom connectivity	PDBbind v2016	2021
NN	NNScore 1.0 [107]	Descriptors of specific interactions and ligand-dependent	MOAD PDBbind	2010
	NNScore 2.0 [108]	Vina empirical terms, protein–ligand atom-type pair counts in predefined distance cutoff	MOAD PDBbind	2011
CNN	AtomNet [109]	Local structure-based 3D grid from protein–ligand structures	DUD-E	2017
	Pafnucy [110]	Atom property-based 3D grid from protein–ligand structures	PDBbind v2016	2017
	Kdeep [111]	Atom type-based 3D grid from protein–ligand structures	PDBbind v2016	2018
	OnionNet [112]	Rotation-free element-pair specific contacts between protein and ligand atoms in different distance ranges	PDBbind v2016	2019
GNN	PotentialNet [113]	Atom node feature and distance matrix	PDBbind v2007	2018
	graphDelta [114]	Atom node features considering local environment and distance matrix	PDBbind v2018	2020
	SIGN [115]	Distance matrix of atom nodes and angle matrix of bond edges	PDBbind v2016	2021

RF-score, the first ML scoring function to outperform classical scoring functions on scoring tasks, was proposed by Ballester and Mitchell in 2010 [99]. It utilized the random forest algorithm with feature selection comprised of protein–ligand atom-wise pair counts at a predefined distance cutoff. In 2013, Zilian and co-workers proposed SFCscore<sup>RF</sup> [100], which also utilized the random forest algorithm, but with feature selection of 63 empirical terms comprised of ligand-dependent descriptors (such as number of rotatable bonds), specific interactions descriptors (such as hydrogen bonds and aromatic interactions) and

surface characteristics (such as polar and hydrophobic contact surfaces). The scoring performance of SFCscore<sup>RF</sup> on two common benchmarks (CASF 2013 and CSAR-NRC HiQ) also significantly outperformed classical scoring functions. However, both RF-score and SFCscore<sup>RF</sup> performed much worse on docking and screening tasks compared to classical scoring functions [116,117].

To address this issue, in 2017, the Zhang group employed a  $\Delta$ -machine learning approach, in which a ML model was employed to parametrize corrections to the Vina score [101]. This strategy enabled the scoring function to have both the excellent docking power of Vina and the accurate scoring performance of the ML method. In their work,  $\Delta_{\text{vina}}\text{RF}_{20}$  was developed using random forest with feature selection comprised of 10 features related to pharmacophore-based solvent-accessible surface area (SASA) and 10 empirical terms from Vina 58 features. As a result,  $\Delta_{\text{vina}}\text{RF}_{20}$  achieved the best performance among a panel of classical scoring functions in all evaluation metrics (scoring, ranking, docking and screening powers) for the CASF 2007 and 2013 benchmarks. In 2019, the same group proposed a subsequent scoring function,  $\Delta_{\text{vina}}\text{XGB}$  [102], that considered explicit water molecules as well as ligand conformation stability and substituted the random forest method with eXtreme Gradient Boosting (XGB). The feature set of  $\Delta_{\text{vina}}\text{XGB}$  consisted of Vina 58 features, 30 SASA features, 3 bridge water features, 2 ligand stability features and 1 metal count term. The training data was enlarged to include both explicitly solvated protein–ligand structures and dry protein–ligand structures, as well as docking decoys. This  $\Delta_{\text{vina}}\text{XGB}$  consistently achieved better performances in scoring, ranking, docking and screening tasks on the CASF-2016 benchmark [102]. In 2022, a newly developed delta ML scoring function,  $\Delta_{\text{LinF9}}\text{XGB}$  [103], used a series of gaussian terms to characterize protein–ligand interactions in different distance ranges and further enlarged the training set to include more weak binders and docking poses.  $\Delta_{\text{LinF9}}\text{XGB}$  achieved superior scoring, ranking and screening performances on the CASF-2016 benchmark. In addition, Nguyen and Wei proposed an algebraic graph theory-based scoring function, AGL-Score [105], which achieved superior scoring, ranking, docking and screening performance on the CASF-2013 benchmark. This method was a gradient boosting trees (GBT) model integrating weighted algebraic subgraph features of protein–ligand complexes.

Recently, customized protein–ligand interaction features became popular in scoring function development, such as ET-score (2021) and ECIF-GBT (2021) [104,106]. ET-score employed protein–ligand interaction features defined by distance-weighted interatomic contacts between atom type pairs of the protein and ligand. ET-Score achieved very good scoring performance (with Pearson R = 0.827) on CASF-2016 benchmark. ECIF-GBT used extended connectivity interaction features (ECIF), which are a set of protein–ligand atom type pair counts that consider each atom's connectivity to define the pairwise types. ECIF-GBT achieved Pearson R of 0.857 on CASF-2016 benchmark. However, both ET-score and ECIF-GBT were trained solely on crystal structures and their performances on docking and screening tasks remain an issue.

Besides the above introduced traditional ML scoring functions, DL models were also applied in protein–ligand scoring function development. Durrant and McCammon proposed two models using neural networks, NNScore 1.0 and NNScore 2.0 [107,108]. NNScore 1.0 employed a simple neural network composed of only one hidden layer with five neurons to classify the active and inactive compounds based on 194 features including both interaction and ligand-dependent terms. Comparatively, NNScore 2.0 included many more interaction terms and estimated the  $\text{pK}_d$  rather than an active/inactive classification. In 2017, Wallach and co-workers introduced AtomNet [109], the first CNN-based scoring function incorporating 3D structural information. The inputs of AtomNet used vectorized 3D grids placed over the protein–ligand interaction interface, with each grid cell storing a value describing the presence of some basic structural features, varying from a simple enumeration of atom types to more complex descriptors. Its network topology was made up of an input layer, followed by four 3D convolutional layers and two fully connected layers, and topped by a logistic-cost layer that assigns probabilities over the active and



inactive classes. AtomNet achieved much better AUC value than Vina (classical scoring function) on the DUD-E test set. Several similar CNN-based scoring functions, such as the CNN model proposed by Ragoza et al. (2017) [118], Pafnucy proposed by Stepniewska-Dziubinska et al. (2017) [110], and Kdeep developed by Jimenez et al. (2018) [111], were published afterward. One of the limitations of these CNN models was the dependency on the coordinate frame. Different orientations of the same protein–ligand structure could generate different representations. In order to address this issue, Zheng and co-workers introduced OnionNet in 2019 [112]. This method used a CNN model with inputs based on rotation-free element-pair specific contacts between protein and ligand in different shells. OnionNet, as well as the subsequent OnionNet-2.0 (2021) [119], achieved excellent scoring performance on the CASF-2016 benchmark.

Deep graph neural network (GNN) methods also became popular in protein–ligand scoring function development. In 2018, Feinberg and co-workers introduced PotentialNet [113], which used a graph convolutional neural network (GCNN) to directly learn protein–ligand structures in terms of both intramolecular and intermolecular interactions. This approach consisted of three major steps to achieve feature learning, including covalent-only propagation, dual noncovalent and covalent propagation, and ligand-based graph gathering. The aggregation of updated ligand atomic vectors was used to predict binding affinity. PotentialNet achieved Pearson R of 0.822 on CASF-2007 benchmark. In 2019, Lim and co-workers proposed a GNN model with distance-aware attention mechanism to differentiate the contribution of each interaction to binding affinity [120]. Their GNN model was also designed to focus on intermolecular interactions rather than memorize certain patterns of ligand molecules. As a result, this GNN model achieved very good performance in terms of both VS and pose prediction. Recently, several novel GNN models, such as graphDelta (2020) [114], graphBAR (2021) and SIGN (2021) [115,121], were published, however, these GNN models did not outperform some of above-mentioned traditional ML scoring functions on the CASF benchmarks.

All the above discussed ML scoring functions were generic scoring functions which aim to perform well on all kinds of target proteins. However, this aim could be hard to achieve due to the particulars of each target. Recently, target-specific ML scoring functions were proposed to focus on a certain target [122–124]. These functions learned from training data of a certain target or family to deal with the special characteristics of the target. A target-specific approach could achieve state-of-the-art performance on well-studied targets with sufficient training data, but it would not be applicable for a novel target with little experimental data available.

#### 4. Structure-Based Virtual Screening

VS is a computational approach used to identify chemical structures that are predicted to have particular properties. In drug discovery, it involves computationally searching large libraries of chemical structures to identify those structures that are most likely to bind to a target protein. Structure-based VS, also known as target-based VS, attempts to predict the best interaction of a ligand against a target protein to form a complex and employs scoring functions to estimate the binding affinity of the protein–ligand complex [125]. As a result, all the ligands are ranked according to their binding scores to the target, and the high scoring ligands are selected for experimental measurement. In recent decades, advances in VS have been made in the following:

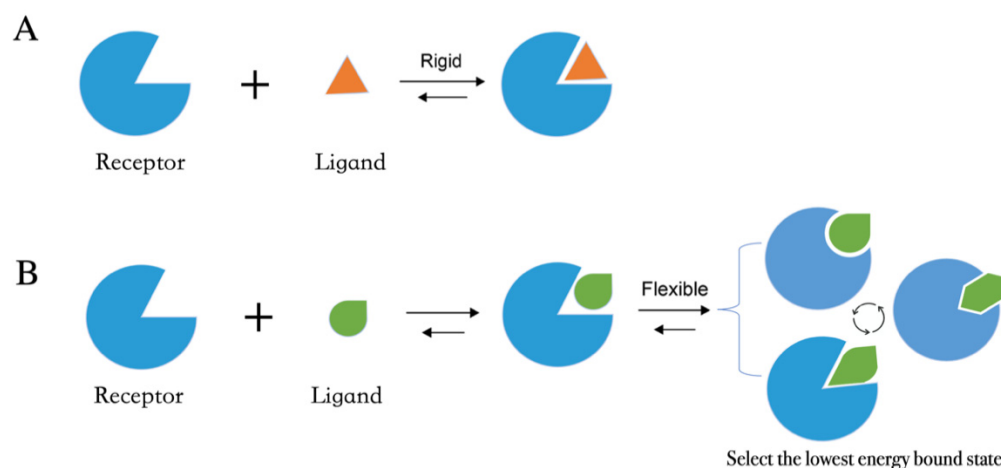
- (1) There have been developments in structure-based VS approaches, including improvements in sampling and scoring methods, that have resulted in significant improvements in docking, scoring and screening performances [46].
- (2) Developments in GPU processing speeds and cloud computing have dramatically increased computational power. Researchers are now able to computationally process vast numbers of compounds in the drug-like chemical space.

- (3) Advancements in structural biology (such as X-ray, NMR and cryo-EM) and computational protein structure prediction (such as AlphaFold2 and RoseTTAFold) [95,126–128] have allowed access to many more 3D structures.
- (4) The number of compounds that are commercially available or can be readily synthesized has grown dramatically in recent years. For example, as of March 2021, the WuXi GalaXi and Enamine REAL Space collections contain 2.1 billion and 17 billion compounds, respectively [129]. In June 2022, the WuXi GalaXi and Enamine REAL Space collections have grown up to 4.4 billion and 22.7 billion compounds, respectively.

The convergence of these breakthroughs has positioned structure-based VS to be a promising direction for the discovery of novel small molecule medicine. With the appropriate computing infrastructure, it becomes practical to virtually screen ultra-large compound library (synthesized or purchasable) to find virtual hit compounds, some of which (usually up to 100 compounds) can be experimentally tested.

#### 4.1. Molecular Docking Protocol

Molecular docking methods predict receptor–ligand interactions at an atomic level and are widely utilized in structure-based VS. The docking process samples the optimal conformation based on the complementarity between the receptor and the ligand. Figure 2A shows the initially proposed “lock-and-key model”, which refers to the rigid docking of receptor and ligand to find the correct orientation for the “key” to open the “lock”. This model emphasizes the importance of geometric complementarity [18]. However, the real binding process is very flexible whereby the receptor and ligand changes their conformation to complement each other well. As shown in Figure 2B, the induced fit model considers structural flexibility and selects the lowest-energy bound state. Currently, major limitations of docking methods include a restricted sampling of both ligand and receptor conformations in pose prediction, as well as the previously discussed limited accuracy of scoring functions in affinity prediction.



**Figure 2.** Two models of molecular docking. (A) A lock-and-key model. (B) Induced fit model.

The methods that improve sampling of ligand conformations can be defined as (i) incremental ligand construction, (ii) multiple conformers generation for docking and (iii) stochastic sampling [130]. In the first approach, the ligands are partitioned into small fragments that are individually docked into the receptor pocket according to the geometric fit. Docked fragments are then incrementally assembled to form an entire ligand within the binding pocket [131]. In the second approach, multiple low-energy conformations of the ligand are generated at first, and then individually docked against the receptor pocket [132].

The third and widely used strategy to account for ligand flexibility are stochastic methods, such as Monte Carlo (MC) or genetic algorithm (GA). MC algorithm, also known as simulated annealing, simulates docking by randomly generating minor changes in

the position, orientation or conformation to generate new poses that are accepted or rejected based on the Metropolis acceptance algorithm [133]. The modeling begins at a high temperature such that there is a high probability of accepting the next conformation sampled. Then, the temperature is progressively decreased to reduce the conformational freedom of the system and to capture the receptor–ligand complex in a low energy state. GA employs a different approach inspired by Darwin’s theory of evolution [134]. The ligand begins as a random population of position, orientation and conformational states modeled as a set of chromosomes. Then, random crossovers and mutations are performed to produce another set of conformations. The conformation with the lowest binding energies with the receptor is accepted and then used to produce a new generation. This cycle is iteratively repeated until the local energy minimum of the receptor–ligand complex has been reached.

Many proteins possess varying degrees of flexibility, which can range from a slight perturbation of the ligand binding pocket to a complete reconstitution of the pocket. Therefore, an inadequate sampling of protein flexibility can result in an increase of both false positives and false negatives in VS experiments. Several approaches have been developed to tackle the issue of protein flexibility in recent years [135]. One common approach, named “ensemble docking”, is to utilize multiple receptor conformations in docking runs and to select the best-scoring conformation for further investigation [136–138]. The receptor conformations are commonly obtained from different X-ray and NMR structures or by sampling structures from molecular dynamics (MD) simulations. For instance, Abagyan and co-workers have investigated strategies for the selection of experimental protein conformations for VS and have found that the use of ensemble conformations of receptors co-crystallized with larger ligands provided the best results [139,140]. However, it has been noted that the use of excessively large numbers of receptor conformers in ensemble docking can lead to an increased number of false positive samples and linearly increased computational costs [135,141]. To alleviate some of these performance issues, ML techniques can be employed to help classify active and inactive compounds following ensemble docking [142]. Chandak and co-workers have tested multiple supervised ML methods trained on the DUD-E database to learn the relationship of a compound’s predicted binding affinities to the classification task.

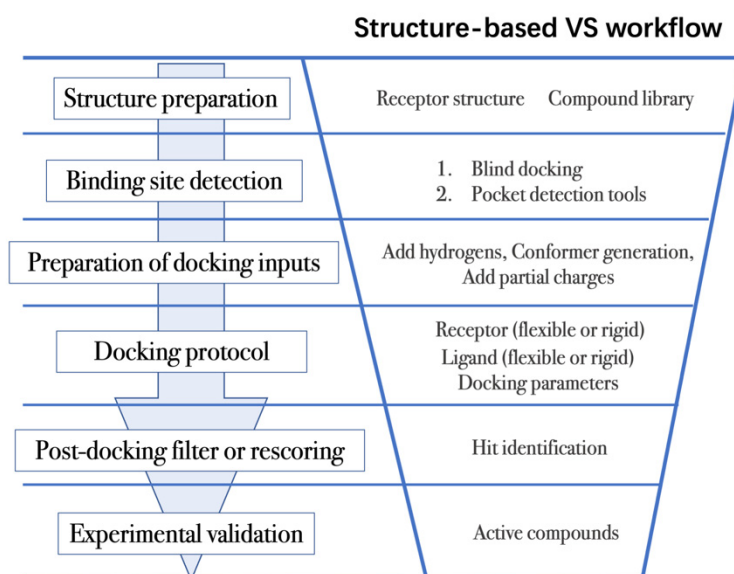
An alternative approach to account for protein flexibility is to employ “soft docking”, where the interactions between the protein amino acid sidechains and the ligand is iteratively changed to allow partial clashing between the atoms of the protein and ligand [143]. For example, Ravindranath and co-workers have proposed a soft docking program, AutoDockFR [144], which simulates sidechain flexibility by sampling a large number of explicitly specified receptor sidechains and searching for energetically favorable binding poses for a given ligand. AutoDockFR optimizes protein–ligand interactions using the AutoDock4 force field and using a GA method combined with a Solis-Wets local search. This soft docking approach has achieved better binding pose prediction compared to rigid protein docking protocols but has also been associated with an increased number of false positive hits in structure-based VS [145].

#### 4.2. Workflow in Virtual Screening

Structure-based VS relies on docking of large collections of compounds into the binding pocket of target protein, and then evaluating whether the protein–ligand contacts will drive binding. As shown in Figure 3, the general VS workflow can be as follows:

- (1) The first step is to obtain the 3D structures of a given target as well as the compound library. Experimental determined structures can be readily retrieved from the Protein Data Bank (PDB) [146], in which more than 120,000 unique protein structures have been determined through an enormous experimental effort. However, this represents a small fraction of the billions of known protein sequences whereby the 3D structure of a novel target is usually not available. In order to overcome this limitation, traditional computational prediction methods (such as homolog modelling and ab initio modelling) [147,148], as well as the recently developed DL methods (such as

AlphaFold2 and RoseTTAFold) [95,127] can be employed to obtain the 3D structures of target proteins. In addition, the compound library or chemical space used in VS is also vital for hit identification.



**Figure 3.** General scheme of a VS workflow.

As discussed above there is a growing number of options to dock to. It is important to note that the selection of which structure to dock to is not trivial. Docking results will differ depending on the conformation, apo/holo status, and quality of structure. One method, screening performance index, can be used to select good structures to use in prospective VS [149]. This index consists of five calculated terms that describe the docking performance of a set of structures on a set of known active compounds. Their testing has generally indicated that co-crystal structures with large ligands bound score well on the index and can be picked for prospective studies. These methods are limited because they require labeled datasets which may not be available for novel targets.

Compound libraries of approved drugs, natural products, already synthesized or purchasable compounds/fragments are commonly used in VS campaigns [29,130,150]. The well-known ZINC database contains over 750 million purchasable compounds, including over 230 million compounds in ready-to-dock 3D formats [151,152]. Recently, Jiankun and coworkers performed docking-based VS using a ultra-large compound library (more than 100 million compounds from ZINC make-on-demand compounds) to discover inhibitors targeting AmpC  $\beta$ -lactamase and D<sub>4</sub> dopamine receptor [29]. Other databases, such as DrugBank [153–155] and Human Metabolome Database (HMDB) [156–159] are used to repurpose the approved drugs or human metabolites to the novel targets.

(2) The next step is to detect the binding site. Typically, the binding pocket on which to focus the docking calculations is known. For example, the binding site is chosen based on the information of co-crystallized ligand/substrate binding site, such as ATP binding site or protein–protein interactions (PPI) interface. However, when the binding site information is missing or a novel binding pocket needs to be explored, there are two commonly employed approaches, “blind docking” simulation [160,161] and pocket prediction algorithms. The first approach uses docking methods to search over the entire target structure to find a favorable ligand binding site, but it has a high computational cost in sampling. For the second approach, several available software can be employed to detect binding pockets, including AlphaSpace [162,163], FTMap [164], MDpocket [165], Fpocket [166], SiteMap [167] etc. These methods detect concave pockets on the protein surface by characterizing the spatial composition of amino acids or using the chemical probe to find favorable hot spots. Since drug

resistance can arise for the orthosteric site of target proteins, these methods can be used to identify additional binding pockets that can be exploited for the design of novel inhibitors, such as allosteric or cryptic pockets [168,169].

- (3) Once the binding site is determined it is important to carefully prepare docking input files to achieve successful VS. The preparation of protein structures starts from the assignment of protonation states for the amino acids, which can be done using software including PROPKA [170], H++ [171], and SPORES [172]. Then hydrogen atoms and partial charges are assigned. A popular software for this task is PDB2PQR [173,174]. In addition, the consideration of water molecules and metal ions can be crucial in certain target structures. Explicit water molecules mediating protein–ligand interactions should be analyzed and can be used to identify water-mediated interactions and avoid incorrect binding poses [175–177]. It is also important to consider coordination interactions between metal ions and ligand molecules for metalloprotein complexes [45,178].

Unlike proteins, most compounds used in VS are stored in line notation, such as Simplified Molecular Input Line Entry Specification (SMILES) string [179]. The 3D atomic coordinates of these compounds can be obtained from the line notation using several opensource softwares, such as RDKit and Openbabel [180–182], or commercial softwares, such as Omega and ConfGen [183–185]. Ligand protonation is also important since it affects the net charge of the molecule and the partial charges of individual atoms. Different docking programs will employ different charge assignment protocols. For example, AutoDock uses Gasteiger-Marsili atomic charges whereas the AutoDock Vina does not require the assignment of atomic charges, since the scoring terms that compose its scoring function are charge-independent [43,186].

- (4) After the input files are created, the appropriate docking protocol must be selected. As has been discussed in Molecular Docking Protocol (Section 4.1), there are many different docking protocols that consider protein and ligand flexibility to enhance the performance of pose prediction. One of the most commonly used protocols is to perform flexible ligand–rigid receptor docking for each docking run, and then dock multiple protein conformations using the ensemble docking strategy [139]. In addition, several docking programs can be combined to avoid the limitations of one algorithm. For instance, Ren and co-workers have explored the effects of using multiple softwares in the pose generation step [187]. They use a RMSD-based criterion to come up with representative poses derived from 3 to 11 different docking programs. The resulting pose prediction achieves better performance than that of each individual docking program.
- (5) Following docking, the results can be rescored or filtered. The computer-generated poses are evaluated based on the ability of the docking protocol to (i) select favorable binding poses for each ligand, and (ii) rank the ligand library to select high scoring hits for experimental measurement. Although the docking calculations are fast enough to process large compound libraries, they suffer from the inherent problem of calculating binding affinities from several simplified scoring terms. One remedy for improving the performance of VS is to employ more rigorous free energy calculations to postprocess docking poses. The main limiting factor in the application of free energy calculations to large chemical libraries is the high computational cost.

In recent years, post-docking filter methods have gained significant interest in drug discovery because they usually provide higher hit rates in VS with low additional computational cost and result in better correlation with experimental data in retrospective benchmarks. Several methods have been designed to eliminate false positive hits obtained from the initial docking experiments. Marcou and co-worker proposed the use of molecular interaction fingerprints (IFP), which are simple bit strings that convert the 3D information of protein–ligand interactions into a 1D vector representation, for the screening of CDK2 inhibitors [188]. The authors demonstrate that using post-docking filters that calculate the Tanimoto similarity of IFP between docked pose and co-crystal pose is more statistically

accurate compared to classical scoring functions in discriminating active compounds from inactive ones. They base this on the assumption that active compounds should have certain specific interactions or contacts with their target to display activity. Bertho and co-workers reported a similar post-docking filtering strategy, namely automatically analyzing poses using self-organizing map (AuPoseSOM) to examine the interatomic contacts between the ligand and the target [189]. This type of approach is target-specific and requires the co-crystal ligand pose as the reference. ML can also be applied to this task. Stafford and co-workers introduced AtomNet PoseRanker, a graph CNN trained on PDBbind v2019 to rerank putative co-crystal poses [149].

Another post-docking strategy is the rescoring of docked poses using a consensus model or an advanced ML scoring function. On one hand, the consensus model uses several different scoring functions to re-assess the docking poses generated from a single docking algorithm. Charifson and co-workers have proposed an approach that takes the intersection of the top-scoring molecules according to two or three different scoring functions. They found it provides a dramatic reduction in the number of false positives identified by individual scoring functions on case studies of p38, IMPDH and HIV protease [190]. On the other hand, advanced ML scoring functions developed in recent years, such as AtomNet [109], vScreenML [191],  $\Delta_{\text{Vina}}\text{RF}_{20}$  [101],  $\Delta_{\text{Vina}}\text{XGB}$  [102], SIEVE-Score [192] and RF-Score-VS [117], outperform classical scoring functions in screening performance comparisons on benchmark test sets. However, there is no guarantee that ML scoring functions can outperform classical scoring functions on novel targets that are largely different from the samples in the training data set [193].

The above (1) to (5) steps summarize the workflow of VS process. Other structure-based approaches, such as MD simulations, have also been widely utilized in combination with docking to improve VS performance. MD simulations are an efficient approach to discover cryptic binding pockets (in step 2, binding site detection) [169,194], to sample multiple receptor conformations in ensemble docking (in step 4, docking protocols) [136], and to evaluate the interactions of the predicted receptor–ligand complexes (in step 5, post-docking analysis) [195,196].

#### 4.3. Case Study

To illustrate one virtual screening approach, we describe the application of  $\Delta_{\text{Lin\_F9}}\text{XGB}$  VS protocol on the LIT-PCBA benchmark dataset [103]. The LIT-PCBA benchmark (discussed in more detail in Section 2.2), contains 15 diverse target proteins and the corresponding curated active/inactive compound library from the PubChem BioAssay database [68]. The target protein has one or several PDB structures, in which the co-crystal ligands are used to determine the docking box. The compound library contains SMILES strings of active and inactive compounds, which are processed with RDKit [181] to generate and protonate low energy 3D conformers for each ligand. Then, flexible ligand-rigid receptor docking is performed using the Smina program with Lin\_F9 scoring function. After docking, the top 5 docking poses were re-scored using  $\Delta_{\text{Lin\_F9}}\text{XGB}$ , and the best-rescored pose was used for VS assessment [103]. Figure 4 illustrates the general workflow of docking-based VS protocol on the LIT-PCBA benchmark.

Multiple groups have evaluated docking programs and protocols on the LIT-PCBA benchmark (Figure 5) [90,103,197,198]. Tran-Nguyen et al. report the best early enrichment across all 15 targets (average EF1% = 7.46) using the IFP post-docking filtering method. Another post-docking filtering method, Rescoring by Interaction Graph-Matching (GRIM) which compares protein–ligand interaction patterns between a docked and a reference (typically X-ray crystal structure) co-complex structure, also performs similarly well [198]. IFP and GRIM outperform classical and ML scoring function methods in this ranking task but are limited in that they are dependent on the selection of the reference structure and do not predict absolute binding free energy. The  $\Delta_{\text{Lin\_F9}}\text{XGB}$  ML scoring function method lead to the greatest number of targets with EF1% > 2 (13/15 targets) and great average early enrichment (average EF1% = 5.55). Overall, the  $\Delta_{\text{Lin\_F9}}\text{XGB}$  has the best performance among

methods that predict binding affinity. In comparison, Zhou et al. and Sunseri et al. report lower early enrichment for their template-based virtual screening methods (FINDSITE-comb2.0 and Fragsite) and their CNN model of GNINA, respectively [90,197]. It should be noted that this comparison of the enrichment results is slightly complicated by the dissimilarity in docking protocols. Sunseri et al. and Yang et al. reported different average EF1% using the Vina docking method likely due to differences in the number of ligand conformers generated, the docking box definition, and the number of PDB templates selected for docking [90,103]. Sunseri et al. used the GNINA software [90], Tran-Nguyen et al. used the Surflex-Dock software [198] and Yang et al. used the Smina software [103].

### Virtual screening on LIT-PCBA benchmark

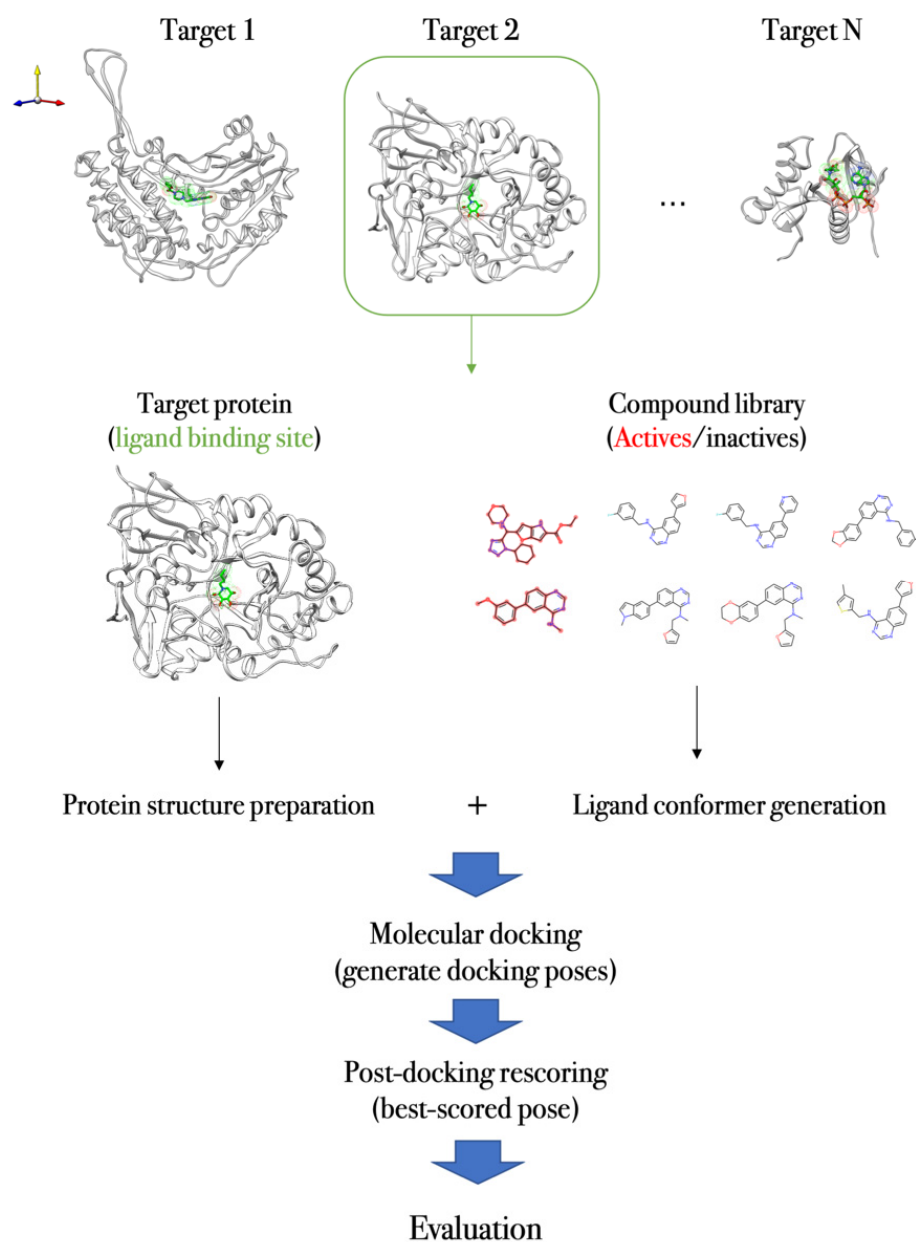
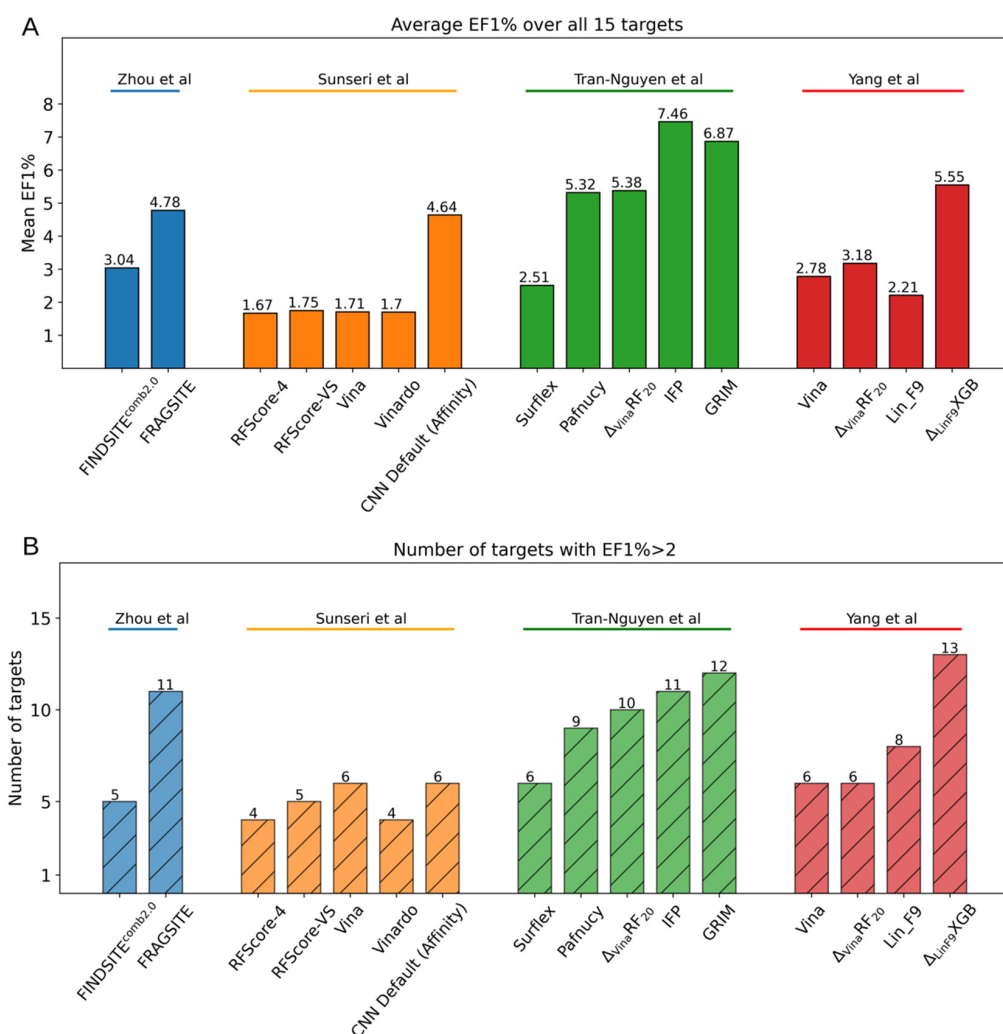


Figure 4. Workflow of docking-based VS protocol on LIT-PCBA benchmark.



**Figure 5.** Collected LIT-PCBA benchmark test results from four different groups (Zhou et al. [197], Sunseri et al. [90], Tran-Nguyen et al. [198] and Yang et al. [103]). (A) Average enrichment factor at top 1% (mean EF1%) is used to evaluate the early hit enrichment performance. (B) Counting number of targets that satisfy the thresholds of EF1% > 2 as a metric to assess the generalizability of the scoring functions on all 15 diverse targets.

## 5. Concluding Remarks and Perspectives

The current era is marked by advanced ML techniques, rapid growth of public data and increase in computing power. These developments in computational tools have advanced ML protein–ligand scoring functions for structure-based VS in early-stage drug discovery. Valuable benchmarks and competitions are developed to blindly evaluate these methods. Representative datasets that contain physiochemical data and guide the training of ML methods are proliferating. ML methods have taken advantage of the improvements in computing power and increase in datasets to outperform classical scoring functions. State of the art deep learning architectures applied in other fields are being successfully applied in drug discovery.

Despite these accomplishments, the applications of ML modeling in drug discovery, especially for deep learning, are still in the preliminary stage. Deep learning methods are commonly critiqued for being a “black-box”, easily over-trained, and lack interpretability. To fully appreciate the results, it is required that the user understands the advantages and limitations of a particular model architecture to associate the underlying molecular features to the prediction [199]. It would be valuable to incorporate informative terms and confidence indices to foster the user’s trust in the prediction and indicate starting



points for improvements. Furthermore, models are at the volition of high quantities of diverse, high-quality, and curated data. Not only does it require immense collaboration to develop these datasets, but also models may not be able to predict novel associations or characteristics that are not represented in the dataset. Therefore, more attention needs to be given in coupling these technological advances with scientific insights.

The evaluation of these methods and integration of these VS methods in a systematic workflow for prospective study is an active field of research. In recent years, some docking programs have been successfully embedded in automated workflows for ultra-large compound library screening [3,200,201]. However, the selection of promising virtual hits (usually less 100 compounds) from many high scoring compounds in the library remains a challenge, since different selection protocols usually lead to different false-positive rates and mixed hit identification results. We anticipate that future work could try to address these practical problems and limitations in prospective VS studies.

Lastly, scoring functions and SBDD protocols can become more practical and informative as techniques improve. The selection of docking structure and binding site should be done in a systematic manner and consider the functional roles of the particular conformation and binding site. Further investigations of specialized scoring functions for other drug technologies, such as PROTACs, macrocycles, covalent inhibitors, antibodies, allosteric inhibitors and drug combinations, are needed. The scope of structure-based docking protocols can be expanded to predict the toxicity and the cellular responses of the compound. It would be valuable to define protocols that would correlate docking of a particular binding site to the perturbation of the molecular pathways or activity. We expect that ML will play a pivotal role in these areas and continue to influence drug discovery research.

**Author Contributions:** All authors contributed to write the review. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the U.S. National Institutes of Health, grant number R35-GM127040.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Kawai, K.; Nagata, N.; Takahashi, Y. De novo design of drug-like molecules by a fragment-based molecular evolutionary approach. *J. Chem. Inf. Model.* **2014**, *54*, 49–56. [[CrossRef](#)] [[PubMed](#)]
2. Lionta, E.; Spyrou, G.; K Vassilatis, D.; Cournia, Z. Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Curr. Top. Med. Chem.* **2014**, *14*, 1923–1938. [[CrossRef](#)] [[PubMed](#)]
3. Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.-F.; Fischer, P.D.; Coote, P.W.; Padmanabha Das, K.M.; Malets, Y.S.; Radchenko, D.S.; Moroz, Y.S.; Scott, D.A. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **2020**, *580*, 663–668. [[CrossRef](#)]
4. Stumpfe, D.; Bajorath, J. Current trends, overlooked issues, and unmet challenges in virtual screening. *J. Chem. Inf. Model.* **2020**, *60*, 4112–4115. [[CrossRef](#)] [[PubMed](#)]
5. Talele, T.T.; Khedkar, S.A.; Rigby, A.C. Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* **2010**, *10*, 127–141. [[CrossRef](#)] [[PubMed](#)]
6. Stein, R.M.; Kang, H.J.; McCorvy, J.D.; Glatfelter, G.C.; Jones, A.J.; Che, T.; Slocum, S.; Huang, X.-P.; Savych, O.; Moroz, Y.S. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **2020**, *579*, 609–614. [[CrossRef](#)]
7. Hartman, G.D.; Egbertson, M.S.; Halczenko, W.; Laswell, W.L.; Duggan, M.E.; Smith, R.L.; Naylor, A.M.; Manno, P.D.; Lynch, R.J. Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J. Med. Chem.* **1992**, *35*, 4640–4642. [[CrossRef](#)]
8. Greer, J.; Erickson, J.W.; Baldwin, J.J.; Varney, M.D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **1994**, *37*, 1035–1054. [[CrossRef](#)]
9. Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284. [[CrossRef](#)]
10. Van Drie, J.H. Computer-aided drug design: The next 20 years. *J. Comput. Aided Mol. Des.* **2007**, *21*, 591–601. [[CrossRef](#)]
11. Abdolmaleki, A.; B Ghasemi, J.; Ghasemi, F. Computer aided drug design for multi-target drug design: SAR/QSAR, molecular docking and pharmacophore methods. *Curr. Drug Targets* **2017**, *18*, 556–575. [[CrossRef](#)] [[PubMed](#)]

12. Acharya, C.; Coop, A.; E Polli, J.; D MacKerell, A. Recent advances in ligand-based drug design: Relevance and utility of the conformationally sampled pharmacophore approach. *Curr. Comput. Aided Drug Des.* **2011**, *7*, 10–22. [[CrossRef](#)] [[PubMed](#)]
13. 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](#)] [[PubMed](#)]
14. De Ruyck, J.; Brysbaert, G.; Blosssey, R.; Lensink, M.F. Molecular docking as a popular tool in drug design, an in silico travel. *Adv. Appl. Bioinform. Chem.* **2016**, *9*, 1. [[CrossRef](#)]
15. Lavecchia, A.; Di Giovanni, C. Virtual screening strategies in drug discovery: A critical review. *Curr. Med. Chem.* **2013**, *20*, 2839–2860. [[CrossRef](#)]
16. Torres, P.H.; Sodero, A.C.; Jofily, P.; Silva-Jr, F.P. Key topics in molecular docking for drug design. *Int. J. Mol. Sci.* **2019**, *20*, 4574. [[CrossRef](#)]
17. Fan, J.; Fu, A.; Zhang, L. Progress in molecular docking. *Quant. Biol.* **2019**, *7*, 83–89. [[CrossRef](#)]
18. Koshland Jr, D.E. The key–lock theory and the induced fit theory. *Angew. Chem. Int. Ed. Engl.* **1995**, *33*, 2375–2378. [[CrossRef](#)]
19. Miteva, M.A.; Lee, W.H.; Montes, M.O.; Villoutreix, B.O. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *J. Med. Chem.* **2005**, *48*, 6012–6022. [[CrossRef](#)]
20. Allen, W.J.; Balias, T.E.; Mukherjee, S.; Brozell, S.R.; Moustakas, D.T.; Lang, P.T.; Case, D.A.; Kuntz, I.D.; Rizzo, R.C. DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36*, 1132–1156. [[CrossRef](#)]
21. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved protein–ligand docking using GOLD. *Proteins: Struct. Funct. Bioinform.* **2003**, *52*, 609–623. [[CrossRef](#)] [[PubMed](#)]
22. 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. 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](#)] [[PubMed](#)]
23. 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](#)]
24. 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](#)]
25. Jorgensen, W.L.; Thomas, L.L. Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876. [[CrossRef](#)]
26. Deflorian, F.; Perez-Benito, L.; Lenselink, E.B.; Congreve, M.; van Vlijmen, H.W.; Mason, J.S.; Graaf, C.d.; Tresadern, G. Accurate prediction of GPCR ligand binding affinity with free energy perturbation. *J. Chem. Inf. Model.* **2020**, *60*, 5563–5579. [[CrossRef](#)]
27. Bhati, A.P.; Wan, S.; Wright, D.W.; Coveney, P.V. Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration. *J. Chem. Theory Comput.* **2017**, *13*, 210–222. [[CrossRef](#)]
28. Genheden, S.; Nilsson, I.; Ryde, U. Binding affinities of factor Xa inhibitors estimated by thermodynamic integration and MM/GBSA. *J. Chem. Inf. Model.* **2011**, *51*, 947–958. [[CrossRef](#)]
29. Lyu, J.; Wang, S.; Balias, T.E.; Singh, I.; Levit, A.; Moroz, Y.S.; O'Meara, M.J.; Che, T.; Alga, E.; Tolmachova, K. Ultra-large library docking for discovering new chemotypes. *Nature* **2019**, *566*, 224–229. [[CrossRef](#)]
30. Huang, S.-Y.; Grinter, S.Z.; Zou, X. Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899–12908. [[CrossRef](#)]
31. Böhm, H.; Stahl, M. The use of scoring functions in drug discovery applications. *Rev. Comput. Chem.* **2003**, *18*, 41–87. [[CrossRef](#)]
32. Talele, T.T.; Arora, P.; Kulkarni, S.S.; Patel, M.R.; Singh, S.; Chudayeu, M.; Kaushik-Basu, N. Structure-based virtual screening, synthesis and SAR of novel inhibitors of hepatitis C virus NS5B polymerase. *Bioorg. Med. Chem.* **2010**, *18*, 4630–4638. [[CrossRef](#)] [[PubMed](#)]
33. Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative assessment of scoring functions: The CASF-2016 update. *J. Chem. Inf. Model.* **2018**, *59*, 895–913. [[CrossRef](#)]
34. Li, Y.; Su, M.; Liu, Z.; Li, J.; Liu, J.; Han, L.; Wang, R. Assessing protein–ligand interaction scoring functions with the CASF-2013 benchmark. *Nat. Protoc.* **2018**, *13*, 666–680. [[CrossRef](#)] [[PubMed](#)]
35. Liu, J.; Wang, R. Classification of current scoring functions. *J. Chem. Inf. Model.* **2015**, *55*, 475–482. [[CrossRef](#)] [[PubMed](#)]
36. Goodsell, D.S.; Morris, G.M.; Olson, A.J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1–5. [[CrossRef](#)]
37. Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein–ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356. [[CrossRef](#)] [[PubMed](#)]
38. Huang, S.Y.; Zou, X. An iterative knowledge-based scoring function to predict protein–ligand interactions: II. Validation of the scoring function. *J. Comput. Chem.* **2006**, *27*, 1876–1882. [[CrossRef](#)]
39. Huang, S.Y.; Zou, X. An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials. *J. Comput. Chem.* **2006**, *27*, 1866–1875. [[CrossRef](#)]
40. Muegge, I.; Martin, Y.C. A general and fast scoring function for protein–ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804. [[CrossRef](#)]
41. Böhm, H.J. A novel computational tool for automated structure-based drug design. *J. Mol. Recognit.* **1993**, *6*, 131–137. [[CrossRef](#)] [[PubMed](#)]

42. Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243–256. [[CrossRef](#)] [[PubMed](#)]
43. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
44. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11–26. [[CrossRef](#)]
45. Yang, C.; Zhang, Y. Lin\_F9: A Linear Empirical Scoring Function for Protein–Ligand Docking. *J. Chem. Inf. Model.* **2021**, *61*, 4630–4644. [[CrossRef](#)]
46. Li, H.; Sze, K.H.; Lu, G.; Ballester, P.J. Machine-learning scoring functions for structure-based virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, e1478. [[CrossRef](#)]
47. Ain, Q.U.; Aleksandrova, A.; Roessler, F.D.; Ballester, P.J. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 405–424. [[CrossRef](#)] [[PubMed](#)]
48. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
49. Liaw, A.; Wiener, M. Classification and regression by random Forest. *R News* **2002**, *2*, 18–22.
50. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
51. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)]
52. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI Open* **2020**, *1*, 57–81. [[CrossRef](#)]
53. Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: Collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977–2980. [[CrossRef](#)]
54. Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-wide collection of binding data: Current status of the PDBbind database. *Bioinformatics* **2015**, *31*, 405–412. [[CrossRef](#)] [[PubMed](#)]
55. Liu, Z.; Su, M.; Han, L.; Liu, J.; Yang, Q.; Li, Y.; Wang, R. Forging the basis for developing protein–ligand interaction scoring functions. *Acc. Chem. Res.* **2017**, *50*, 302–309. [[CrossRef](#)] [[PubMed](#)]
56. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
57. Berman, H.M. The protein data bank: A historical perspective. *Acta Crystallogr. A* **2008**, *64*, 88–95. [[CrossRef](#)]
58. Smith, R.D.; Dunbar Jr, J.B.; Ung, P.M.-U.; Esposito, E.X.; Yang, C.-Y.; Wang, S.; Carlson, H.A. CSAR benchmark exercise of 2010: Combined evaluation across all submitted scoring functions. *J. Chem. Inf. Model.* **2011**, *51*, 2115–2131. [[CrossRef](#)]
59. Dunbar Jr, J.B.; Smith, R.D.; Yang, C.-Y.; Ung, P.M.-U.; Lexa, K.W.; Khazanov, N.A.; Stuckey, J.A.; Wang, S.; Carlson, H.A. CSAR benchmark exercise of 2010: Selection of the protein–ligand complexes. *J. Chem. Inf. Model.* **2011**, *51*, 2036–2046. [[CrossRef](#)]
60. Damm-Ganamet, K.L.; Smith, R.D.; Dunbar Jr, J.B.; Stuckey, J.A.; Carlson, H.A. CSAR benchmark exercise 2011–2012: Evaluation of results from docking and relative ranking of blinded congeneric series. *J. Chem. Inf. Model.* **2013**, *53*, 1853–1870. [[CrossRef](#)]
61. Dunbar Jr, J.B.; Smith, R.D.; Damm-Ganamet, K.L.; Ahmed, A.; Esposito, E.X.; Delproposto, J.; Chinnaswamy, K.; Kang, Y.-N.; Kubish, G.; Gestwicki, J.E. CSAR data set release 2012: Ligands, affinities, complexes, and docking decoys. *J. Chem. Inf. Model.* **2013**, *53*, 1842–1852. [[CrossRef](#)]
62. Smith, R.D.; Damm-Ganamet, K.L.; Dunbar Jr, J.B.; Ahmed, A.; Chinnaswamy, K.; Delproposto, J.E.; Kubish, G.M.; Tinberg, C.E.; Khare, S.D.; Dou, J. CSAR benchmark exercise 2013: Evaluation of results from a combined computational protein design, docking, and scoring/ranking challenge. *J. Chem. Inf. Model.* **2016**, *56*, 1022–1031. [[CrossRef](#)] [[PubMed](#)]
63. Carlson, H.A.; Smith, R.D.; Damm-Ganamet, K.L.; Stuckey, J.A.; Ahmed, A.; Convery, M.A.; Somers, D.O.; Kranz, M.; Elkins, P.A.; Cui, G. CSAR 2014: A benchmark exercise using unpublished data from pharma. *J. Chem. Inf. Model.* **2016**, *56*, 1063–1077. [[CrossRef](#)] [[PubMed](#)]
64. Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V.A.; Walters, W.P.; Kuhn, B.; Rudolph, M.G. 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](#)]
65. 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. D3R Grand Challenge 3: Blind prediction of protein–ligand poses and affinity rankings. *J. Comput. Aided Mol. Des.* **2019**, *33*, 1–18. [[CrossRef](#)] [[PubMed](#)]
66. Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J.A.; Kang, Y.N.; Delproposto, J.; Kubish, G.; Dunbar, J.B.; Carlson, H.A. D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. *J. Comput. Aided Mol. Des.* **2016**, *30*, 651–668. [[CrossRef](#)] [[PubMed](#)]
67. Parks, C.D.; Gaieb, Z.; Chiu, M.; Yang, H.; Shao, C.; Walters, W.P.; Jansen, J.M.; McGaughey, G.; Lewis, R.A.; Bembenek, S.D. D3R grand challenge 4: Blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **2020**, *34*, 99–119. [[CrossRef](#)]
68. Tran-Nguyen, V.-K.; Jacquemard, C.; Rognan, D. LIT-PCBA: An unbiased data set for machine learning and virtual screening. *J. Chem. Inf. Model.* **2020**, *60*, 4263–4273. [[CrossRef](#)]
69. Huang, N.; Shoichet, B.K.; Irwin, J.J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801. [[CrossRef](#)]

70. 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](#)]
71. Rohrer, S.G.; Baumann, K. Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model.* **2009**, *49*, 169–184. [[CrossRef](#)]
72. Wang, Y.; Suzek, T.; Zhang, J.; Wang, J.; He, S.; Cheng, T.; Shoemaker, B.A.; Gindulyte, A.; Bryant, S.H. PubChem bioassay: 2014 update. *Nucleic Acids Res.* **2014**, *42*, D1075–D1082. [[CrossRef](#)] [[PubMed](#)]
73. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [[CrossRef](#)] [[PubMed](#)]
74. Butkiewicz, M.; Lowe, E.W.; Mueller, R.; Mendenhall, J.L.; Teixeira, P.L.; Weaver, C.D.; Meiler, J. Benchmarking ligand-based virtual High-Throughput Screening with the PubChem database. *Molecules* **2013**, *18*, 735–756. [[CrossRef](#)] [[PubMed](#)]
75. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: A web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201. [[CrossRef](#)]
76. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [[CrossRef](#)]
77. Chen, X.; Liu, M.; Gilson, M.K. BindingDB: A Web-Accessible Molecular Recognition Database. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 719–725. [[CrossRef](#)]
78. Nicola, G.; Liu, T.; Hwang, L.; Gilson, M. BindingDB: A protein–ligand database for drug discovery. *Biophys. J.* **2012**, *102*, 61a. [[CrossRef](#)]
79. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [[CrossRef](#)]
80. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Motow, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [[CrossRef](#)]
81. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Motow, P.; Nowotka, M. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [[CrossRef](#)]
82. Bento, A.P.; Gaulton, A.; Hersey, A.; Bellis, L.J.; Chambers, J.; Davies, M.; Krüger, F.A.; Light, Y.; Mak, L.; McGlinchey, S. The ChEMBL bioactivity database: An update. *Nucleic Acids Res.* **2014**, *42*, D1083–D1090. [[CrossRef](#)] [[PubMed](#)]
83. Liebeschuetz, J.W.; Cole, J.C.; Korb, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J. Comput. Aided Mol. Des.* **2012**, *26*, 737–748. [[CrossRef](#)] [[PubMed](#)]
84. Bell, E.W.; Zhang, Y. DockRMSD: An open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism. *J. Cheminform.* **2019**, *11*, 40. [[CrossRef](#)]
85. Meli, R.; Biggin, P.C. spyrmsd: Symmetry-corrected RMSD calculations in Python. *J. Cheminform.* **2020**, *12*, 1–7. [[CrossRef](#)] [[PubMed](#)]
86. Allen, W.J.; Rizzo, R.C. Implementation of the Hungarian algorithm to account for ligand symmetry and similarity in structure-based design. *J. Chem. Inf. Model.* **2014**, *54*, 518–529. [[CrossRef](#)]
87. Brozell, S.R.; Mukherjee, S.; Balias, T.E.; Roe, D.R.; Case, D.A.; Rizzo, R.C. Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J. Comput. Aided Mol. Des.* **2012**, *26*, 749–773. [[CrossRef](#)]
88. Forli, S.; Huey, R.; Pique, M.E.; Sanner, M.F.; Goodsell, D.S.; Olson, A.J. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **2016**, *11*, 905–919. [[CrossRef](#)]
89. Ashtawy, H.M.; Mahapatra, N.R. A comparative assessment of ranking accuracies of conventional and machine-learning-based scoring functions for protein–ligand binding affinity prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2012**, *9*, 1301–1313. [[CrossRef](#)]
90. Sunseri, J.; Koes, D.R. Virtual Screening with Gnina 1.0. *Molecules* **2021**, *26*, 7369. [[CrossRef](#)]
91. Ackloo, S.; Al-awar, R.; Amaro, R.E.; Arrowsmith, C.H.; Azevedo, H.; Batey, R.A.; Bengio, Y.; Betz, U.A.; Bologna, C.G.; Chodera, J.D. CACHE (Critical Assessment of Computational Hit-finding Experiments): A public–private partnership benchmarking initiative to enable the development of computational methods for hit-finding. *Nat. Rev. Chem.* **2022**, 1–9. [[CrossRef](#)]
92. Goh, G.B.; Hodas, N.O.; Vishnu, A. Deep learning for computational chemistry. *J. Comput. Chem.* **2017**, *38*, 1291–1307. [[CrossRef](#)] [[PubMed](#)]
93. Li, H.; Sze, K.H.; Lu, G.; Ballester, P.J. Machine-learning scoring functions for structure-based drug lead optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, e1465. [[CrossRef](#)]
94. Ramakrishnan, R.; von Lilienfeld, O.A. Machine learning, quantum chemistry, and chemical space. *Rev. Comput. Chem.* **2017**, *30*, 225–256. [[CrossRef](#)]
95. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
96. Méndez-Lucio, O.; Ahmad, M.; del Rio-Chanona, E.A.; Wegner, J.K. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nat. Mach. Intell.* **2021**, *3*, 1033–1039. [[CrossRef](#)]
97. Wu, Z.; Ramsundar, B.; Feinberg, E.N.; Gomes, J.; Geniesse, C.; Pappu, A.S.; Leswing, K.; Pande, V. MoleculeNet: A benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513–530. [[CrossRef](#)]
98. Shen, C.; Ding, J.; Wang, Z.; Cao, D.; Ding, X.; Hou, T. From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, e1429. [[CrossRef](#)]

99. Ballester, P.J.; Mitchell, J.B. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* **2010**, *26*, 1169–1175. [[CrossRef](#)]
100. Zilian, D.; Sottriffer, C.A. Sfcscorerf: A random forest-based scoring function for improved affinity prediction of protein–ligand complexes. *J. Chem. Inf. Model.* **2013**, *53*, 1923–1933. [[CrossRef](#)]
101. Wang, C.; Zhang, Y. Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest. *J. Comput. Chem.* **2017**, *38*, 169–177. [[CrossRef](#)]
102. Lu, J.; Hou, X.; Wang, C.; Zhang, Y. Incorporating explicit water molecules and ligand conformation stability in machine-learning scoring functions. *J. Chem. Inf. Model.* **2019**, *59*, 4540–4549. [[CrossRef](#)] [[PubMed](#)]
103. Yang, C.; Zhang, Y. Delta Machine Learning to Improve Scoring-Ranking-Screening Performances of Protein–Ligand Scoring Functions. *J. Chem. Inf. Model.* **2022**, *62*, 2696–2712. [[CrossRef](#)] [[PubMed](#)]
104. Rayka, M.; Karimi-Jafari, M.H.; Firouzi, R. ET-score: Improving Protein-ligand Binding Affinity Prediction Based on Distance-weighted Interatomic Contact Features Using Extremely Randomized Trees Algorithm. *Mol. Inform.* **2021**, *40*, 2060084. [[CrossRef](#)] [[PubMed](#)]
105. Nguyen, D.D.; Wei, G.-W. Agl-score: Algebraic graph learning score for protein–ligand binding scoring, ranking, docking, and screening. *J. Chem. Inf. Model.* **2019**, *59*, 3291–3304. [[CrossRef](#)]
106. Sánchez-Cruz, N.; Medina-Franco, J.L.; Mestres, J.; Barril, X. Extended connectivity interaction features: Improving binding affinity prediction through chemical description. *Bioinformatics* **2021**, *37*, 1376–1382. [[CrossRef](#)]
107. Durrant, J.D.; McCammon, J.A. NNScore: A neural-network-based scoring function for the characterization of protein–Ligand complexes. *J. Chem. Inf. Model.* **2010**, *50*, 1865–1871. [[CrossRef](#)]
108. Durrant, J.D.; McCammon, J.A. NNScore 2.0: A neural-network receptor–ligand scoring function. *J. Chem. Inf. Model.* **2011**, *51*, 2897–2903. [[CrossRef](#)]
109. Wallach, I.; Dzamba, M.; Heifets, A. AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv* **2015**, arXiv:1510.02855.
110. Stepniewska-Dziubinska, M.M.; Zielenkiewicz, P.; Siedlecki, P. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics* **2018**, *34*, 3666–3674. [[CrossRef](#)]
111. Jiménez, J.; Skalic, M.; Martínez-Rosell, G.; De Fabritiis, G. K deep: Protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *J. Chem. Inf. Model.* **2018**, *58*, 287–296. [[CrossRef](#)]
112. Zheng, L.; Fan, J.; Mu, Y. Onionnet: A multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. *ACS Omega* **2019**, *4*, 15956–15965. [[CrossRef](#)] [[PubMed](#)]
113. Feinberg, E.N.; Sur, D.; Wu, Z.; Husic, B.E.; Mai, H.; Li, Y.; Sun, S.; Yang, J.; Ramsundar, B.; Pande, V.S. PotentialNet for molecular property prediction. *ACS Cent. Sci.* **2018**, *4*, 1520–1530. [[CrossRef](#)] [[PubMed](#)]
114. Karlov, D.S.; Sosnin, S.; Fedorov, M.V.; Popov, P. graphDelta: MPNN scoring function for the affinity prediction of protein–ligand complexes. *ACS Omega* **2020**, *5*, 5150–5159. [[CrossRef](#)]
115. Li, S.; Zhou, J.; Xu, T.; Huang, L.; Wang, F.; Xiong, H.; Huang, W.; Dou, D.; Xiong, H. Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; pp. 975–985.
116. Ashtawy, H.M.; Mahapatra, N.R. Task-specific scoring functions for predicting ligand binding poses and affinity and for screening enrichment. *J. Chem. Inf. Model.* **2018**, *58*, 119–133. [[CrossRef](#)]
117. Wójcikowski, M.; Ballester, P.J.; Siedlecki, P. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci. Rep.* **2017**, *7*, 1–10. [[CrossRef](#)] [[PubMed](#)]
118. Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D.R. Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* **2017**, *57*, 942–957. [[CrossRef](#)]
119. Wang, Z.; Zheng, L.; Liu, Y.; Qu, Y.; Li, Y.-Q.; Zhao, M.; Mu, Y.; Li, W. Onionnet-2: A convolutional neural network model for predicting protein-ligand binding affinity based on residue-atom contacting shells. *Front. Chem.* **2021**, 913. [[CrossRef](#)]
120. Lim, J.; Ryu, S.; Park, K.; Choe, Y.J.; Ham, J.; Kim, W.Y. Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation. *J. Chem. Inf. Model.* **2019**, *59*, 3981–3988. [[CrossRef](#)]
121. Son, J.; Kim, D. Development of a graph convolutional neural network model for efficient prediction of protein-ligand binding affinities. *PLoS ONE* **2021**, *16*, e0249404. [[CrossRef](#)]
122. Wang, Y.; Li, L.; Zhang, B.; Xing, J.; Chen, S.; Wan, W.; Song, Y.; Jiang, H.; Jiang, H.; Luo, C. Discovery of novel disruptor of silencing telomeric 1-like (DOT1L) inhibitors using a target-specific scoring function for the (S)-adenosyl-l-methionine (SAM)-dependent methyltransferase family. *J. Med. Chem.* **2017**, *60*, 2026–2036. [[CrossRef](#)]
123. Shen, C.; Weng, G.; Zhang, X.; Leung, E.L.-H.; Yao, X.; Pang, J.; Chai, X.; Li, D.; Wang, E.; Cao, D. Accuracy or novelty: What can we gain from target-specific machine-learning-based scoring functions in virtual screening? *Brief. Bioinform.* **2021**, *22*, bbaa410. [[CrossRef](#)] [[PubMed](#)]
124. Yang, Y.; Lu, J.; Yang, C.; Zhang, Y. Exploring fragment-based target-specific ranking protocol with machine learning on cathepsin S. *J. Comput. Aided Mol. Des.* **2019**, *33*, 1095–1105. [[CrossRef](#)] [[PubMed](#)]
125. Maia, E.H.B.; Assis, L.C.; De Oliveira, T.A.; Da Silva, A.M.; Taranto, A.G. Structure-based virtual screening: From classical to artificial intelligence. *Front. Chem.* **2020**, *8*, 343. [[CrossRef](#)]

126. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [[CrossRef](#)]
127. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [[CrossRef](#)] [[PubMed](#)]
128. Baek, M.; Baker, D. Deep learning and protein structure modeling. *Nat. Methods* **2022**, *19*, 13–14. [[CrossRef](#)]
129. Frye, L.; Bhat, S.; Akinsanya, K.; Abel, R. From computer-aided drug discovery to computer-driven drug discovery. *Drug Discover. Today Technol.* **2021**, *39*, 111–117. [[CrossRef](#)] [[PubMed](#)]
130. Ma, D.-L.; Chan, D.S.-H.; Leung, C.-H. Drug repositioning by structure-based virtual screening. *Chem. Soc. Rev.* **2013**, *42*, 2130–2141. [[CrossRef](#)]
131. Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins: Struct. Funct. Bioinform.* **1999**, *37*, 228–241. [[CrossRef](#)]
132. Kearsley, S.K.; Underwood, D.J.; Sheridan, R.P.; Miller, M.D. Flexibases: A way to enhance the use of molecular docking methods. *J. Comput. Aided Mol. Des.* **1994**, *8*, 565–582. [[CrossRef](#)]
133. Hart, T.N.; Read, R.J. A multiple-start Monte Carlo docking method. *Proteins: Struct. Funct. Bioinform.* **1992**, *13*, 206–222. [[CrossRef](#)] [[PubMed](#)]
134. Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662. [[CrossRef](#)]
135. Wong, C.F. Flexible receptor docking for drug discovery. *Expert Opin. Drug Discov.* **2015**, *10*, 1189–1200. [[CrossRef](#)] [[PubMed](#)]
136. Tian, S.; Sun, H.; Pan, P.; Li, D.; Zhen, X.; Li, Y.; Hou, T. Assessing an ensemble docking-based virtual screening strategy for kinase targets by considering protein flexibility. *J. Chem. Inf. Model.* **2014**, *54*, 2664–2679. [[CrossRef](#)] [[PubMed](#)]
137. Korb, O.; Olsson, T.S.; Bowden, S.J.; Hall, R.J.; Verdonk, M.L.; Liebeschuetz, J.W.; Cole, J.C. Potential and limitations of ensemble docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262–1274. [[CrossRef](#)] [[PubMed](#)]
138. Amaro, R.E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J.A.; Miao, Y.; Smith, J.C. Ensemble docking in drug discovery. *Biophys. J.* **2018**, *114*, 2271–2278. [[CrossRef](#)] [[PubMed](#)]
139. Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184. [[CrossRef](#)]
140. Rueda, M.; Bottegoni, G.; Abagyan, R. Recipes for the selection of experimental protein conformations for virtual screening. *J. Chem. Inf. Model.* **2010**, *50*, 186–193. [[CrossRef](#)]
141. Mohammadi, S.; Narimani, Z.; Ashouri, M.; Firouzi, R.; Karimi-Jafari, M.H. Ensemble learning from ensemble docking: Revisiting the optimum ensemble size problem. *Sci. Rep.* **2022**, *12*, 1–15. [[CrossRef](#)]
142. Chandak, T.; Mayginnis, J.P.; Mayes, H.; Wong, C.F. Using machine learning to improve ensemble docking for drug discovery. *Proteins: Struct. Funct. Bioinform.* **2020**, *88*, 1263–1270. [[CrossRef](#)]
143. Huang, S.-Y.; Zou, X. Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016–3034. [[CrossRef](#)] [[PubMed](#)]
144. Ravindranath, P.A.; Forli, S.; Goodsell, D.S.; Olson, A.J.; Sanner, M.F. AutoDockFR: Advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comp. Biol.* **2015**, *11*, e1004586. [[CrossRef](#)] [[PubMed](#)]
145. Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into protein–ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **2016**, *17*, 144. [[CrossRef](#)] [[PubMed](#)]
146. Burley, S.K.; Berman, H.M.; Kleywegt, G.J.; Markley, J.L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The single global macromolecular structure archive. *Protein Crystallogr.* **2017**, 627–641. [[CrossRef](#)]
147. Lee, J.; Freddolino, P.L.; Zhang, Y. Ab initio protein structure prediction. In *From Protein Structure to Function with Bioinformatics*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 3–35.
148. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)]
149. Stafford, K.A.; Anderson, B.M.; Sorenson, J.; van den Bedem, H. AtomNet PoseRanker: Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High-Throughput Screens. *J. Chem. Inf. Model.* **2022**, *62*, 1178–1189. [[CrossRef](#)]
150. Rollinger, J.M.; Stuppner, H.; Langer, T. Virtual screening for the discovery of bioactive natural products. *Nat. Compd. Drugs Vol. 1* **2008**, 211–249. [[CrossRef](#)]
151. Sterling, T.; Irwin, J.J. ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. [[CrossRef](#)]
152. Irwin, J.J.; Shoichet, B.K. ZINC— a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182. [[CrossRef](#)]
153. 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. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [[CrossRef](#)]
154. Cuesta, S.A.; Mora, J.R.; Márquez, E.A. In silico screening of the DrugBank database to search for possible drugs against SARS-CoV-2. *Molecules* **2021**, *26*, 1100. [[CrossRef](#)] [[PubMed](#)]
155. 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](#)] [[PubMed](#)]

156. Wishart, D.S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A.C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S. HMDB: The human metabolome database. *Nucleic Acids Res.* **2007**, *35*, D521–D526. [[CrossRef](#)] [[PubMed](#)]
157. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [[CrossRef](#)] [[PubMed](#)]
158. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [[CrossRef](#)]
159. Sardanelli, A.M.; Isgrò, C.; Palese, L.L. SARS-CoV-2 main protease active site ligands in the human metabolome. *Molecules* **2021**, *26*, 1409. [[CrossRef](#)]
160. Liu, Y.; Grimm, M.; Dai, W.-t.; Hou, M.-c.; Xiao, Z.-X.; Cao, Y. CB-Dock: A web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacol. Sin.* **2020**, *41*, 138–144. [[CrossRef](#)]
161. Zhang, W.; Bell, E.W.; Yin, M.; Zhang, Y. EDock: Blind protein–ligand docking by replica-exchange monte carlo simulation. *J. Cheminform.* **2020**, *12*, 1–17. [[CrossRef](#)]
162. Rooklin, D.; Wang, C.; Katigbak, J.; Arora, P.S.; Zhang, Y. AlphaSpace: Fragment-centric topographical mapping to target protein–protein interaction interfaces. *J. Chem. Inf. Model.* **2015**, *55*, 1585–1599. [[CrossRef](#)]
163. Katigbak, J.; Li, H.; Rooklin, D.; Zhang, Y. AlphaSpace 2.0: Representing Concave Biomolecular Surfaces Using  $\beta$ -Clusters. *J. Chem. Inf. Model.* **2020**, *60*, 1494–1508. [[CrossRef](#)]
164. Ngan, C.H.; Bohnuud, T.; Mottarella, S.E.; Beglov, D.; Villar, E.A.; Hall, D.R.; Kozakov, D.; Vajda, S. FTMAP: Extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.* **2012**, *40*, W271–W275. [[CrossRef](#)] [[PubMed](#)]
165. Schmidtke, P.; Bidon-Chanal, A.; Luque, F.J.; Barril, X. MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011**, *27*, 3276–3285. [[CrossRef](#)]
166. Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tuffièry, P. Fpocket: Online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.* **2010**, *38*, W582–W589. [[CrossRef](#)] [[PubMed](#)]
167. Halgren, T.A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389. [[CrossRef](#)] [[PubMed](#)]
168. Wagner, J.R.; Lee, C.T.; Durrant, J.D.; Malmstrom, R.D.; Feher, V.A.; Amaro, R.E. Emerging computational methods for the rational discovery of allosteric drugs. *Chem. Rev.* **2016**, *116*, 6370–6390. [[CrossRef](#)]
169. Oleinikovas, V.; Saladino, G.; Cossins, B.P.; Gervasio, F.L. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *JACS* **2016**, *138*, 14257–14263. [[CrossRef](#)]
170. Bas, D.C.; Rogers, D.M.; Jensen, J.H. Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins: Struct. Funct. Bioinform.* **2008**, *73*, 765–783. [[CrossRef](#)]
171. Anandkrishnan, R.; Aguilar, B.; Onufriev, A.V. H++ 3.0: Automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541. [[CrossRef](#)]
172. Ten Brink, T.; Exner, T.E. pKa based protonation states and microspecies for protein–Ligand docking. *J. Comput. Aided Mol. Des.* **2010**, *24*, 935–942. [[CrossRef](#)]
173. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667. [[CrossRef](#)]
174. Dolinsky, T.J.; Czodrowski, P.; Li, H.; Nielsen, J.E.; Jensen, J.H.; Klebe, G.; Baker, N.A. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525. [[CrossRef](#)] [[PubMed](#)]
175. Lie, M.A.; Thomsen, R.; Pedersen, C.N.; Schiøtt, B.; Christensen, M.H. Molecular docking with ligand attached water molecules. *J. Chem. Inf. Model.* **2011**, *51*, 909–917. [[CrossRef](#)] [[PubMed](#)]
176. Kumar, A.; Zhang, K.Y. Investigation on the effect of key water molecules on docking performance in CSARdock exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1880–1892. [[CrossRef](#)] [[PubMed](#)]
177. Murphy, R.B.; Repasky, M.P.; Greenwood, J.R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N.A.; Schmitz, C.D.; Abel, R.; Farid, R. WScore: A flexible and accurate treatment of explicit water molecules in ligand–Receptor docking. *J. Med. Chem.* **2016**, *59*, 4364–4384. [[CrossRef](#)] [[PubMed](#)]
178. Santos-Martins, D.; Forli, S.; Ramos, M.J.; Olson, A.J. AutoDock4Zn: An improved AutoDock force field for small-molecule docking to zinc metalloproteins. *J. Chem. Inf. Model.* **2014**, *54*, 2371–2379. [[CrossRef](#)]
179. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. [[CrossRef](#)]
180. Landrum, G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. 2013.
181. Bento, A.P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L.J.; De Veij, M.; Leach, A.R. An open source chemical structure curation pipeline using RDKit. *J. Cheminform.* **2020**, *12*, 1–16. [[CrossRef](#)]
182. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 1–14. [[CrossRef](#)]
183. Hawkins, P.C.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584. [[CrossRef](#)]

184. Hawkins, P.C.; Nicholls, A. Conformer generation with OMEGA: Learning from the data set and the analysis of failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919–2936. [[CrossRef](#)]
185. Watts, K.S.; Dalal, P.; Murphy, R.B.; Sherman, W.; Friesner, R.A.; Shelley, J.C. ConfGen: A conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534–546. [[CrossRef](#)] [[PubMed](#)]
186. Huey, R.; Morris, G.M. Using AutoDock 4 with AutoDocktools: A tutorial. *Scrpps Res. Inst. USA* **2008**, *8*, 54–56.
187. Ren, X.; Shi, Y.-S.; Zhang, Y.; Liu, B.; Zhang, L.-H.; Peng, Y.-B.; Zeng, R. Novel consensus docking strategy to improve ligand pose prediction. *J. Chem. Inf. Model.* **2018**, *58*, 1662–1668. [[CrossRef](#)] [[PubMed](#)]
188. Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195–207. [[CrossRef](#)]
189. Bouvier, G.; Evrard-Todeschi, N.; Girault, J.-P.; Bertho, G. Automatic clustering of docking poses in virtual screening process using self-organizing map. *Bioinformatics* **2010**, *26*, 53–60. [[CrossRef](#)]
190. 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](#)]
191. Adeshina, Y.O.; Deeds, E.J.; Karanicolas, J. Machine learning classification can reduce false positives in structure-based virtual screening. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 18477–18488. [[CrossRef](#)]
192. 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](#)]
193. Su, M.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Tapping on the black box: How is the scoring power of a machine-learning scoring function dependent on the training set? *J. Chem. Inf. Model.* **2020**, *60*, 1122–1136. [[CrossRef](#)]
194. Kuzmanic, A.; Bowman, G.R.; Juarez-Jimenez, J.; Michel, J.; Gervasio, F.L. Investigating cryptic binding sites by molecular dynamics simulations. *Acc. Chem. Res.* **2020**, *53*, 654–661. [[CrossRef](#)]
195. Sgobba, M.; Caporuscio, F.; Anighoro, A.; Portioli, C.; Rastelli, G. Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. *Eur. J. Med. Chem.* **2012**, *58*, 431–440. [[CrossRef](#)] [[PubMed](#)]
196. Kumar, K.; Anbarasu, A.; Ramaiah, S. Molecular docking and molecular dynamics studies on  $\beta$ -lactamases and penicillin binding proteins. *Mol. Biosyst.* **2014**, *10*, 891–900. [[CrossRef](#)] [[PubMed](#)]
197. Zhou, H.; Cao, H.; Skolnick, J. FRAGSITE: A fragment-based approach for virtual ligand screening. *J. Chem. Inf. Model.* **2021**, *61*, 2074–2089. [[CrossRef](#)] [[PubMed](#)]
198. Tran-Nguyen, V.-K.; Bret, G.; Rognan, D. True Accuracy of Fast Scoring Functions to Predict High-Throughput Screening Data from Docking Poses: The Simpler the Better. *J. Chem. Inf. Model.* **2021**, *61*, 2788–2797. [[CrossRef](#)] [[PubMed](#)]
199. Gawehn, E.; Hiss, J.A.; Schneider, G. Deep learning in drug discovery. *Mol. Inform.* **2016**, *35*, 3–14. [[CrossRef](#)]
200. Labbé, C.M.; Rey, J.; Lagorce, D.; Vavruša, M.; Becot, J.; Sperandio, O.; Villoutreix, B.O.; Tufféry, P.; Miteva, M.A. MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Res.* **2015**, *43*, W448–W454. [[CrossRef](#)]
201. Gentile, F.; Yaacoub, J.C.; Gleave, J.; Fernandez, M.; Ton, A.-T.; Ban, F.; Stern, A.; Cherkasov, A. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **2022**, *17*, 672–697. [[CrossRef](#)]