Utrecht University

Structure-based Virtual Screening: Challenges and Future Directions in Scoring Functions

Tim van den Nobelen[1]

Supervisor: Dr. Ed Moret[1]

Affiliations: [1]Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, PO Box 80082, 3508 TB Utrecht, The Netherlands

Layman's summary

Virtual screening and molecular docking are important techniques used in computer-aided drug design. They help find potential drugs by using computer programs to simulate how different molecules interact with proteins in the body. However, the accuracy of these methods depends on the algorithms and scoring systems used. Search algorithms are like search engines that look for the best-fitting molecule and protein shapes. Scoring functions, on the other hand, help evaluate how well a molecule binds to a protein and how strong that binding is. They help determine which molecules are most likely to be effective as drugs. Despite a lot of research in this area, accurately predicting how molecules and proteins interact is still a challenge. This review looks at the basic elements of the databases of molecules, the search algorithms, and the scoring systems used in virtual screening and molecular docking. It also discusses the difficulties scientists face in this field and explores ideas for future research to improve these methods.

Table of Contents

Abstract

Virtual screening and molecular docking play crucial roles in computer-aided drug design. The reliability of these methods relies on the precision of the search algorithms and scoring functions employed. Search algorithms are responsible for identifying the most appropriate ligand and protein conformations, while scoring functions determine the binding mode, site, and affinity of the ligand, enabling the identification of potential drug leads. Despite extensive research, accurately and rapidly predicting ligand-protein interactions remains a challenge. Therefore, this review examines the fundamental aspects of libraries, search algorithms, and scoring functions employed in virtual screening and molecular docking. Additionally, it discusses the challenges encountered and explores potential future research directions in this field.

Introduction

Since its development in the early 1990s, high-throughput screening (HTS) has played a crucial role in drug discovery and screening processes (Putatunda et al., 2023). HTS enables the characterization of 10,000 to 100,000 compounds based on their biological activity, selectivity, bioavailability, and toxicity (Mayr & Bojanic, 2009). However, identifying effective drugs requires significant time and financial investment, limiting the discovery and optimization of lead compounds (Suay-García et al., 2022). To mitigate these challenges, computer-aided drug design (CADD) has emerged as a practical approach, significantly reducing costs and time in the drug discovery process. One prominent CADD method is virtual screening (VS), which employs in silico screening of compound libraries to identify potential drug leads for specific targets. The first publication on VS dates back to 1997 (Horvath, 1997) and has since contributed to the approval of at least 69 commercial drugs (Sabe et al., 2021).

VS can be broadly classified into two categories based on the type of information utilized. Ligand-based virtual screening relies on the concept that chemically similar structures exhibit similar biological effects (Lill, 2013). The second category is structure-based virtual screening, which involves docking multiple ligands onto a target protein to assess their affinity (Zhu et al., 2022). Consequently, a 3D structure of the target protein is required, and the ligands are ranked based on their affinity with the protein. This process entails selecting ligands from a library and employing algorithms to determine the optimal conformation for both the ligand and the protein. Subsequently, molecular docking software utilizes scoring functions to evaluate the binding affinity of the ligand-protein complex (Sabe et al., 2021). While VS significantly contributes to drug development, it still has certain limitations. Libraries may suffer from incomplete data sets, ligands and proteins employed in molecular docking are often limited to a single rigid conformation, and scoring functions may yield false positives and negatives.

This review overviews various databases/libraries, algorithms, and scoring functions utilized in molecular docking. Furthermore, future directions and potential enhancements to the existing molecular docking workflow will be discussed.

Structure-based Virtual Screening

Structure-based virtual screening (SBVS) uses hundreds of ligands from a dataset and determines the affinity with a specified protein in a 3D environment. To start, 3D structures of the target proteins and ligands must be available. The ligands and proteins need to be in the correct configuration, and lastly, algorithms determine the binding affinity between the ligand and protein. The following sections discuss the most used libraries and databases, configuration algorithms, and docking/scoring functions.

Libraries and Databases

As mentioned previously, a necessary condition is access to a 3D structure of the target protein, and the ligands are being docked (N. Cavasotto, 2011). There are some commercially available databases, but most are accessible without cost. Below is a summary of existing databases primarily used for SBVS:

- Drugbank (Wishart et al., 2018): Drugbank started in 2006 at Dr. David Wishart's lab at the University of Alberta to help academic researchers get detailed drug information. Drugbank has since grown into an online database covering detailed drug data with comprehensive drug target information. The database contains over 11,900 drug entries, including FDA-approved small molecule and biotechnology drugs and investigational and nutraceuticals.

- ZINC (Irwin et al., 2020): Maintained by Irwin and Shoichet Laboratories, Zinc is a free database of commercially and annotated compounds for VS. Together, it has nearly 2 billion compounds, which can be searched on explicit atomic-level graph-based methods.

- PubChem (Kim et al., 2023): Maintained by the National Institutes of Health (NIH), Pubchem is a public database where scientific data can be stored for others to use. It contains information about 114,823,599 unique chemical compounds, 186,035 protein targets, and more(*Statistics - PubChem*, n.d.).

- Protein Data Bank (PDB) (Burley et al., 2023): The Protein Data Bank is an extensive archive of 3D structure data for proteins, DNA, and RNA. It is one of the most widely used libraries fundamental for research and education in health, biology, and biotechnology. PDB grows by 10% each year, going from 48,169 entries in 2008 to 204,826 admissions at the beginning of 2023 (*PDB Statistics: PDB Data Distribution by Experimental Method and Molecular Type*, n.d.).

- ChEMBL (Mendez et al., 2019): The European Institute of Bioinformatics sustains a database of bioactive molecules with medicinal properties. Over the years, they collected around 2.3 million compounds with 15.2 million known biological activities.

- PDBBind Database (Z. Liu et al., 2015): The PDBBind database, developed by Prof. Shaomeng Wang's group at the University of Michigan, is valuable. Its purpose is to comprehensively compile experimentally measured binding affinity data for all biomolecular complexes available in the Protein Data Bank. The latest release (version 2020) encompasses binding affinity data for a total of 23,496 biomolecular complexes.

- ChemSpider (Pence & Williams, 2010): Owned by the Royal Society of Chemistry, ChemSpider has a chemical structure database of over 100 million structures with associated properties. Just like PubChem, ChemSpider collects data from high-quality data sources. Up to 1000 structure downloads are allowed daily; further contact is needed for more downloads, thus not 100% free of charge (Maia et al., 2020).

Search Algorithms

Once a target protein and ligands are selected from a library, search algorithms are employed to explore the orientation and conformation of the ligands within the docking site. In rigid docking, search algorithms explore different positions of the ligands by considering translational and rotational degrees of freedom. In the case of flexible docking, additional conformational degrees of freedom are introduced (Maia et al., 2020). Various techniques are utilized to predict the ligands' correct conformation and can be can be classified into three distinct classes: meta-heuristic algorithms, machine learning-based algorithms, and other technique-based algorithms (Table 1).

- Meta-heuristic algorithms belong to the class of optimization algorithms specifically designed to tackle complex optimization problems that are challenging or infeasible to solve using traditional mathematical approaches. These algorithms prove particularly useful when dealing with situations where only partial or incomplete information is available. They often possess a probabilistic nature and are characterized by their efficiency in exploring and exploiting the search space, effectively striking a balance between exploration and exploitation to discover optimal solutions (Abdel-Basset et al., 2018; Bianchi et al., 2009).

- Machine learning-based algorithms leverage the vast amount of available molecular structural information and biological activity datasets. These algorithms utilize the datasets to learn, make decisions, and recognize patterns. The significant advantage of machine learning lies in its ability to leverage acquired knowledge during the learning process, eliminating the need for computationally expensive simulations. By leveraging existing data, machine learning algorithms can predict and infer interactions between ligands and proteins, leading to efficient and accurate predictions (Oliveira et al., 2023).

- Other technique-based algorithms encompass approaches that do not fit neatly into the meta-heuristic or machine learning-based classifications. These algorithms often incorporate a range of diverse methodologies and techniques to address specific challenges in molecular docking. These may include approaches based on statistical analysis, physics-based models, or hybrid methods that combine elements from multiple algorithmic strategies. The distinguishing characteristic of these techniques is their unique and specialized nature, tailored to tackle specific aspects or problems encountered in the molecular docking process(Maia et al., 2020).

Overall, these three classes of algorithms provide a diverse toolkit for tackling different aspects of molecular docking, each bringing its strengths and advantages to the table.

Table 1 Algorithms used in virtual screening

Class	Type	Algorithm
Meta-Heuristic Algorithms:	**Evolutionary Algorithms:**	• Genetic Algorithms (Xia et al., 2017) • Differential evolution (Friesner et al., 2004), • Ant Colony Optimization (Korb et al., 2006) • Tuba search (Baxter et al., 1998) • Particle Swarm Optimization (Gowthaman et al., 2015) • PSOVina (Ng et al., 2015)
	Statistical Methods:	• Simulated Annealing (SA) (Doucet & Pelletier, 2007), Hatmal and Taha (Hatmal & Taha, 2017) • Conformational Space Annealing (CSA) (Shin et al., 2011)
Machine Learning-based Algorithms:	**Machine Learning:**	• Artificial Neural Networks (ANNs) (Ashtawy & Mahapatra, 2018) • Support Vector Machines (Sengupta & Bandyopadhyay, 2012) • Bayesian Techniques (Abdo et al., 2010) • Decision Tree (Tin Kam Ho, 1998) • k-Nearest Neighbor (kNN) (Peterson et al., 2009) • Kohonen;'s SOMs and Counterpropagation ANNs (Schneider et al., 2009) • Ensemble Methods using Machine Learning (Korkmaz et al., 2015)
Other Techniques-Based Algorithms	**Statistical Methods:**	• Monte Carlo (Harrison et al., 2010)
	Similarity-based Algorithms:	• Based on Substructures (Tresadern et al., 2009) • Pharmacochemical (Cruz-Monteagudo et al., 2014) • Overlapping Volumes (Leach et al., 2010) • Molecular Interaction Fields (MIFs) (Willett, 2006) • Hybrid Approach (H. Haga et al., 2016; Morris et al., 2009)
	Incremental Construction:	• FlexX complex construct algorithm (Rarey et al., 1996) • Surflex-Dock approach (Spitzer & Jain, 2012)
	Local Search:	• Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (Trott & Olson, 2009)
	Exhaustive Search:	• eHiTS (Zsoldos et al., 2007)
	Linear programming method:	• Simplex Method (Ruiz-Carmona et al., 2014)

Scoring Functions

During the docking process, while search algorithms explore various conformations of each ligand from the compound library, scoring functions play a crucial role in assessing the quality of these conformations. Scoring functions aim to achieve three primary objectives: determining the binding mode or site of a ligand within a protein, predicting the absolute binding affinity, and identifying potential drug leads for the specific protein target (J. Li et al., 2019).

Scoring functions can be broadly classified into four major classes, as illustrated in Figure 1: Physics-based, Empirical, Knowledge-based, and machine learning-based scoring functions. The first three classes, often called "classical" scoring functions, primarily employ linear regression methods for their calculations. On the other hand, the machine learning-based scoring functions utilize non-linear regression methods.

In the subsequent chapters, we will provide a concise overview of each of these scoring functions, highlighting their distinctive characteristics and methodologies.

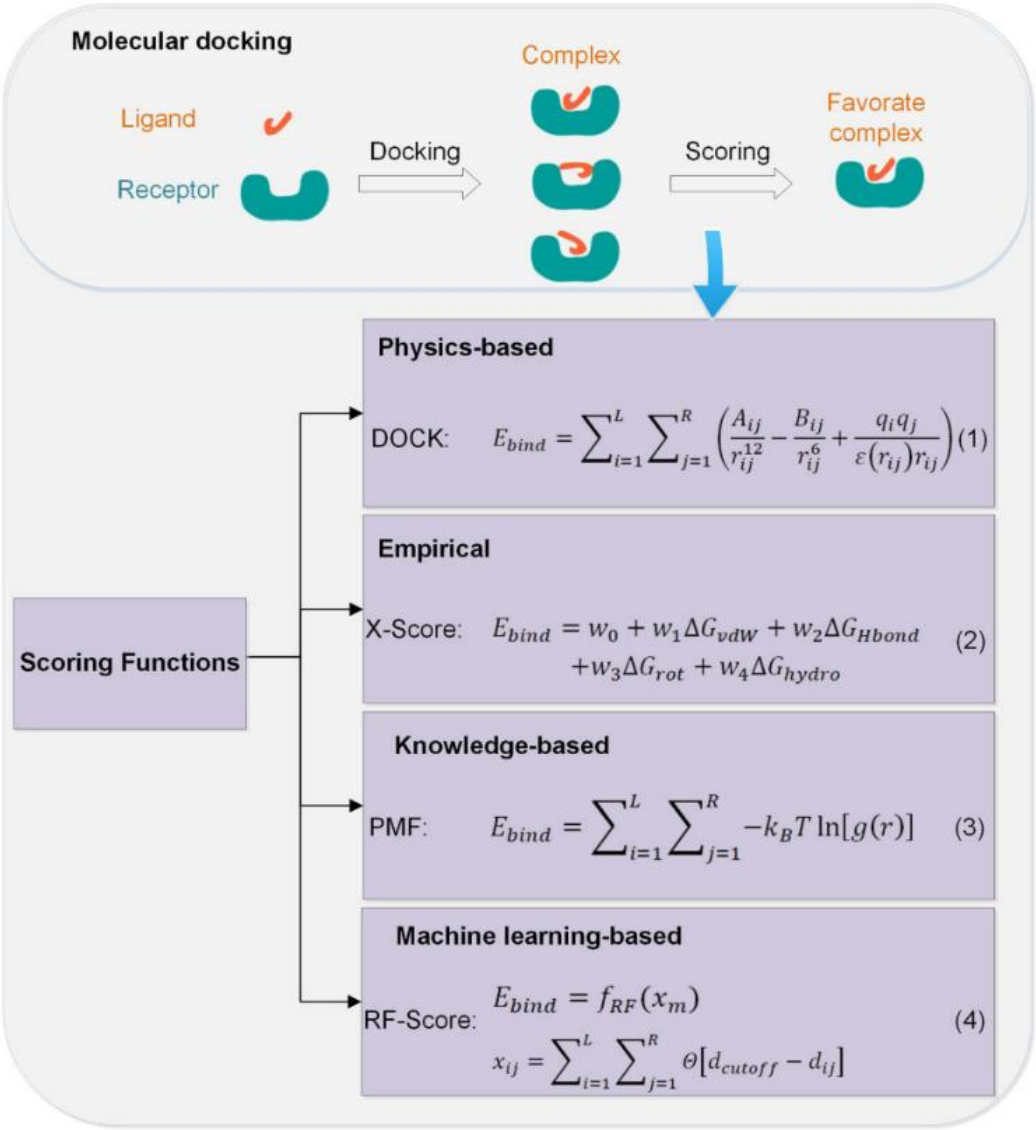

Figure 1 Four categories of scoring functions for structure-based virtual screening (obtained from (J. Li et al., 2019)**)**

Classical-based Scoring Functions

Physics-based

Physics-based scoring functions rely on the intermolecular interactions between the ligand's atoms and the target. These interactions encompass Van Der Waals forces, electrostatic interactions, bond stretching, bending, and torsional forces. The parameters for these interactions are obtained from experimental data and developed following the principles of molecular mechanics (Ferreira et al., 2015). However, the accuracy of predicting the binding energy is heavily dependent on the functional form of the potential energy and its associated parameters, which can be challenging to determine due to the nature of physics-based scoring functions. In response to the challenges posed by covalent interactions, polarization, and charge transfer in docking (Raha et al., 2007; Senn & Thiel, 2009), recent studies have introduced a new scoring function based on quantum mechanics. However, it should be noted that quantum mechanics-based scoring functions offer higher accuracy but have increased computational costs compared to physics-based scoring functions. A hybrid quantum mechanical/molecular mechanics approach has been developed to balance computational efficiency and predictive accuracy. This approach aims to provide a compromise solution for addressing these challenges (Chaskar et al., 2014).

Physics-based scoring functions have the advantage of directly calculating interactions between protein and ligand atoms. They are particularly suitable for estimating binding free energy and offer higher predictive accuracy than other scoring functions. This is because physics-based scoring functions consider enthalpy, solvation, and entropy factors.

Empirical

Empirical scoring functions are utilized to estimate the binding affinity of a protein-ligand complex by considering essential energetic factors such as hydrogen bonds, hydrophobic effects, and steric clashes (Eldridge et al., 1997; Friesner et al., 2006; Zheng & Merz, 2011). These scoring functions are typically optimized by employing a training set with known binding affinities, and the weights of the energetic factors are determined through linear regression analysis (Kadukova & Grudinin, 2017). Empirical scoring functions involve two main research directions. The first direction focuses on utilizing large, high-quality training datasets to optimize protein-ligand structures. This consists in employing comprehensive data to enhance the accuracy of the scorings function. The second direction involves selecting suitable energy terms through stepwise variables and systematic selection specific to the target protein (Fornabaio et al., 2004; Kerzmann et al., 2006). Considering these research directions, empirical scoring functions have become widely adopted in protein-ligand docking programs. These scoring functions are crucial in facilitating the study of protein-ligand interactions and have found extensive use in the field.

Empirical scoring functions share the characteristic of decomposing protein-ligand binding affinities into single energy terms, similar to physics-based scoring functions. However, empirical scoring functions often adopt a flexible and intuitive functional form, deviating from the well-established models utilized by physics-based scoring functions. The simplicity of energy terms in empirical scoring functions enables efficient prediction of binding affinity, ligand pose, and virtual screening, all while maintaining low computational costs (Y. Li, Liu, et al., 2014).

Knowledge-based

In knowledge-based scoring functions, the binding affinity is determined by adding the interactions between the atoms of a protein and the molecular target (Ferreira et al., 2015). These functions leverage statistical observations from large databases and utilize pairwise energy potentials derived from known ligand-receptor complexes, using the inverse Boltzmann statistical principle, to create a general scoring function (Muegge & Martin, 1999). The underlying principle is that intermolecular interactions occurring near specific types of atoms or functional groups, which are more frequently observed, are more likely to contribute positively to the binding affinity. The final score is computed

as the sum of individual interaction scores. The major advantage of knowledge-based scoring functions is their ability to balance computational cost and predictive accuracy compared to physics-based and empirical scoring functions. However, locating the reference state poses a challenge for knowledge-based scoring functions. The reference state represents the expected distribution or behavior of atomic pairs in a given system, serving as a baseline for comparison in the scoring process. By comparing the observed atomic pairs in the target system with the reference state, the scoring function can evaluate the compatibility and likelihood of the observed interactions. Currently, two classical strategies are employed for determining the reference state. One approach involves approximating the reference state using a random distribution of atomic pairs in the training set (Velec et al., 2005). The other approach introduces correction methods based on the first strategy to enhance the accuracy of knowledge-based scoring functions (Huang & Zou, 2006).

Since the training sets for knowledge-based scoring functions solely rely on structural information and are independent of experimental binding affinity data, they avoid potential ambiguities associated with experimental conditions that could influence binding affinity. This suggests that knowledge-based scoring functions are more suitable for predicting binding poses than precise binding affinities (Zheng & Merz, 2013).

Machine Learning-Based Scoring Functions

The use of predictive models based on machine learning has become increasingly prevalent in virtual screening due to the availability of large databases containing molecular information and the advantages offered by machine learning techniques, such as accuracy, expanded chemical libraries, new molecular descriptors, and similarity search techniques (Hönig et al., 2023). Machine learning algorithms utilize datasets to learn and, based on the knowledge acquired, make decisions, predictions, and recognize patterns. Additionally, the performance of a machine learning system is expected to improve over time and adapt to changes (Leguizamón, 2011).

In contrast to classical scoring functions with assumed mathematical functional forms, machine-learning-based scoring functions utilize various machine-learning algorithms such as support vector machines, random forests, neural networks, and deep learning. While machine-learning-based scoring functions have shown superior performance to classical scoring functions (Ma et al., 2013), they are typically not directly integrated into docking software but instead employed for rescoring purposes (Zhang et al., 2017). Machine-learning-based scoring functions rely heavily on the training dataset (Zhang & Zhang, 2017). By docking the protein and ligand using classical docking software and subsequently rescoring the docked structure with machine-learning-based scoring functions, the overall accuracy of the process can be improved.

Scoring Function Evaluation

Various search algorithms and scoring functions have been devised and integrated into VS software applications, as outlined in Table 2. Due to the multitude of available methods, it is crucial to conduct a comparative evaluation of these scoring functions. In 2009, Renxiao Wang and colleagues introduced the comparative assessment of scoring functions (CASF) benchmarks as a standardized framework (Cheng et al., 2009). These benchmarks utilize the PDBbind database and have evolved, starting with CASF-2007 (based on PDBbind version 2007) and progressing to CASF-2013 and CASF-2016 (Y. Li, Han, et al., 2014; Y. Li, Liu, et al., 2014; Su et al., 2019). The scoring process is dissociated from the conformational sampling process to assess the performance of scoring functions. Evaluating scoring function effectiveness encompasses multiple aspects, including scoring power, ranking power, docking power, and screening power. By employing this comprehensive approach, the scoring functions can be rigorously scrutinized and compared

Table 2 Overview of the critical features of widely utilized VS software options. The first column contains the software used and its reference. The second column contains the type of license of the software. The last two columns list the algorithm used for docking and the scoring function. (adapted from (Maia et al., 2020)**)**

Software	License	Protein Flexibility	Docking Algorithm	Scoring Function
AutoDock4 (Morris et al., 2009)	Free for academic use	Yes	Genetic Algorithm Simulated Annealing	Hybrid (Physics-based and Empirical
Autodock Vina (Trott & Olson, 2009)	Open-source	Yes	Genetic Algorithm Simulated Annealing Local Search Particle Swarm	Hybrid (Empirical and Knowledge-based)
DOCK 6 (Allen et al., 2015)	Free for academic use	Yes	Shape Fitting (Sphere Sets) Lowest energy binding	Physics-based
SwissDock/EADock DSS (Grosdidier et al., 2011)	Free for academic use	No	Stochastic (Tabu search based) Local Search Combination of broad and local search of the conformational space	Empirical
eHiTS (Zsoldos et al., 2007)	Freeware for academic use	No	Exhaustive Search	Physics-based
FITTED (Corbeil et al., 2007)	Commercial	Yes	Genetic Algorithm	Hybrid (Empirical and Knowledge-based)
FlexX (Rarey et al., 1996)	Commercial	No	Incremental Construction	Physics-based
FLIPDock (Zhao & Sanner, 2007)	Freeware for academic use	Yes	Genetic Algorithm	Empirical
Fred (McGann, 2011)	Free for academic use	No	Exhaustive Search Algorithm	Physics-based
GalaxyDock2 (Shin et al., 2013)	Freeware	Yes	Conformational Analysis Genetic Algorithm	Hybrid
GeauxDock (Fang et al., 2016)	Open-source	Yes	Monte Carlo	Physics-based
GlamDock (Tietze & Apostolakis, 2007)	Freeware	No	Monte Carlo Simulated Annealing Local Search Conformational Analysis	Hybrid (Empirical and Knowledge-based)
Glide (Friesner et al., 2004)	Commercial	Yes	Conformational Analysis Monte Carlo Sampling	Empirical
Gold (Verdonk et al., 2003)	Commercial	Yes	Genetic Algorithm	Empirical
ICM (Abagyan et al., 1994)	Commercial	Yes	Monte Carlo Minimization	Physics-based
iGEMDOCK/GEMDOCK (Hsu et al., 2011)	Freeware	Yes	Genetic Algorithm	Physics-based
LigandFit (Montes et al., 2007)	Commercial	Yes	Monte Carlo	Empirical
MOE (Vilar et al., 2008)	Commercial	Yes	Conformational Analysis	Hybrid (Empirical and Physics-based)
ParaDockS (Meier et al., 2010)	Freeware	No	Genetic Algorithm	Hybrid (Empirical and Physics-based)
rDOCK (Ruiz-Carmona et al., 2014)	Open-source	Yes	Genetic Algorithm Monte Carlo Simplex Minimization	Hybrid (Empirical and Knowledge-based)
SLIDE (Schnecke & Kuhn, 2000)	Free for academic use	Yes	Conformational Analysis	Empirical
Surflex (Spitzer & Jain, 2012)	Commercial	Yes	Incremental Construction	Empirical
Sybyl-X (*Certara Software | Accelerated Drug Process through Software*, n.d.)	Commercial	Yes	Incremental Construction	Physics-based
vLifeDock (Chopade, 2015)	Commercial	Yes	Genetic Algorithm	Empirical

In a study by Su et al. in 2019, the CASF-2016 benchmark was employed to evaluate a panel of 25 scoring functions utilized in various VS software (Table 3). This assessment aims to analyze the scoring functions' "average" performance using a test set comprising 285 diverse protein-ligand complexes with well-established crystal structures and binding affinities (Su et al., 2019). Consistent processing methods were applied to all protein-ligand complex structures and the decoy ligand binding poses. The evaluation of the scoring functions was based on four distinct criteria:

- Scoring power: This criterion measures the capability of a scoring function to generate binding scores that exhibit a linear correlation with experimental binding data.

- Ranking power: The ranking power criterion focuses on the ability of a scoring function to accurately rank the known ligands of a specific target protein based on their binding affinities, assuming precise knowledge of the binding poses of these ligands.

- Docking power: The docking power criterion assesses the capacity of a scoring function to identify the native binding pose of a ligand among a set of computer-generated decoys. Ideally, the native binding pose should be ranked as the top-scoring pose.

- Screening power: This criterion evaluates the ability of a scoring function to discern true binders to a given target protein from a pool of random molecules during the screening process.

This study utilizes Δ*SAS* as the reference model for comparing other scoring functions. Figure 2 illustrates the overall performance of the scoring functions across four distinct criteria. The results reveal good docking power among the tested scoring functions, with several achieving success rates exceeding 70%. However, even the top-ranked scoring functions (excluding $\Delta_{Vina}RF_{20}$) only yield moderate correlation coefficients of approximately 0.60 in both the scoring power and ranking power tests. The screening power of the scoring function is notably weaker, with the highest-ranked functions achieving success rates of around 40%. While most scoring functions excel in one or two specific aspects, certain functions, such as $\Delta_{Vina}RF_{20}$ and ChemPLP@GOLD, exhibit a more balanced performance across all evaluated aspects. Highlighting $\Delta_{Vina}RF_{20}$ utilizes a combination of the empirical scoring function in Autodock Vina and a machine-learning model.

Table 3 Summary of scoring functions tested in CASF-2016 (Adapted from (Su et al., 2019)**)**

Scoring Function	Source	Classification	Reference
Jain	Discovery Studio (Version 4.1)	Empirical scoring function	(Jain, 1996)
LigScore1/Ligscore2		Empirical scoring function	(Krammer et al., 2005)
PMF/PMF04		Knowledge-based potential	(Muegge, 2006; Muegge & Martin, 1999)
LUDI1/LUDI2/LUDI3		Empirical scoring function	(H. J. Böhm, 1998; H.-J. Böhm, 1994)
PLP1/PLP2		Empirical scoring function	(G. Verkhivker et al., 1995; G. M. Verkhivker et al., 2000)
GoldScore	GOLD (Version 5.2)	Physics-based function	(Jones et al., 1997)
ChemScore		Empirical scoring function	(Baxter et al., 1998; Eldridge et al., 1997)
ChemPLP		Empirical scoring function	(Korb et al., 2009)
ASP		Knowledge-based potential	(Mooij & Verdonk, 2005)
G-Score	SYBYL (Version 8.1)	Physics-based function	(Jones et al., 1997)
PMF		Knowledge-based potential	(Muegge, 2006; Muegge & Martin, 1999)
D-Score		Physics-based function	(Meng et al., 1992)
ChemScore		Empirical scoring function	(Baxter et al., 1998; Eldridge et al., 1997)
GlideScore-SP	Schrodinger (Version 2016)	Empirical scoring function	(Friesner et al., 2004, 2006; Halgren et al., 2004)
GlideScore-XP		Empirical scoring function	(Friesner et al., 2004, 2006; Halgren et al., 2004)
London-dG	MOE (Version 2015)	Empirical scoring function	MOE User Manual
ASE		Empirical scoring function	
Affinity		Empirical scoring function	
Alpha-HB		Empirical scoring function	
GBVI-WSA-dG		Physics-based function	
Autodock vina	Autodock Vina (Version 1.1.12)	Empirical scoring function	(Trott & Olson, 2009)
X-Score (HP/HM/HS)	X-Score (Version 1.3) from the author	Empirical scoring function	(R. Wang et al., 2002)
$\Delta_{Vina}RF_{20}$	From the author	Descriptor-based Machine-learning model	(C. Wang & Zhang, 2017)
DrugScore2018	From the author	Knowledge-based potential	(Dittrich et al., 2019)
$DrugScore^{CSD}$		Knowledge-based potential	(Velec et al., 2005)
ΔSAS	In-0house software	Single descriptor	NA.

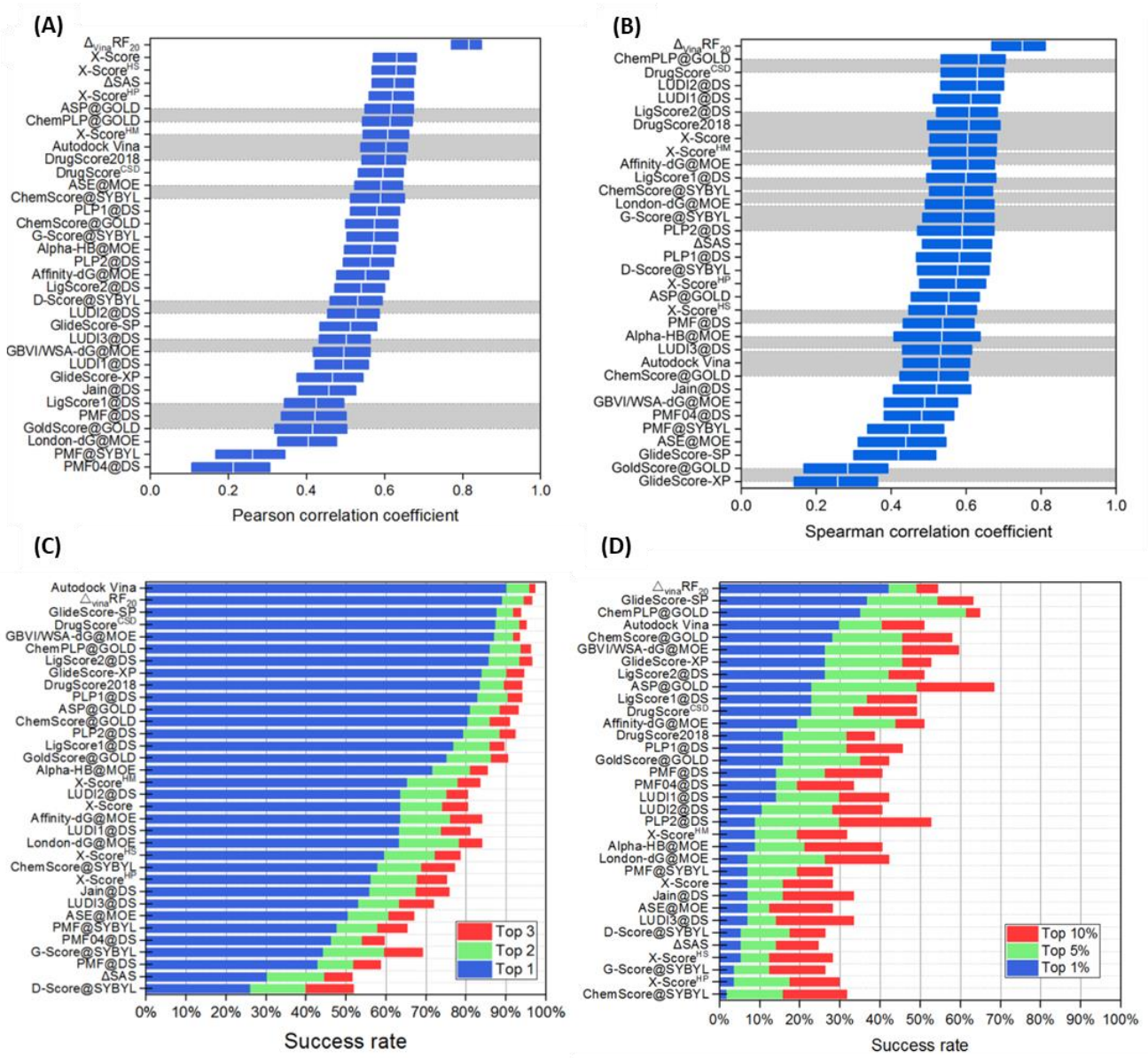

Figure 2 Performance evaluation of scoring functions in various tests. (A) Scoring Power Test: Pearson correlation coefficients (represented by short white lines) and their 90% confidence intervals (blue bars) for each scoring function. (B) Ranking Power Test: Average Spearman correlation coefficients (ρ) and their 90% confidence intervals (blue bars) were obtained on 57 target proteins. (C) Docking Power Test: Success rates of each scoring function for detecting native ligand binding poses (RMSD < 2.0 Å). Blue, green, and red bars represent success rates for the top one, top two, and top three binding poses, respectively. (D) Forward Screening Test: Success rates of each scoring function in detecting the highest-affinity ligand for a given target protein. Blue, green, and red bars represent success rates for considering the top 1%, 5%, and 10% candidates in screening, respectively. Scoring functions are ranked in descending order in all figures. (Figures obtained from (Su et al., 2019)**)**

Future Directions

The utilization of CADD has significantly enhanced the drug development process. However, scoring functions still give crude approximations even with the advancements made in VS and molecular docking. Therefore, further development is needed to improve the outcome of scoring functions. Bellow, advancements and possible suggestions are given to improve scoring functions and mitigate false positives/negatives in VS and molecular docking.

Molecular Dynamics and DFT Calculations

As mentioned above, physics-based scoring functions offer higher accuracy than other scoring functions. Molecular dynamics (MD) and density functional theory (DFT) calculations can further improve the outcome of physics-based scoring functions.

MD simulations are crucial calculations that come after virtual screening simulations. They should be viewed as an advanced and complementary technique to docking. Additionally, they can be employed before docking to explore the conformational dynamics of a protein molecule and facilitate the sampling and clustering of different conformations relevant to ligand binding (Menchon et al., 2018). Overall, molecular dynamics plays a vital role in the discovery and development of commercial drugs by offering valuable insights into the dynamic behavior of biological molecules and the interactions between drugs and their targets.

DFT has emerged as a successful and promising approach. Extensive evidence supports the accuracy of DFT in describing the electronic and structural properties of small molecules through the computation of their electronic structures. DFT makes it possible to calculate orbital energies, such as the highest occupied molecular orbital and lowest unoccupied molecular orbital (Sakkiah & Lee, 2012).

MD and DFT calculations are applied on a limited scale in VS and molecular docking. The main obstacle in using these calculations is computational cost and time (J. Li et al., 2019; Maia et al., 2020). Quantum computers may be the solution to these limitations. Using superpositions, entanglement, and quantum gates, quantum computers can cut the computational cost by the square root a traditional supercomputer would take (H. Liu et al., 2022). Although quantum computers are not fully there yet, the development of quantum hardware and algorithms promises positive prospects for virtual screening and molecular docking (H. Liu et al., 2022).

Ligand Conformations and Protein Flexibility

Within contemporary docking software, search algorithms are employed to determine the optimal conformation of a ligand by minimizing its energy state. Similarly, protein targets are often treated as rigid entities, utilizing known conformations. However, relying solely on scoring the ligand and protein in these static states introduces a biased assumption. Ligands will not always adopt their optimal configuration, and target proteins exhibit inherent flexibility. Consequently, it is imperative to account for the dynamic nature of ligands and proteins to avoid potential limitations and inaccuracies in docking studies.

One potential approach to address this challenge is to evaluate a diverse range of ligand conformations, calculate their respective scores, and then average them. The Boltzmann distribution can be employed to enhance this strategy, assigning higher weights to conformations with greater probabilities. By incorporating this probabilistic weighting, a more realistic representation of the ligand's activity can be achieved, considering the dynamic nature of ligand binding. This refined approach holds the potential to provide improved insights into ligand behavior and enhance the accuracy of activity predictions in docking studies.

Moreover, to account for the flexibility of protein targets, docking programs have incorporated techniques such as soft potentials or ensemble docking. Soft potentials are modifications of the

Lennard-Jones potential, commonly used to describe the intermolecular interactions between atoms in docking calculations. The Lennard-Jones potential comprises short-range attractive and long-range repulsive terms (van der Waals interactions) (Adams, 2001). The repulsive term prevents atoms from overlapping and accounts for steric clashes. However, strict adherence to the Lennard-Jones potential in protein-ligand interactions can result in overly rigid docking poses and unrealistic structures. To address this limitation, soft potentials are introduced to modify the repulsive term of the Lennard-Jones potential. Soft potentials allow for limited overlaps between the ligand and protein molecules, providing more flexibility and accommodating minor distortions or deformations in the binding interface. By incorporating soft potentials, docking programs can generate more realistic representations of molecular interactions and improve the accuracy of ligand binding predictions. This modification enables exploring a broader range of binding modes and increases the chances of capturing relevant binding poses (Guedes et al., 2018).

Ensemble docking, on the other hand, takes into consideration the flexibility of the protein by docking the ligand onto multiple protein conformations. This approach explores a range of potential binding modes and accounts for the conformational variability of the protein (Guedes et al., 2018). Furthermore, a flexible docking method has been developed, incorporating experimentally derived protein conformations and integrating Boltzmann-weighted energy penalties associated with protein flexibility into the scoring function. This strategy enables a more comprehensive assessment of ligand binding and protein-ligand interactions by considering the dynamic nature of both the ligand and the protein target (Fischer et al., 2014).

While these implementations enhance scoring outcomes, they also necessitate higher computational costs. However, there are potential solutions to address this challenge, namely through advancements in computational power or quantum computing.

Consensus Scoring and Tailored Scoring Functions

As illustrated in Table 2, there exists a multitude of docking software encompassing various algorithms and scoring functions, each with varying performance characteristics. Ideally, a universal program capable of accurately predicting all ligand-protein interactions would be desirable. To foster advancements in CADD, the Drug Design Data Resource (D3R) has organized four Grand Challenges to date. These challenges provide a platform for participants to benchmark and compete with their docking tools against a diverse set of pertinent protein targets (Parks et al., 2020). By participating in these challenges, researchers are driven to enhance existing methods, pushing the boundaries of innovation and vying to establish their tools as leading contenders in the field of CADD.

To enhance the accuracy of scoring outcomes, consensus scoring approaches have been employed, wherein multiple scoring functions are combined using statistical methods or voting schemes. Wang and Wang (2001) provided a theoretical foundation for the efficacy of affinity predictions, demonstrating that the mean value of repeated samplings converges toward the actual value (R. Wang & Wang, 2001). This concept suggests that by aggregating multiple scoring functions, a more reliable estimation of ligand-protein affinity can be achieved, resulting in improved predictive performance in virtual screening and drug discovery endeavors.

Lastly, target-specific scoring functions have led to significant advancements in VS and docking studies. Numerous investigations have been conducted to evaluate the performance of docking software under specific conditions, such as charged binding pockets (Deng & Verlinde, 2008) or DNA-ligand systems (de Oliveira et al., 2022), as well as overall performance assessments (Onodera et al., 2007; Z. Wang et al., 2016). By employing specific data training sets, scoring functions can better account for the specific interactions associated with a particular target class.

While target-specific scoring functions have demonstrated promising outcomes, it is essential to acknowledge the potential limitations and sources of inaccuracy that arise from the requirement for a

substantial training set to develop a robust scoring function. The necessity for an extensive training set can present challenges, mainly when the availability of experimental structures is limited. To overcome this limitation, protein-ligand conformations obtained from docking experiments are frequently employed for training target-specific scoring functions. This approach enables the utilization of a more extensive dataset, facilitating the derivation of reliable scoring functions. Consequently, the issue of limited experimental data is mitigated, thereby supporting the development of accurate and effective scoring functions for virtual screening and drug discovery endeavors.

Conclusion

The accurate prediction of protein-ligand binding affinity is of utmost importance in CADD. Developing precise search algorithms and scoring functions has been a critical focus in this field. In recent years, the availability of an expanding array of data sources comprising measured binding affinities and datasets containing active, decoy, and true inactive compounds has facilitated the derivation of more effective scoring functions. Although significant progress has been made, further enhancements are necessary to achieve a comprehensive and robust search algorithm and scoring function suitable for hit-to-lead optimization and *de novo* design studies. By continuously refining these computational tools, CADD researchers aim to enhance the accuracy and reliability of predicting protein-ligand binding affinities, ultimately facilitating the discovery and optimization of novel drug candidates.

Finally, The CADD tools necessitate a diverse range of expertise from researchers to navigate through the various steps of the process successfully. This includes selecting and preparing targets and ligands, analyzing the results, and comprehensively understanding computational methods, chemistry, and biology. Therefore, researchers with multidisciplinary expertise are essential in harnessing the full potential of CADD tools. Their proficiency in various fields enables them to make informed decisions, optimize the drug discovery process, and ultimately contribute to advancing pharmaceutical research and development.

Bibliography

Abagyan, R., Totrov, M., & Kuznetsov, D. (1994). ICM?A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*, *15*(5), 488–506. https://doi.org/10.1002/jcc.540150503

Abdel-Basset, M., Abdel-Fatah, L., & Sangaiah, A. K. (2018). Metaheuristic algorithms: A comprehensive review. In *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications* (pp. 185–231). Elsevier. https://doi.org/10.1016/B978-0-12-813314-9.00010-4

Abdo, A., Chen, B., Mueller, C., Salim, N., & Willett, P. (2010). Ligand-Based Virtual Screening Using Bayesian Networks. *Journal of Chemical Information and Modeling*, *50*(6), 1012–1020. https://doi.org/10.1021/ci100090p

Adams, J. B. (2001). Bonding Energy Models. In *Encyclopedia of Materials: Science and Technology* (pp. 763–767). Elsevier. https://doi.org/10.1016/B0-08-043152-6/00146-7

Allen, W. J., Balius, T. E., Mukherjee, S., Brozell, S. R., Moustakas, D. T., Lang, P. T., Case, D. A., Kuntz, I. D., & Rizzo, R. C. (2015). DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, *36*(15), 1132–1156. https://doi.org/10.1002/jcc.23905

Ashtawy, H. M., & Mahapatra, N. R. (2018). Task-Specific Scoring Functions for Predicting Ligand Binding Poses and Affinity and for Screening Enrichment. *Journal of Chemical Information and Modeling*, *58*(1), 119–133. https://doi.org/10.1021/acs.jcim.7b00309

Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R., & Eldridge, M. D. (1998). Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins*, *33*(3), 367–382.

Bianchi, L., Dorigo, M., Gambardella, L. M., & Gutjahr, W. J. (2009). A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, *8*(2), 239–287. https://doi.org/10.1007/s11047-008-9098-4

Böhm, H. J. (1998). Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *Journal of Computer-Aided Molecular Design*, *12*(4), 309–323. https://doi.org/10.1023/a:1007999920146

Böhm, H.-J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, *8*(3), 243–256. https://doi.org/10.1007/BF00126743

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chao, H., Chen, L., Craig, P. A., Crichlow, G. V, Dalenberg, K., Duarte, J. M., Dutta, S., Fayazi, M., Feng, Z., Flatt, J. W., Ganesan, S., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., … Zardecki, C. (2023). RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Research*, *51*(D1), D488–D508. https://doi.org/10.1093/nar/gkac1077

Certara Software | Accelerated drug process through software. (n.d.). Retrieved 19 May 2023, from https://www.certara.com/software/

Chaskar, P., Zoete, V., & Röhrig, U. F. (2014). Toward On-The-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: Development and Benchmark of a Scoring Function. *Journal of Chemical Information and Modeling*, *54*(11), 3137–3152. https://doi.org/10.1021/ci5004152

Cheng, T., Li, X., Li, Y., Liu, Z., & Wang, R. (2009). Comparative assessment of scoring functions on a diverse test set. *Journal of Chemical Information and Modeling*, *49*(4), 1079–1093. https://doi.org/10.1021/ci9000053

Chopade, A. R. (2015). Molecular Docking Studies of Phytocompounds From Phyllanthus Species as Potential Chronic Pain Modulators. *Scientia Pharmaceutica*, *83*(2), 243–267. https://doi.org/10.3797/scipharm.1408-10

Corbeil, C. R., Englebienne, P., & Moitessier, N. (2007). Docking Ligands into Flexible and Solvated Macromolecules. 1. Development and Validation of FITTED 1.0. *Journal of Chemical Information and Modeling*, *47*(2), 435–449. https://doi.org/10.1021/ci6002637

Cruz-Monteagudo, M., Medina-Franco, J. L., Pérez-Castillo, Y., Nicolotti, O., Cordeiro, M. N. D. S., & Borges, F. (2014). Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today*, *19*(8), 1069–1080. https://doi.org/10.1016/j.drudis.2014.02.003

de Oliveira, T. A., Medaglia, L. R., Maia, E. H. B., Assis, L. C., de Carvalho, P. B., da Silva, A. M., & Taranto, A. G. (2022). Evaluation of Docking Machine Learning and Molecular Dynamics Methodologies for DNA-Ligand Systems. *Pharmaceuticals*, *15*(2), 132. https://doi.org/10.3390/ph15020132

Deng, W., & Verlinde, C. L. M. J. (2008). Evaluation of Different Virtual Screening Programs for Docking in a Charged Binding Pocket. *Journal of Chemical Information and Modeling*, *48*(10), 2010–2020. https://doi.org/10.1021/ci800154w

Dittrich, J., Schmidt, D., Pfleger, C., & Gohlke, H. (2019). Converging a Knowledge-Based Scoring Function: DrugScore ^{2018}. *Journal of Chemical Information and Modeling*, *59*(1), 509–521. https://doi.org/10.1021/acs.jcim.8b00582

Doucet, N., & Pelletier, J. N. (2007). Simulated annealing exploration of an active-site tyrosine in TEM-1β-lactamase suggests the existence of alternate conformations. *Proteins: Structure, Function, and Bioinformatics*, *69*(2), 340–348. https://doi.org/10.1002/prot.21485

Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, *11*(5), 425–445. https://doi.org/10.1023/A:1007996124545

Fang, Y., Ding, Y., Feinstein, W. P., Koppelman, D. M., Moreno, J., Jarrell, M., Ramanujam, J., & Brylinski, M. (2016). GeauxDock: Accelerating Structure-Based Virtual Screening with Heterogeneous Computing. *PLOS ONE*, *11*(7), e0158898. https://doi.org/10.1371/journal.pone.0158898

Ferreira, L., dos Santos, R., Oliva, G., & Andricopulo, A. (2015). Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, *20*(7), 13384–13421. https://doi.org/10.3390/molecules200713384

Fischer, M., Coleman, R. G., Fraser, J. S., & Shoichet, B. K. (2014). Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nature Chemistry*, *6*(7), 575–583. https://doi.org/10.1038/nchem.1954

Fornabaio, M., Spyrakis, F., Mozzarelli, A., Cozzini, P., Abraham, D. J., & Kellogg, G. E. (2004). Simple, Intuitive Calculations of Free Energy of Binding for Protein−Ligand Complexes. 3. The Free Energy Contribution of Structural Water Molecules in HIV-1 Protease Complexes. *Journal of Medicinal Chemistry*, *47*(18), 4507–4516. https://doi.org/10.1021/jm030596b

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., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739–1749. https://doi.org/10.1021/jm0306430

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. https://doi.org/10.1021/jm051256o

Gowthaman, R., Lyskov, S., & Karanicolas, J. (2015). DARC 2.0: Improved Docking and Virtual Screening at Protein Interaction Sites. *PLOS ONE*, *10*(7), e0131612. https://doi.org/10.1371/journal.pone.0131612

Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, *39*(suppl), W270–W277. https://doi.org/10.1093/nar/gkr366

Guedes, I. A., Pereira, F. S. S., & Dardenne, L. E. (2018). Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. In *Frontiers in Pharmacology* (Vol. 9, Issue SEP). Frontiers Media S.A. https://doi.org/10.3389/fphar.2018.01089

H. Haga, J., Ichikawa, K., & Date, S. (2016). Virtual Screening Techniques and Current Computational Infrastructures. *Current Pharmaceutical Design*, *22*(23), 3576–3584. https://doi.org/10.2174/1381612822666160414142530

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, *47*(7), 1750–1759. https://doi.org/10.1021/jm030644s

Harrison, R. L., Granja, C., & Leroy, C. (2010). *Introduction to Monte Carlo Simulation*. 17–21. https://doi.org/10.1063/1.3295638

Hatmal, M. M., & Taha, M. O. (2017). Simulated annealing molecular dynamics and ligand–receptor contacts analysis for pharmacophore modeling. *Future Medicinal Chemistry*, *9*(11), 1141–1159. https://doi.org/10.4155/fmc-2017-0061

Hönig, S. M. N., Lemmen, C., & Rarey, M. (2023). Small molecule superposition: A comprehensive overview on pose scoring of the latest methods. *WIREs Computational Molecular Science*, *13*(2). https://doi.org/10.1002/wcms.1640

Horvath, D. (1997). *A Virtual Screening Approach Applied to the Search for Trypanothione Reductase Inhibitors*.

Hsu, K.-C., Chen, Y.-F., Lin, S.-R., & Yang, J.-M. (2011). iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *BMC Bioinformatics*, *12*(S1), S33. https://doi.org/10.1186/1471-2105-12-S1-S33

Huang, S.-Y., & Zou, X. (2006). An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials. *Journal of Computational Chemistry*, *27*(15), 1866–1875. https://doi.org/10.1002/jcc.20504

Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of Chemical Information and Modeling*, *60*(12), 6065–6073. https://doi.org/10.1021/acs.jcim.0c00675

Jain, A. N. (1996). Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *Journal of Computer-Aided Molecular Design*, *10*(5), 427–440. https://doi.org/10.1007/BF00124474

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. https://doi.org/10.1006/jmbi.1996.0897

Kadukova, M., & Grudinin, S. (2017). Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. *Journal of Computer-Aided Molecular Design*, *31*(10), 943–958. https://doi.org/10.1007/s10822-017-0068-8

Kerzmann, A., Neumann, D., & Kohlbacher, O. (2006). SLICK − Scoring and Energy Functions for Protein−Carbohydrate Interactions. *Journal of Chemical Information and Modeling*, *46*(4), 1635–1642. https://doi.org/10.1021/ci050422y

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research*, *51*(D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956

Korb, O., Stützle, T., & Exner, T. E. (2006). *PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design* (pp. 247–258). https://doi.org/10.1007/11839088_22

Korb, O., Stützle, T., & Exner, T. E. (2009). Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, *49*(1), 84–96. https://doi.org/10.1021/ci800298z

Korkmaz, S., Zararsiz, G., & Goksuluk, D. (2015). MLViS: A Web Tool for Machine Learning-Based Virtual Screening in Early-Phase of Drug Discovery and Development. *PLOS ONE*, *10*(4), e0124600. https://doi.org/10.1371/journal.pone.0124600

Krammer, A., Kirchhoff, P. D., Jiang, X., Venkatachalam, C. M., & Waldman, M. (2005). LigScore: a novel scoring function for predicting binding affinities. *Journal of Molecular Graphics and Modelling*, *23*(5), 395–407. https://doi.org/10.1016/j.jmgm.2004.11.007

Leach, A. R., Gillet, V. J., Lewis, R. A., & Taylor, R. (2010). Three-Dimensional Pharmacophore Methods in Drug Discovery. *Journal of Medicinal Chemistry*, *53*(2), 539–558. https://doi.org/10.1021/jm900817u

Leguizamón, G. (2011). Arthur K. Kordon: Applying computational intelligence: how to create value. *Genetic Programming and Evolvable Machines*, *12*(1), 85–86. https://doi.org/10.1007/s10710-010-9120-3

Li, J., Fu, A., & Zhang, L. (2019). An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. In *Interdisciplinary Sciences – Computational Life Sciences* (Vol. 11, Issue 2, pp. 320–328). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s12539-019-00327-w

Li, Y., Han, L., Liu, Z., & Wang, R. (2014). Comparative assessment of scoring functions on an updated benchmark: 2. evaluation methods and general results. *Journal of Chemical Information and Modeling*, *54*(6), 1717–1736. https://doi.org/10.1021/ci500081m

Li, Y., Liu, Z., Li, J., Han, L., Liu, J., Zhao, Z., & Wang, R. (2014). Comparative Assessment of Scoring Functions on an Updated Benchmark: 1. Compilation of the Test Set. *Journal of Chemical Information and Modeling*, *54*(6), 1700–1716. https://doi.org/10.1021/ci500080q

Lill, M. (2013). Virtual screening in drug design. *Methods in Molecular Biology*, *993*, 1–12. https://doi.org/10.1007/978-1-62703-342-8_1

Liu, H., Low, G. H., Steiger, D. S., Häner, T., Reiher, M., & Troyer, M. (2022). Prospects of quantum computing for molecular sciences. *Materials Theory*, *6*(1). https://doi.org/10.1186/s41313-021-00039-z

Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., Nie, W., Liu, Y., & Wang, R. (2015). PDB-wide collection of binding data: Current status of the PDBbind database. *Bioinformatics*, *31*(3), 405–412. https://doi.org/10.1093/bioinformatics/btu626

Ma, D.-L., Chan, D. S.-H., & Leung, C.-H. (2013). Drug repositioning by structure-based virtual screening. *Chemical Society Reviews*, *42*(5), 2130. https://doi.org/10.1039/c2cs35357a

Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Frontiers in Chemistry*, *8*. https://doi.org/10.3389/fchem.2020.00343

Mayr, L. M., & Bojanic, D. (2009). Novel trends in high-throughput screening. In *Current Opinion in Pharmacology* (Vol. 9, Issue 5, pp. 580–588). https://doi.org/10.1016/j.coph.2009.08.004

McGann, M. (2011). FRED Pose Prediction and Virtual Screening Accuracy. *Journal of Chemical Information and Modeling*, *51*(3), 578–596. https://doi.org/10.1021/ci100436p

Meier, R., Pippel, M., Brandt, F., Sippl, W., & Baldauf, C. (2010). <scp>ParaDockS</scp> : A Framework for Molecular Docking with Population-Based Metaheuristics. *Journal of Chemical Information and Modeling*, *50*(5), 879–889. https://doi.org/10.1021/ci900467x

Menchon, G., Maveyraud, L., & Czaplicki, G. (2018). *Molecular Dynamics as a Tool for Virtual Ligand Screening* (pp. 145–178). https://doi.org/10.1007/978-1-4939-7756-7_9

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., Magariños, M. P., Mosquera, J. F., Mutowo, P., Nowotka, M., Gordillo-Marañón, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C. J., Segura-Cabrera, A., … Leach, A. R. (2019). ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research*, *47*(D1), D930–D940. https://doi.org/10.1093/nar/gky1075

Meng, E. C., Shoichet, B. K., & Kuntz, I. D. (1992). Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*, *13*(4), 505–524. https://doi.org/10.1002/jcc.540130412

Montes, M., Miteva, M. A., & Villoutreix, B. O. (2007). Structure-based virtual ligand screening with LigandFit: Pose prediction and enrichment of compound collections. *Proteins: Structure, Function, and Bioinformatics*, *68*(3), 712–725. https://doi.org/10.1002/prot.21405

Mooij, W. T. M., & Verdonk, M. L. (2005). General and targeted statistical potentials for protein-ligand interactions. *Proteins*, *61*(2), 272–287. https://doi.org/10.1002/prot.20588

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

Muegge, I. (2006). PMF Scoring Revisited. *Journal of Medicinal Chemistry*, *49*(20), 5895–5902. https://doi.org/10.1021/jm050038s

Muegge, I., & Martin, Y. C. (1999). A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *Journal of Medicinal Chemistry*, *42*(5), 791–804. https://doi.org/10.1021/jm980536j

N. Cavasotto, C. (2011). Homology Models in Docking and High-Throughput Docking. *Current Topics in Medicinal Chemistry*, *11*(12), 1528–1534. https://doi.org/10.2174/156802611795860951

Ng, M. C. K., Fong, S., & Siu, S. W. I. (2015). PSOVina: The hybrid particle swarm optimization algorithm for protein–ligand docking. *Journal of Bioinformatics and Computational Biology*, *13*(03), 1541007. https://doi.org/10.1142/S0219720015410073

Oliveira, T. A. de, Silva, M. P. da, Maia, E. H. B., Silva, A. M. da, & Taranto, A. G. (2023). Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods. *Drugs and Drug Candidates*, *2*(2), 311–334. https://doi.org/10.3390/ddc2020017

Onodera, K., Satou, K., & Hirota, H. (2007). Evaluations of molecular docking programs for virtual screening. *Journal of Chemical Information and Modeling*, *47*(4), 1609–1618. https://doi.org/10.1021/ci7000378

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., Ameriks, M. K., Mirzadegan, T., Burley, S. K., Amaro, R. E., & Gilson, M. K. (2020). D3R grand challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *Journal of Computer-Aided Molecular Design*, *34*(2), 99–119. https://doi.org/10.1007/s10822-020-00289-y

PDB Statistics: PDB Data Distribution by Experimental Method and Molecular Type. (n.d.). Retrieved 23 May 2023, from https://www.rcsb.org/stats/summary

Pence, H. E., & Williams, A. (2010). ChemSpider: An Online Chemical Information Resource. *Journal of Chemical Education*, *87*(11), 1123–1124. https://doi.org/10.1021/ed100697w

Peterson, Y. K., Wang, X. S., Casey, P. J., & Tropsha, A. (2009). Discovery of Geranylgeranyltransferase-I Inhibitors with Novel Scaffolds by the Means of Quantitative Structure−Activity Relationship Modeling, Virtual Screening, and Experimental Validation. *Journal of Medicinal Chemistry*, *52*(14), 4210–4220. https://doi.org/10.1021/jm8013772

Putatunda, C., Solanki, P., Pathania, S., Kumar, A., & Walia, A. (2023). Chapter 2 - Screening strategies. In A. K. Bhatt, R. K. Bhatia, & T. C. Bhalla (Eds.), *Basic Biotechniques for Bioprocess and Bioentrepreneurship* (pp. 23–46). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-816109-8.00002-7

Raha, K., Peters, M. B., Wang, B., Yu, N., Wollacott, A. M., Westerhoff, L. M., & Merz, K. M. (2007). The role of quantum mechanics in structure-based drug design. *Drug Discovery Today*, *12*(17–18), 725–731. https://doi.org/10.1016/j.drudis.2007.07.006

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology*, *261*(3), 470–489. https://doi.org/10.1006/jmbi.1996.0477

Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., Barril, X., Hubbard, R. E., & Morley, S. D. (2014). rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Computational Biology*, *10*(4), e1003571. https://doi.org/10.1371/journal.pcbi.1003571

Sabe, V. T., Ntombela, T., Jhamba, L. A., Maguire, G. E. M., Govender, T., Naicker, T., & Kruger, H. G. (2021). Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. In *European Journal of Medicinal Chemistry* (Vol. 224). Elsevier Masson s.r.l. https://doi.org/10.1016/j.ejmech.2021.113705

Sakkiah, S., & Lee, K. W. (2012). Pharmacophore-based virtual screening and density functional theory approach to identifying novel butyrylcholinesterase inhibitors. *Acta Pharmacologica Sinica*, *33*(7), 964–978. https://doi.org/10.1038/aps.2012.21

Schnecke, V., & Kuhn, L. A. (2000). Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design*, *20*(1), 171–190. https://doi.org/10.1023/A:1008737207775

Schneider, P., Tanrikulu, Y., & Schneider, G. (2009). Self-Organizing Maps in Drug Discovery: Compound Library Design, Scaffold-Hopping, Repurposing. *Current Medicinal Chemistry*, *16*(3), 258–266. https://doi.org/10.2174/092986709787002655

Sengupta, S., & Bandyopadhyay, S. (2012). APPLICATION OF SUPPORT VECTOR MACHINES IN VIRTUAL SCREENING. *International Journal for Computational Biology*, *1*(1), 56. https://doi.org/10.34040/IJCB.1.1.2012.20

Senn, H. M., & Thiel, W. (2009). QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition*, *48*(7), 1198–1229. https://doi.org/10.1002/anie.200802019

Shin, W.-H., Heo, L., Lee, J., Ko, J., Seok, C., & Lee, J. (2011). LigDockCSA: Protein-ligand docking using conformational space annealing. *Journal of Computational Chemistry*, *32*(15), 3226–3232. https://doi.org/10.1002/jcc.21905

Shin, W.-H., Kim, J.-K., Kim, D.-S., & Seok, C. (2013). GalaxyDock2: Protein-ligand docking using beta-complex and global optimization. *Journal of Computational Chemistry*, *34*(30), 2647–2656. https://doi.org/10.1002/jcc.23438

Spitzer, R., & Jain, A. N. (2012). Surflex-Dock: Docking benchmarks and real-world application. *Journal of Computer-Aided Molecular Design*, *26*(6), 687–699. https://doi.org/10.1007/s10822-011-9533-y

Statistics - PubChem. (n.d.). Retrieved 23 May 2023, from https://pubchem.ncbi.nlm.nih.gov/docs/statistics

Su, M., Yang, Q., Du, Y., Feng, G., Liu, Z., Li, Y., & Wang, R. (2019). Comparative Assessment of Scoring Functions: The CASF-2016 Update. *Journal of Chemical Information and Modeling*, *59*(2), 895–913. https://doi.org/10.1021/acs.jcim.8b00545

Suay-García, B., Bueso-Bordils, J. I., Falcó, A., Antón-Fos, G. M., & Alemán-López, P. A. (2022). Virtual Combinatorial Chemistry and Pharmacological Screening: A Short Guide to Drug Design. In *International Journal of Molecular Sciences* (Vol. 23, Issue 3). MDPI. https://doi.org/10.3390/ijms23031620

Tietze, S., & Apostolakis, J. (2007). GlamDock: Development and Validation of a New Docking Tool on Several Thousand Protein−Ligand Complexes. *Journal of Chemical Information and Modeling*, *47*(4), 1657–1672. https://doi.org/10.1021/ci7001236

Tin Kam Ho. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(8), 832–844. https://doi.org/10.1109/34.709601

Tresadern, G., Bemporad, D., & Howe, T. (2009). A comparison of ligand based virtual screening methods and application to corticotropin releasing factor 1 receptor. *Journal of Molecular Graphics and Modelling*, *27*(8), 860–870. https://doi.org/10.1016/j.jmgm.2009.01.003

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. https://doi.org/10.1002/jcc.21334

Velec, H. F. G., Gohlke, H., & Klebe, G. (2005). DrugScore ^{CSD} Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *Journal of Medicinal Chemistry*, *48*(20), 6296–6303. https://doi.org/10.1021/jm050436v

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, *52*(4), 609–623. https://doi.org/10.1002/prot.10465

Verkhivker, G., Appelt, K., Freer, S. T., & Villafranca, J. E. (1995). Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Engineering Design and Selection*, *8*(7), 677–691. https://doi.org/10.1093/protein/8.7.677

Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., Freer, S. T., Larson, V., Luty, B. A., Marrone, T., & Rose, P. W. (2000). Deciphering common failures in molecular docking of ligand-protein complexes. *Journal of Computer-Aided Molecular Design*, *14*(8), 731–751. https://doi.org/10.1023/a:1008158231558

Vilar, S., Cozza, G., & Moro, S. (2008). Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. *Current Topics in Medicinal Chemistry*, *8*(18), 1555–1572. https://doi.org/10.2174/156802608786786624

Wang, C., & Zhang, Y. (2017). Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest. *Journal of Computational Chemistry*, *38*(3), 169–177. https://doi.org/10.1002/jcc.24667

Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, *16*(1), 11–26. https://doi.org/10.1023/a:1016357811882

Wang, R., & Wang, S. (2001). How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *Journal of Chemical Information and Computer Sciences*, *41*(5), 1422–1426. https://doi.org/10.1021/ci010025x

Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S., & Hou, T. (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics*, *18*(18), 12964–12975. https://doi.org/10.1039/C6CP01555G

Willett, P. (2006). Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*, *11*(23–24), 1046–1053. https://doi.org/10.1016/j.drudis.2006.10.005

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., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Xia, J., Feng, B., Shao, Q., Yuan, Y., Wang, X., Chen, N., & Wu, S. (2017). Virtual Screening against Phosphoglycerate Kinase 1 in Quest of Novel Apoptosis Inhibitors. *Molecules*, *22*(6), 1029. https://doi.org/10.3390/molecules22061029

Zhang, L., Ai, H.-X., Li, S.-M., Qi, M.-Y., Zhao, J., Zhao, Q., & Liu, H.-S. (2017). Virtual screening approach to identifying influenza virus neuraminidase inhibitors using molecular docking combined with machine-learning-based scoring function. *Oncotarget*, *8*(47), 83142–83154. https://doi.org/10.18632/oncotarget.20915

Zhang, L., & Zhang, S. (2017). Using game theory to investigate the epigenetic control mechanisms of embryo development. *Physics of Life Reviews*, *20*, 140–142. https://doi.org/10.1016/j.plrev.2017.01.007

Zhao, Y., & Sanner, M. F. (2007). FLIPDock: Docking flexible ligands into flexible receptors. *Proteins: Structure, Function, and Bioinformatics*, *68*(3), 726–737. https://doi.org/10.1002/prot.21423

Zheng, Z., & Merz, K. M. (2011). Ligand Identification Scoring Algorithm (LISA). *Journal of Chemical Information and Modeling*, *51*(6), 1296–1306. https://doi.org/10.1021/ci2000665

Zheng, Z., & Merz, K. M. (2013). Development of the Knowledge-Based and Empirical Combined Scoring Algorithm (KECSA) To Score Protein–Ligand Interactions. *Journal of Chemical Information and Modeling*, *53*(5), 1073–1083. https://doi.org/10.1021/ci300619x

Zhu, H., Zhang, Y., Li, W., & Huang, N. (2022). A Comprehensive Survey of Prospective Structure-Based Virtual Screening for Early Drug Discovery in the Past Fifteen Years. *International Journal of Molecular Sciences*, *23*(24), 15961. https://doi.org/10.3390/ijms232415961

Zsoldos, Z., Reid, D., Simon, A., Sadjad, S. B., & Johnson, A. P. (2007). eHiTS: A new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics and Modelling*, *26*(1), 198–212. https://doi.org/10.1016/j.jmgm.2006.06.002