



Review

Key Topics in Molecular Docking for Drug Design

Pedro H. M. Torres ¹, Ana C. R. Sodero ², Paula Jofily ³ and Floriano P. Silva-Jr ^{4,*}

¹ Department of Biochemistry, University of Cambridge, Cambridge, CB2 1GA, UK

² Department of Drugs and Medicines; School of Pharmacy; Federal University of Rio de Janeiro, Rio de Janeiro 21949-900, RJ, Brazil

³ Laboratório de Modelagem e Dinâmica Molecular, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21949-900, RJ, Brazil

⁴ Laboratório de Bioquímica Experimental e Computacional de Fármacos, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro 21949-900, RJ, Brazil

* Correspondence: floriano@ioc.fiocruz.br; Tel.: +55-21-38658248

Received: 3 June 2019; Accepted: 10 July 2019; Published: 15 September 2019

Abstract: Molecular docking has been widely employed as a fast and inexpensive technique in the past decades, both in academic and industrial settings. Although this discipline has now had enough time to consolidate, many aspects remain challenging and there is still not a straightforward and accurate route to readily pinpoint true ligands among a set of molecules, nor to identify with precision the correct ligand conformation within the binding pocket of a given target molecule. Nevertheless, new approaches continue to be developed and the volume of published works grows at a rapid pace. In this review, we present an overview of the method and attempt to summarise recent developments regarding four main aspects of molecular docking approaches: (i) the available benchmarking sets, highlighting their advantages and caveats, (ii) the advances in consensus methods, (iii) recent algorithms and applications using fragment-based approaches, and (iv) the use of machine learning algorithms in molecular docking. These recent developments incrementally contribute to an increase in accuracy and are expected, given time, and together with advances in computing power and hardware capability, to eventually accomplish the full potential of this area.

Keywords: computer-aided drug design; structure-based drug design; benchmarking sets; consensus methods; fragment-based; machine learning

1. Introduction

Molecular docking is a method which analyses the conformation and orientation (referred together as the “pose”) of molecules into the binding site of a macromolecular target. Searching algorithms generate possible poses, which are ranked by scoring functions [1]. Several software were developed during the last decades, amongst which are some well-known examples, such as AutoDock [2], AutoDock Vina [3], DockThor [4,5], GOLD [6,7], FlexX [8] and Molegro Virtual Docker [9].

The first step in a docking calculation is to obtain the target structure, which commonly consists of a large biological molecule (protein, DNA or RNA) [10] (Figure 1). The structures of these macromolecules can be readily retrieved from the Protein Data Bank (PDB) [11], which provides access to 3D atomic coordinates obtained by experimental methods. However, it is not unusual that the experimental 3D structure of the target is not available. In order to overcome this issue, computational prediction methods, such as comparative and ab initio modelling can be used to obtain the three-dimensional structure of proteins [1].

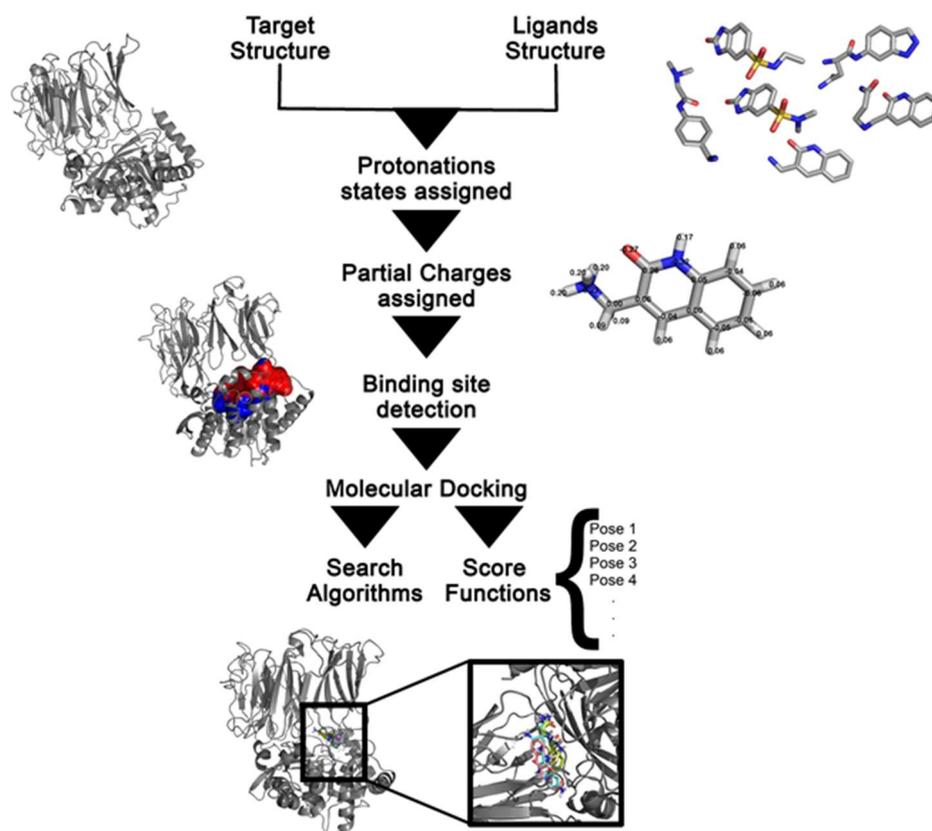


Figure 1. General workflow of molecular docking calculations. The approaches normally start by obtaining 3D structures of target and ligands. Then, protonation states and partial charges are assigned. If not previously known, the target binding site is detected, or a blind docking simulation may be performed. Molecular docking calculations are carried out in two main steps: posing and scoring, thus generating a ranked list of possible complexes between target and ligands.

Usually, the binding site location on which to focus the docking calculations is known. However, when the binding region information is missing, there are two commonly employed approaches: either the most probable binding sites are algorithmically predicted or a “blind docking” simulation is carried out. The latter has a high computational cost, since the search covers all the target structure [12]. Several available software can be used to detect binding sites. MolDock [9], for example, uses an integrated cavity detection algorithm to identify potential binding sites. DoGSiteScorer is an algorithm that determines possible pockets and their druggability scores, which describe the potential of the binding site to interact with a small drug-like molecule [13]. Fragment Hotspot Maps [14] uses small molecular probes to identify surface regions in the receptor that are prone to interact with small molecules. These predicted interaction sites can then be provided as the centre of the sampling space.

Moreover, information derived from such hotspots or even from previous experimental knowledge (e.g., NMR, mass spectrometry) can be used to generate distance restraints, which is known to greatly increase protein-small molecule docking accuracy [15].

During docking calculations, a common strategy is to employ a grid representation that includes precalculated potential energies for interaction within the target binding site [16]. This approach speeds up the docking runs and basically consists of the discretisation of the binding site [17]. Then, at each grid point, interactions related to the Lennard–Jones and electrostatic potentials are calculated.

Ligand structure is also required and can be obtained from small molecules databases, such as ZINC [18] and PubChem [19]. These online databases facilitate the retrieval of a large number of compounds for subsequent virtual screening. If not directly available, the 3D atomic coordinates of

these compounds can be obtained from the 2D structures (or even from simpler representation schemes, such as SMILES) using several available software, such as ChemSketch (Advanced Chemistry Development, Inc., Toronto, On, Canada, www.acdlabs.com, 2019), ChemDraw (PerkinElmer Informatics), Avogadro [20] and Concord [21]. It is worth noting that for small molecule ligands all that is needed initially is a stereochemically defined geometry with the correct relevant protonation state, since conformations will be explored by the docking software in the context of the target's binding site.

Charges are usually assigned through algorithms that distribute the net charge of a molecule among its constituent atoms as partial atom-centred charges. Furthermore, most docking methods assume that a particular protonation state and charge distribution in the molecules do not change between their bound and unbound states [3]. Nevertheless, it is crucial for successful docking to evaluate free torsions, protonation states and charge assignments. The protonation states of the target's amino acid residues can be critical to ligand interactions and, consequently, to the binding affinity prediction. There are several software available to evaluate the pKa of the amino acid residues, such as PropKa [22] and H++ [23].

Ligand protonation is also important since it affects the net charge of the molecule and the partial charges of individual atoms. Nonetheless, each docking program will employ a different charge assignment protocol [1]. For example, in the MolDock program, the protein and the ligands are automatically prepared (charges and protonation states assigned) and simplified charge and protonation schemes are used, as described by Thomsen and Christensen (2006). AutoDock uses Gasteiger–Marsili atomic charges whereas the closely-related AutoDock Vina does not require the assignment of atomic charges, since the terms that compose its scoring function are charge-independent [3,24]. The DockThor algorithm, as implemented in the homonymous web portal, automatically generates the topology files (i.e., atom types and partial charges) for the protein, ligand and cofactors according to the MMFF94S force field [4,5,25].

Two aspects are crucial to docking programs: search algorithms and scoring functions. The search algorithm analyses and generates ligand pose at a target's binding site, taking into consideration the roto-translational and internal degrees of freedom of the ligand [10].

Search strategies are often classified as systematic, stochastic or deterministic [16]. Systematic search algorithms explore each ligand's degree of freedom incrementally. As the number of free rotatable bonds increases, the number of evaluations can undergo a combinatorial explosion [16,26,27]. This class of search algorithms can be subdivided in exhaustive, incremental construction (which relies on the fragmentation of the ligand) and conformational ensemble [26]. FlexX [8] and eHits [28], for example, employ fragment-based approaches with systematic algorithms (incremental construction and graph matching, respectively).

A number of algorithms were also developed to use information from protein and ligand pharmacophores. Those algorithms try to match the distances between each of the ligand's and protein's pharmacophoric points [29]. The software FLEXX-PHARM, for example, is an extended version of FLEXX and applies pharmacophoric features as constraints into a docking calculation [30].

Stochastic search algorithms perform random changes in the ligand's degrees of freedom. However, this kind of algorithm does not guarantee convergence to the best solution. To improve it, an iterative process can be performed. Monte Carlo, Evolutionary Algorithms (including genetic), Tabu Search and Swarm Optimisation are some of the most common stochastic algorithm implementations [26]. Several software use stochastic algorithms as search methods, such as AutoDock [2], GOLD [6], DockThor [4,5,25] and MolDock [9] (Table 1).

Table 1. Molecular docking software.

Software	Posing	Scoring	Availability	Reference
Vina	Iterated Local Search + BFGS Local Optimiser	Empirical/Knowledge-Based	Free (Apache License)	Trott, 2010 [3]
AutoDock4	Lamarckian Genetic Algorithm, Genetic Algorithm or Simulated Annealing	Semiempirical	Free (GNU License)	Morris, 2009; Huey, 2007 [31,32]
Molegro/MolDock	Differential Evolution (Alternatively Simplex Evolution and Iterated Simplex)	Semiempirical	Commercial	Thomsen, 2006 [9]
Smina	Monte Carlo stochastic sampling + local optimisation	Empirical (customisable)	Free (GNU License)	Koes, 2013 [33]
Plants	Ant Colony Optimisation	Empirical	Academic License	Korb, 2007; Korb, 2009 [34,35]
ICM	Biased Probability Monte Carlo + Local Optimisation	Physics-Based	Commercial	Abagyan, 1993; Abagyan, 1994 [36,37]
Glide	Systematic search + Optimisation (XP mode also uses anchor-and-grow)	Empirical	Commercial	Friesner, 2004 [38]
Surflex	Fragmentation and alignment to idealised molecule (Protomol) + BFGS optimisation	Empirical	Commercial	Jain, 2003; Jain 2007 [39,40]
GOLD	Genetic Algorithm	Physics-based (GoldScore), Empirical (ChemScore, ChemPLP) and Knowledge-based (ASP)	Commercial	Jones, 1997; Verdonk 2003 [6,7]
GEMDOCK	Generic Evolutionary Algorithm	Empirical (includes pharmacophore potential)	Free (for non-commercial research)	Yang, 2004 [41]
Dock6	Anchor-and-grow incremental construction	Physics-based (several other options)	Academic License	Allen, 2015 [42]
GAsDock	Entropy-based multi-population genetic algorithm	Physics-based	*	Li, 2004 [43]
FlexX	Fragment-Based Pattern-recognition (Pose Clustering) + Incremental Growth	Empirical	Commercial	Rarey, 1996; Rarey, 1996b [8,44]
Fred	Conformer generation + Systematic rigid body search	Empirical (defaults to Chemgauss3)	Commercial	McGann, 2011 [45]
DockThor	Steady-state genetic algorithm (with Dynamic Modified Restricted Tournament Selection method)	Physics-based + Empirical	Free (Webserver)	De Magalhães, 2014[4,25]

*Availability is unclear.

In deterministic search, the orientation and conformation of the ligand in each iteration is determined by the previous state, and the new state has equal or lower energy value than the previous one [16,26]. However, this kind of algorithm has higher computational cost and often leads to the undesired trapping of the resulting conformations to a local energy minimum [16]. Examples are energy minimisation methods and molecular dynamics (MD) simulations.

The overall size of the ligand, especially if it contains a large number of rotatable bonds impacts most docking algorithms in a negative way, both in terms of computational cost of each individual docking run and in terms of docking accuracy [46]. That is the case because each new rotatable bond inherently increases the ligand's degrees of freedom, thus increasing the number of possible conformations. The enhanced conformational space is therefore much more complex to explore, rendering less accurate results, usually even with increased sampling steps. The magnitude of this effect is distinct for different algorithms [3,47] and fragment-based ones seem to exhibit superior performance in such cases [46].

Some algorithms can combine different search strategies, and often MD simulations are used to analyse the time-resolved trajectory of the ligand-bound system and to further pinpoint the best docking solutions [48–51].

After the generation of thousands of ligand orientations, additional scoring functions may be used to rank the conformations. They may be based on binding energy, free energy, or a qualitative numerical measure to approximate interaction energies [52]. Currently, scoring functions are grouped into three major types: force field, empirical and knowledge-based [26,27,53].

Force field-based functions consist of a sum of energy terms [26]. The potential energy usually accounts for bonded (bond length, angle, dihedrals) and nonbonded (van der Waals, electrostatic) terms. This type of function usually neglects solvent effects and entropies [16]. The DockThor program [4], for example, employs a scoring function for pose prediction based on the MMFF94S force field composed of three energy terms [54], i.e., the torsional term for bonded interactions, the electrostatic potential and the Buffered-14-7 term for the van der Waals potential (Equation (1)):

$$E_{MMFF94} = \sum 0.5(V_1(1 + \cos \Phi) + V_2(1 - \cos 2\Phi) + V_3(1 + \cos 3\Phi)) + \sum \frac{332.0716q_iq_j}{\varepsilon(R_{ij} + \delta_{elec})} + \sum \varepsilon_{ij} \left(\frac{(1 + \delta_{vdW})R_{ij}^*}{R_{ij} + \delta_{vdW}R_{ij}^*} \right)^7 \left(\frac{(1 + \gamma)R_{ij}^{*7}}{R_{ij}^7 + \gamma R_{ij}^{*7}} - 2 \right), \quad (1)$$

where V_1 , V_2 and V_3 are constants dependent on the types of the atoms i and j , ϕ is the i-j-k-l torsion angle, q_i and q_j are the partial charges of atoms i and j , ε is the dielectric constant given by a distance-dependent sigmoidal dielectric function [55], R_{ij} is the internuclear separation between atoms i and j , and δ_{elec} is the electrostatic buffering. Repulsion at short distances and van der Waals interactions are calculated by the last term, the Buf-14-7 potential [56]. In this term, ε_{ij} is the well depth, R_{ij}^* is the minimum-energy separation (Å) that depends on the MMFF94S types of the atoms i and j , and $\delta_{vdW} = 0.07$ and $\gamma = 0.12$ are the buffering constants.

Empirical scoring functions are derived from quantitative structure-activity relationships which were first idealised by Hansh and Fujita [16,57]. The goal is to predict binding affinity with high accuracy by using known experimental binding affinity data [26]. ChemScore [58] and GlideScore [59] are examples of empirical scoring functions.

Knowledge-based functions are based on frequency of atom pairs interactions observed in experimentally determined 3D structures of ligand-target complexes [16,26]. DrugScore of FlexX program [60] and PMF [61] are examples of knowledge-based functions.

Binding affinity prediction is still a major challenge for docking programs and most approaches rely upon consensus scoring schemes and rescoring approaches [16,26,27]. Consensus scoring for improving molecular docking accuracy is an ever-evolving research topic and will be addressed further in this review.

1.1. Molecular Docking in Drug Design

Molecular docking is a key component of the Computer-aided Drug Design toolbox. It is part of the so-called “structure-based drug design” methods and was first developed in the middle 80s through early 90s for predicting the binding mode of known active compounds and virtually screening large digital compound libraries to reduce costs and speed up drug discovery [62]. Docking tools have also been used in the hit-to-lead optimisation process. The latter application imposes the biggest challenge as predicting relative binding affinities for a series of related compounds has been the Achilles heel of most docking software since the very beginning of their development. Nevertheless, docking can still be used in hit-to-lead optimisation by indicating if the designed analogues of a hit compound present improved molecular interactions with the target.

Another widely known shortcoming of traditional docking methodologies is the poor modelling of receptor flexibility [63–65]. Some docking algorithms are able to partially mitigate this issue by allowing side-chain movement of active-site residues. Nevertheless, larger conformational changes might be triggered upon ligand-binding or might be a prerequisite to the binding event itself. A strategy, usually referred to as Receptor Ensemble Docking (or simply Ensemble Docking) is the most frequently used to model those scenarios. It is based on the concept of Conformational Selection and consists in using multiple conformations of the receptor molecule, that can be obtained via different methods, such as MD simulations [66,67], Normal Mode Analysis [68], and even by using alternative experimentally-determined receptor conformations [69]. It is worth noting that some software, such as GOLD and Glide have implemented functionality to execute this type of analysis.

The main limitations and challenges in the docking methodology have been identified nearly two decades ago [16] but they are still the subject of a very active research field. As described earlier here, two key components of the docking methodology are the conformational search algorithm and the scoring function. The former can suffer dramatically in performance when dealing with longer and flexible ligands, especially for shallow and chemically featureless binding sites, such as in polymer binding proteins (e.g., peptidases and glycosidases). Force-field based scoring functions suffer from the inherent problem of calculating binding affinities from the simplified interaction energies necessary to keep the docking calculations fast enough to process large compound libraries. Although binding affinities can be more accurately predicted from calculated binding free energies the latter also suffers from a problem of subtraction of large numbers (interaction energy between the ligand and protein on one hand and the cost of bringing the two molecules out of solvent and into an intimate complex on the other hand), which are often calculated with sub-optimal accuracy, and yield a small number as a result of the calculation [70].

In the following sections, we will review and discuss a selection of the main topics in the literature for molecular docking in drug design, all of which intend to address the above discussed limitations and advances in the methodology.

2. Benchmarking Sets

When using computational methods for molecular docking, it is paramount to assess the performance and accuracy of the programs to be employed. This not only allows one to know the degree of credibility that can be expected in the results, but also helps choosing the method or program better suited to the task at hand. To that end, there are many benchmarking databases that provide targets and ligands for docking, along with additional information such as true binding affinity, experimental binding pose, and actives/inactives distinction. Experimental information can then be compared to the docking program's predictions through different statistical metrics, which allows the assessment of its performance.

2.1. Benchmarking Sets for Pose Prediction and Binding Affinity Calculations

The development of either empirical parametric or nonparametric regression models for docking pose and binding affinity predictions must be based on experimental data so that their functions may be properly parameterised (or inferred) and thus better represent reality. Moreover, the performance of these models must also be evaluated on such data. In light of this demand, there are many benchmarking datasets which aim to group as much high-quality data as possible [71–74].

The most widely employed of these is PDDBind [71]. This database is a result of an effort to screen the entire Protein Data Bank (PDB) [11] for experimentally determined 3D structures of protein-ligand complexes and collect their experimentally measured binding affinities. There is also a refined set of complexes [75] and a core set derived from it [76], which has become the standard set for benchmarking scoring functions (SFs). It is noteworthy that PDDBind is also widely used in training machine learning SFs for binding affinity predictions [77–79].

There are also benchmarking databases which encompass specific complexes or purposes, such as protein-protein complexes [80], membrane protein-protein complexes [81], and a blind set based on PDDBind for testing machine learning SFs [82].

Accuracy for pose prediction can be assessed by root mean square deviation (RMSD) calculations comparing predicted pose and experimental pose. To compare binding affinity predictions with experimentally determined affinities for a set of multiple data points, one can too calculate RMSD for the values, but also the Pearson correlation coefficient (R_p) and the Spearman rank-correlation (R_s) [83].

2.2. Benchmarking Sets for Virtual Screening

Benchmarking databases for virtual screening (VS) consist on datasets with selected known active ligands and inactive decoys for a single protein target [84]. Since information on inactive molecules is scarce in comparison to active ones, most decoys are not selected based on experimental data but are instead putative inactive compounds [85], whose selection must be made carefully so as to avoid artificial enrichment [86]. This scarcity occurs because active molecules are better described and documented, however, the opposite asymmetry is observed in nature: from a varied set of molecules which come in contact with a given protein, only a few specific ones will be active against it. Therefore, VS programs must be capable of identifying active compounds amidst a large pool of inactive ones, thus, benchmarking sets mirror this natural asymmetry by providing many putative decoys for a single known active molecule. In order to prevent bias, the active and decoys sets' characteristics must be equally balanced: one set must not be more structurally complex or diverse than the other [87,88]; both sets should not cover small chemical spaces [84]; and there must not be any actual binders among the decoys (Latent Actives in the Decoy Set, LADS) [89]. Datasets are therefore curated in order to avoid bias as well as provide as much useful data as possible; the most widely used are described as follows.

The Directory of Useful Decoys (DUD) was created based on the principle that decoys must resemble the physical properties of the actives but be sufficiently chemically distinct to be in fact nonbinders [90]. DUD then became the gold standard benchmark for VS [91]. It was later improved into the Directory of Useful Decoys-Enhanced (DUD-E) [92], which selects decoys based on more physicochemical properties, adds more targets, and provides a tool for decoy generation based on user-input actives.

The Demanding Evaluation Kits for Objective in Silico Screening (DEKOIS) [89] was created with special attention to avoiding poorly embedded actives and LADS. A new version, DEKOIS 2.0 [93], was released two years later with additional physicochemical properties for matching decoys and an enhanced elimination of LADS.

The Maximum Unbiased Validation (MUV) [94] datasets were curated with special care for the chemical diversity of the actives set, in order to avoid over-representation of chemical entities and thus avert overestimation of performance. An exclusion of potentially unspecific active compounds was also implemented, as well as removal of actives devoid of decoys in its chemical space.

There are also databases for assessing virtual screening with specific targets: G-Protein-Coupled Receptor (GPCR) Ligand Library (GLL) and GPCR Decoy Database (GDD) [95], NRLiSt BDB for nuclear receptors [96] and MUBD-HDACs for histone deacetylases [97].

It is noteworthy that it is also possible to generate decoys for specific compounds when the target of interest is not available. User-input ligands must be provided in SMILES format, and a decoy set is curated based on their molecular properties. DecoyFinder [98] was the first application to provide this tool, searching the ZINC database for molecules similar to actives by comparing chemical

descriptors. At about the same time DecoyFinder was published, DUD was upgraded to DUD-E, which also allows searching the ZINC database for decoys utilising the same search method employed to construct the database's new target subsets. In 2017, Wang et al. [99] argued that these tools lacked computational speed for large active sets and flexible input options to avoid bias in the user-specified active set. To address these issues, they created RADER (RAPid DEcoy Retriever), which selects decoys from four different databases, including ZINC.

2.3. Evaluation Metrics

The most widely used metrics to assess ranking performance in VS are receiver operating characteristic (ROC) curves and enrichment factors (EF). The ROC method plots the rank's specificity and sensitivity into a curve whose area (area under the curve, AUC) ranges from 0 (worst performance) to 1 (best performance), where 0.5 reflects a randomly distributed ranking order. The calculations are made based on cut-offs throughout the whole rank, and therefore ROC reflects only overall performance [100,101]. However, when evaluating VS performance, the enrichment at the top of the rank is most important (i.e., the early recognition problem), since there can be found the molecules identified by the SF as the most probable actives [102]. EF can be used to calculate the enrichment at an early single cut-off [83] or at many cut-offs [101], which addresses the early recognition problem, however its main setback resides in the fact that its maximum value depends on the active/inactive ratio on the dataset [101,103].

It is noteworthy that by calculating Youden's index (sensitivity + specificity - 1) for all cut-offs made in the ROC curve, one can determine the optimal threshold (i.e., the cut-off with the highest index) through which continuous binding predictions of a particular SF can be converted into to binary active/inactive classification [104].

Other metrics have been suggested and applied to better address the early recognition problem. For instance, the Robust Initial Enhancement (RIE) metric [105] applies weight to the active molecules. The active will weigh closer to 1 the better ranked it is, and its weight will fall as the rank increases. A RIE value of 1 indicates a random distribution of the rank, and its maximum value depends on the active/inactive ratio, similarly to EF. The Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic (BEDROC) [102] incorporates the RIE weighing strategy into ROC curves: performance is measured in a 0 to 1 range and advantage is given to better ranked actives. One drawback of the BEDROC approach is that the magnitude of this advantage is controlled by a single parameter, which can frustrate performance comparisons between different studies [103].

No single benchmarking set or metric can be considered to be best overall for molecular docking. Rather, they are chosen differently depending on the inquiry, as well as carefully, in order to avoid biasing issues. Erroneous estimations of performance negatively impact studies and are also very hard to detect based on benchmarking results alone. Nonetheless, benchmarking datasets provide invaluable means for quality assessment of computational methods in drug discovery.

3. Consensus Methods

With the continued development of new scoring functions (SFs) and the improvement of well-established ones, the use of docking strategies that combine two or more SFs has become increasingly common. That is especially interesting because the various available functions perform differently across the spectrum of potential interactions, and presumably, in an ideal combination, the shortcomings of a particular function may be compensated by the others.

This strategy was first suggested by Charifson and co-workers in a study in which they benchmarked several SFs, both individually and in combination, using p38, IMPDH and HIV protease as model systems. Their approach involved taking the intersection of the top-scoring molecules according to two or three different functions available at the time and they found it provided a "dramatic reduction in the number of false positives identified by individual SFs" [106].

A consensus-docking protocol will generally differ in three major aspects: (i) the means by which the poses are obtained, (ii) the selection of the SFs, and (iii) the algorithm used to achieve the consensus. Realistically, the number of possible procedures is overwhelming, and, to date, no single

protocol has been proven remarkably superior to the others. Nevertheless, it is absolutely clear that consensus methods perform consistently better when compared to individual SFs (c.f. referenced papers in Table 2).

The theoretical rationale for this was explored in 2001, soon after the first approaches, in a work in which the authors simulated an idealised computer experiment where scores were generated for a hypothetical set of 5000 compounds and the effects of consensus strategies were evaluated. The authors suggest that the improvement is largely due to the fact that the mean value of repeated samplings tends to be closer to the true value than any single sampling [107].

Although some initiatives have been explored to come up with composite scoring schemes that are applied simultaneously during the posing procedure [108], in most cases, the consensus is achieved after the conformational sampling. Moreover, it is widely accepted that the conformational sampling is not the major bottleneck in the docking process [109,110] therefore, a greater fraction of the developed methods generate the docking poses using a single algorithm and subsequently use a different set of SFs to re-assess them (Tables 2 and 3). Nevertheless, several groups have focused on obtaining more reliable poses, for example, Ren and co-workers have explored the effects of using multiple software in the pose generation step [111]. They used a RMSD-based criterion to come up with a representative pose derived from a minimum of three and a maximum of 11 docking programs. A pose representative was selected for all possible combinations and their method achieved an increase in the success rate (pose-to-reference RMSD < 2.0 Å) of approximately 5% when compared to the best independent program.

Additionally, the concept of “consensus level” has been explored in recent works [112,113], and similarly to the previously described approach, it uses a combination of docking software to generate ligand poses, which are then clustered and the number of software that predict the same pose is taken as the consensus level. This metric can then be used to reject compounds that fail to attain a certain level and true ligands are less likely to being rejected, which, in turn increases the enrichment factors.

Another consensus posing strategy is to reject a given pose if it two or more programs fail to “converge” to that conformation. Houston and Walkinshaw have demonstrated that the success rate can be increased from ~60% to ~80% by simply rejecting a molecule if the RMSD between the poses calculated by two programs (AutoDock and VINA) is greater than 2.0 Å. The idea behind this approach is that a correct pose is more likely to be predicted by more than one algorithm, thus eliminating the misleading orientations (which could be considered false positives) [114].

Some initiatives combine consensus posing and scoring, as is the case of the VoteDock approach (and two correlated functions), proposed by Plewczynski et al., in which they combine cross-software pose conformation agreement, in the form of a voting system, with a composite scoring obtained via multivariate linear regression with results performing consistently better than individual SFs [115].

Besides consensus posing, many groups have focused their efforts on creating consensus scoring schemes. Very recently, Perez-Castillo and co-workers have applied the Genetic Algorithm to devise the best combination from a total of 15 SFs (or 87 scoring components) that maximises either the enrichment factor or the BEDROC value. Their results suggest that combining scoring components, instead of SFs themselves is a more effective strategy. Their algorithm, CompScore, is made available as a webserver [116].

Other reported strategies for achieving scoring function consensus are sequential docking [117,118], linear regression [119], rank-by-rank, rank-by-number, rank-by-vote [86,107,120] and standard deviation consensus [121]. Combinations of consensus docking strategies and ligand-based approaches have also been suggested [122,123].

Table 2. Consensus docking methods.

Source	T ^a	Posing ^b	F ^c	Consensus Strategy	Analysis	Ref.
DUD-E/ PDB	102/3	4	4	Standard Deviation Consensus (SDC), Variable SDC (vSDC)	Rank/Score curves Hit recovery count	Chaput, 2016 [121]
DUD-E	21	8	8	Gradient Boosting	EF, ROCAUC	Ericksen, 2017 [124]
PDBBind DUD	228/1	Vina, AutoDock	2	Compound rejection if pose RMSD > 2.0 Å	Success rate	Houston, 2013 [114]
PDB	3	GAsDock	2	Multi-Objective Scoring Function Optimisation	EF	Kang, 2019 [108]
mTOR ^d Inhibitors	1	Glide	26	Linear Combination	BEI Correlation	Li, 2018 [119]
PDB	220	FlexX	9	Several ^e	Compression and Accuracy	Oda, 2006 [120]
DUD-E	102	Dock 3.6	15	Genetic Algorithm used to combine SF components	EF, BEDROC	Perez-Castillo, 2019 [116]
PDBBind	1300	7	7	RMSD-based pose consensus, multivariate linear regression	Success rate	Plewczynski, 2011 [115]
DUD	35	10	10	Compound rejection based on RMSD consensus level	EF	Poli, 2016 [112]
PDBBind	3535	11	11	Selection of representative pose with minimum RMSD	Success rate	Ren, 2018 [111]
PDB	100	AutoDock	11	Supervised Learning (Random Forests), Rank-by-rank	Average RMSD, Success rate	Teramoto, 2007 [125]
PDB DUD	130/3	10	10	Compound rejection based on RMSD consensus level	EF, ROCAUC	Tuccinardi (2014) [113]
PDBBind CSAR	421	Glide	7	Support Vector Rank Regression	Top pose /Top Rank	Wang, 2013 [126]
PDB	4	GEMDOCK GOLD	2	Rank-by-rank, Rank-by-score	Rank/Score curve, GH Score, CS index	Yang, 2005 [127]

^aTotal number of targets used in the assay; ^bPosing software used. If more than two software were used, than only the number is indicated; ^cNumber of scoring functions used; ^dIn this study, the dataset was composed of 25 mammalian target of rapamycin (mTOR) kinase inhibitors retrieved from the literature and six mTOR crystal structures retrieved from PDB; ^eThe purpose of this study was to evaluate several different consensus strategies (e.g., rank-by-vote, rank-by-number, etc).

Table 3. Recent works using consensus docking approaches.

Target	Lig.	Posing	F ^a	Consensus Strategy	Hits/Test	Best Activity (IC ₅₀)	Ref.
EBOV Glycoprotein	3.57 × 10 ⁷	VINA, FlexX	2	Sequential Docking	-	-	Onawole, 2018 [117]
β-secretase (BACE1)	1.13e5	Surflex	12	Z-scaled rank-by-number Principal Component Analysis Sequential Docking	2/20	51.6 μM	Liu, 2012 [128]
c-Met Kinase	738	2	2	Compound rejection if pose RMSD > 2.0 Å	-	-	Aliebrahimi, 2017 [118]
Acetylcholinesterase	14,758	4	4	vSDC[121]	12/14	47.3 nM	Mokrani, 2019 [129]
PIN1	32,500	10	10	Compound rejection based on RMSD consensus level	1/10	13.4 μM 53.9 μM ^c	Spena, 2019 [130]
Akt1	47	LigandFit	5	Support Vector Regression Compound rejection based on RMSD consensus level	6/6 ^b	7.7 nM	Zhan, 2014 [123]
Monoacylglycerol Lipase (MAGL)	4.80 × 10 ⁵	4	4	Compound rejection based on RMSD consensus level	1/3	6.1 μM	Mouawad, 2019 [131]

^aNumber of scoring functions used; ^bThis work consisted of a Quantitative Structure-Activity Relationship (QSAR) model using consensus docking as descriptors. Six compounds were designed, synthesised and tested, exhibiting IC₅₀ values between 7.7 nM and 4.3 μM; ^cFirst IC₅₀ value: inhibitory activity against PIN1 isomerisation. Second IC₅₀ value: inhibitory effects on ovarian cancer cell lines.

Machine learning algorithms have also been employed in the determination of the consensus in recent developments. Early efforts used Random Forest algorithms to achieve consensus for 11 different SFs, outperforming the regular rank-by-rank approach in about 5%–10% and individual SFs by a far greater margin [125]. Support Vector Rank Regression (SVRR) has been suggested as a possible tool to combine seven distinct SFs (Glide-Score, EmodelScore, EnergyScore, GoldScore, ChemScore, ASPScore and PLPScore) computed using GLIDE and GOLD docking programs, and was shown to improve correct top pose prediction (RMSD < 2.0 Å) by 12.1% and correct top ligand selection by 46.3% [126]. In another study, Ericksen and collaborators used gradient boosting to derive a consensus score and benchmarked this approach using 21 targets selected from DUD-E, gradient boosting was shown to outperform traditional consensus methods (maximum, median and mean scores) and as well as the mean-variance consensus [124]. A summary of the aforementioned works can be found in Table 2.

Although molecular docking was first applied over three decades ago, it is apparent, given the virtually endless protocols, that there is still much improvement to be made in the field. In this sense, initiatives such as the Community Structure-Activity Resource (CSAR active from 2010 to 2014) [73,132] and the Drug Design Data Resource (D3R) [133,134] are invaluable as they promote the standardisation of validation datasets and metrics, as well as serve as a repository for the knowledge accumulated in the field.

A simple comparison made with a keyword search software in the SCOPUS database for the years 1995 until 2018 (“TITLE-ABS-KEY (*software* AND docking) AND PUBYEAR > 1994 AND PUBYEAR < 2019” where the word *software* is replaced by several of the mostly employed docking programs) shows the relative prevalence of these software. Substituting the term *software* for

consensus, shows that consensus methods, in spite of consistently showing superior results, are less frequently mentioned in the literature than some of the more common docking programs (at least in the searched fields, i.e., title, abstract and keywords) (Figure 2). While one could argue that this could be due to the fact that the fraction of works that indeed use consensus methods also mention other software, Figure 3, which contains the ratio of (research and conference) papers mentioning “molecular docking” OR “ligand docking” to the ones mentioning (“molecular docking” OR “ligand docking”) AND consensus, shows that the discrepancy is even more pronounced (an average of 88.36 works that cite molecular docking per each work that mentions the word consensus—Figure 3).

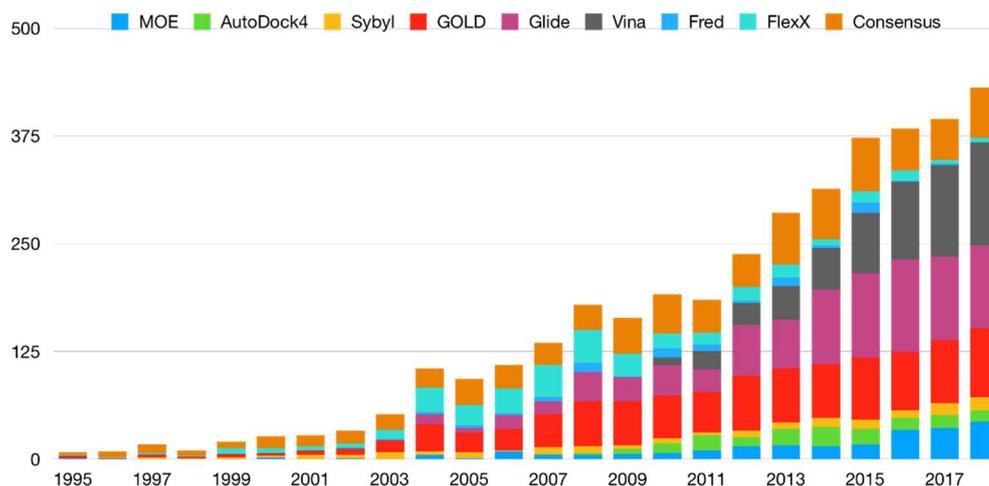


Figure 2. Scopus search results for the query “TITLE-ABS-KEY (*software* AND docking) AND PUBYEAR > 1994 AND PUBYEAR < 2019” where the word *software* is substituted for one of the eight most common docking software or by the word *consensus*.

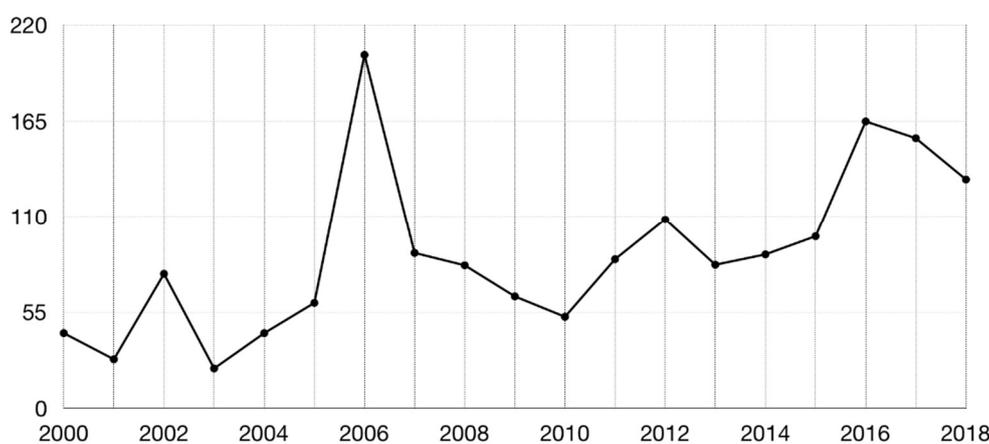


Figure 3. Ratio of the numbers of papers containing either the expression “molecular docking” or “ligand docking” to the number of papers containing either of the two expressions AND the word *consensus*.

There is also clear disparity in the level of elaborateness between the protocols used by the groups that develop and the ones that implement these methods. As a result, the virtual screening protocols used by the latter (such as sequential docking, rank-by-number and RMSD-based pose rejection) are often less involved than the ones suggested by the former. Table 3 summarises recent works that employed consensus docking in their screening methodologies, along with the best experimentally-determined activity. Despite using more straightforward methodologies to achieve consensus, these studies show the importance of combining distinct SFs, since they have still been able to find relatively potent ligands. It appears that, easy-to-use, carefully designed and validated

docking pipelines which include consensus posing and/or scoring are called for and could be widely adopted in structure-based drug design studies, both in academic and industrial settings.

4. Efficient Exploration of Chemical Space: Fragment-Based Approaches

4.1. The Chemical Space

Since it was first described in the late 1990s [135], fragment-based drug (or, less frequently, lead [136]) discovery (FBD/LD) has gained a lot of attention and many drug candidates developed with the use of such approaches have reached clinical trials [137]. The fundamental aspect that fosters its popularity is that it allows an efficient exploration of the chemical space with relatively small sampling, i.e., by combining smaller fragments that show high ligand efficiency, it is possible to design very potent ligands which would, otherwise, be dispersed in a vast pool of possible molecules. Additionally, it has been demonstrated that the probability of a given interaction between a given ligand and receptor is inversely proportional to the ligand complexity [138,139], suggesting that higher hit-rates could be achieved by screening less complex molecules.

In 2007, researchers from Reymond's group at the University of Berne used a graph-based approach to generate all possible topologies for chemically-stable compounds presenting up to 11 atoms, and they generated a database containing near 26.4 million (2.64×10^7) molecules (GDB11) [140]. Since then, they have created new sets of increasingly larger molecules, namely containing up to 13 heavy atoms (GDB13— 9.7×10^8 molecules) and up to 17 heavy atoms (GDB17— 1.66×10^{11}). These numbers might seem overwhelming, but not if compared to an astonishing 10^{60} estimated drug-like molecules (with up to 30 heavy atoms) [62].

Very recently, researchers from UCSF (University of California, San Francisco) have completed ultra-large campaigns, screening approximately 99 million compounds for AmpC β -lactamase and 138 million compounds for the D4 dopamine receptor, ultimately finding 30 compounds with sub-micromolar activity, including one with picomolar activity (180 pM) [141]. Endeavours of such magnitude have not been customarily undertaken, since they require great use of computational resources, therefore, fragment-based approaches can help efficiently explore the chemical space since (i) they have a small amount of degrees of freedom, leading to faster spatial sampling, (ii) they can be combined to create larger, more potent ligands, requiring reduced screening libraries to achieve comparable chemical space coverage, and (iii) the reduced complexity of fragments should lead to increased hit-rates.

Experimentally, due to the reduced affinities, these fragments must be screened using more sensitive biophysical assays, such as Fluorescence-Based Thermal Shift, NMR Spectroscopy and Surface Plasmon Resonance [142]. Molecular docking can also be an invaluable tool for the detection of potentially interacting fragments and several examples will be discussed below. Candidate fragments detected by experimental or computational approaches are then usually evaluated through X-ray Crystallography [142] or even High Throughput X-ray Crystallography (HTX), where protein crystals are soaked in high concentrations of one or more fragments and the structure of the complex is subsequently determined [143].

4.2. Fragment Libraries

Some aspects must be taken into consideration when tailoring fragment libraries in order to optimise fragment-based drug design (FBDD) outcomes. For instance, because fragments are smaller, they tend to bind less tightly to the protein targets, exhibiting lower potency values. Therefore, it is advantageous to use size-normalised parameters, such as Ligand Efficiency (LE) [144], Binding Efficiency Index (BEI) [145] or Fit Quality (FQ) [146], to prioritise the evaluated molecules. These can then serve as objective parameters to a successful subsequent lead optimisation [147]. Secondly, Harren Jhoti's group has suggested an adjusted set of rules [148] (or guidelines [149]), termed Rule of Three (RO3), derived from hits obtained via High Throughput X-ray Crystallography (HTX) and inspired by Lipinski's Rule of Five [150]. These stemmed from the observation that successful hits customarily present molecular weight under 300 Da, three or fewer hydrogen bond donors, three or

fewer hydrogen bond acceptors, clogP under three and, additionally, three or fewer rotatable bonds and a polar surface area under 60 Å². These guidelines can help filtering fragment libraries for efficient screening, both experimentally and computationally. A third matter worth noting is the reported “lack of tri-dimensionality” in fragment libraries, which can hinder the development of ligands with high affinity for certain classes of targets [151].

Fragment libraries can be generic or generated ad hoc (targeted, or focused libraries). Many of the generic libraries are commercially available on demand, and thus may be readily used in experimental screens, and the compound chemical structures are usually also available as Structure-Data Files (SDF), which can be straightforwardly converted to other structural formats, such as MOL2, PDB and PDBQT and used for virtual screening (cf. Verheij [152] work on lead-likeness for sources of such libraries). Fragment libraries usually contain 10² to 10⁴ molecules, which are generally compliant with the RO3 and are idealised to maximise attributes such as solubility, chemical stability, scaffold complexity, tri-dimensionality and tractability [151,153,154]. Tractability-guided fragmentation algorithms and pipelines can be used to generate specialised fragment libraries starting from collections such as the World Drug Index (which has been fragmented using the RECAP algorithm [155]) or natural products libraries [156].

The combination of fragments into a larger molecule has been classified into four distinct categories, namely Merging, Linking, Growing and “SAR by catalogue” [153]. In fragment merging, two fragments occupying an overlapping site are joined together to obtain a larger molecule with higher affinity. Conversely, in fragment linking, the fragments are usually bound to two distinct binding pockets (or sub-pockets) and are joined together via the construction of a linker fragment, that ideally allows the maintenance of the initial orientation of the fragments. Fragment growing consists of the design and incorporation of new functional groups that are expected to form new interactions with the receptor, thus increasing the binding affinity. Finally, the “SAR by catalogue” is particularly interesting from the virtual screening angle due to its simplicity; in this approach, a fragment initially detected (and ideally confirmed by experimental techniques) is then used as an “anchor” to query a database for larger molecules that contain the original fragment. Thus, effectively, this strategy is largely used to create more focused libraries.

4.3. Molecular Docking in FBDD

Many groups have used FBDD to idealise potent ligands for disease-modifying protein targets with extensive use of molecular docking and virtual screening approaches. In a study developed by Chen and Shoichet, a fragment-based approach was used as an alternative to a lead-like virtual screen campaign, obtaining increased hit rates for β-Lactamase inhibitors, and ultimately yielding hits in the low μM range [157]. This indicates that even using similar docking protocols, fragment-based approaches can yield more accurate initial hits when compared to lead-like molecules screening.

These computationally-driven works reflect some of the experimental strategies discussed above, since the initial screens for promising fragments are usually followed by a fragment-joining step, which can be accomplished in a manual [158] or automated [159] way. Recently, Park et al. have been able to design nanomolar-range inhibitors for the protein Glycogen Synthase Kinase-3 β, using AutoDock [32] as the initial tool to perform virtual screening of fragment libraries in three independent subsites and LigBuilder [160] as the tool to connect a series of selected fragments [159].

Employing the “SAR by catalogue” method, Zhao and co-workers, after initial filtering of the ZINC database, have used an in-house docking solution to prioritise anchor fragments that bind the BRD4 bromodomain, which were then used to further interrogate the database and retrieve compounds containing the selected moieties, ultimately finding compounds with activity in the low micromolar range (7.0–7.5 μM) [161]. Using a similar “anchor-based” analogue search approach, Rudling and co-workers have used Dock3.6 to find inhibitors in the low micromolar range for MTH1 protein, an interesting cancer target, and in a second round of prospection for commercially available analogues, they managed to further optimise the initial hits to achieve IC₅₀ values as low as 9 nM [162].

Hernandez et al. have suggested non-nucleoside inhibitors of flaviviral methyltransferases (Zika virus and Dengue virus NS5MTase) presenting $IC_{50} \sim 20 \mu M$ by screening a focused library constructed using a knowingly binding core substructure, encoded organic chemistry rules and commercially available building blocks. The authors refer to this approach as fragment-growing [163].

The successful combination of fragment-based virtual screening and NMR screening has also been reported. Fjellström et al. have identified Activated Factor XI inhibitors using Glide to prioritise 1800 molecules (out of 6.5×10^3 from AstraZeneca screening collection with molecular weight (MW) < 250 g/mol) for NMR fragment screening. Subsequent structure-based expansion and re-scoring of 13 NMR hits yielded a compound with activity of 1.0 nM [158]. Using an inverted approach, Akabayov and co-workers used an initial NMR screen of a library containing 1000 fragments to identify moieties that bind T7 DNA primase, the two most promising hits were then used to query ZINC database, once more reflecting the “SAR by catalogue” approach, and the selected molecules (approximately 3000 per scaffold) were docked to DNA primase structure, using Autodock4. About half of the 16 selected compounds showed inhibitory activities [164].

Amaning and co-workers prospected for MEK1 inhibitors carrying out a virtual screening campaign of approximately 10^4 molecules, used to prioritise fragments to be further characterised by differential scanning fluorimetry (DSF), surface plasmon resonance and X-ray crystallography. Interestingly, a parallel biochemical screening of the same library showed that the 5% of the best scoring molecules in the virtual screening contained 30% of the biochemical hits and, according to the authors, this indicates that the VS–DSF combination can be used to ‘jump-start’ a project in an early phase when a biochemical or other biophysical assays are not available [165]. Additionally, it has been suggested that characteristics such as novelty and potency are likely to differ considerably between hits determined by experimental screening and those determined by virtual screening [166].

Besides prospecting for new molecules in fragment-based VS campaigns, molecular docking is extensively used to hypothesise interaction modes and better characterise the ligand-receptor interactions [167–169], and remains an invaluable asset in the drug development toolkit.

5. Machine Learning-Based Approaches

Scoring and ranking candidate molecules through binding affinity prediction is the most challenging aspect of molecular docking and VS. Classical SFs must simplify and generalise many aspects of the receptor-ligand interaction in order to maintain efficiency, approachability and accessibility [27]. Moreover, these SFs employ linear regression models: parametric supervised learning methods, which assume a specific predetermined functional form [170]. In other words, parametric methods fit the input variables (such as van der Waals and electrostatic energy terms) to the output (binding energy score) into a function whose form is already specified, and which is adjusted during the development of the SF in a theory-inspired fashion [77]. This rigid scheme often results in unadaptable SFs which fail to capture intrinsic nonlinearities in the data and therefore underperform in situations not accounted for in their formulation [77,171].

Alternatively, nonparametric machine learning (ML) algorithms (often referred to as just “machine learning”) can be used to replace [77,172–174] or improve [82,175–178] predetermined functional forms in classical SFs for binding affinity predictions. They have also been successfully applied in binders/nonbinders identification in virtual screening [175,179–181] and native pose prediction [126,172,182].

ML methods are divided into two broad groups: supervised and unsupervised learning. Unsupervised learning algorithms are employed to model the training data when there is no output available. Thus, these algorithms are commonly used for clustering data based on the degree of similarity between their features, for detecting associations between the data points, and for density estimations. In supervised learning, however, the output variables are known and provided to the algorithm along with the input for training. In nonparametric supervised learning, no functional form is assumed. It is then possible to infer the correlations between input and output from the training data itself and utilise it to predict the output for datasets of which the outcomes are unknown [170].

This allows for more diverse and accurate SFs: more features from the docked complex can be accounted for implicitly, therefore skirting modelling assumptions and necessary generalisations of classical SFs [77,82,171]. Moreover, the adeptness of the ML algorithm can be adjusted by tailoring the training dataset. For instance, increasing the diversity of the training complexes results in ML SFs with greater comprehensiveness. In fact, it has been shown that increasing the size of the training set boosts the scoring function's performance [82,172,183].

This contrasts greatly with classical SFs, whose parametric behaviour remains unable to improve performance with larger training datasets [82]. On the other hand, increasing the level of feature detail in training sets comprising of similar complexes may provide greater discrimination power when studying such data [183,184].

5.1. Protein Target Types: Generic and Family-Specific

Machine learning SFs can be considered family-specific or generic. It has been shown that family-specific SFs can outperform most accurate generic ones at said protein family's predictions [183,184]. Until recently, however, it was not clear whether a family-specific SF carried any advantages over generic ones whose training includes all complexes and features utilised in training the former [83]. It was later shown that random forest trained with family-specific data only slightly outperformed the universal model. This outperformance grew, however, when predicting more difficult targets with less active ligands [185]. In a 2018 study with deep learning neural networks, Imrie et al. [183] showed that family-specific models trained with a subset of the entire dataset outperformed universally trained models, and that only limited family data was required for this outperformance to occur. For each different protein family, the importance of the features used to describe the data varies [184], therefore, specific SFs are able to better assimilate these characteristics as a result of dealing with less broad and more nuanced data [183–185].

Machine learning SFs have been regarded both as knowledge-based [186,187] and empirical [188]. However, it is important to note that this categorisation has extensively been used in regard to classical SFs, and therefore it should not obscure the fact that there is a more fundamental difference between them: the former consists of nonparametric and the latter of parametric learning (Figure 4).

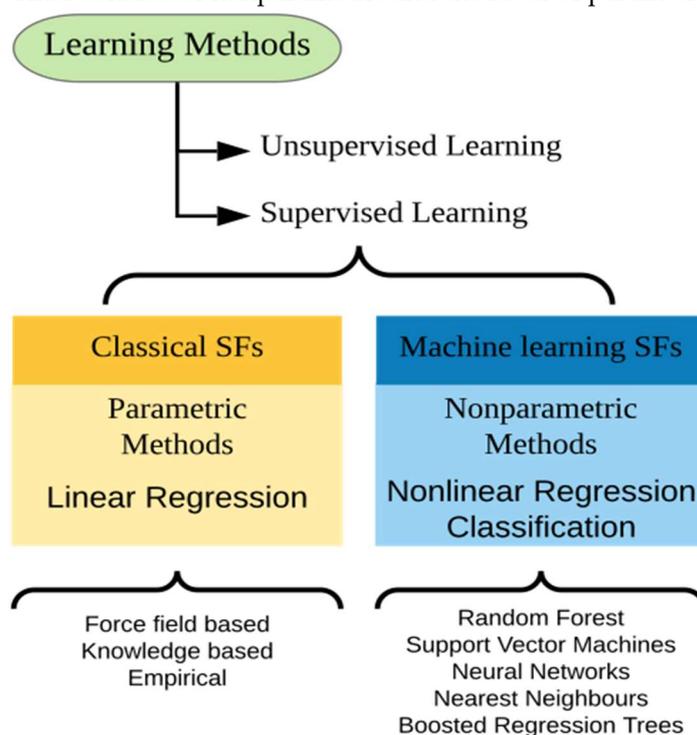


Figure 4. Learning methods can be broadly divided into supervised learning, when there is data available for training and parameterisation; and unsupervised learning, when there is no such data. Unsupervised learning cannot be used for binding affinity predictions and virtual screening.

Supervised learning, on the other hand, can be divided into parametric and nonparametric learning. Parametric learning assumes a predetermined functional form, as observed in linear regression, and is the method employed in classical scoring functions. Nonparametric learning, or just machine learning, does not presume a predetermined functional form, which is instead inferred from the data itself. It can yield continuous output, as in nonlinear regression, or discrete output, for classification problems such as binders/nonbinders identification.

5.2. Experiment Types: Binding Affinity Prediction and Virtual Screening

SFs designed for binding affinity predictions can also be used for virtual screening experiments, as long as the predicted results are ordered from best to worst binding score. If a binary active/inactive distinction is desired, one can establish an optimal activity threshold score by analysing the SF's performance on a benchmarking dataset (c.f. Benchmark Datasets section). However, ML classifiers built for VS may present better discrimination since their training utilises datasets specific for portraying virtual screening circumstances i.e., they are often trained on data derived from in silico approaches (as opposed to crystal structures of complexes) which do not always represent the correct binding mode, and the features from docked decoy molecules are also used for training [189].

5.3. Algorithms and Feature Selection

Feature selection plays an important role in the development of ML methods. Selecting a subset of features which are appropriate and effective for characterising the data not only improves prediction performance, but also reduces computational expense and facilitates the understanding of the intrinsic patterns underlying the data [190].

The first ML SF to outperform classical SFs [83], random forest (RF)-Score [77], utilised the random forest (RF) algorithm with intermolecular interaction features comprised of the number of a particular protein-ligand atom type pair interacting within a certain distance range [77]. Other descriptors such as energy terms from classical SFs, solvent accessible surface area, entropy, hydrophobic interactions and chemical descriptors have been applied by works such as those of Springer et al. (PostDOCK) [181], Pereira et al. (DeepVS) [177], Jiménez et al. (Kdeep) [78], Durrant et al. (NNScore) [79], Koppisetty et al. [191] and Liu et al. (B2BScore) [192] with various degrees of success. It has been shown that richer and more precise chemical descriptors do not generally result in more accurate predictions [193], and that different SFs have very different responses to an increase in the number of features [171].

Other ways of describing the data have been explored. For instance, Kundu et al. [194] utilised fundamental molecular descriptors for the proteins and the ligands, without any intermolecular interaction features, which circumvents the need for binding pose information. Srinivas et al. [195] utilised collaborative filtering, an algorithm extensively employed for recommendation systems (i.e., predicting appropriate online customer recommendations), to bypass the explicit definition of receptor and ligand features. The similarities in the data are inferred only based on the results of the recorded binding assays.

5.4. Deep Learning

Deep learning neural networks have recently been applied to pose prediction and ranking [78,173,177,183,196]. Convolutional neural networks, which are known to present outstanding image recognition capabilities [197], in molecular docking, have been explored mainly by featurising the protein-ligand complexes as three-dimensional grids. Deep learning SFs have yielded state-of-the-art results [78,183,196], comparable to and even surpassing those achieved by random forest, support vector machines, and boosted regression trees, the other non-neural network algorithms reported to be the most accurate for protein-ligand scoring [171,198].

5.5. Recent Applications and Perspectives

It is noteworthy that although the current ML techniques already promise to advance computational drug discovery, some limitations still need to be addressed. For instance, larger amounts of data are still required to reach optimal deep learning performance, and it is not clear whether at some point learning saturation can occur [183]. Furthermore, complex nonparametric learning models can be difficult to interpret. Sieg et al. [199] very recently pointed out that bias is being implicitly learned from standard benchmarking sets, and suggested guidelines to avoid fallacious models.

ML SFs for molecular docking have only recently been introduced. Naturally, most studies are dedicated to assessing and improving their predictive powers, and not as many have applied them in drug discovery and repurposing experiments. Nonetheless, existing prospective studies show positive results (Table 4). In 2011, Kinnings et al. [175] created a support vector machine-based SF to improve binding affinity prediction from classical SFs and used it to identify that phosphodiesterase inhibitors could potentially be repurposed towards *Mycobacterium tuberculosis* protein InhA. One year later, Zhan et al. [123] used support vector machine to integrate classical docking scores, interaction profiles and molecular descriptors to identify six novel Akt1 inhibitors. Durrant et al. (2015) used NNScore, a neural network SF, to describe 39 novel oestrogen-receptor ligands, whose activities were experimentally confirmed [200].

Among the ML SFs mentioned in this section, those readily accessible for use are the following: RF-Score; NN-Score; Ragoza et al.'s final optimised model architecture; DLScore; and kDEEP. These are available as downloadable standalone programs, with the exception of Kdeep, which can be found at playmolecule.org. If online docking is desired, CSM-lig [201] (for binding affinity predictions) is also available as a web-server. To the best of our knowledge, none of these SFs have been integrated into docking programs such as the ones summarised in Table 4.

Machine learning methods have shown positive results, as well as promising room for more enhancement. In addition, the availability of benchmarking data for training and testing is likely to be further expanded, which will consequently improve the predictive power of these techniques. Therefore, nonparametric machine learning is potentially the next step to drastically improve molecular docking predictiveness and accuracy.

Table 4. Recent developments using machine learning (ML) algorithms in molecular docking.

SF Name	ML Algorithm	Training Database	Best Performance	Generic or Family Specific	Type of Docking Study	Reference
RF-Score	RF ^a	PDBbind	Rp ^b = 0.776	Generic	BAP ^c	Ballester 2010 [77]
B2BScore	RF	PDBbind	Rp = 0.746	Generic	BAP	Liu 2013 [192]
SFCScore ^{RF}	RF	PDBbind	Rp = 0.779	Generic	BAP	Zilian, 2013 [202]
PostDOCK	RF	Constructed from PDB	92% accuracy	Generic	VS ^d	Springer, 2005 [181]
-	SVM ^e	DUD	-	Both	VS	Kinnings, 2011 [175]
ID-Score	SVR ^f	PDBbind	Rp = 0.85	Generic	BAP	Li, 2013 [203]
NNScore	NN ^g	PDB; MOAD; PDBbind-CN	EF = 10.3	Generic	VS	Durrant, 2010 [79]
CScore	NN	PDBbind	Rp = 0.7668 (gen.) Rp = 0.8237 (fam. spec.)	Both	BAP	Ouyang, 2011 [174]
-	Deep NN	CSAR, DUD-E	ROCAUC = 0.868	Generic	VS	Ragoza, 2017 [196]
-	Deep NN	DUD-E	ROCAUC = 0.92	Both	VS	Imrie, 2018 [183]

DLScore	Deep NN	PDBbind	Rp = 0.82	Generic	BAP	Hassan, 2018 [173]
DeepVS	Deep NN	DUD	ROCAUC = 0.81	Generic	VS	Pereira, 2016 [177]
Kdeep	Deep NN	PDBbind	Rp = 0.82	Generic	BAP	1. Jiménez z, 2018 [78]

^aRandom Forest; ^bPearson's Correlation Coefficient; ^cBinding Affinity Prediction; ^dVirtual Screening; ^eSupport Vector Machine; ^fSupport Vector Regression; ^gNeural Network.

5. Conclusions

Molecular docking has been established as a pivotal technique among the computational tools for structure-based drug discovery. Here we addressed key aspects of the methodology and discussed recent trends in the literature for advancing and employing the technique for successful drug design. Benchmarking sets and the various metrics available are crucial for validating performance gains achieved by new docking software but must be carefully chosen since no single one can be regarded as the absolute best for molecular docking. A significant improvement in the performance of all docking software can be achieved by employing multiple SFs for consensus posing and/or scoring. As reviewed here, there is a plethora of protocols for consensus docking to be explored by the user.

FBDD emerged as a successful paradigm for developing new drugs, combining the serendipity of target-based high throughput screening with the rationality of structure-based drug design approaches. Molecular docking has important roles in FBDD, from planning and prioritisation of fragment library composition to finding analogues with improved binding affinities through large-scale VS of compound libraries.

ML is a branch of artificial intelligence that has gained much attention in diverse fields of science and technology and molecular docking methods are also taking advantage of this pulsating area. Although recent, the flexibility of ML in modelling data has already rendered more diverse and accurate SFs implicitly accounting for more features from the docked complex.

Funding: FPSJr is a productivity fellow from the National Council of Technological and Scientific Development (CNPq) and holds a Newton Advanced Fellowship from the United Kingdom Academy of Medical Sciences. PHMT research is funded by The Cystic Fibrosis Trust (SRC 010 - RG92232). PJ is a M.Sc. grantee supported by the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES - Brazil).

Acknowledgments: The authors thank the Brazilian National Council for Research and Development (CNPq), the State of Rio de Janeiro Research Foundation (FAPERJ), the Oswaldo Cruz Foundation for general financial support. The authors also thank the researchers from the Blundell Group from the University of Cambridge for helpful discussions and Dr. Isabella Alvim Guedes from the National Laboratory for Scientific Computation (LNCC, Petrópolis, Brazil) for insightful clarifications regarding DockThor's algorithm.

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.

Abbreviations

ML	Machine Learning
RF	Random Forest
MD	Molecular Dynamics
SF	Scoring Function
FBDD	Fragment-Based Drug Design
VS	Virtual Screening
MW	Molecular Weight
SAR	Structure-Activity Relationship
QSAR	Quantitative Structure-Activity Relationship

EF	Enrichment Factor
ROC	Receiver Operating Characteristic
PDB	Protein Data Bank
RMSD	Root-Mean-Square Deviation
MUV	Maximum Unbiased Validation
DUD	Directory of Useful Decoys
GPCR	G-Protein-Coupled Receptor
LADS	Latent Actives in the Decoy Set
BEDROC	Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic
AUC	Area Under the Curve
RIE	Robust Initial Enhancement
DUD-E	Directory of Useful Decoys, Enhanced
DEKOIS	Demanding Evaluation Kits for Objective in Silico Screening
HTX	High Throughput X-ray Crystallography
RO3	Rule of Three
DSF	Differential Scanning Fluorimetry
R _p	Pearson correlation coefficient
R _s	Spearman rank-correlation
BFGS	Broyden–Fletcher–Goldfarb–Shanno

References

1. Liu, Y.; Zhang, Y.; Zhong, H.; Jiang, Y.; Li, Z.; Zeng, G.; Chen, M.; Shao, B.; Liu, Z.; Liu, Y. Application of molecular docking for the degradation of organic pollutants in the environmental remediation: A review. *Chemosphere* **2018**, *203*, 139–150.
2. 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.
3. 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.* **2009**, *28*, 455–461.
4. De Magalhães, C.S.; Almeida, D.M.; Barbosa, H.J.C.; Dardenne, L.E. A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Inf. Sci. (Ny)* **2014**, *289*, 206–224.
5. de Magalhães, C.S.; Barbosa, H.J.C.; Dardenne, L.E. Selection-Insertion Schemes in Genetic Algorithms for the Flexible Ligand Docking Problem. *Lect. Notes Comput. Sci.* **2004**, *3102*, 368–379.
6. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R.; Uk, K.B.R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
7. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved protein-ligand docking using GOLD. *Proteins Struct. Funct. Genet.* **2003**, *52*, 609–623.
8. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
9. Thomsen, R.; Christensen, M.H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49*, 3315–3321.
10. Gioia, D.; Bertazzo, M.; Recanatini, M.; Masetti, M.; Cavalli, A. Dynamic docking: A paradigm shift in computational drug discovery. *Molecules* **2017**, *22*, 2029.
11. Berman, H.M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
12. Hetényi, C.; Van Der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* **2006**, *580*, 1447–1450.
13. Volkamer, A.; Kuhn, D.; Grombacher, T.; Rippmann, F.; Rarey, M. Combining global and local measures for structure-based druggability predictions. *J. Chem. Inf. Model.* **2012**, *52*, 360–372.
14. Radoux, C.J.; Olsson, T.S.G.; Pitt, W.R.; Groom, C.R.; Blundell, T.L. Identifying Interactions that Determine Fragment Binding at Protein Hotspots. *J. Med. Chem.* **2016**, *59*, 4314–4325.
15. Fu, D.Y.; Meiler, J. Predictive Power of Different Types of Experimental Restraints in Small Molecule Docking: A Review. *J. Chem. Inf. Model.* **2018**, *58*, 225–233.
16. Brooijmans, N.; Kuntz, I.D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.

17. Meng, E.C.; Shoichet, B.K.; Kuntz, I.D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
18. Irwin, J.J.; Shoichet, B.K. ZINC—A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2006**, *45*, 177–182.
19. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.
20. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17.
21. Pearlman, R.S. Rapid Generation of High Quality Approximate 3-dimension Molecular Structures. *Chem. Des. Auto. News* **1987**, *2*, 1–7.
22. McCammon, J.A.; Nielsen, J.E.; Baker, N.A.; Dolinsky, T.J. PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.
23. Anandakrishnan, R.; Aguilar, B.; Onufriev, A.V. H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541.
24. 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.
25. Dardenne, L.E.; Barbosa, H.J.C.; De Magalhães, C.S.; Almeida, D.M.; da Silva, E.K.; Custódio, F.L.; Guedes, I.A. DockThor Portal. Available online: <https://dockthor.lncc.br/v2/> (accessed on 22 March 2019).
26. Guedes, I.A.; de Magalhães, C.S.; Dardenne, L.E. Receptor-ligand molecular docking. *Biophys. Rev.* **2014**, *6*, 75–87.
27. Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. DOCKING and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935.
28. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B.S.; Johnson, A.P. eHiTS: An Innovative Approach to the Docking and Scoring Function Problems. *Curr. Protein Pept. Sci.* **2006**, *7*, 421–435.
29. Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C.R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153*, 7–26.
30. Hindle, S.A.; Rarey, M.; Buning, C.; Lengauer, T. Flexible docking under pharmacophore type constraints. *J. Comput. Aided Mol. Des.* **2002**, *16*, 129–149.
31. Huey, R.; Morris, G.M.; Olson, A.J.; Goodsell, D.S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152.
32. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
33. Koes, D.R.; Baumgartner, M.P.; Camacho, C.J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904.
34. Korb, O.; Stütze, T.; Exner, T.E. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96.
35. Korb, O.; Stütze, T.; Exner, T.E. An ant colony optimization approach to flexible protein–ligand docking. *Swarm Intell.* **2007**, *1*, 115–134.
36. Abagyan, R.; Kuznetsov, D.; Totrov, M. ICM-New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from. *J. Comput. Chem.* **1994**, *15*, 488–506.
37. Abagyan, R.; Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.
38. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
39. Jain, A.N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.
40. 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.
41. Yang, J.M.; Chen, C.C. GEMDOCK: A Generic Evolutionary Method for Molecular Docking. *Proteins Struct. Funct. Bioinform.* **2004**, *55*, 288–304.

42. 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.
43. Li, H.; Li, C.; Gui, C.; Luo, X.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. GAsDock: A new approach for rapid flexible docking based on an improved multi-population genetic algorithm. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4671–4676.
44. Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput. Aided Mol. Des.* **1996**, *10*, 41–54.
45. McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578–596.
46. Plewczynski, D.; Łażniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, *32*, 742–755.
47. Chang, M.W.; Ayeni, C.; Breuer, S.; Torbett, B.E. Virtual screening for HIV protease inhibitors: A comparison of AutoDock 4 and Vina. *PLoS ONE* **2010**, *5*, e11955.
48. Capoferri, L.; Leth, R.; ter Haar, E.; Mohanty, A.K.; Grootenhuys, P.D.J.; Vottero, E.; Commandeur, J.N.M.; Vermeulen, N.P.E.; Jørgensen, F.S.; Olsen, L.; et al. Insights into regioselective metabolism of mefenamic acid by cytochrome P450 BM3 mutants through crystallography, docking, molecular dynamics, and free energy calculations. *Proteins Struct. Funct. Bioinform.* **2016**, *84*, 383–396.
49. Feng, Z.; Pearce, L.V.; Xu, X.; Yang, X.; Yang, P.; Blumberg, P.M.; Xie, X.-Q. Structural Insight into Tetrameric hTRPV1 from Homology Modeling, Molecular Docking, Molecular Dynamics Simulation, Virtual Screening, and Bioassay Validations. *J. Chem. Inf. Model.* **2015**, *55*, 572–588.
50. Vadloori, B.; Sharath, A.K.; Prabhu, N.P.; Maurya, R. Homology modelling, molecular docking, and molecular dynamics simulations reveal the inhibition of Leishmania donovani dihydrofolate reductase-thymidylate synthase enzyme by Withaferin-A. *BMC Res. Notes* **2018**, *11*, 246.
51. Yadav, D.K.; Kumar, S.; Misra, S.; Yadav, L.; Teli, M.; Sharma, P.; Chaudhary, S.; Kumar, N.; Choi, E.H.; Kim, H.S.; et al. Molecular Insights into the Interaction of RONS and Thieno [3, 2-c]pyran Analogs with SIRT6/COX-2: A Molecular Dynamics Study. *Sci. Rep.* **2018**, *8*, 4777.
52. Makhouri, F.R.; Ghasemi, J.B. Combating Diseases with Computational Strategies Used for Drug Design and Discovery. *Curr. Top. Med. Chem.* **2019**, *18*, 2743–2773.
53. Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964–12975.
54. Halgren, T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
55. Hingerty, B.E.; Ritchie, R.H.; Ferrell, T.L.; Turner, J.E. Dielectric effects in biopolymers: The theory of ionic saturation revisited. *Biopolymers* **1985**, *24*, 427–439.
56. Halgren, T.A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.
57. Hansch, C.; Fujita, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.
58. Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425–445.
59. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
60. Velec, H.F.G.; Gohlke, H.; Klebe, G. 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. *J. Med. Chem.* **2005**, *48*, 6296–6303.
61. Muegge, I. PMF scoring revisited. *J. Med. Chem.* **2006**, *49*, 5895–5902.
62. Bohacek, R.S.; McMartin, C.; Guida, W.C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **1996**, *16*, 3–50.
63. 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.

64. Korb, O.; Olsson, T.S.G.; 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.
65. Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.
66. De Paris, R.; Vahl Quevedo, C.; Ruiz, D.D.; Gargano, F.; de Souza, O.N. A selective method for optimizing ensemble docking-based experiments on an InhA Fully-Flexible receptor model. *BMC Bioinform.* **2018**, *19*, 235.
67. De Paris, R.; Frantz, F.A.; Norberto de Souza, O.; Ruiz, D.D.A. wFReDoW: A Cloud-Based Web Environment to Handle Molecular Docking Simulations of a Fully Flexible Receptor Model. *BioMed Res. Int.* **2013**, *2013*, 469363.
68. Cavasotto, C.N.; Kovacs, J.A.; Abagyan, R.A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.
69. Damm, K.L.; Carlson, H.A. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* **2007**, *129*, 8225–8235.
70. Leach, A.R.; Shoichet, B.K.; Peishoff, C.E. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.
71. Wang, R.; Fang, X.; Lu, Y. The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures—Journal of Medicinal Chemistry (ACS Publications). *J. Med. Chem.* **2004**, *47*, 2977–2980.
72. Ahmed, A.; Smith, R.D.; Clark, J.J.; Dunbar, J.B., Jr.; Carlson, H.A. Recent improvements to Binding MOAD: A resource for protein-ligand Binding affinities and structures. *Nucleic Acids Res.* **2015**, *43*, D465–D469.
73. Smith, R.D.; Ung, P.M.-U.; Esposito, E.X.; Wang, S.; Carlson, H.A.; Dunbar, J.B.; Yang, C.-Y. CSAR Benchmark Exercise of 2010: Combined Evaluation Across All Submitted Scoring Functions. *J. Chem. Inf. Model.* **2011**, *51*, 2115–2131.
74. Block, P. AffinDB: A freely accessible database of affinities for protein-ligand complexes from the PDB. *Nucleic Acids Res.* **2006**, *34*, D522–D526.
75. Wang, R.; Fang, X.; Lu, Y.; Yang, C.Y.; Wang, S. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.
76. Zhao, Z.; Liu, J.; Wang, R.; Liu, Z.; Liu, Y.; Han, L.; Li, Y.; Nie, W.; Li, J. PDB-wide collection of binding data: Current status of the PDBbind database. *Bioinformatics* **2014**, *31*, 405–412.
77. Ballester, P.J.; Mitchell, J.B.O. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* **2010**, *26*, 1169–1175.
78. Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G. KDEEP: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 287–296.
79. 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.
80. Vreven, T.; Moal, I.H.; Vangone, A.; Pierce, B.G.; Kastriitis, P.L.; Torchala, M.; Chaleil, R.; Jiménez-García, B.; Bates, P.A.; Fernandez-Recio, J.; et al. Updates to the Integrated Protein–Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. *J. Mol. Biol.* **2015**, *427*, 3031–3041.
81. Koukos, P.I.; Faro, I.; van Noort, C.W.; Bonvin, A.M.J.J. A Membrane Protein Complex Docking Benchmark. *J. Mol. Biol.* **2018**, *430*, 5246–5256.
82. Li, H.; Leung, K.S.; Wong, M.H.; Ballester, P.J. Improving autodock vina using random forest: The growing accuracy of binding affinity prediction by the effective exploitation of larger data sets. *Mol. Inform.* **2015**, *34*, 115–126.
83. 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.
84. Irwin, J.J. Community benchmarks for virtual screening. *J. Comput. Aided Mol. Des.* **2008**, *22*, 193–199.
85. Kirchmair, J.; Markt, A.P.; Distinto, S.; Wolber, A.G. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *J. Comput. Aided Mol. Des.* **2008**, *22*, 213–228.
86. Verdonk, M.L.; Berdini, V.; Hartshorn, M.J.; Mooij, W.T.M.; Murray, C.W.; Taylor, R.D.; Watson, P. Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.

87. Good, A.C.; Oprea, T.I. Optimization of CAMD techniques 3. Virtual screening enrichment studies: A help or hindrance in tool selection? *J. Comput. Aided Mol. Des.* **2008**, *22*, 169–178.
88. Lovell, T.; Chen, H.; Lyne, P.D.; Giordanetto, F.; Li, J. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. [*J. Chem. Inf. Model.* **46**, 401–415 (2006)] by. *J. Chem. Inf. Model.* **2008**, *48*, 246–246.
89. Vogel, S.M.; Bauer, M.R.; Boeckler, F.M. DEKOIS: Demanding evaluation kits for objective in silico screening—A versatile tool for benchmarking docking programs and scoring functions. *J. Chem. Inf. Model.* **2011**, *51*, 2650–2665.
90. Huang, N.; Shoichet, B.K.; Irwin, J.J. Benchmarking Sets for Molecular Docking Benchmarking Sets for Molecular Docking. *Society* **2006**, *49*, 6789–6801.
91. Wallach, I.; Lilien, R. Virtual decoy sets for molecular docking benchmarks. *J. Chem. Inf. Model.* **2011**, *51*, 196–202.
92. 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.
93. Bauer, M.R.; Ibrahim, T.M.; Vogel, S.M.; Boeckler, F.M. Evaluation and optimization of virtual screening workflows with DEKOIS 2.0—A public library of challenging docking benchmark sets. *J. Chem. Inf. Model.* **2013**, *53*, 1447–1462.
94. 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.
95. Gatica, E.A.; Cavasotto, C.N. Ligand and decoy sets for docking to G protein-coupled receptors. *J. Chem. Inf. Model.* **2012**, *52*, 1–6.
96. Lagarde, N.; Ben Nasr, N.; Jérémie, A.; Guillemain, H.; Laville, V.; Labib, T.; Zagury, J.F.; Montes, M. NRLiSt BDB, the manually curated nuclear receptors ligands and structures benchmarking database. *J. Med. Chem.* **2014**, *57*, 3117–3125.
97. Xia, J.; Tilahun, E.L.; Kebede, E.H.; Reid, T.E.; Zhang, L.; Wang, X.S. Comparative modeling and benchmarking data sets for human histone deacetylases and sirtuin families. *J. Chem. Inf. Model.* **2015**, *55*, 374–388.
98. Cereto-Massagué, A.; Guasch, L.; Valls, C.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. DecoyFinder: An easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics* **2012**, *28*, 1661–1662.
99. Wang, L.; Pang, X.; Li, Y.; Zhang, Z.; Tan, W. RADER: A RAPID DEcoy RETRIEVER to facilitate decoy based assessment of virtual screening. *Bioinformatics* **2017**, *33*, 1235–1237.
100. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874.
101. Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual Screening Workflow Development Guided by the “Receiver Operating Characteristic” Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2547.
102. Truchon, J.F.; Bayly, C.I. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
103. Empereur-Mot, C.; Guillemain, H.; Latouche, A.; Zagury, J.F.; Viallon, V.; Montes, M. Predictiveness curves in virtual screening. *J. Cheminform.* **2015**, *7*, doi:10.1186/s13321-015-0100-8.
104. Alghamedy, F.; Bopaiah, J.; Jones, D.; Zhang, X.; Weiss, H.L.; Ellingson, S.R. Incorporating Protein Dynamics Through Ensemble Docking in Machine Learning Models to Predict Drug Binding. *AMIA Jt. Summits Transl. Sci.* **2018**, *2017*, 26–34.
105. Sheridan, R.P.; Singh, S.B.; Fluder, E.M.; Kearsley, S.K. Protocols for Bridging the Peptide to Nonpeptide Gap in Topological Similarity Searches. *J. Chem. Inf. Comput. Sci.* **2002**, *41*, 1395–1406.
106. 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.
107. Wang, R.; Wang, S. How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422–1426.
108. Kang, L.; Li, H.; Jiang, H.; Wang, X.; Zheng, M.; Luo, J.; Zhang, H.; Liu, X. An effective docking strategy for virtual screening based on multi-objective optimization algorithm. *BMC Bioinform.* **2009**, *10*, 58.

109. Nguyen, D.D.; Cang, Z.; Wu, K.; Wang, M.; Cao, Y.; Wei, G.W. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. *J. Comput. Aided Mol. Des.* **2018**, *33*, 71–82.
110. Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
111. 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.
112. Poli, G.; Martinelli, A.; Tuccinardi, T. Reliability analysis and optimization of the consensus docking approach for the development of virtual screening studies. *J. Enzyme Inhib. Med. Chem.* **2016**, *31*, 167–173.
113. Tuccinardi, T.; Poli, G.; Romboli, V.; Giordano, A.; Martinelli, A. Extensive consensus docking evaluation for ligand pose prediction and virtual screening studies. *J. Chem. Inf. Model.* **2014**, *54*, 2980–2986.
114. Houston, D.R.; Walkinshaw, M.D. Consensus docking: Improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model.* **2013**, *53*, 384–390.
115. Plewczynski, D.; Łażniewski, M.; Grotthuss, M. Von; Rychlewski, L.; Ginalski, K. VoteDock: Consensus docking method for prediction of protein-ligand interactions. *J. Comput. Chem.* **2011**, *32*, 568–581.
116. Perez-castillo, Y.; Sotomayor-burneo, S.; Jimenes-vargas, K.; Gonzalez-, M. CompScore: Boosting structure-based virtual screening performance by incorporating docking scoring functions components into consensus scoring. *BioRxiv* **2019**, doi:10.1101/550590.
117. Onawole, A.T.; Kolapo, T.U.; Sulaiman, K.O.; Adegoke, R.O. Structure based virtual screening of the Ebola virus trimeric glycoprotein using consensus scoring. *Comput. Biol. Chem.* **2018**, *72*, 170–180.
118. Aliebrahimi, S.; Karami, L.; Arab, S.S.; Montasser Kouhsari, S.; Ostad, S.N. Identification of Phytochemicals Targeting c-Met Kinase Domain using Consensus Docking and Molecular Dynamics Simulation Studies. *Cell Biochem. Biophys.* **2017**, *76*, 135–145.
119. Li, D.D.; Meng, X.F.; Wang, Q.; Yu, P.; Zhao, L.G.; Zhang, Z.P.; Wang, Z.Z.; Xiao, W. Consensus scoring model for the molecular docking study of mTOR kinase inhibitor. *J. Mol. Graph. Model.* **2018**, *79*, 81–87.
120. Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J. Chem. Inf. Model.* **2006**, *46*, 380–391.
121. Chaput, L.; Martinez-Sanz, J.; Quiniou, E.; Rigolet, P.; Saettel, N.; Mouawad, L. VSDC: A method to improve early recognition in virtual screening when limited experimental resources are available. *J. Cheminform.* **2016**, *8*, doi:10.1186/s13321-016-0112-z.
122. Mavrogeni, M.E.; Pronios, F.; Zareifi, D.; Vasilakaki, S.; Lozach, O.; Alexopoulos, L.; Meijer, L.; Myriantopoulos, V.; Mikros, E. A facile consensus ranking approach enhances virtual screening robustness and identifies a cell-active DYRK1 α inhibitor. *Future Med. Chem.* **2018**, *10*, 2411–2430.
123. Zhan, W.; Li, D.; Che, J.; Zhang, L.; Yang, B.; Hu, Y.; Liu, T.; Dong, X. Integrating docking scores, interaction profiles and molecular descriptors to improve the accuracy of molecular docking: Toward the discovery of novel Akt1 inhibitors. *Eur. J. Med. Chem.* **2014**, *75*, 11–20.
124. Ericksen, S.S.; Wu, H.; Zhang, H.; Michael, L.A.; Newton, M.A.; Hoffmann, F.M.; Wildman, S.A. Machine Learning Consensus Scoring Improves Performance Across Targets in Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2017**, *57*, 1579–1590.
125. Teramoto, R.; Fukunishi, H. Supervised consensus scoring for docking and virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 526–534.
126. Wang, W.; He, W.; Zhou, X.; Chen, X. Optimization of molecular docking scores with support vector rank regression. *Proteins Struct. Funct. Bioinform.* **2013**, *81*, 1386–1398.
127. Yang, J.M.; Hsu, D.F. Consensus scoring criteria in structure-based virtual screening. *Emerg. Inf. Technol. Conf. 2005* **2005**, *2005*, 165–167.
128. Liu, S.; Fu, R.; Zhou, L.-H.; Chen, S.-P. Application of Consensus Scoring and Principal Component Analysis for Virtual Screening against β -Secretase (BACE-1). *PLoS ONE* **2012**, *7*, e38086.
129. Mokrani, E.H.; Bensegueni, A.; Chaput, L.; Beauvineau, C.; Djeghim, H.; Mouawad, L. Identification of New Potent Acetylcholinesterase Inhibitors Using Virtual Screening and In Vitro Approaches. *Mol. Inform.* **2019**, *38*, 1800118.
130. Russo Spena, C.; De Stefano, L.; Poli, G.; Granchi, C.; El Boustani, M.; Ecce, F.; Grassi, G.; Grassi, M.; Canzonieri, V.; Giordano, A.; et al. Virtual screening identifies a PIN1 inhibitor with possible antiovarian cancer effects. *J. Cell. Physiol.* **2019**, doi:10.1002/jcp.28224.

131. Mouawad, N.; Jha, V.; Poli, G.; Granchi, C.; Rizzolio, F.; Caligiuri, I.; Minutolo, F.; Lapillo, M.; Tuccinardi, T.; Macchia, M. Computationally driven discovery of phenyl(piperazin-1-yl) methanone derivatives as reversible monoacylglycerol lipase (MAGL) inhibitors. *J. Enzyme Inhib. Med. Chem.* **2019**, *34*, 589–596.
132. Damm-Ganamet, K.L.; Dunbar, J.B.; Ahmed, A.; Esposito, E.X.; Stuckey, J.A.; Gestwicki, J.E.; Chinnaswamy, K.; Delproposito, J.; Smith, R.D.; Carlson, H.A.; et al. CSAR Data Set Release 2012: Ligands, Affinities, Complexes, and Docking Decoys. *J. Chem. Inf. Model.* **2013**, *53*, 1842–1852.
133. Walters, W.P.; Liu, S.; Chiu, M.; Shao, C.; Rudolph, M.G.; Burley, S.K.; Gilson, M.K.; Feher, V.A.; Gaieb, Z.; Kuhn, B.; et al. D3R Grand Challenge 2: Blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **2017**, *32*, 1–20.
134. Nevins, N.; Yang, H.; Walters, W.P.; Ameriks, M.K.; Parks, C.D.; Gilson, M.K.; Gaieb, Z.; Lambert, M.H.; Shao, C.; Chiu, M.; et al. D3R Grand Challenge 3: Blind prediction of protein–ligand poses and affinity rankings. *J. Comput. Aided Mol. Des.* **2019**, *33*, 1–18.
135. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.
136. Romasanta, A.K.S.; van der Sijde, P.; Hellsten, I.; Hubbard, R.E.; Keseru, G.M.; van Muijlwijk-Koezen, J.; de Esch, I.J.P. When fragments link: A bibliometric perspective on the development of fragment-based drug discovery. *Drug Discov. Today* **2018**, *23*, 1596–1609.
137. Erlanson, D.A. Introduction to fragment-based drug discovery. *Top. Curr. Chem.* **2012**, *317*, 1–32.
138. Hann, M.M.; Leach, A.R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.
139. Leach, A.R.; Hann, M.M. Molecular complexity and fragment-based drug discovery: Ten years on. *Curr. Opin. Chem. Biol.* **2011**, *15*, 489–496.
140. Fink, T.; Raymond, J.L. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: Assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *J. Chem. Inf. Model.* **2007**, *47*, 342–353.
141. Lyu, J.; Irwin, J.J.; Roth, B.L.; Shoichet, B.K.; Levit, A.; Wang, S.; Tolmacheva, K.; Singh, I.; Tolmachev, A.A.; Che, T.; et al. Ultra-large library docking for discovering new chemotypes. *Nature* **2019**, *566*, 224.
142. Scott, D.E.; Coyne, A.G.; Hudson, S.A.; Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **2012**, *51*, 4990–5003.
143. Blundell, T.L.; Jhoti, H.; Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* **2002**, *1*, 45–54.
144. Hopkins, A.L.; Groom, C.R.; Alex, A. Ligand efficiency: A useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430–431.
145. Abad-Zapatero, C.; Metz, J.T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* **2005**, *10*, 464–469.
146. Reynolds, C.H.; Bembenek, S.D.; Tounge, B.A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4258–4261.
147. Schultes, S.; De Graaf, C.; Haaksma, E.E.J.; De Esch, I.J.P.; Leurs, R.; Krämer, O. Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discov. Today Technol.* **2010**, *7*, 157–162.
148. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A ‘Rule of Three’ for fragment-based lead discovery? Recent. *Drug Discov. Today* **2003**, *8*, 876–877.
149. Jhoti, H.; Williams, G.; Rees, D.C.; Murray, C.W. The “rule of three” for fragment-based drug discovery: Where are we now? *Nat. Rev. Drug Discov.* **2013**, *12*, 644–644.
150. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.
151. Morley, A.D.; Pugliese, A.; Birchall, K.; Bower, J.; Brennan, P.; Brown, N.; Chapman, T.; Drysdale, M.; Gilbert, I.H.; Hoelder, S.; et al. Fragment-based hit identification: Thinking in 3D. *Drug Discov. Today* **2013**, *18*, 1221–1227.
152. Verheij, H.J. Leadlikeness and structural diversity of synthetic screening libraries. *Mol. Divers.* **2006**, *10*, 377–388.
153. Fischer, M.; Hubbard, R.E. Fragment-based ligand discovery. *Mol. Interv.* **2009**, *9*, 22–30.
154. Schuffenhauer, A.; Ruedisser, S.; Jahnke, W.; Marzinzik, A.; Selzer, P.; Jacoby, E. Library Design for Fragment Based Screening. *Curr. Top. Med. Chem.* **2005**, *5*, 751–762.

155. Lewell, X.Q.; Judd, D.B.; Watson, S.P.; Hann, M.M. RECAP—Retrosynthetic Combinatorial Analysis Procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.
156. Prescher, H.; Koch, G.; Schuhmann, T.; Ertl, P.; Bussenault, A.; Glick, M.; Dix, I.; Petersen, F.; Lizos, D.E. Construction of a 3D-shaped, natural product like fragment library by fragmentation and diversification of natural products. *Bioorg. Med. Chem.* **2017**, *25*, 921–925.
157. Chen, Y.; Shoichet, B.K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358–364.
158. Fjellström, O.; Akkaya, S.; Beisel, H.G.; Eriksson, P.O.; Erixon, K.; Gustafsson, D.; Jurva, U.; Kang, D.; Karis, D.; Knecht, W.; et al. Creating novel activated factor XI inhibitors through fragment based lead generation and structure aided drug design. *PLoS ONE* **2015**, *10*, e0113705.
159. Park, H.; Shin, Y.; Kim, J.; Hong, S. Application of Fragment-Based de Novo Design to the Discovery of Selective Picomolar Inhibitors of Glycogen Synthase Kinase-3 Beta. *J. Med. Chem.* **2016**, *59*, 9018–9034.
160. Wang, R.; Gao, Y.; Lai, L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *J. Mol. Model.* **2004**, *6*, 498–516.
161. Zhao, H.; Gartenmann, L.; Dong, J.; Spiliotopoulos, D.; Caflisch, A. Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2493–2496.
162. Rudling, A.; Gustafsson, R.; Almlöf, I.; Homan, E.; Scobie, M.; Warpman Berglund, U.; Helleday, T.; Stenmark, P.; Carlsson, J. Fragment-Based Discovery and Optimization of Enzyme Inhibitors by Docking of Commercial Chemical Space. *J. Med. Chem.* **2017**, *60*, 8160–8169.
163. Hernandez, J.; Hoffer, L.; Coutard, B.; Querat, G.; Roche, P.; Morelli, X.; Decroly, E.; Barral, K. Optimization of a fragment linking hit toward Dengue and Zika virus NS5 methyltransferases inhibitors. *Eur. J. Med. Chem.* **2019**, *161*, 323–333.
164. Akabayov, S.R.; Richardson, C.C.; Arthanari, H.; Akabayov, B.; Ilic, S.; Wagner, G. Identification of DNA primase inhibitors via a combined fragment-based and virtual screening. *Sci. Rep.* **2016**, *6*, 36322.
165. Amaning, K.; Lowinski, M.; Vallee, F.; Steier, V.; Marcireau, C.; Ugolini, A.; Delorme, C.; Foucalt, F.; McCort, G.; Derimay, N.; et al. The use of virtual screening and differential scanning fluorimetry for the rapid identification of fragments active against MEK1. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 3620–3626.
166. Barelier, S.; Eidam, O.; Fish, I.; Hollander, J.; Figaroa, F.; Nachane, R.; Irwin, J.J.; Shoichet, B.K.; Siegal, G. Increasing chemical space coverage by combining empirical and computational fragment screens. *ACS Chem. Biol.* **2014**, *9*, 1528–1535.
167. Adams, M.; Kobayashi, T.; Lawson, J.D.; Saitoh, M.; Shimokawa, K.; Bigi, S.V.; Hixon, M.S.; Smith, C.R.; Tatamiya, T.; Goto, M.; et al. Fragment-based drug discovery of potent and selective MKK3/6 inhibitors. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1086–1089.
168. Darras, F.H.; Pockes, S.; Huang, G.; Wehle, S.; Strasser, A.; Wittmann, H.J.; Nimczick, M.; Sottriffer, C.A.; Decker, M. Synthesis, biological evaluation, and computational studies of Tri- and tetracyclic nitrogen-bridgehead compounds as potent dual-acting AChE inhibitors and h H3 receptor antagonists. *ACS Chem. Neurosci.* **2014**, *5*, 225–242.
169. He, Y.; Guo, X.; Yu, Z.H.; Wu, L.; Gunawan, A.M.; Zhang, Y.; Dixon, J.E.; Zhang, Z.Y. A potent and selective inhibitor for the UBLCP1 proteasome phosphatase. *Bioorg. Med. Chem.* **2015**, *23*, 2798–2809.
170. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2007; ISBN 978-0-387-31073-2.
171. Ashtawy, H.M.; Mahapatra, N.R. A comparative assessment of predictive accuracies of conventional and machine learning scoring functions for protein-ligand binding affinity prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, *12*, 335–347.
172. Ashtawy, H.M.; Mahapatra, N.R. Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins. *BMC Bioinform.* **2015**, *16*, doi:10.1186/1471-2105-16-S6-S3.
173. Hassan, M.; Mogollón, D.C.; Fuentes, O. DLSCORE: A Deep Learning Model for Predicting Protein-Ligand Binding Affinities. *ChemRxiv* **2018**, *13*, 53.
174. Ouyang, X.; Handoko, S.D.; Kwok, C.K. Cscore: A Simple Yet Effective Scoring Function for Protein–Ligand Binding Affinity Prediction Using Modified Cmac Learning Architecture. *J. Bioinform. Comput. Biol.* **2011**, *9*, 1–14.

175. Kinnings, S.L.; Liu, N.; Tonge, P.J.; Jackson, R.M.; Xie, L.; Bourne, P.E. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J. Chem. Inf. Model.* **2011**, *51*, 408–419.
176. Hsin, K.Y.; Ghosh, S.; Kitano, H. Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology. *PLoS ONE* **2013**, *8*, e83922.
177. Pereira, J.C.; Caffarena, E.R.; Dos Santos, C.N. Boosting Docking-Based Virtual Screening with Deep Learning. *J. Chem. Inf. Model.* **2016**, *56*, 2495–2506.
178. Pason, L.P.; Sottriffer, C.A. Empirical Scoring Functions for Affinity Prediction of Protein-ligand Complexes. *Mol. Inform.* **2016**, *35*, 541–548.
179. Silva, C.G.; Simoes, C.J.V.; Carreiras, P.; Brito, R.M.M. Enhancing Scoring Performance of Docking-Based Virtual Screening Through Machine Learning. *Curr. Bioinform.* **2016**, *11*, 408–420.
180. Korkmaz, S.; Zararsiz, G.; Goksuluk, D. MLViS: A web tool for machine learning-based virtual screening in early-phase of drug discovery and development. *PLoS ONE* **2015**, *10*, e0124600.
181. Springer, C.; Adalsteinsson, H.; Young, M.M.; Kegelmeyer, P.W.; Roe, D.C. PostDOCK: A Structural, Empirical Approach to Scoring Protein Ligand Complexes. *J. Med. Chem.* **2005**, *48*, 6821–6831.
182. 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.
183. Imrie, F.; Bradley, A.R.; Van Der Schaar, M.; Deane, C.M. Protein Family-Specific Models Using Deep Neural Networks and Transfer Learning Improve Virtual Screening and Highlight the Need for More Data. *J. Chem. Inf. Model.* **2018**, *58*, 2319–2330.
184. Wang, Y.; Guo, Y.; Kuang, Q.; Pu, X.; Ji, Y.; Zhang, Z.; Li, M. A comparative study of family-specific protein-ligand complex affinity prediction based on random forest approach. *J. Comput. Aided Mol. Des.* **2015**, *29*, 349–360.
185. Wójcikowski, M.; Ballester, P.J.; Siedlecki, P. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci. Rep.* **2017**, *7*, 46710.
186. Cao, Y.; Li, L. Improved protein-ligand binding affinity prediction by using a curvature-dependent surface-area model. *Bioinformatics* **2014**, *30*, 1674–1680.
187. Yuriev, E.; Ramsland, P.A. Latest developments in molecular docking: 2010–2011 in review. *J. Mol. Recognit.* **2013**, *26*, 215–239.
188. Guedes, I.A.; Pereira, F.S.S.; Dardenne, L.E. Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Front. Pharmacol.* **2018**, *9*, 1089.
189. Li, L.; Wang, B.; Meroueh, S.O. Support Vector Regression Scoring of Receptor–Ligand Complexes for Rank-Ordering and Virtual Screening of Chemical Libraries. *J. Chem. Inf. Model.* **2011**, *51*, 2132–2138.
190. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2011**, *3*, 1157–1182.
191. Koppisetty, C.A.K.; Frank, M.; Kemp, G.J.L.; Nyholm, P.G. Computation of binding energies including their enthalpy and entropy components for protein-ligand complexes using support vector machines. *J. Chem. Inf. Model.* **2013**, *53*, 2559–2570.
192. Liu, Q.; Kwok, C.K.; Li, J. Binding affinity prediction for protein-ligand complexes based on β contacts and B factor. *J. Chem. Inf. Model.* **2013**, *53*, 3076–3085.
193. Ballester, P.J.; Schreyer, A.; Blundell, T.L. Does a More Precise Chemical Description of Protein–Ligand Complexes Lead to More Accurate Prediction of Binding Affinity? *J. Chem. Inf. Model.* **2014**, *54*, 944–955.
194. Kundu, I.; Paul, G.; Banerjee, R. A machine learning approach towards the prediction of protein–ligand binding affinity based on fundamental molecular properties. *RSC Adv.* **2018**, *8*, 12127–12137.
195. Srinivas, R.; Klimovich, P.V.; Larson, E.C. Implicit-descriptor ligand-based virtual screening by means of collaborative filtering. *J. Cheminform.* **2018**, *10*, 56.
196. 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.
197. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Convolutional Neural Networks ImageNet Classification with Deep Convolutional Neural Network. *Commun. ACM* **2017**, *60*, doi:10.1145/3065386.
198. Khamis, M.A.; Goma, W.; Ahmed, W.F. Machine learning in computational docking. *Artif. Intell. Med.* **2015**, *63*, 135–152.
199. Sieg, J.; Flachsenberg, F.; Rarey, M. In the need of bias control: Evaluation of chemical data for Machine Learning Methods in Virtual Screening. *J. Chem. Inf. Model.* **2019**, *59*, 947–961.

200. Durrant, J.D.; Carlson, K.E.; Martin, T.A.; Offutt, T.L.; Mayne, C.G.; Katzenellenbogen, J.A.; Amaro, R.E. Neural-Network Scoring Functions Identify Structurally Novel Estrogen-Receptor Ligands. *J. Chem. Inf. Model.* **2015**, *55*, 1953–1961.
201. Pires, D.E.V.; Ascher, D.B. CSM-lig: A web server for assessing and comparing protein-small molecule affinities. *Nucleic Acids Res.* **2016**, *44*, W557–W561.
202. Zilian, D.; Sotriffer, C.A. SFCscore RF: A Random Forest-Based Scoring Function for Improved Affinity Prediction of Protein–Ligand Complexes. *J. Chem. Inf. Model.* **2013**, *53*, 1923–1933.
203. Li, G.-B.; Yang, L.-L.; Wang, W.-J.; Li, L.-L.; Yang, S.-Y. ID-Score: A New Empirical Scoring Function Based on a Comprehensive Set of Descriptors Related to Protein–Ligand Interactions. *J. Chem. Inf. Model.* **2013**, *53*, 592–600.



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).