Review

Charting a Path to Success in Virtual Screening

Stefano Forli

Molecular Graphics Laboratory, Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA; E-Mail: forli@scripps.edu; Tel.: +1-858-784-2055; Fax: +1-858-784-2860

Abstract: Docking is commonly applied to drug design efforts, especially high-throughput virtual screenings of small molecules, to identify new compounds that bind to a given target. Despite great advances and successful applications in recent years, a number of issues remain unsolved. Most of the challenges and problems faced when running docking experiments are independent of the specific software used, and can be ascribed to either improper input preparation or to the simplified approaches applied to achieve high-throughput speed. Being aware of approximations and limitations of such methods is essential to prevent errors, deal with misleading results, and increase the success rate of virtual screening campaigns. In this review, best practices and most common issues of docking and virtual screening will be discussed, covering the journey from the design of the virtual experiment to the hit identification.

Keywords: docking; drug design; molecular modeling; virtual screening

1. Introduction

The goal of a virtual screen is to provide chemical insights and inspirations for designing a drug [1]. Automated docking of small molecules is applied in many drug design efforts, and due to the ubiquitous availability of large computational resources, high throughput virtual screenings are now routinely applied in campaigns to assess druggability of novel targets where often no binders are known. A large number of successful applications have been reported using a variety of docking techniques [2–11]. However, despite the improvements in recent years, a number of issues remain unsolved. The binding event is a fairly intricate process which imposes at the same time dramatic and subtle changes on the components forming the complex, where ostensible conformational changes are associated with

invisible perturbations in the energetic equilibria. Modeling this event, where enthalpy and entropy components are deeply entangled, is a non-trivial task that can require a considerable computational effort. Indeed, most of the challenges faced when running docking experiments are due to the simplified approaches applied to achieve high-throughput speed and evaluate large numbers of ligands. Being aware of approximations and limitations of the method is essential to prevent errors, deal with misleading results, and increase the success rate of virtual screening campaigns.

At the dawn of computer history, Charles Babbage, the creator of the first programmable computer was asked: "*Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?*" [12]. Almost two centuries later, the answer is still the same: starting from wrong data will inevitably compromise the results, regardless of the accuracy of the calculation performed. Therefore, the very first rule of any scientific simulation should be to ensure the highest possible quality of input data, which, specifically for docking, refers to ligand and target input structures. This will not guarantee success by all means, but will definitely reduce chances of failure. Best practices and most common issues of docking and virtual screening will be discussed, covering the journey from the design of the virtual experiment to the hit identification (Figure 1).
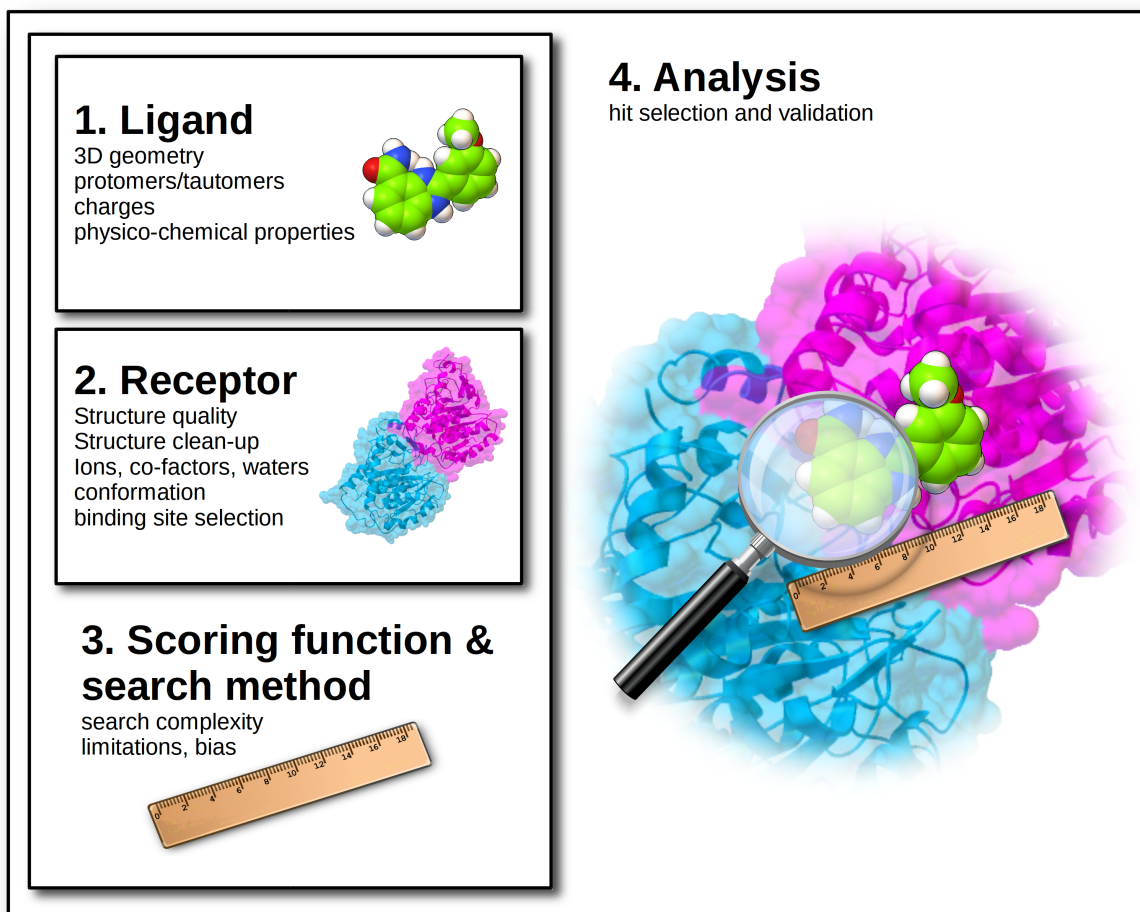

Figure 1. Overview of the main issues to be addressed when designing a virtual screening experiment.

2. Ligand Structures

The choice of the library of ligands to be screened is one of the most stringent criteria, *de facto* determining the chances of finding a potential binder for the target of interest. The following guidelines should be considered when generating in-house virtual libraries, while the use of repositories such as ZINC [13] will provide high quality ligand structures ready to be docked.

2.1. Accurate 3D Geometries

In order to cut computation times, the vast majority of docking programs do not alter bond angles or bond lengths during the calculation, but sample only torsion angles. For these reason, it is a matter of paramount importance to generate proper input geometries with optimal bond angles and lengths (although, in some cases ligands may present distorted geometries when bound to enzymes, which specifically induce structural strains to facilitate the chemical reactions they catalyze) [14]. Beside energy minimization of the coordinates, it is also important to handle the absolute chirality of the stereogenic centers properly, or enumerate all enantiomers whenever it is not specified. This is often the case when a ligand library is generated from SMILES strings available from chemical vendors. Conformationally challenging ligands can also be modeled by pre-generating a large number of conformations to be docked rigidly [15]. Ring conformations, and macrocycles in particular, also need dedicated pre-processing because their conformation cannot be sampled effectively during docking. In this case, it is common to use protocols to generate multiple low-energy conformations prior to docking [16,17] and dock them independently; alternatively, specialized methods to simulate their flexibility during docking are available [18]. Several excellent tools have been developed over the years to generate high-quality coordinates, and many of them are available for free for academics. There can be used as standalone programs, such as OpenBabel [19] (and the Avogadro GUI [20]), RDKit [21], ChemAxon [22] or accessed as web servers, such as Frog3D [23], CACTVS [24]. For an excellent review on freely available conformer generators, see reference [25]. Finally, before accepting a ligand for a docking, it is important to check that the docking software to be used has parameters and proper atom types for each ligand atom. Some programs refuse to run if unsupported types are found, while other could silently assign a default type (with resulting low-quality parameters) and issue only a warning message in the log files. If the unparametrized element is essential to establish interactions with the target structure this can be a major issue. In fact, even in case of a successful docking run, the interaction energy could likely be not estimated correctly. There have been cases reported where wrong input geometries [26] generated very reactive species (accordingly to Sayle's definition) [27].

2.2. Tautomers and Protonation States

The definition of tautomers as "isomers of organic compounds that readily interconvert, usually by the migration of hydrogen from one atom to another" [27] describes exactly what the vast majority of docking programs does not: account for changes in hybridization or hydrogen count on the ligand during docking. Specifically, the movement of hydrogens around a molecule (prototropic tautomerism) depends on several conditions such as pH, solvent and temperature [28,29]. One of the consequences of

prototropic tautomerism is the dramatic modification of the hydrogen bond pattern (Figure 2a), or the change of chirality (Figure 2b). These events can determine the (in)success of a docking if (mis)matching with the target counterpart (Figure 2c).
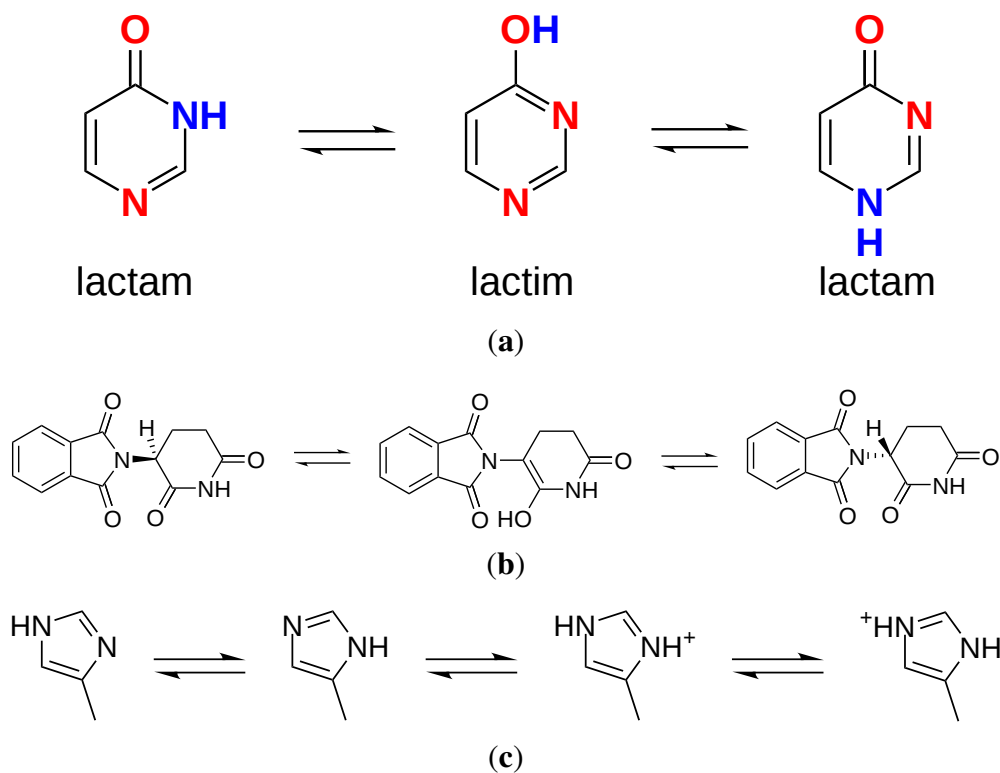

Figure 2. Examples of the effects of tautomerization and protonation states on molecular structures. (**a**) Tautomers of pyrimidin-4-one and hydrogen bond patterns [30] (*red*: acceptor; *blue*: donor); (**b**) Racemization of thalidomide via keto-enol tautomerisation [31]; (**c**) Histidine tautomers and protonation states [27].

The same can be true for protonation states. Depending on the pH of either experimental or physiological conditions being simulated, chemical moieties on the ligand can be protonated or deprotonated, acquiring charged groups that are not present in the neutral gas phase form of the molecule. In general, predicted stable protomers are generated at physiological pH 7.4, but catalytic sites might present micro-conditions which can deviate considerably from this value. This is often due to either the presence of acid residues (like those found in HIV-1 protease [32] or BACE-1 [33] sites) or basic metals like zinc (as found in carbonic anhydrase [34] or HDAC enzymes [35]).

In some well-studied cases, experimental data or accurate calculations demonstrated the key role of tautomers and protonation states in ligand binding (Figure 3). In the case of the anticancer drug methotrexate, for example, neutron diffraction studies [36] identified which of the several stable protomers [37] is the most probable in the complex with DHFR: the two ligand carboxylic groups are deprotonated, while N1 nitrogen must be protonated to establish two hydrogen bonds with the aspartic acid (Figure 3a). If wrong protomers are docked (e.g., N1 is not protonated), the full hydrogen bond interaction network would not be restored, preventing the identification the correct docked pose or resulting in inaccurate score.
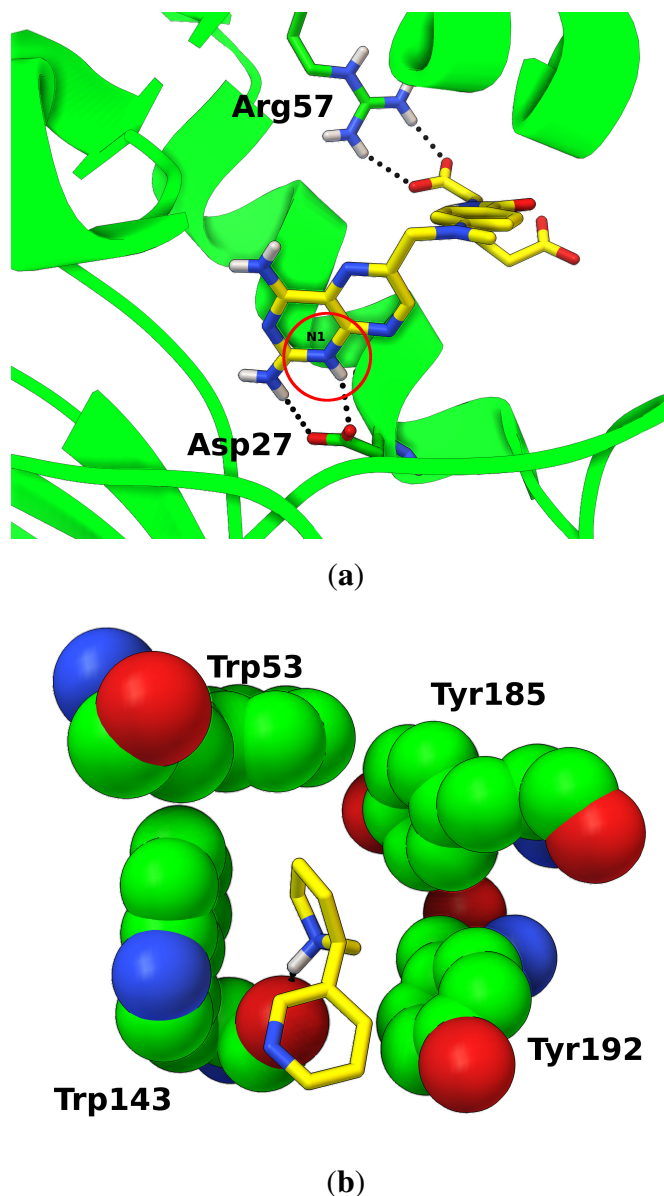
Figure 3. Examples of specific protonation states in ligand binding. (**a**) Methotrexate bound to *E.* coli DHFR; protein structure (*green*) is shown as secondary structure ribbons, and interacting residues as sticks; the ligand is shown as sticks (*yellow*); hydrogen bonds are shown as spheres (*black*); protonated N1 nitrogen is circled in red. (PDB:2inq) [36]; (**b**) Protonated nicotine bound to the acetylcholine binding protein (AChBP) of *L. stagnalis*. The ligand is shown as sticks (*yellow*); residues forming the aromatic cage are shown as spheres (*green*); hydrogen bonds are shown as spheres (*black*) (PDB:1uw6).

Another example is the positively charged nicotine bound to nicotinic acetylcholine receptors (nAChR) [38]. The charged feature of the the pyrrolidine ring nitrogen is an essential component of one of the oldest pharmacophores studied [39]. The receptor structure has a characteristic aromatic cage (Figure 3b) specialized in interacting and stabilizing the cation moieties in endogenous ligands.

Computational methods are available to predict tautomers and protomers [30], but their success is limited by the lack of availability of large experimental datasets on tautomer distributions. Moreover, enzyme binding is known to stabilize rare tautomers in ligands [30] that would be unlikely to be

predicted. To date, ensemble dockings have been proven to be the most effective way to model multiple tautomers and protomers [40].

2.3. Charges

Charges are another aspect of ligand-receptor interaction that require special care. The formal charge of a molecule is a function of the chemical moieties it contains and the environment pH. In the context of a receptor structure, different local conditions that deviate from physiological conditions can induce protomers with different net charges than the ones predicted in solution.

Once formal charges are assigned, partial charges can be calculated. Partial charges are a simplified representation of the distribution of electrons in a molecule across its single atoms, and are the only way to model polarity (*i.e.*, dipoles) in classical force field simulations [41] (with the exception of computationally expensive and intricate polarizable force fields) [42,43].

Partial charges can be calculated using quantum mechanics (QM) [44–46], semi-emipirical [47–49], force field-based [50–52], or fully empirical [53] methods. QM and semi-empirical methods are the most accurate but also the most computationally demanding. Moreover, charges calculated with these methods depend on the molecular conformation, so several protocols have been proposed to overcome this limitation [54–56]. Force field-based or fully empirical methods, on the other hand, are orders of magnitude faster, conformationally independent, and provide acceptable results for most drug-like molecules. They are not suitable for inorganic compounds, metal-coordinating species and unstable/radical moieties, for which they lack appropriate parameters.

In principle, the choice of which method to use should be evaluated on the basis of the target considered, especially for systems where charge distribution is known to play a role [57,58]. However, scoring functions are usually calibrated using a specific charge set, therefore a consistency criterion should prevail (unless specific target properties cannot be properly described). Charges calculated with different methods will present a fairly large variability, and inevitably can have an influence on the results [59]. If different charge sets are applied to ligand and target structures, mismatching charge pairing can occur, likely increasing the false negative hit rate. Moreover, the absolute value of formal charges in large ligands should be assessed prior to docking because in strongly charged molecules, the Coulomb contribution to the scoring function will likely overcome other weak interactions, and increase the rate of false positives [55]. Most docking scoring functions cannot handle such highly charged systems unless specifically calibrated for that purpose. Conversely, it must be noted also that there are scoring functions which ignore entirely the electrostatic terms [60].

2.4. Physicochemical Properties

Intrinsic ligand properties can be used to pre-filter libraries to enrich them and limit the number of compounds to dock. One of the most popular criteria is the "rule of five" (Ro5) derived by Lipinski *et al.* [61] from a retrospective analysis of more than 2000 drugs. A similar filter, namely the 'rule of three', was suggested for fragments with thresholds adapted to their smaller size [62]. Both correlate the number of hydrogen bonds, molecular weight, and octanol-water partition coefficient (logs*P*) to oral bioavailability and drug likeness. While criticism [63,64] from their indiscriminate

enforcement as literal "rules" is well-deserved, their use as guidelines is recommended to provide support in prioritizing hits.

Another useful criterion to analyze a library is the identification of Pan Assay Interference Compounds (PAINS). A series of chemical groups has been found to be responsible for the promiscuity of certain compounds in many high throughput screenings [65]. Compounds containing these groups have been reported to be found in a suspiciously large number of assays, therefore they should be at least flagged as problematic, if not removed. More stringent criteria should be applied when filtering the library to exclude compounds containing reactive groups (e.g., acrylamide, chloroacetamide), which would establish undesired irreversible interactions with the target structure.

The size of a library, and with it the computation time, may also be reduced by using a diversity subset [66] on the assumption that similar ligands bind in a similar fashion. Hits obtained by screening the diversity set can be then used to generate a focused library containing ligands structurally similar to the hits, and this library can be used to sample the local chemical space for improved hits. A limitation of the approach is that the aforementioned assumption is not always valid. In tight cavities, for example, where shape complementarity criteria are more stringent, valuable results can be easily missed. Fragment-based drug design (FBDD) [67,68] is another efficient approach that provides high coverage of chemical space with relatively small libraries (1k–10k compounds) [68], often with relatively high hit rates [68]. Fragments are defined as small molecules with MW of $\leq$300 Da [69]. (compared to drug-like molecules with MW around 500 Da). In virtue of their small size, fragments are likely to require fewer structural adjustments on the target counterpart, therefore their binding can be easier to model. The main limitations of this approach are the issues of hit detection (see Section 5), the strong role of waters in fragment binding (see Section 3.4), and the added complexity of growing hit fragments into drug-like molecules [70].

Information about the target can also be used to tailor a more focused library. For example, if the library is built by strictly adhering to drug-like properties (*i.e.*, MW $\geq$ 450 Da.), it is unlikely to succeed in screenings on targets with very small pockets [71]. Conversely, very large binding sites are able to bind larger ligands, accommodating multiple binding modes [72] or even more than one ligand at the same time [73]. These structures are unlikely to be targeted effectively by libraries containing mostly fragment-like molecules.

3. Target Structure

The second crucial factor necessary for the proper outcome of a virtual screening campaign is the selection of the target structure to be used. Different criteria can be used to guide this selection, especially with structures obtained by crystallography [74,75].

3.1. Structure Quality

For structures obtained via X-ray crystallography, R-factor and resolution can provide an indication on the overall quality (low R-factor and high resolution should be preferred), but they do not provide details about specific regions [75]. More accurate information for single atoms or groups of atoms can be obtained by checking the B-factor (or temperature factor, or Debye-Waller factor) [76], which

provides a measure of their dynamic behavior within the crystal (essentially, the lower the value, the lower the uncertainty on atom positions). Occupancy is another property used to define the quality of atom positions, providing a measure of the consistency of placement across the molecules in the crystal. Side chains that are solvent-exposed or that sample more than one conformation tend to have low occupancy values. In cases when the structure presents disordered regions (*i.e.*, regions where there is not enough electron density information to build the atomic models) the final structure could present either incomplete features (*i.e.*, sequence gaps, missing residues or side chains), or residues that are modeled in 'reasonable' conformations, but have no experimental data to support them [75]. This modeled data is usually reported in the paper describing the structure and should be easily recognizable by its very high B-factor, but this is not always the case. Conversely, some regions of the density maps could be left uninterpreted, with undefined density blobs that can correspond to waters, crystallization buffer molecules, or even unidentified ligands [75]. Cases have been reported in which an unclear combination of human errors and software issues led to highly inaccurate structures [77,78]. When no experimentally determined structures are available, it is possible to use homology modeling techniques [79] to build reasonable models of the desired protein, provided that enough template structures with high sequence and structure homology are available. The degree of success in using these structures for structure-based drug design is variable and depends strongly on the specific type of proteins considered [80–82].

3.2. Structure Clean-Up

No structure should be used directly in a docking calculation without a certain degree of prior inspection and clean-up. Structures should be analyzed to check for the presence of artifacts such as improper bond lengths and angles, and overall conformation assessment of the backbone and side chains should be performed.

The amount of pre-processing required depends on the quality of the structure and the experimental conditions used for resolving it. Structures resolved by crystallography often contain salts and other molecules from the buffer, or additives used to induce the crystal formation. All these molecules should be removed prior to docking. Also, even in structures with acceptable quality, a certain degree of uncertainty is associated with the orientation of asparagine, glutamine and histidine side chains, since electron density is not sufficient to resolve the difference between nitrogen and oxygen atoms [83]. In the absence of structural hints suggesting the preference of one conformation over the other, their crystallographic conformation should not be trusted completely. Flipped conformations can also be assigned if the ligand structure is not considered during the structural refinement (see Figure 4). Structures solved via NMR usually contain an ensemble of models, so a single model should be extracted and processed separately. Structures obtained by neutron diffraction [85] usually have resolution high enough to distinguish hydrogen atoms. However, fully deuterated crystals (*i.e.*, all ^{1}H atoms are replaced by ^{2}H) [86] are often used to increase the signal-to-noise ratio and enhance the visibility of the molecular structure [87]. Many docking programs do not recognize deuterium properly, which has to be replaced by hydrogen when preparing these structures. The protein sequence used to solve the structures could differ from the wild-type or the one used during the assays. The reasons for such differences are various. Mutations can be inserted to study the insurgence of drug resistance [88,89] or inhibit the

autolytic activity in an enzyme [90]. Non-natural amino acids can be inserted to address the phase problem in crystallographic structure determination [91]; recombinant proteins can be expressed using a medium containing selenium to replace the sulfur-containing residues cysteine and methionine with their corresponding selenium derivatives, selenomethionine [92] and selenocysteine [93]. In some cases, modifications can be substantial to the point where the protein sequence is altered to include entire domains that are not present in the wild-type. An example is provided by the chimeric GPCR structures engineered to create soluble (hence crystallizable) versions of the transmembrane proteins [94]. These modifications should be reverted when required (*i.e.*, to study the active form of an enzyme), or accounted for otherwise. Whenever missing atoms or missing residues are encountered, structures should be amended and discontinuities filled or capped. Specialized tools [95,96] may be used to rebuild incomplete residues or even entire missing loops [97]. Over the years, efforts have been made to prevent issues with structural models deposited in the PDB [98] by designing dedicated validation software [99–101]. While these tools are now applied routinely to check new PDB entries, it would be a good practice to run them on input structures prior to docking to check for errors that could be introduced during the preparation steps.
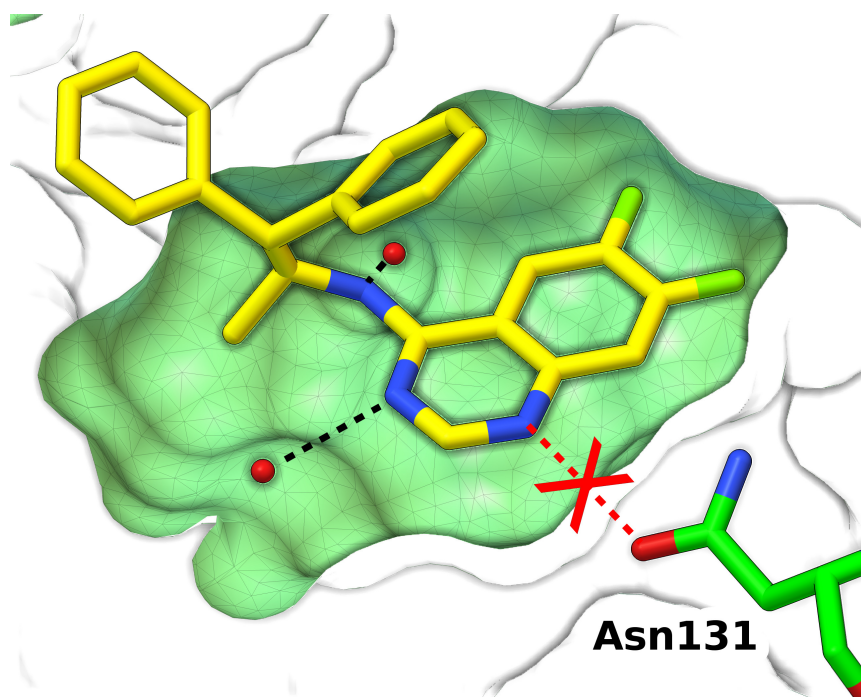

Figure 4. Wrong conformation of Asn131 side chain in the crystal structure of a scytalone dehydratase inhibitor prevents the formation a key hydrogen bond. Protein buried cavity is shown as surface (*green*); ligand (*yellow*) and Asn131 (*green*) are shown as sticks; waters are shown as spheres (*red*). Hydrogen bonds are shown as dashed lines (*black*) (PDB:5std) [84].

3.3. Protonation States

Similar to what discussed for the ligand, protonation states and tautomers play an important role also in the target structure. For histidine side chains, for example, four states are possible (Figure 2c). When more than one key residue can present multiple states, modeling only the most probable ones

for each residue is not sufficient. For example, simulations [102] and experimental evidence [103] suggest that when a ligand is bound to the two catalytic Asp25 in the asymmetric homodimer of HIV-1 protease [32], the two carboxylic groups present an asymmetric ionization. Analogous conditions are found in β-secretase (BACE1) [104], where the protonation state of the catalytic dyad formed by Asp32 and Asp228 is essential to screen for inhibitors [105] Dockings of known inhibitors against eight distinct states of the dyad show that different binders prefer different states [33].

To date, the best way to model state multiplicity in the target is to dock against protomer and tautomer ensembles.

3.4. Coordinating Metal Ions, Co-Factors and Waters

Special care should be taken when dealing with structural features such as catalytic metals, co-factors and structurally conserved waters. Enzymes in which metal ions play a catalytic role (metalloenzymes) are widely distributed in all cells and involved in a large number of metabolic processes [106]. The modeling of ligand-metal coordination is problematic for most scoring functions [107–109], due to the difficulty in describing the partial covalent bond nature of the interaction [110]. For this reason, when docking in systems containing catalytic metals, the best choice would be to use docking programs that have been validated on the specific metal ion to be considered [110,111], or have dedicated terms in the scoring function to describe metal coordination [112,113].

As a general guideline, if a co-factor is bound to the structure, (e.g., NADP, heme, . . .) it should be conserved in the target structure used for docking. If removed, it would leave a favorable cavity unoccupied: docking results will be biased toward it, but it is unlikely that any ligand would be able to actually engage it when competing with the affinity and the concentration of an endogenous binder. Exceptions are of course possible, as when ligands are designed to explicitly to compete with a natural substrate. An example is the drug chloroquine, which binds in the active site of lactate dehydrogenate of *Plasmodium* falciparum displacing the NADH cofactor [114]. Methods to model competitive binding by docking multiple ligands at the same time have been developed, but they have not been extensively tested [115].

For waters, the common practice is to remove them entirely from the structure and dock into a "dry" model [116]. However, this may not be the optimal choice when conserved waters are known to have direct involvement in ligand binding [39,117], which is particularly frequent with fragments [118]. Several docking programs handle waters explicitly during docking with different degrees of approximation [47,116,118–122]. Whenever information on the possible involvement of waters is available, these methods should be preferred, especially when structurally conserved waters (*i.e.*, very low B-factors) have been identified in the cavities considered in the docking. In the case of low resolution structures, where no waters have been resolved, predictive methods can be tested [118,123].

3.5. Structure Conformation

Proteins in solution exist in a variety of conformational states distributed in an complex energy landscape [124]. Some of these conformations represent discrete states which are associated with specific biological functions that have been extensively investigated, like the active/inactive states in protein

kinases [125,126], and GPCRs [127–129]. On the other hand, the energetic components regulating the dynamic transitions between states is complex and still poorly understood [124,130]. Binding ligands can either stabilize existing conformations [125,128] or trigger specific conformational changes (Figure 5) [131]. Moreover, different ligands induce small re-arrangements in the binding site (induced fit) [132]. The scenario can be roughly summarized by saying that the receptor that binds with a predicted ligand is rarely what we expect it to be, so the choice of the structure to use must take in to account the structural variability of the target. In general dockings using *holo* structures provide higher enrichment rates than those performed using *apo* structures [133], likely because *holo* structures present some degree of induced fit due to the presence of a ligand.
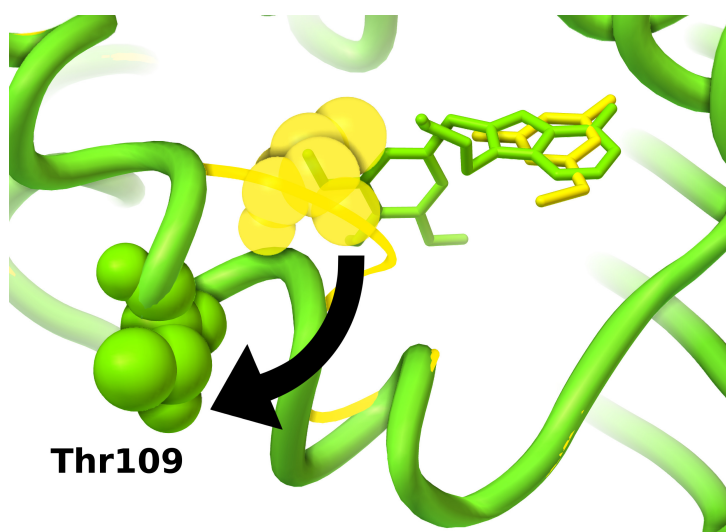

Figure 5. Conformational re-arrangements in HSP90 induced by ATP site inhibitors. Protein secondary structures are shown as ribbons; conformations of key residue Thr109 are shown as spheres; ligands are shown as sticks. Small fragments can bind into a protein conformation similar to the *apo* conformation (*yellow*, PDB:2wi2) [134], in which an unwinded loop covers part of the binding site; larger ligands induce conformational rearrangements folding the loop in a helix conformation and uncovering a larger binding site (*green*, PDB:1uy6) [131].

It is important to take into account limitations and biases of the experimental method used to determine the structure. In X-ray crystallography, for example, structures are commonly solved at very low temperatures (100 K or lower) [135] and in conditions of supersaturation [136]. The formation of the crystal lattice during packing can add an extra bias, perturb backbone and side-chain conformations, and induce improper hinge-like motions [137]. Crystallographic conformations in general provide a biased representation of the flexible behavior of a protein. For example, for targets that display a lot of structural flexibility in solution, such as HIV-1 protease, a single NMR ensemble can present more structural variability than a hundred crystal structures [74]. Depending on the goal, a possible strategy could be to target highly structurally conserved regions [138], or focus on a well-defined specific conformation [125]. When this is not possible, conformational variability must be taken into account in docking experiments. There are different ways in which target flexibility can be modeled, as discussed in detail in recent reviews [139,140]. Essentially, two main classes of motions are modeled using

orthogonal and complementary approaches: ligand induced-fit involving small and local re-arrangements of a few amino acids can be modeled with sufficient accuracy by simulating *flexible side chains* during docking; large conformational changes such as hinge and domain movements, loop rearrangements, and changes involving the protein backbone in general, can be modeled by running *ensemble dockings* on discrete pre-calculated conformations.

Depending on the software, the number of side chains that can be made flexible is limited by the increase in search complexity and exacerbation of scoring function limitations, such as simplified solvation models and unbalanced energy contributions between ligand and flexible receptor. Therefore, only a small set of residues should be made flexible, and the choice can be based on occupancy, B-factor values, or variability across multiple crystal structures (in complex with different ligands, if possible).

For ensemble dockings, multiple receptor conformations can be selected from experimental structures [141] or generated through computational methods such as molecular dynamics simulations [142], or normal mode analysis [143]. Ensemble docking has been reported to improve the overall success rate in virtual screenings [9,141,144–146], although requirements of computation time and power can increase considerably, scaling linearly with the number of target structures considered. Massive virtual screening campaigns, involving large conformational ensembles of flexible proteins, have been made possible with the support from IBM and the computer time donated by volunteers participating to the World Community Grid initiative [147].

3.6. Binding Site Selection

If a known binding site is going to be targeted, all available information should be used to define search boundaries. The search should encompass the regions occupied by known ligands, but not be limited to them, since alternative binding modes could engage residues that have not been characterized yet. However, too large search boundaries may reduce the accuracy, require longer computation times, and increase false positive rates. Known binders such as endogenous molecules should be added to the ligand library to be screened, especially if resolved in complex with the protein. Including them provides a validation of the screening protocol ("*are search parameters thorough enough to dock ligands similar to the known binder?*") and the scoring function ("*can the scoring function recognize known binders as hits?*"). These docking results will also provide a score reference value to be used when evaluating the virtual screening hits.

When no binders are known, it is possible to run dockings considering the entire protein volume in a blind docking [148] experiment. While this method has shown a considerable degree of success [149,150], higher accuracy and efficiency can be achieved by running multiple focused dockings on regions predicted to have high affinity interactions [151]. In the identification of a site, any structural information available should be exploited. Crystallization additives and buffer molecules can provide insights [75] on potential interaction loci and preferred chemical moieties. Even structurally consistent waters (*i.e.*, with low B-factor and high occupancy) can hint at regions with favorable hydrogen bond interactions. A variety of software is also available to identify energetically favorable sites that are likely to accommodate ligands [152].

4. Scoring Function and Search Method

The scoring function is the key for a successful screening. A good scoring function should have an acceptable balance between speed and accuracy. Since the goal of a virtual screening is to identify new potential binders while reducing the number of unsuccessful assays, the ideal scoring function should also have a low rate of false positive hits. Scoring functions differ by their design and can be classified in three main categories: [153] empirical, knowledge-based and forcefield-based. Their quality is bound to the quality of the data sets used to build and calibrate them [75,154]. A great variability in success rate has been reported in comparisons between a number of docking programs [4,153–155] and the common agreement is that no software is the absolute best across all the targets. Extreme cases like strongly charged sites [156] or these with hydrophobic/hydrophilic dominant components [157] are known to be more problematic for general purpose scoring functions. Indeed, it has been suggested that generalized scoring functions should be replaced by target-specific models, which would likely yield better results [158]. Therefore, whenever possible, the scoring function should be validated for the particular target to be considered in the screening, using known ligands when available. Two main issues are common to most scoring functions. One is the additive nature of the score when estimating enthalpy. The different interaction terms composing a scoring function are calculated independently and the final score sums up these components. For this reason, larger molecules are likely to have higher scores because of their ability to establish more interactions [159] (see Section 5). The other is the poor description of the entropic component associated with losses in vibrational and conformational degrees of freedom [160].

The scoring function provides the description of the energetic landscape that the search algorithm will sample. The complexity of this landscape depends on the energetic terms composing the scoring function, and the degrees of freedom associated with any moving atoms in ligand and target structures. If the complexity gets beyond the capabilities of the search method, *i.e.*, very large and flexible ligands, too many flexible side chains, docking volume too large, *etc.*, the screening is bound to fail, or it will require a much larger computational effort to generate reliable poses. In order to maximize the chances of success, choices resulting in increased search complexity should be used with parsimony, while superfluous complexity should be avoided independently of the efficiency of the search algorithm. Conversely, any experimental data that could be used to simplify the search should be used. Finally, most of the search algorithms are stochastic [161], therefore, prior to running a full virtual screen, it is a good practice to test the search protocol with a representative subset of ligands from the library (even better: including ligands for which experimental coordinates are available). This will ensure that the protocol is robust enough, and that the search parameters are sufficient to guarantee a certain degree of convergence and consistency in the results.

5. Results and Assay

When docking calculations are completed, virtual screening results need to be processed and analyzed to identify a set of hits to be assayed. A series of criteria can be used to filter and prioritize results to be evaluated by either visual inspection or rescoring schemes.

5.1. Hit Selection

The goal of a virtual screening is to provide new chemical scaffolds as bioactive molecules for a given target [162], and a hit is defined as "*a primary active compound, with non-promiscuous binding behaviour, exceeding a certain threshold value in a given assay*" [163]. Screening hits are meant to provide as many and diverse starting points as possible for the identification of improved molecules (*i.e.*, leads), and the event of identifying of a high-affinity binder is rather unlikely, to the point that this expectation has been deemed as one of the pitfalls of virtual screening [162]. In fact, the accuracy of current scoring functions is known to be higher in reproducing the correct pose of a given ligand than in scoring multiple ligands [164]. Hence, some degree of chemical expertise can definitely help in the hit identification to discriminate between interesting compounds, reliable interactions, poses and predicted affinities. As a general guideline, achiral compounds are preferable to compounds with stereogenic centers: chiral compounds are in fact more complex to synthesize, while chemical vendors often provide only racemic mixtures (or charge extra for enantiomerically pure compounds). The first criterion for selecting hits is the rank (or score) provided by the docking, but it is rarely the only one. Ligand efficiency [165] calculated on the docking score can also be a useful metric to identify promising compounds [166], especially fragment-like molecules [167]. Ligand efficiency can also help mitigate the bias of scoring functions toward large ligands (see Section 4), because effective binders usually engage most of their atoms in specific high-affinity interactions. A common approach to prune results [168] is test different combinations of variable parameters (e.g., energy, ligand efficiency) and interaction filters to reduce the number of potential hits to a size that is suitable for visual inspection.

Knowledge-driven criteria should be considered whenever experimental data is available. Patterns such as interactions with key residues and chemical feature overlap with known binders [169] are useful to bias hit selection, while interactions with exposed backbone can be favored to avoid resistance from mutations [170].

Structural diversity should also be emphasized in order to sample a large portion of the chemical space. During the selection phase, experience and intuition (often in the form of rather arbitrary choices) [162] tend to compensate for a lack of absolute criteria for identifying hits, scoring function weaknesses, and model approximation.

Whereas relying on a single scoring function can bias the results, combining rankings generated with multiple docking programs (consensus scoring) has been used successfully to enrich results [171,172]. Also, once the number of potential hits has been reduced by orders of magnitude, it is possible to apply more computationally expensive scoring methods that provide better estimates of the free energy of binding by combining molecular mechanics and quantum mechanics [173], taking into account entropy terms and solvation effects [174–178] or using empirical methods [168].

5.2. Hit Validation

The caveat of any virtual screening shuold be: " *Your computation is only as good as your experimental follow up*" (E.R. Zartler) [179]. Clearly, it is important that hits selected in the virtual screening are actually accessible for testing, either through commercial vendors or via feasible synthetic pathways. Ligand physico-chemical properties can affect the assay outcome. Among them,

unpredictable poor water solubility is a common issue, while fluorescence can interfere with several common assays [180,181]. The choice of the assay itself can also impact the outcome of a screening, especially as a consequence of its minimum sensitivity threshold for low affinity binders. Therefore the choice of the assay should not overlook the size and expected affinity of compounds to be tested. This is particularly relevant for protein-protein interface inhibitors and fragments, which are expected to show relatively low affinities [182]. Interestingly, different experimental methodologies can identify different hits even when testing the same fragment library [183]. In general, assays that have been successfully applied and validated on the specific target should be preferred.

6. Conclusions

Virtual screening involves a large number of steps, and each step comes with its own pitfalls. While many issues are intrinsic to the simplified models used in the modeling, others are the result of improper pre-processing of the input or misinterpretation of the results. No single best recipe is available, and success rate may vary considerably depending on the target and the expertise of the user. Software is not (yet) meant to replace chemical intuition or deep knowledge of the biological target, which are essential to the identification of hits. Expertise is achieved over time, and as Niels Bohr said: "*An expert is a man who has made all the mistakes which can be made in a very narrow field*". Luckily, thanks to software improvements and the massive computational power available today, it is possible to make in a few days the number of mistakes that not too long ago would have taken years.

Virtual screening is one of the many drug design tools available to scientists, and when properly used can be an invaluable support to research. Knowing how a method can fail is as useful as knowing how it works, so it is fundamental to maintain a critical attitude when dealing with results. A wise combination of rigorous approaches and clever workarounds can reduce chances of failure. Carefully pondering docking results and interpreting them as informed suggestions can significantly boost chances of success, and the rest is computation overhead.

Acknowledgments

Author Contributions

S.F. wrote the manuscript.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Walters, W.P.; Namchuk, M. Designing screens: How to make your hits a hit. *Nat. Rev. Drug Discov.* **2003**, *2*, 259–266.

2. Perola, E.; Xu, K.; Kollmeyer, T.M.; Kaufmann, S.H.; Prendergast, F.G.; Pang, Y.P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.* **2000**, *43*, 401–408.

3. Grüneberg, S.; Stubbs, M.T.; Klebe, G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: Strategy and experimental confirmation. *J. Med. Chem.* **2002**, *45*, 3588–3602.

4. 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*, 935–949.

5. Shoichet, B.K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.

6. Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* **2005**, *48*, 1088–1097.

7. Kolb, P.; Ferreira, R.S.; Irwin, J.J.; Shoichet, B.K. Docking and chemoinformatic screens for new ligands and targets. *Curr. Opin. Biotechnol.* **2009**, *20*, 429–436.

8. 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.

9. Fischer, M.; Coleman, R.G.; Fraser, J.S.; Shoichet, B.K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* **2014**, *6*, 575–583.

10. Al Olaby, R.R.; Cocquerel, L.; Zemla, A.; Saas, L.; Dubuisson, J.; Vielmetter, J.; Marcotrigiano, J.; Khan, A.G.; Catalan, F.V.; Perryman, A.L.; *et al.* Identification of a novel drug lead that inhibits HCV infection and cell-to-cell transmission by targeting the HCV E2 glycoprotein. *PLoS ONE* **2014**, *9*, e111333.

11. Perryman, A.L.; Yu, W.; Wang, X.; Ekins, S.; Forli, S.; Li, S.G.; Freundlich, J.S.; Tonge, P.J.; Olson, A.J. A Virtual Screen Discovers Novel, Fragment-Sized Inhibitors of Mycobacterium tuberculosis InhA. *J. Chem. Inf. Model.* **2015**, *55*, 645–659.

12. Babbage, C. *Passages from the Life of a Philosopher*; Cambridge University Press: Cambridge, UK, 2011.

13. Irwin, J.J.; Shoichet, B.K. ZINC-a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

14. Uitdehaag, J.C.; Mosi, R.; Kalk, K.H.; van der Veen, B.A.; Dijkhuizen, L.; Withers, S.G.; Dijkstra, B.W. X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. *Nat. Struct. Mol. Biol.* **1999**, *6*, 432–436.

15. Malloci, G.; Vargiu, A.V.; Serra, G.; Bosin, A.; Ruggerone, P.; Ceccarelli, M. A Database of Force-Field Parameters, Dynamics, and Properties of Antimicrobial Compounds. *Molecules* **2015**, *20*, 13997–14021.

16. Poulsen, A.; William, A.; Blanchard, S.; Nagaraj, H.; Williams, M.; Wang, H.; Lee, A.; Sun, E.; Teo, E.L.; Tan, E.; *et al.* Structure-based design of nitrogen-linked macrocyclic kinase inhibitors leading to the clinical candidate SB1317/TG02, a potent inhibitor of cyclin dependant kinases (CDKs), Janus kinase 2 (JAK2), and Fms-like tyrosine kinase-3 (FLT3). *J. Mol. Model.* **2013**, *19*, 119–130.

17. Watts, K.S.; Dalal, P.; Tebben, A.J.; Cheney, D.L.; Shelley, J.C. Macrocycle conformational sampling with MacroModel. *J. Chem. Inf. Model.* **2014**, *54*, 2680–2696.

18. Forli, S.; Botta, M. Lennard-Jones potential and dummy atom settings to overcome the AutoDock limitation in treating flexible ring systems. *J. Chem. Inf. Model.* **2007**, *47*, 1481–1492.

19. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33, doi:10.1186/1758-2946-3-33.

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, doi:10.1186/1758-2946-4-17.

21. Landrum, G. RDKit: Open-Source Cheminformatics, 2013. Available online: http://www.rdkit.org (accessed on 14 July 2015).

22. ChemAxon. Standardizer, JChem 15.7.13.0, 2015. Available online: http://www.chemaxon.com (accessed on 14 July 2015).

23. Miteva, M.A.; Guyon, F.; Tufféry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*, W622–W627.

24. CACTVS-Server. Online SMILES Translator and Structure File Generator, 2010. Available online: http://cactus.nci.nih.gov/translate/ (accessed on 20 July 2015).

25. Ebejer, J.P.; Morris, G.M.; Deane, C.M. Freely available conformer generation methods: How good are they? *J. Chem. Inf. Model.* **2012**, *52*, 1146–1158.

26. Sivakumar, P.M.; Vignesh, N.; Kumar, G.R.; Doble, M. Computational approaches to enhance activity of taxanes as antimitotic agent. *Med. Chem. Res.* **2012**, *21*, 2557–2570.

27. Sayle, R.A. So you think you understand tautomerism? *J. Comput.-Aided Mol. Des.* **2010**, *24*, 485–496.

28. Sitzmann, M.; Ihlenfeldt, W.D.; Nicklaus, M.C. Tautomerism in large databases. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 521–551.

29. Smith, M.B.; March, J. *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*; John Wiley & Sons: New York, NY, USA, 2007.

30. Martin, Y.C. Let's not forget tautomers. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 693–704.

31. Tian, C.; Xiu, P.; Meng, Y.; Zhao, W.; Wang, Z.; Zhou, R. Enantiomerization mechanism of thalidomide and the role of water and hydroxide ions. *Chem.-A Eur. J.* **2012**, *18*, 14305–14313.

32. Brik, A.; Wong, C.H. HIV-1 protease: Mechanism and drug discovery. *Org. Biomol. Chem.* **2003**, *1*, 5–14.

33. Barman, A.; Prabhakar, R. Protonation states of the catalytic dyad of β-secretase (BACE1) in the presence of chemically diverse inhibitors: A molecular docking study. *J. Chem. Inf. Model.* **2012**, *52*, 1275–1287.

34. Stams, T.; Chen, Y.; Christianson, D.W.; Boriack-Sjodin, P.; Hurt, J.D.; Laipis, P.; Silverman, D.N.; Liao, J.; May, J.A.; Dean, T. Structures of murine carbonic anhydrase IV and human carbonic anhydrase II complexed with brinzolamide: Molecular basis of isozyme-drug discrimination. *Protein Sci.* **1998**, *7*, 556–563.

35. Vannini, A.; Volpari, C.; Filocamo, G.; Casavola, E.C.; Brunetti, M.; Renzoni, D.; Chakravarty, P.; Paolini, C.; de Francesco, R.; Gallinari, P.; *et al*. Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15064–15069.

36. Bennett, B.; Langan, P.; Coates, L.; Mustyakimov, M.; Schoenborn, B.; Howell, E.E.; Dealwis, C. Neutron diffraction studies of *Escherichia coli* dihydrofolate reductase complexed with methotrexate. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18493–18498.

37. Cocco, L.; Roth, B.; Temple, C.; Montgomery, J.A.; London, R.E.; Blakley, R.L. Protonated state of methotrexate, trimethoprim, and pyrimethamine bound to dihydrofolate reductase. *Arch. Biochem. Biophys.* **1983**, *226*, 567–577.

38. Hurst, R.; Rollema, H.; Bertrand, D. Nicotinic acetylcholine receptors: From basic science to therapeutics. *Pharmacol. Ther.* **2013**, *137*, 22–54.

39. Blum, A.P.; Lester, H.A.; Dougherty, D.A. Nicotinic pharmacophore: The pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13206–13211.

40. Park, M.S.; Gao, C.; Stern, H.A. Estimating binding affinities by docking/scoring methods using variable protonation states. *Proteins Struct. Funct. Bioinform.* **2011**, *79*, 304–314.

41. Heinz, H.; Suter, U.W. Atomic charges for classical simulations of polar systems. *J. Phys. Chem. B* **2004**, *108*, 18341–18352.

42. Ponder, J.W.; Wu, C.; Ren, P.; Pande, V.S.; Chodera, J.D.; Schnieders, M.J.; Haque, I.; Mobley, D.L.; Lambrecht, D.S.; DiStasio, R.A., Jr.; *et al*. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

43. Baker, C.M. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2015**, *5*, 241–254.

44. Slater, J.C. A simplification of the Hartree-Fock method. *Phys. Rev.* **1951**, *81*, 385.

45. Mulliken, R.S. Electronic population analysis on LCAO–MO molecular wave functions. I. *J. Chem. Phys.* **1955**, *23*, 1833–1840.

46. Jones, T.A.; Zou, J.Y.; Cowan, S.T.; Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1991**, *47*, 110–119.

47. Dewar, M.J.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

48. Stewart, J.J. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.

49. Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. Universal quantum mechanical model for solvation free energies based on gas-phase geometries. *J. Phys. Chem. B* **1998**, *102*, 3257–3271.

50. Halgren, T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

51. Jorgensen, W.L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

52. Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187217.

53. Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.

54. Basma, M.; Sundara, S.; Çalgan, D.; Vernali, T.; Woods, R.J. Solvated ensemble averaging in the calculation of partial atomic charges. *J. Comput. Chem.* **2001**, *22*, 1125–1137.

55. Marenich, A.V.; Olson, R.M.; Kelly, C.P.; Cramer, C.J.; Truhlar, D.G. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033.

56. Woods, R.; Chappelle, R. Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. *J. Mol. Struct. THEOCHEM* **2000**, *527*, 149–156.

57. Pierpont, C.G. Studies on charge distribution and valence tautomerism in transition metal complexes of catecholate and semiquinonate ligands. *Coord. Chem. Rev.* **2001**, *216*, 99–125.

58. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

59. Mittal, R.R.; Harris, L.; McKinnon, R.A.; Sorich, M.J. Partial charge calculation method affects CoMFA QSAR prediction accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 704–709.

60. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.

61. 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. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.

62. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A "rule of three" for fragment-based lead discovery? *Drug Discov. Today* **2003**, *8*, 876–877.

63. Abad-Zapatero, C. A Sorcerer's apprentice and the rule of five: From rule-of-thumb to commandment and beyond. *Drug Discov. Today* **2007**, *12*, 995–997.

64. Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011**, *54*, 7784–7796.

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

66. Su, A.I.; Lorber, D.M.; Weston, G.S.; Baase, W.A.; Matthews, B.W.; Shoichet, B.K. Docking molecules by families to increase the diversity of hits in database screens: Computational strategy and experimental evaluation. *Proteins Struct. Funct. Bioinform.* **2001**, *42*, 279–293.

67. Huang, D.; Caflisch, A. Library screening by fragment-based docking. *J. Mol. Recognit.* **2010**, *23*, 183–193.

68. Chen, Y.; Shoichet, B.K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358–364.

69. Feyfant, E.; Cross, J.B.; Paris, K.; Tsao, D.H. Fragment-based drug design. In *Chemical Library Design*; Humana Press: Totowa, NJ, USA, 2011; pp. 241–252.

70. Siegal, G.; Eiso, A.; Schultz, J. Integration of fragment screening and library design. *Drug Discov. Today* **2007**, *12*, 1032–1039.

71. Graves, A.P.; Shivakumar, D.M.; Boyce, S.E.; Jacobson, M.P.; Case, D.A.; Shoichet, B.K. Rescoring docking hit lists for model cavity sites: Predictions and experimental testing. *J. Mol. Biol.* **2008**, *377*, 914–934.

72. Bisignano, P.; Doerr, S.; Harvey, M.J.; Favia, A.D.; Cavalli, A.; de Fabritiis, G. Kinetic characterization of fragment binding in AmpC β-lactamase by high-throughput molecular simulations. *J. Chem. Inf. Model.* **2014**, *54*, 362–366.

73. Schumacher, M.A.; Miller, M.C.; Brennan, R.G. Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein. *Embo J.* **2004**, *23*, 2923–2930.

74. Davis, A.M.; St-Gallay, S.A.; Kleywegt, G.J. Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discov. Today* **2008**, *13*, 831–841.

75. Cooper, D.R.; Porebski, P.J.; Chruszcz, M.; Minor, W. X-ray crystallography: Assessment and validation of protein-small molecule complexes for drug discovery. *Expert Opin. Drug Discov.* **2011**, *6*, 771–782.

76. Debye, P. Interferenz von Röntgenstrahlen und Wärmebewegung. *Ann. Phys.* **1913**, *348*, 49–92.

77. Miller, G. Scientific publishing. A scientist's nightmare: Software problem leads to five retractions. *Science* **2006**, *314*, 1856–1857.

78. Matthews, B.W. Five retracted structure reports: Inverted or incorrect? *Protein Sci.* **2007**, *16*, 1013–1016.

79. Hillisch, A.; Pineda, L.F.; Hilgenfeld, R. Utility of homology models in the drug discovery process. *Drug Discov. Today* **2004**, *9*, 659–669.

80. Evers, A.; Klebe, G. Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J. Med. Chem.* **2004**, *47*, 5381–5392.

81. Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of g-protein coupled receptors suitable targets? *Proteins Struct. Funct. Bioinform.* **2003**, *50*, 5–25.

82. Kufareva, I.; Rueda, M.; Katritch, V.; GPCR Dock 2010 Participants; Stevens, R.C.; Abagyan, R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **2011**, *19*, 1108–1126.

83. Carter, C.W.; Sweet, R.M. *Macromolecular Crystallography*; Gulf Professional Publishing: Oxford, UK, 2003; Volume 374.

84. Wawrzak, Z.; Sandalova, T.; Steffens, J.J.; Basarab, G.S.; Lundqvist, T.; Lindqvist, Y.; Jordan, D.B. High-resolution structures of scytalone dehydratase-inhibitor complexes crystallized at physiological pH. *Proteins Struct. Funct. Bioinform.* **1999**, *35*, 425–439.

85. Piccoli, P.M.; Koetzle, T.F.; Schultz, A.J. Single crystal neutron diffraction for the inorganic chemist—A practical guide. *Comments Inorg. Chem* **2007**, *28*, 3–38.

86. Blakeley, M.P.; Langan, P.; Niimura, N.; Podjarny, A. Neutron crystallography: Opportunities, challenges, and limitations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 593–600.

87. Meilleur, F.; Weiss, K.L.; Myles, D.A. Deuterium labeling for neutron structure-function-dynamics analysis. In *Micro and Nano Technologies in Bioanalysis*; Humana Press: Totowa, NJ, USA, 2009; pp. 281–292.

88. Kear, J.L.; Blackburn, M.E.; Veloro, A.M.; Dunn, B.M.; Fanucci, G.E. Subtype polymorphisms among HIV-1 protease variants confer altered flap conformations and flexibility. *J. Am. Chem. Soc.* **2009**, *131*, 14650–14651.

89. Heaslet, H.; Rosenfeld, R.; Giffin, M.; Lin, Y.C.; Tam, K.; Torbett, B.E.; Elder, J.H.; McRee, D.E.; Stout, C.D. Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *63*, 866–875.

90. Tiefenbrunn, T.; Forli, S.; Baksh, M.M.; Chang, M.W.; Happer, M.; Lin, Y.C.; Perryman, A.L.; Rhee, J.K.; Torbett, B.E.; Olson, A.J.; *et al*. Small molecule regulation of protein conformation by binding in the flap of HIV protease. *ACS Chem. Biol.* **2013**, *8*, 1223–1231.

91. Strub, M.P.; Hoh, F.; Sanchez, J.F.; Strub, J.M.; Böck, A.; Aumelas, A.; Dumas, C. Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* **2003**, *11*, 1359–1367.

92. Hendrickson, W.A. Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron Radiat.* **1999**, *6*, 845–851.

93. Johansson, L.; Gafvelin, G.; Arnér, E.S. Selenocysteine in proteins—Properties and biotechnological use. *Biochim. Biophys. Acta (Bba)-Gen. Subj.* **2005**, *1726*, 1–13.

94. Jaakola, V.P.; Griffith, M.T.; Hanson, M.A.; Cherezov, V.; Chien, E.Y.; Lane, J.R.; Ijzerman, A.P.; Stevens, R.C. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, *322*, 1211–1217.

95. Guex, N.; Peitsch, M.C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714–2723.

96. Šali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

97. Fiser, A.; Do, R.K.G.; Šali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773.

98. Read, R.J.; Adams, P.D.; Arendall, W.B.; Brunger, A.T.; Emsley, P.; Joosten, R.P.; Kleywegt, G.J.; Krissinel, E.B.; Lütteke, T.; Otwinowski, Z.; *et al*. A new generation of crystallographic validation tools for the protein data bank. *Structure* **2011**, *19*, 1395–1412.

99. Chen, V.B.; Arendall, W.B.; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *66*, 12–21.

100. Word, J.M.; Lovell, S.C.; Richardson, J.S.; Richardson, D.C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.

101. Vriend, G. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **1990**, *8*, 52–56.

102. Harte, W.E., Jr.; Beveridge, D.L. Prediction of the protonation state of the active site aspartyl residues in HIV-1 protease-inhibitor complexes via molecular dynamics simulation. *J. Am. Chem. Soc.* **1993**, *115*, 3883–3886.

103. Wang, Y.X.; Freedberg, D.I.; Yamazaki, T.; Wingfield, P.T.; Stahl, S.J.; Kaufman, J.D.; Kiso, Y.; Torchia, D.A. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry* **1996**, *35*, 9945–9950.

104. Luo, Y.; Bolon, B.; Kahn, S.; Bennett, B.D.; Babu-Khan, S.; Denis, P.; Fan, W.; Kha, H.; Zhang, J.; Gong, Y.; *et al*. Mice deficient in BACE1, the Alzheimer's β-secretase, have normal phenotype and abolished β-amyloid generation. *Nat. Neurosci.* **2001**, *4*, 231–232.

105. Polgár, T.; Keserü, G.M. Virtual screening for β-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.* **2005**, *48*, 3749–3755.

106. Vallee, B.L.; Williams, R. Metalloenzymes: The entatic nature of their active sites. *Proc. Natl. Acad. Sci. USA* **1968**, *59*, 498–505.

107. Englebienne, P.; Fiaux, H.; Kuntz, D.A.; Corbeil, C.R.; Gerber-Lemaire, S.; Rose, D.R.; Moitessier, N. Evaluation of docking programs for predicting binding of Golgi α-mannosidase II inhibitors: A comparison with crystallography. *Proteins Struct. Funct. Bioinform.* **2007**, *69*, 160–176.

108. Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7–S26.

109. Hu, X.; Balaz, S.; Shelver, W.H. A practical approach to docking of zinc metalloproteinase inhibitors. *J. Mol. Graph. Model.* **2004**, *22*, 293–307.

110. Irwin, J.J.; Raushel, F.M.; Shoichet, B.K. Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry* **2005**, *44*, 12316–12328.

111. Santos-Martins, D.; Forli, S.; Ramos, M.J.; Olson, A.J. AutoDock4Zn: An Improved Autodock Force Field for Small-Molecule Docking to Zinc Metalloproteins. *J. Chem. Inf. Model.* **2014**, *54*, 2371–2379.

112. Seebeck, B.; Reulecke, I.; Kämper, A.; Rarey, M. Modeling of metal interaction geometries for protein–ligand docking. *Proteins Struct. Funct. Bioinform.* **2008**, *71*, 1237–1254.

113. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

114. Read, J.A.; Wilkinson, K.W.; Tranter, R.; Sessions, R.B.; Brady, R.L. Chloroquine Binds in the Cofactor Binding Site ofPlasmodium falciparum Lactate Dehydrogenase. *J. Biol. Chem.* **1999**, *274*, 10213–10218.

115. Li, H.; Li, C. Multiple ligand simultaneous docking: Orchestrated dancing of ligands in binding sites of protein. *J. Comput. Chem.* **2010**, *31*, 2014–2022.

116. Villacanas, O.; Madurga, S.; Giralt, E.; Belda, I. Explicit treatment of water molecules in protein-ligand docking. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 145–154.

117. Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Liu, B.; Topol, I.A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J.W. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* **1995**, *3*, 581–590.

118. Forli, S.; Olson, A.J. A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *J. Med. Chem.* **2012**, *55*, 623–638.

119. Huang, N.; Shoichet, B.K. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51*, 4862–4865.

120. Verdonk, M.L.; Chessari, G.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Nissink, J.W.M.; Taylor, R.D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.

121. De Beer, S.; Vermeulen, N.P.; Oostenbrink, C. The role of water molecules in computational drug design. *Curr. Top. Med. Chem.* **2010**, *10*, 55–66.

122. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A.B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R.E.; Morley, S.D. rDock: A fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.* **2014**, *10*, e1003571.

123. Guimaraes, C.R.; Mathiowetz, A.M. Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations. *J. Chem. Inf. Model.* **2010**, *50*, 547–559.

124. Fenimore, P.W.; Frauenfelder, H.; McMahon, B.; Young, R. Bulk-solvent and hydration-shell fluctuations, similar to α-and β-fluctuations in glasses, control protein motions and functions. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14408–14413.

125. Liu, Y.; Gray, N.S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358–364.

126. Levinson, N.M.; Kuchment, O.; Shen, K.; Young, M.A.; Koldobskiy, M.; Karplus, M.; Cole, P.A.; Kuriyan, J. A Src-like inactive conformation in the abl tyrosine kinase domain. *PLoS Biol.* **2006**, *4*, 753, doi:10.1371/journal.pbio.0040144.

127. Dror, R.O.; Arlow, D.H.; Borhani, D.W.; Jensen, M.Ø.; Piana, S.; Shaw, D.E. Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4689–4694.

128. Standfuss, J.; Edwards, P.C.; D'Antona, A.; Fransen, M.; Xie, G.; Oprian, D.D.; Schertler, G.F. The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **2011**, *471*, 656–660.

129. Gouldson, P.R.; Kidley, N.J.; Bywater, R.P.; Psaroudakis, G.; Brooks, H.D.; Diaz, C.; Shire, D.; Reynolds, C.A. Toward the active conformations of rhodopsin and the β2-adrenergic receptor. *Proteins Struct. Funct. Bioinform.* **2004**, *56*, 67–84.

130. Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P.W.; Jansson, H.; McMahon, B.H.; Stroe, I.R.; Swenson, J.; Young, R.D. A unified model of protein dynamics. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5129–5134.

131. Wright, L.; Barril, X.; Dymock, B.; Sheridan, L.; Surgenor, A.; Beswick, M.; Drysdale, M.; Collier, A.; Massey, A.; Davies, N.; *et al.* Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. *Chem. Biol.* **2004**, *11*, 775–785.

132. Koshland, D.E. The key-lock theory and the induced fit theory. *Angew. Chem. Int. Ed. Engl.* **1995**, *33*, 2375–2378.

133. McGovern, S.L.; Shoichet, B.K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895–2907.

134. Brough, P.A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N.G.; Davis, B.; Drysdale, M.J.; Dymock, B.; Eccles, S.A.; Garcia-Echeverria, C.; *et al.* Combining hit identification strategies: Fragment-based and in silico approaches to orally active 2-aminothieno [2,3-*d*] pyrimidine inhibitors of the Hsp90 molecular chaperone. *J. Med. Chem.* **2009**, *52*, 4794–4809.

135. Weik, M.; Colletier, J.P. Temperature-dependent macromolecular X-ray crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 437–446.

136. McPherson, A. Introduction to protein crystallization. *Methods* **2004**, *34*, 254–265.

137. Eyal, E.; Gerzon, S.; Potapov, V.; Edelman, M.; Sobolev, V. The limit of accuracy of protein modeling: Influence of crystal packing on protein structure. *J. Mol. Biol.* **2005**, *351*, 431–442.

138. Kornev, A.P.; Haste, N.M.; Taylor, S.S.; Ten Eyck, L.F. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17783–17788.

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

140. Yuriev, E.; Holien, J.; Ramsland, P.A. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *J. Mol. Recognit.* **2015**, *28*, 581–604.

141. Craig, I.R.; Essex, J.W.; Spiegel, K. Ensemble docking into multiple crystallographically derived protein structures: An evaluation based on the statistical analysis of enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 511–524.

142. Lin, J.H.; Perryman, A.L.; Schames, J.R.; McCammon, J.A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633.

143. 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.

144. Rao, S.; Sanschagrin, P.C.; Greenwood, J.R.; Repasky, M.P.; Sherman, W.; Farid, R. Improving database enrichment through ensemble docking. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 621–627.

145. Henriksen, S.T.; Liu, J.; Estiu, G.; Oltvai, Z.; Wiest, O. Identification of novel bacterial histidine biosynthesis inhibitors using docking, ensemble rescoring, and whole-cell assays. *Bioorg. Med. Chem.* **2010**, *18*, 5148–5156.

146. Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.

147. IBM. World Community Grid, 2007. Available online: http://www.worldcommunitygrid.org/about_us/viewAboutUs.do (accessed on 20 July 2015).

148. Hetényi, C.; van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **2002**, *11*, 1729–1737.

149. 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.

150. Davis, I.W.; Raha, K.; Head, M.S.; Baker, D. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Sci.* **2009**, *18*, 1998–2002.

151. Ghersi, D.; Sanchez, R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites. *Proteins: Struct. Funct. Bioinform.* **2009**, *74*, 417–424.

152. Ghersi, D.; Sanchez, R. Beyond structural genomics: Computational approaches for the identification of ligand binding sites in protein structures. *J. Struct. Funct. Genomics* **2011**, *12*, 109–117.

153. Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.

154. Warren, G.L.; Andrews, C.W.; Capelli, A.M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Senger, S.; *et al.* A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.

155. Huang, S.Y.; Grinter, S.Z.; Zou, X. Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899–12908.

156. Moitessier, N.; Westhof, E.; Hanessian, S. Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **2006**, *49*, 1023–1033.

157. Perola, E.; Walters, W.P.; Charifson, P.S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins Struct. Funct. Bioinform.* **2004**, *56*, 235–249.

158. Ross, G.A.; Morris, G.M.; Biggin, P.C. One size does not fit all: The limits of structure-based models in drug discovery. *J. Chem. Theory Comput.* **2013**, *9*, 4266–4274.

159. Schulz-Gasch, T.; Stahl, M. Scoring functions for protein–ligand interactions: A critical perspective. *Drug Discov. Today Technol.* **2004**, *1*, 231–239.

160. Chia-en, A.C.; Chen, W.; Gilson, M.K. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1534–1539.

161. Merlitz, H.; Wenzel, W. Comparison of stochastic optimization methods for receptor–ligand docking. *Chem. Phys. Lett.* **2002**, *362*, 271–277.

162. Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J.L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D.K. Recognizing pitfalls in virtual screening: A critical review. *J. Chem. Inf. Model.* **2012**, *52*, 867–881.

163. Bleicher, K.H.; Böhm, H.J.; Müller, K.; Alanine, A.I. Hit and lead generation: Beyond high-throughput screening. *Nat. Rev. Drug Discov.* **2003**, *2*, 369–378.

164. Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.

165. Hopkins, A.L.; Groom, C.R.; Alex, A. Ligand efficiency: A useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430–431.

166. Cosconati, S.; Forli, S.; Perryman, A.L.; Harris, R.; Goodsell, D.S.; Olson, A.J. Virtual screening with AutoDock: Theory and practice. *Expert Opin. Drug Discov.* **2010**, *5*, 597–607.

167. Carr, R.A.; Congreve, M.; Murray, C.W.; Rees, D.C. Fragment-based lead discovery: Leads by design. *Drug Discov. Today* **2005**, *10*, 987–992.

168. Balius, T.E.; Mukherjee, S.; Rizzo, R.C. Implementation and evaluation of a docking-rescoring method using molecular footprint comparisons. *J. Comput. Chem.* **2011**, *32*, 2273–2289.

169. Perryman, A.L.; Santiago, D.N.; Forli, S.; Santos-Martins, D.; Olson, A.J. Virtual screening with AutoDock Vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: Participation in the SAMPL4 protein–ligand binding challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 429–441.

170. Ghosh, A.K.; Chapsal, B.D.; Weber, I.T.; Mitsuya, H. Design of HIV protease inhibitors targeting protein backbone: An effective strategy for combating drug resistance. *Acc. Chem. Res.* **2007**, *41*, 78–86.

171. 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.

172. Plewczynski, D.; Łażniewski, M.; Grotthuss, M.V.; Rychlewski, L.; Ginalski, K. VoteDock: Consensus docking method for prediction of protein–ligand interactions. *J. Comput. Chem.* **2011**, *32*, 568–581.

173. Gleeson, M.P.; Gleeson, D. QM/MM as a tool in fragment based drug discovery. A cross-docking, rescoring study of kinase inhibitors. *J. Chem. Inf. Model.* **2009**, *49*, 1437–1448.

174. Rastelli, G.; Rio, A.D.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797–810.

175. Gallicchio, E.; Deng, N.; He, P.; Wickstrom, L.; Perryman, A.L.; Santiago, D.N.; Forli, S.; Olson, A.J.; Levy, R.M. Virtual screening of integrase inhibitors by large scale binding free energy calculations: The SAMPL4 challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 475–490.

176. Slynko, I.; Scharfe, M.; Rumpf, T.; Eib, J.; Metzger, E.; Schüle, R.; Jung, M.; Sippl, W. Virtual screening of PRK1 inhibitors: Ensemble docking, rescoring using binding free energy calculation and QSAR model development. *J. Chem. Inf. Model.* **2014**, *54*, 138–150.

177. De Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N.P. Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308–2318.

178. Deng, N.; Forli, S.; He, P.; Perryman, A.; Wickstrom, L.; Vijayan, R.; Tiefenbrunn, T.; Stout, D.; Gallicchio, E.; Olson, A.J.; *et al.* Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease. *J. Phys. Chem. B* **2014**, *119*, 976–988.

179. Zartler, E.R. Quantum Tessera Consulting, 2013. Available online: http://www.quantumtessera.com/your-computation-is-only-as-good-as-your-experimental-follow-up/ (accessed on 14 July 2015).

180. Lo, M.C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **2004**, *332*, 153–159.

181. Lea, W.A.; Simeonov, A. Fluorescence polarization assays in small molecule screening. *Expert Opin. Drug Discov.* **2011**, *6*, 17–32.

182. Dias, D.M.; Van Molle, I.; Baud, M.G.; Galdeano, C.; Geraldes, C.F.; Ciulli, A. Is NMR fragment screening fine-tuned to assess druggability of protein–protein interactions? *ACS Med. Chem. Lett.* **2013**, *5*, 23–28.

183. Wielens, J.; Headey, S.J.; Rhodes, D.I.; Mulder, R.J.; Dolezal, O.; Deadman, J.J.; Newman, J.; Chalmers, D.K.; Parker, M.W.; Peat, T.S.; *et al.* Parallel Screening of Low Molecular Weight Fragment Libraries Do Differences in Methodology Affect Hit Identification? *J. Biomol. Screen.* **2013**, *18*, 147–159.

184. Sanner, M.F. Python: A programming language for software integration and development. *J. Mol. Graph. Model.* **1999**, *17*, 57–61.