Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Protein structure-based drug design: from docking to molecular dynamics

Paweł Śledź and Amedeo Caflisch

Recent years have witnessed rapid developments of computer-aided drug design methods, which have reached accuracy that allows their routine practical applications in drug discovery campaigns. Protein structure-based methods are useful for the prediction of binding modes of small molecules and their relative affinity. The high-throughput docking of up to 10^6 small molecules followed by scoring based on implicit-solvent force field can robustly identify micromolar binders using a rigid protein target. Molecular dynamics with explicit solvent is a low-throughput technique for the characterization of flexible binding sites and accurate evaluation of binding pathways, kinetics, and thermodynamics. In this review we highlight recent advancements in applications of ligand docking tools and molecular dynamics simulations to ligand identification and optimization.

Address
Department of Biochemistry, University of Zurich, Winterthurerstr. 190, 8057 Zürich, Switzerland

Corresponding authors: Śledź, Paweł (p.sledz@bioc.uzh.ch), Caflisch, Amedeo (caflisch@bioc.uzh.ch)

Current Opinion in Structural Biology 2018, **48**:93–102

This review comes from a themed issue on **Folding and binding**

Edited by **Amedeo Caflisch** and **Stefano Gianni**

https://doi.org/10.1016/j.sbi.2017.10.010

Introduction

Computational methods have played pivotal role in drug discovery efforts for many years [1]. Development of several approved drugs including early examples of captopril [2], saquinavir, ritonavir, indinavir [3], and tirofiban [4], has benefited substantially from the use of computer-aided drug design (CADD), which nowadays constitutes an essential part of the discovery pipeline at pharmaceutical companies [5•,6•]. The CADD tools are commonly classified into ligand-based (two-dimensional, 2D) and protein structure-based (3D). In this review we will focus on the 3D methods and discuss their potential and limitations. Their principles and implementations have evolved together with the concepts of molecular recognition on protein surface. In particular, the historic 'lock and key' mechanism that served as a textbook explanation of substrate recognition at the enzyme active site has gradually developed into 'hand and glove' concept to account for protein flexibility and mutual adaptability of receptor and ligand.

Structure-based CADD supports hit identification and medicinal chemistry optimization by addressing two major tasks: predicting how small molecules bind to the protein target, and estimating (relative) binding affinity. We first review docking, originally inspired by the lock and key concept, which is used for both tasks. We then present a fragment-based method for high-throughput docking based on molecular mechanics and transferable force field. Finally, we discuss molecular dynamics (MD) protocols, which provide atomistic details of hand and glove-like association events. The use of MD simulation-based methods is increasing steadily as they are most adequate for the analysis of thermodynamics and kinetics of ligand binding and unbinding. The section on fragment docking focuses on the methods and programs developed in the group of the last author, while the review of MD simulations of binding is more general.

Docking of small molecules to proteins

Automatic docking is concerned with the determination of the optimal position(s) and orientation(s) of a small molecule in a protein target. It has been reported that while the success of the approach is target-dependent and software suite-dependent, it poorly correlates with the binding affinity but rather depends on the quality of interactions that the ligand makes to the protein [7•]. Quality of protein–ligand interactions can be to some extent expressed by the ligand efficiency (LE), the average binding energy per non-hydrogen (or heavy) atom of the ligand. However, it should be noted that most studies of the predictive ability of docking are biased toward the molecules that bind the protein target with detectable affinity and available crystal structure. A study of about 300 kinase inhibitors has shown that a simple scoring function (van der Waals energy only) outperforms total energy (i.e. van der Waals and electrostatics) in fitting binding affinity values but has poor predictive power (i.e. lower enrichment than ranking by total energy) for *in silico* screening by high-throughput docking [8]. The real challenge of *in silico* screening is the calculation of relative binding energies with sufficient accuracy such that there are as many true positives as possible among the final selection of compounds for *in vitro* testing. In turn, the successful evaluation of binding energy relies on the

accurate prediction of the binding mode. Recent studies have reported high success rate of fragment screening by docking using transferable force fields with implicit solvent treatment of electrostatics desolvation effects [9,10].

Virtual screening by high-throughput docking

The principle of virtual screening is to evaluate the library of molecules for possibility of binding to the protein, and to shortlist the ones that are most likely to bind with the highest affinity. As mentioned above, the main challenge is not to identify the few nanomolar binders in the small-molecule library (if any at all) but rather to reduce the number of false negatives in the subset of compounds that are selected for validation by *in vitro* assays. There are few studies that systematically analyze the success ratio of docking campaigns (also called the hit rate), that is, the percentage of compounds correctly predicted to bind the protein target. While many papers report very good hit rates [11–13,14•,15], the criteria defining a hit are always subjective and study-dependent. The most stringent criterion is to consider as validated only those hits confirmed by the crystal structure of target-ligand complex. In this context it has to be noted that even for millimolar binders it is possible to obtain the crystal structure of the complex with the target protein. On the other hand, it can be very difficult and sometimes impossible to solve the crystal structure of complex with a potent ligand (e.g. nanomolar affinity) because the binding site can be either occluded by crystal contacts or not accessible to the ligand due to the tight packing of the protein molecules in the crystal (which mainly affects soaking experiments). Most commonly used criteria for the hit rate are based on affinity as measured in biochemical assays or biophysical experiments *in vitro* (typically K_D or IC_{50} below 100 μM) or semi-quantitative data, for example, from ligand-based NMR spectroscopy [16]. Such success stories have to be approached with caution, as it is clear that the selection process often involves visual inspection and examination through users with significant expertise and can be biased toward scaffolds disclosed previously in the literature. To properly benchmark the performance of different software suites, common criteria should be introduced and human intervention should be minimized which is not simple because of the complexity of the analysis of binding poses [17] and/or costs related to the *in vitro* validation.

Computer programs for flexible ligand docking

There is a plethora of software suites developed for the automatic docking of flexible small molecules into (mainly rigid) protein structures. On the other hand, only very few docking programs have gained broad recognition and are used by a large community [18]. These include Dock [19], GOLD [20], and AutoDock [21]. These solutions have gained high popularity due to their pioneering role in the field and thanks to extensive developments, which have turned them into user-friendly computer programs. More recently, rDock has emerged as an efficient docking tool distributed as open source code [22•]. The most popular docking tools share similar sampling procedures (genetic algorithms-based optimization in the conformational space of the rotatable bonds or grid-based searches) and some of them use force field-based evaluation of the binding energy. A high degree of convergence toward the same pose in multiple docking runs of the same ligand (with different initial random populations of the genetic algorithm) was reported as necessary condition for successful prediction of the binding mode [23], despite being frequently neglected. Importantly, the probability of successful prediction of the binding mode decreases substantially as the intrinsic flexibility of the ligand grows [23], and depends on high-quality interactions made with the receptor [7•]. Thus predictive ability has been validated for rigid fragments [9,24], while docking of peptides with more than a dozen rotatable bonds (backbone φ and ψ angles and side chain χ angles) is considered speculative.

Fragment docking

Nearly 20 years ago, the group of the last author developed a program for high-throughput docking of rigid fragments called SEED (Solvation Energy for Exhaustive Docking) [25]. SEED performs an exhaustive search in a discrete space defined by rotations around individual protein/fragment hydrogen bonds and/or hydrophobic contacts (Figure 1). This way, the essential feature of fragment-based drug discovery — making the high quality interactions with the protein [26] — is considered as a prerequisite and allows to reduce the complexity of search in the conformational space, and to enrich the docked poses in positives. A very efficient evaluation of bad contacts for filtering out poses with steric clashes and a two-step evaluation of the binding energy make the execution of SEED extremely rapid (about 1s per fragment). In both steps the energy evaluation is based on a transferable force field with continuum dielectric treatment of desolvation effects. The first step filters out the majority of the poses by the rapid evaluation of the van der Waals and Coulombic interactions on a 3D grid with a crude and very efficient approximation of desolvation effects [27]. In the second step the nonbonding interactions are calculated without grid-based approximation, and desolvation penalties are evaluated by the generalized Born equation with numerical calculation of the Born radii [28]. Importantly, the SEED binding energy does not require any fitting parameter and thus SEED can be used also for protein targets for which inhibitors have not been reported.

Successful high-throughput docking campaigns with SEED have been published for proteases, kinases, and bromodomains [9,24,29•]. In a recent application SEED was used to screen for the CREBBP bromodomain a library of nearly 1500 fragments, which took less than one hour on a commodity computer, and resulted in a 50%