Archives of Biochemistry and Biophysics 583 (2015) 105–119

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/ locate/yabbi

Review article

Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description

Francesca Spyrakis [a], Claudio N. Cavasotto [b, *]

[a] Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 183, 41125, Modena, Italy
[b] Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA) — CONICET — Partner Institute of the Max Planck Society (IBioBA-MPSP), Godoy Cruz 2390, C1425FQD, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form
3 August 2015
Accepted 3 August 2015
Available online 10 August 2015

Keywords:
Structure-based drug discovery
Homology modeling
Ligand docking
Virtual screening
Protein flexibility
Active site water molecules

ABSTRACT

Structure-based virtual screening is currently an established tool in drug lead discovery projects. Although in the last years the field saw an impressive progress in terms of algorithm development, computational performance, and retrospective and prospective applications in ligand identification, there are still long-standing challenges where further improvement is needed. In this review, we consider the conceptual frame, state-of-the-art and recent developments of three critical "structural" issues in structure-based drug lead discovery: the use of homology modeling to accurately model the binding site when no experimental structures are available, the necessity of accounting for the dynamics of intrinsically flexible systems as proteins, and the importance of considering active site water molecules in lead identification and optimization campaigns.

1. Introduction

The formation of non-covalent complexes between macromolecules and small-molecules is essential for the proper functioning of cellular processes, such as enzyme catalysis and signal transduction. With the aim of designing chemical modulators of therapeutic relevant targets, the pharmaceutical industry basically relied on high-throughput screening, a costly strategy involving the experimental screening of chemical libraries against a specific target. Since about 40 years ago, three-dimensional (3D) structures of protein-ligand complexes have been used to guide the optimization of drug leads in terms of potency and selectivity [1], thus incorporating structural knowledge into the drug discovery process. Later on, computational methods became also available to model protein-ligand interaction, and more recently, to *in silico* screen large chemical libraries against a biomolecular target. Structure-based virtual screening (SBVS) strategies thus became the *in silico* counterpart of older high-throughput screening approaches [2–7]. Since then, SBVS has experienced a continuous improvement in terms of algorithm development, computational performance, and retrospective and prospective applications in drug lead identification [8,9]. The dramatic surge of CPU power, and the advent of GPUs, also significantly increased the feasibility of computational

Abbreviations: GPCR, G Protein-coupled receptor; HM, homology modeling; HTD, high-throughput docking; LSHM, Ligand-steered homology modeling; MD, molecular dynamics; MRC, multiple receptor conformations; PDB, Protein Data Bank; PMF, potential of mean force; REMD, replica-exchange molecular dynamics; SAR, structure-activity relationships; SBVS, structure-based virtual screening; SDM, site-directed mutagenesis; VS, virtual screening.

* Corresponding author.
E-mail addresses: cnc@cavasotto-lab.net, ccavasotto@ibioba-mpsp-conicet.gov.ar (C.N. Cavasotto).

http://dx.doi.org/10.1016/j.abb.2015.08.002
0003-9861/

simulations: screening million-compound libraries towards pharmaceutically relevant receptors is now quite feasible even with standard computational resources.

There are four main components in an SBVS protocol (Fig. 1):

i) *The target structure*. A crucial element in any SBVS campaign is the availability of the 3D structure of the macromolecular target of interest. The rate of X-ray structures solved and deposited in the Protein Data Bank (PDB) is constantly increasing with thousands of proteins being added each year. At the moment this review is written (August 2015) ~110000 biological macromolecular structures are present in the PDB [10]. The probability of successfully crystallizing a new target depends on many factors, such as the sequence length, the inherent flexibility, the presence of transmembrane helices, and the net charge, among others [11]. Thus, not all pharmaceutically relevant targets can be easily crystallized. NMR also represents a great source of protein structural information, even though somehow limited by protein solubility and molecular weight. The inclusion of water molecules in the structure should be also decided at this stage (cf. section "Modeling waters within the active site"). Target structures should be carefully prepared to ensure structural integrity, assign the correct residue protonation and tautomer states, and inspect Asn and Gln flips. When no experimental structure is available, or easily obtainable, comparative or homology modeling (HM) can furnish reliable target models to SBVS (cf. section "*In silico* target models in structure-based virtual screening").

ii) *The compound library*. Compound libraries to be screened against the aforementioned target also need to be carefully prepared, taking into account the most representative tautomers, protomers and stereoisomers at the pH of interest (usually 7.4) [12]. In cases of retrospective protein-ligand docking, the availability of ligand activity information for the considered target, and an un-biased set of inactive molecules, may also contribute to obtain meaningful results from docking-based screening [13,14].

iii) *The docking strategy*. Each molecule from the chemical library has to be placed within the target binding site optimizing protein-ligand interactions, and retaining the most favorable poses; each pose has to be scored by a native or external scoring function, according to which the library of screened molecules is ranked. Many docking algorithms and scoring functions have been developed and implemented along these years, in order to properly estimate the free energy of binding, or to maximize the separation of potential ligands and decoys, and thus to place the potential binders at the top of the hit list [15–18]. In the choice of the docking or virtual screening strategy, different aspects should be considered. The level of flexibility for both the target and the molecule library has to be specified. During docking, although both entities could be considered rigid (as in the original "lock and key" model),

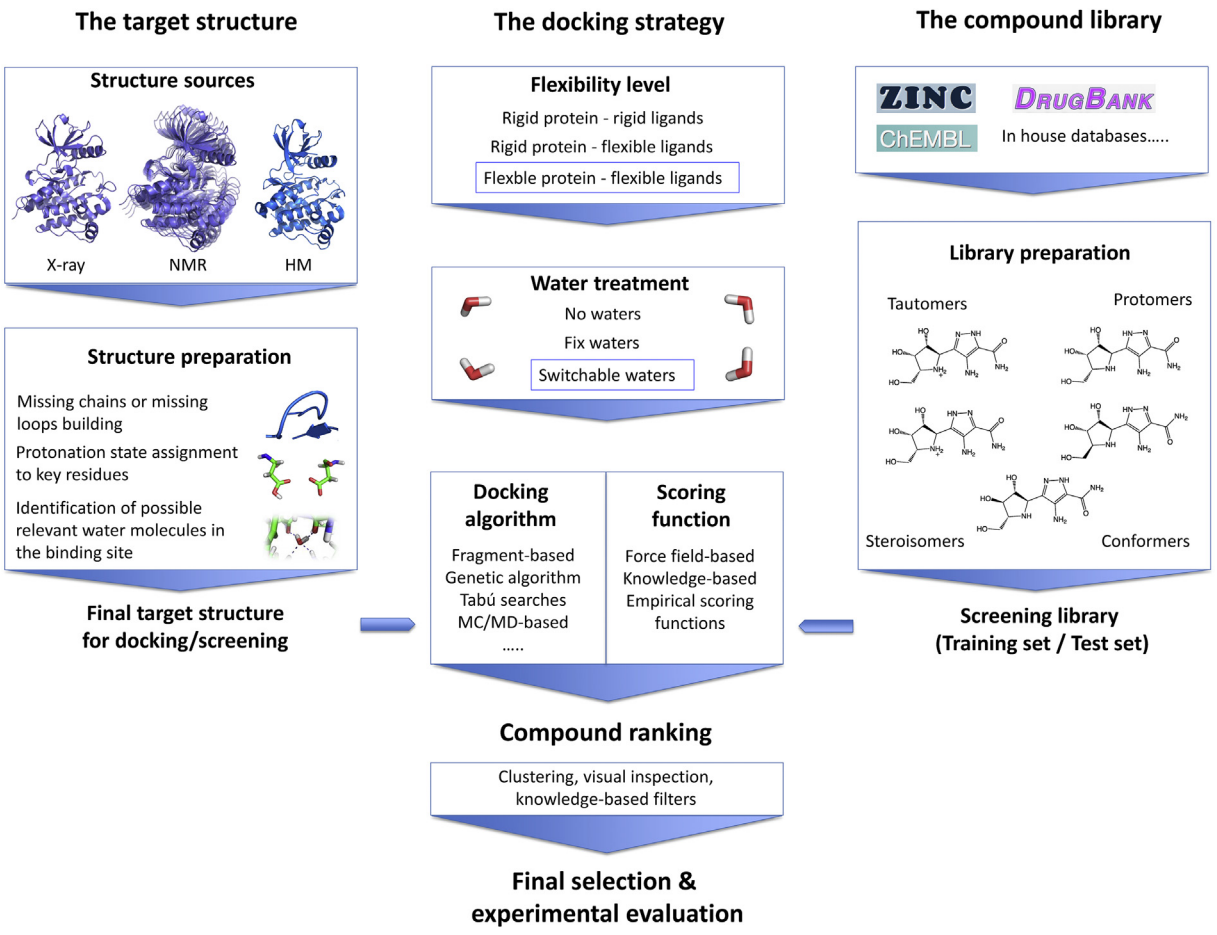

Fig. 1. Structure-based virtual screening workflow. The principal steps to be followed to perform virtual screening experiments and the available strategies are shown.