Bioorganic & Medicinal Chemistry 24 (2016) 4961–4969

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

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

Challenges of docking in large, flexible and promiscuous binding sites

Martin Kotev [a], Robert Soliva [b], Modesto Orozco [a,c,*]

[a] Joint BSC-IRB Program in Computational Biology. Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona 08028, Spain
[b] Joint BSC-IRB Program in Computational Biology, Barcelona Supercomputing Center, Barcelona 08031, Spain
[c] Department of Biochemistry and Molecular Biology, University of Barcelona (UB), Barcelona 08029, Spain

ARTICLE INFO

Article history:
Received 11 June 2016
Revised 5 August 2016
Accepted 6 August 2016
Available online 6 August 2016

Keywords:
Induced-fit docking
PELE simulations
Soluble Epoxide Hydrolase
Benchmark

ABSTRACT

After decades of work, the correct determination of the binding mode of a small molecule into a target protein is still a challenging problem, whose difficulty depends on: (i) the sizes of the binding site and the ligand; (ii) the flexibility of both interacting partners, and (iii) the differential solvation of bound and unbound partners. We have evaluated the performance of standard rigid(receptor)/flexible(ligand) docking approaches with respect to last-generation fully flexible docking methods to obtain reasonable poses in a very challenging case: soluble Epoxide Hydrolase (sEH), a flexible protein showing different binding sites. We found that full description of the flexibility of both protein and ligand and accurate description of solvation leads to significant improvement in the ability of docking to reproduce well known binding modes, and at the same time capture the intrinsic binding promiscuity of the protein.

1. Introduction

Accuracy in predicting how a small molecule binds to its target is an essential requirement for structure-based drug design (SBDD). The correct pose prediction of a small molecule in a binding site allows gaining insight on which chemical features are essential for achieving its biological activity,[1,2] and using this information to improve the properties of the lead compound. In principle, the binding mode of ligands can be obtained by means of experimental techniques, but the time-restrictions in practical drug design projects are often incompatible with X-ray and NMR time scales, forcing the use of computational approaches, particularly of docking algorithms.[3]

Current docking methods are typically based on a rigid receptor around which a ligand is rotated, translated and flexed until the best fit is found. Although docking has been proven to be an extremely useful tool,[4] it yields suboptimal results in cross-docking experiments, i.e., when one ligand is docked to the image of the receptor that is either unbound, or bound to a significantly different ligand.[5] Bearing in mind that most macromolecular targets undergo some kind of induced-fit upon binding,[6] the number of cases where standard docking procedures find problems is not negligible.

A simple strategy to model the flexibility of the macromolecular receptor is to submit it to a molecular dynamics (MD) simulation, from which a series of relevant conformations for rigid docking are derived.[2,3] Although this might mitigate the negative impact of the rigid receptor approximation, plain MD does not usually provide sufficient protein conformational sampling, and there are not clear rules to extract a subset of representative configurations from the crude ensemble.[7] Similar or even larger problems arise when protein models for docking are obtained from experimental ensembles.[8] Alternative strategies to ensemble docking arise from the use of Induced-Fit Docking algorithms. One such approach is based on the iterative combination of rigid receptor docking plus protein structure prediction for certain areas of the active site.[9] This method has proven to be especially relevant for detecting significantly large conformational transitions upon ligand binding, such as the kinase DFG-in/out conformational change.[9] However, in general, solving the flexibility problem requires direct coupling between conformational exploration and binding, which in principle can be achieved by brute force molecular dynamics algorithms.[10] Unfortunately, these calculations are still in the proof-of-concept stage requiring highly specialized computer resources which are rarely accessible to standard drug-design laboratories. Use of biased MD simulations can reduce the cost of calculations, but they are still too expensive and requires a previous knowledge of the expected binding mode,[11] something uncommon in the practical drug design scenario.

* Corresponding author.
E-mail address: modesto.orozco@irbbarcelona.org (M. Orozco).

http://dx.doi.org/10.1016/j.bmc.2016.08.010
0968-0896/

A few intermediate methods aiming to couple docking with conformational transitions, but escaping from the complexity of brute force MD simulations have been suggested. For example, in the essential dynamics molecular dynamics (ED/MD) approach the protein dynamics in the essential deformation space is coupled with sampling of ligand degrees of freedom in the Cartesian space[12,13] allowing a coupled sampling of protein and ligand flexibilities. Medusadock,[14] models receptor and ligand flexibility simultaneously by using sets of discrete rotamers and RosettaLigand[15] allows docking ligands to receptors by using a Monte Carlo minimization procedure in which the rigid body position and orientation of the small molecule and the protein sidechain conformations are optimized simultaneously.[16] Finally, PELE (Protein Energy Landscape Exploration)[17–22] implements a Metropolis Monte Carlo algorithm, where new trial configurations are produced for both ligand and protein while docking is performed. The program implements side chain prediction algorithms, specific ligand rotamer libraries, implicit solvent model,[23] and all-atom OPLS2005 force field[24] to compute receptor–ligand interactions.

We explore here the performance of two state-of-the-art docking methods: GLIDE, based on the traditional paradigm of a rigid receptor and a flexible ligand, and PELE, a fully flexible approach. Of note, for the rigid receptor approximation, both standard docking against a series of experimental X-ray structures and docking against two MD-generated ensembles are tested. The benchmark system selected was soluble Epoxide Hydrolase (sEH, Fig. 1a and b), an enzyme that metabolizes epoxyeicosanoic acids (EETs) generating products that have been linked to a variety of therapeutic areas, such as inflammation, pain and hypertension.[25] This protein was chosen, not only for its potential pharmacological interest, but because of its novelty (structural data on this protein were not available when GLIDE or PELE were developed) and its complexity, since it shows[26,27] a complex binding site characterized by a huge hourglass-shaped binding cavity with the catalytic residues in a center channel separating a left-hand site (LHS) from a right-hand site (RHS), both able to bind ligands.[26,27] The combination of high flexibility and complexity of the binding cavity (divided in 3 different sub-cavities) makes sEH an extremely challenging benchmark for docking approaches.

Results presented here confirm the power of standard docking approaches (exemplified by GLIDE) but also their caveats. It is found that generating an ensemble of conformations via MD can in some instances improve the results as compared to using only an X-ray structure. However, for this challenging system it is found that representing both small molecule and protein flexibility simultaneously is crucial in order to have reasonable poses, to discriminate between alternative binding sites and to derive poses that can be useful for further lead optimization studies. In this sense, the use of fully flexible methods as PELE provides much improved results.

2. Methods and computational details

More than 90 structures of human sEH are deposited in the PDB. A recently deposited one (PDB entry 5AHX) corresponds to the *apo* structure, while a variety of other crystals correspond to complexes with ligands of different sizes bound to different subcavities. For our docking experiment we selected 20 of these ligands based on diversity criteria (see Fig. 2) and cross-docked them to 6 protein structures corresponding to the unbound *apo* structure (5AHX) and 5 bound states (see Fig. 2) selected to cover the different binding modes of the protein (5AI0, 5AIC, 5AKE, 5ALX and 5ALO). Additionally, for the MD-ensemble approach, a series of protein conformations were generated by standard molecular dynamics simulations for the *apo* structure (5AHX) and also for the complex of sEH when bound to Fulvestrant (4J03), an anti-cancer agent that was serendipitously found to be a potent sEH inhibitor (see Fig. 2). All the protein models were prepared for further calculations using the Protein Preparation Wizard (PPW) tool of Schrödinger software[28,29] thus building any incomplete sidechains, optimizing the hydrogen bond network and predicting the protonation states of all His, Asp, Glu, Lys, and Arg sidechains. Solvent molecules were removed in all cases, as well as any other molecules that might be present from the crystallization buffers.

2.1. CMIP calculations

Classical molecular interaction potentials (CMIP using Na^+ and CH_4 as probes;[30]) were used to define the recognition characteristics of sEH binding site. The electrostatic term in CMIP was obtained by solving the non-linear Poisson Boltzmann equation[31] (with standard protein and solvent dielectric definition, while the steric terms were accounted for by van der Waals formalism using standard AMBER parameters.[30]

2.2. Standard docking

Standard docking calculations (rigid target, flexible ligand) were performed using GLIDE, one of the most validated docking programs in the field.[32–34] GLIDE grids were centered on catalytic triad

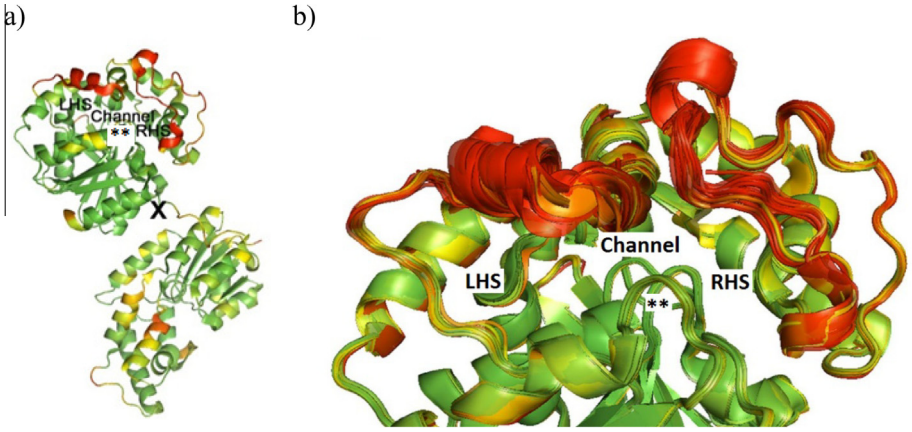

Figure 1. Structure of sEH. (a) Catalytic (upper) and lid (lower) domains depicted with a B-factor ribbon (green-yellow-red spectra). The 'x' denotes the site at which the lid region was cut to simplify the system. The ** sign is the approximate starting position for the ligands in the PELE simulations. (b) Binding site of sEH, highlighting the LHS, RHS, and Channel subsites and B-factor ribbon marking the most flexible loops for 56 PDB entries. The color coding is from green (rigid) to red (flexible).