Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Pharmacology

Predicting how drug molecules bind to their protein targets

Moira M Rachman[1], Xavier Barril[1,2] and Roderick E Hubbard[3,4]

There have been substantial advances in the application of molecular modelling and simulation to drug discovery in recent years, as massive increases in computer power are coupled with continued development in the underlying methods and understanding of how to apply them. Here, we survey recent advances in one particular area — predicting how a known ligand binds to a particular protein. We focus on the four contributing classes of calculation: predicting where a binding site is on a protein; characterizing where chemical functional groups will bind to that site; molecular docking to generate a binding mode for a ligand and dynamics simulations to refine that pose and allow for protein conformation change. Examples of successful application are provided for each class.

Addresses
[1] Facultat de Farmàcia and Institut de Biomedicina, Universitat de Barcelona, Av. Joan XXIII, 27-31, 08028 Barcelona, Spain
[2] Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain
[3] YSBL, University of York, Heslington, York YO10 5DD, UK
[4] Vernalis (R&D) Ltd, Granta Park, Abington, Cambridge CB21 6GB, UK

Corresponding author: Hubbard, Roderick E (roderick.hubbard@york.ac.uk)

Current Opinion in Pharmacology 2018, **42**:34–39

This review comes from a themed issue on **New technologies**

Edited by **Antoine Bril**, **Oliver Nosjean** and **Patrick Genissel**

https://doi.org/10.1016/j.coph.2018.07.001

Introduction

The majority of drug discovery projects begin with identification of a small molecule compound which binds to a defined site on a specific biological molecule (usually a protein), affecting the function of that target protein. This initial hit is then optimized to incorporate adequate drug-like properties (affinity, selectivity, efficacy, ADME, etc.) into a candidate compound that generates the desired therapeutic effect and is suitable for clinical trials.

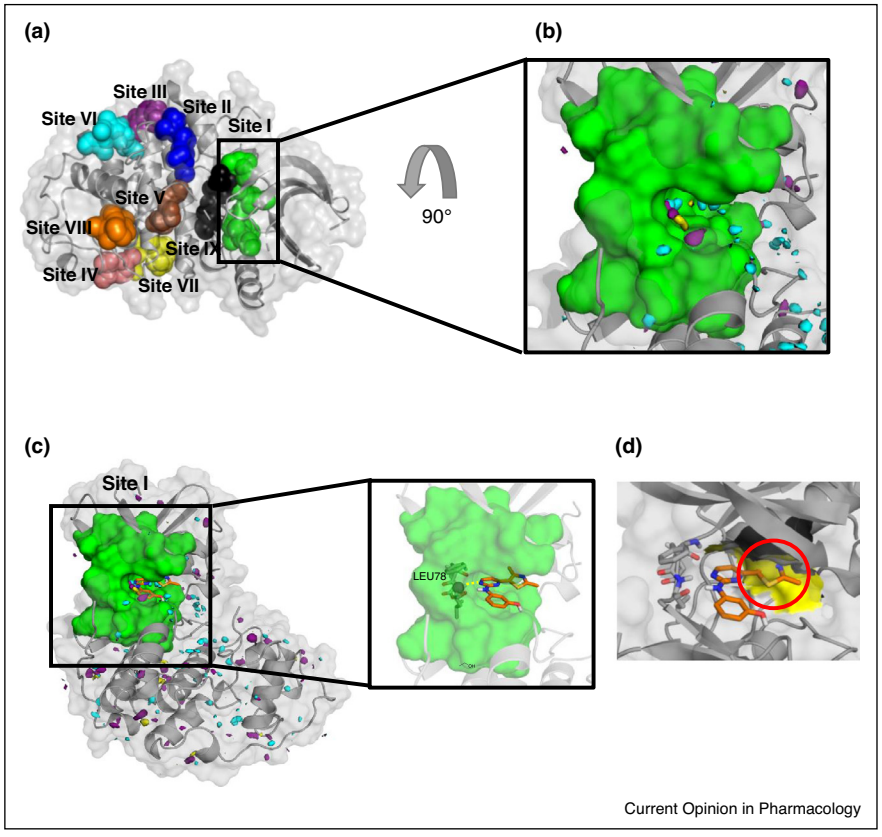
Over the past thirty years, there has been a steady increase in the use of structure-based methods in this drug discovery process where models of how compounds bind to the target can allow rational design of the required improvements in the compounds. For some targets, experimental methods can provide structural information with sufficient throughput and speed to interactively guide the structure-based design. For example, X-ray crystallography provides an atomic level picture of how compounds bind and NMR spectroscopy can provide varying levels of information on interactions between the compound and the protein, such as whether a compound binds, where it is binding to and (in some limited cases) a structure of the compound binding to the target. However, it is often not possible to generate such structures with sufficient speed to inform decisions about compound optimization.

In this review, we survey recent developments in the computational methods that predict how compounds bind to their protein target using either an experimentally determined structure of the target or a model based on sequence homology. Some of the methods can be used to screen compound libraries (real or virtual) for initial hits; in addition, the methods can help to guide optimization of compounds in structure-based design. These applications are not discussed in detail here. What we focus on are the methods that, once a compound is demonstrated to bind, can be used to predict the position and orientation or 'pose' of the compound binding. As summarized in Figure 1, we have loosely divided these methods into four categories: (1) identifying binding sites; (2) characterizing the potential of a binding site to bind chemical matter; (3) predicting the position and orientation (or pose) of compound binding and (4) dynamic docking to explore both the energetics of binding and conformational change to refine the pose. Before summarizing these in turn, we first survey some history and the issue that underpins all molecular modelling — the ability to estimate energy of interaction.

Origins of the methods

A more detailed description of the origins of structure-based design methods is provided elsewhere [1] but there are three influential developments that should be highlighted — CHARMM [2], GRID [3] and DOCK [4]. The Karplus group developed molecular dynamics (MD) simulations of a protein in 1977 [2], which led to the development of the CHARMM [5] (Chemistry at Harvard Molecular Mechanics) program which became a central platform for many molecular simulation methods over the following decades. One of the most influential developments for structure-based drug discovery in the 1980s was the program GRID from Goodford [3]. This introduced the idea of

Figure 1



Methods for predicting ligand binding modes illustrated through an example of calculations for the kinase, CDK2 (protein structure taken from the PDB code: 1CKP). **(a)** Binding site prediction by fpocket [15] (default settings) by clustering solvent inaccessible spheres and disregarding solvent exposed spheres. A 'druggability' score is assigned to each predicted pocket. In this example, Site I, obtained the highest score (0.8), while the remaining eight pockets score very low (<0.1). **(b)** Polar hot spots identified through mixed solvent MD simulations using MDMix [20]. Ethanol and water were used to probe the binding pocket, from which high and low energy areas are identified. The low energy areas probed by ethanol (deep purple), help to identify donor or acceptor features that could be exploited by ligand binding. Water (cyan) and hydrophobic (yellow) sites are also probed. **(c)** These hot spots were then used to guide docking of the ligand from PDB structure 1PXM. Docking was performed with rDock [31], using a donor as a pharmacophoric restraint (sphere) to interact with the backbone of LEU78 (yellow dashed line) in the CDK2 structure. **(d)** This was followed by pose refinement using MD [54] to explore the flexibility of the pocket; for example, the yellow surface indicates possibility for a clash between ligand and protein.

characterizing what types of chemical functionality would bind to a binding site by calculating the energy of interaction between the protein and a functional group at each point on a grid. Finally, there is the DOCK program from the Kuntz group [4] which was the first widely used program for computationally docking compounds into the structure of a protein. Although some of the ideas within these programs were built on the work of others, the programs (and their authors) became major promoters of the ideas of using computational methods to characterize and predict how compounds can bind to proteins and formed the foundation of the current generation of methods.

Predicting the energy of interaction between protein and ligand — scoring functions

All structure-based design methods critically rely on an estimate of the energy of interaction between a ligand (or probe) and the protein. Most approaches still rely on the rather simplistic treatment established in the early methods [2–4] where the non-covalent interactions are treated with simple Coulombic (for electrostatics) or Lennard-Jones (for van der Waals) interaction potentials but there is increasing use of more sophisticated treatments. The theoretical bases for these more advanced calculations were established a long time ago. What has changed in recent years is the relentless increase in computer power allowing these methods to be applied within a realistic timeframe. There are three main areas to highlight.

The first are the perturbation methods [6] which calculate changes in free energy by performing extensive MD while transforming (in this case the compound) from one chemical structure to another. The second is a number of approaches for more extensive treatment of