Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Accelerating drug discovery through tight integration of expert molecular design and predictive scoring

Robert Abel[1], Sayan Mondal[1], Craig Masse[2], Jeremy Greenwood[1], Geraldine Harriman[2], Mark A Ashwell[2], Sathesh Bhat[1], Ronald Wester[2], Leah Frye[1], Rosana Kapeller[2] and Richard A Friesner[3]

Modeling protein–ligand interactions has been a central goal of computational chemistry for many years. We here review recent progress toward this goal, and highlight the role free energy calculation methods and computational solvent analysis techniques are now having in drug discovery. We further describe recent use of these methodologies to advance two separate drug discovery programs targeting acetyl-CoA carboxylase and tyrosine kinase 2. These examples suggest that tight integration of sophisticated chemistry teams with state-of-the-art computational methods can dramatically improve the efficiency of small molecule drug discovery.

Addresses
[1] Schrodinger Inc., New York, NY 10036, United States
[2] Nimbus Therapeutics, Cambridge, MA 02141, United States
[3] Department of Chemistry, Columbia University, New York, NY 10027, United States

Corresponding author: Friesner, Richard A (raf8@columbia.edu)

Current Opinion in Structural Biology 2017, **43**:38–44

This review comes from a themed issue on **Theory and simulation**

Edited by **Carol Post** and **Ronald Levy**

http://dx.doi.org/10.1016/j.sbi.2016.10.007

Introduction

Modeling protein–ligand interactions has been a central goal of computational chemistry for many years. The prediction of the structure and binding affinity of protein–ligand complexes are essential tasks if computational methods are to facilitate structure-based drug design. Calculations of sufficient accuracy and robustness have the potential to substantially reduce costs and timelines in both the lead discovery and lead optimization phases of a drug discovery project, and dramatically expand the chemical space of ligands that can be evaluated as potential drug candidate molecules.

Over the past 30 years, three distinct types of computational technologies have evolved to address modeling of protein–ligand binding. The fastest methods involve docking the ligand into the receptor site, using conformational search methods to determine the structure of the complex, and an empirical scoring function, suitable for application to diverse ligand chemistries, to evaluate binding affinity [1–4]. Current rigid receptor docking programs require a few seconds to a few minutes per ligand, enabling virtual screening of millions of candidate ligands, an approach that is suitable for lead discovery applications.

A second approach employs molecular mechanics potential energy functions along with a continuum description of aqueous solvation, typically employing Poisson–Boltzmann (PB) or generalized Born (GB) solvation models in conjunction with a surface area (SA) term. The MM-GBSA approach, requiring only a few minutes of computation time per ligand, has had considerable success in approximately rank ordering congeneric series, and is widely used in the pharmaceutical industry in lead optimization [5,6].

The third approach is to carry out all-atom, explicit solvent molecular dynamics (MD) simulations [7]. In principle, with sufficient accuracy of the force field and adequate sampling of phase space, high accuracy in both structure and binding affinity prediction can be achieved. Only recently has there been progress in unbiased structural prediction via MD simulation, using brute force simulation methods, enhanced sampling algorithm based on metadynamics, and other techniques [4,8,9]. However, when modeling ligands in a congeneric series, it is usually possible to generate a reasonably accurate initial binding mode for new ligands in the series from a combination of project crystallography, known SAR, and docking calculations.

A rigorous statistical mechanical method first described more than 60 years ago, free energy perturbation theory (FEP), can be used to efficiently calculate relative binding affinities of congeneric ligands [10]. In this approach, the initial ligand is 'alchemically' transformed to a different target ligand, by progressive modification of the Hamiltonian of the system. The first FEP calculations

of protein–ligand binding affinity were carried out in the early 1980s by a number of groups [11]. At that time, computing power was inadequate to converge the calculations, and force fields for both protein and ligand exhibited significant deficiencies. Over the next 30 years, the combination of improvements in force fields with the exponential growth of computing capabilities has enabled significant progress to be made [12,13]. Most recently, these significant advances have enabled free energy calculations to perform well in blind tests and active discovery projects [14••,15,16•,17••,18••,19•,20].

A second, quite useful application of MD simulation is to identify sites of high water occupancy in the receptor active site and assign approximate thermodynamics properties (e.g., displacement free energies) to these sites [21,22]. The results of such a calculation, for example performed using the WaterMap methodology, can be employed to heuristically guide ligand design to displace and/or replace certain water molecules based on their thermodynamic profile. More recently this information been incorporated into docking calculations [3].

Along with these technological advances, there comes a human challenge: how to effectively integrate these related yet distinct multi-disciplines for idea generation, compound optimization and prioritization for synthesis in the context of an active drug discovery project. Clearly there are two major goals of such an interface; firstly, early elimination of synthetic target molecules with little chance of achieving the desired potency and drug-like properties and secondly, to focus synthetic effort on the highest expected value target molecules, perhaps in the face of significant synthetic challenges.

In the present paper, we describe the state-of-the-art application of these technologies to several collaborative drug discovery projects involving a closely integrated partnership between Schrodinger Inc. (a computational chemistry company) and Nimbus Therapeutics (a biotechnology company). In these projects, docking and WaterMap calculations are the predominant approaches to identify initial lead compounds, and WaterMap, MM-GBSA, and FEP calculations are extensively employed to prioritize compounds for synthesis in lead optimization. The commitment in these projects to allow computation to drive ligand design is significantly larger than is common in industry, enabling an assessment of their effectiveness under practical discovery project conditions.

Design of an allosteric inhibitor of acetyl-CoA carboxylase

Acetyl-CoA carboxylase (ACC) is the rate limiting enzyme in de-novo fatty acid synthesis [23,24]. There are two isozymes of this protein, ACC1 and ACC2. ACC1, which is located in the cystosol, catalyzes the initial step in fatty acid synthesis in lipogenic tissues. ACC2 which is bound to the mitochondrial membrane, regulates fatty acid oxidation in oxidative tissue by allosteric modulation of carnitine palmitoyltransferase via malonyl CoA production [23,25]. An extensive series of biological studies using animal models has shown that inhibition of both ACC1 and ACC2 can have a substantial impact on a wide variety of disease physiology, including cancer metabolism, accumulation of fat in the liver, and development of diabetes [24]. Furthermore, mice in which the ACC2 gene is knocked out, or ACC1 gene knockout is performed in a suitable tissue dependent manner, do not exhibit major adverse effects, suggesting that inhibition of ACC is a clinically acceptable approach for treating the relevant disease indications [26–32]. Consequently, ACC has been a high priority target for drug discovery for over a decade.

Both ACC isozymes are composed of two domains, the biotin carboxylase (BC) and carboxyltransferase (CT) domains. Initial efforts to inhibit ACC focused on designing ligands which bind to the active site in the CT domain [23,25,33–37]. However, the active site of this domain is very hydrophobic, often resulting in drug candidates with poor pharmaceutical properties [33,38–40]. An alternative is to target the BC domain by preventing dimerization, which is necessary for ACC's enzymatic activity. The potent ACC inhibitor and natural product, Soraphen, has been shown to bind to at the BC dimer interface, in a shallow hydrophilic pocket [33,41–43].

Our work to develop a more satisfactory drug-like molecule that might be able to bind the Soraphen pocket located at the BC dimer interface is extensively detailed in Ref. [44]. An analysis of the WaterMap results of the BC dimer interface was crucial in the decision to initiate a discovery campaign for this target. The WaterMap results indicated the presence of multiple adjacent high energy (easily displaceable) waters not displaced by Soraphen might provide an opportunity to achieve improved potency in a drug-like small molecule Soraphen analogue. With this knowledge in-hand, a highly successful virtual screen was pursued to identify compounds that bind to the Soraphen pocket by displacing these key water sites.

Several hundred compounds were prioritized from this protocol for purchase. One of the compounds tested, ND-022, with a potency of 3.9 μM against ACC1, was advanced into lead optimization. Optimization was performed using WaterMap and MM-GBSA calculations, based on the crystal structure of ND-022 complexed with the BC domain. Figure 1a shows the co-crystal structure of ND-022 complexed with hACC2 BC over-laid with the structure of hACC2 BC complexed with Soraphen A. At the time this project was active, a robust and accurate FEP methodology was not available, so this approach was not used in ACC lead optimization. Nevertheless, an exceptional Development Candidate was identified after