Free-energy calculations in structure-based drug design

Michael R. Shirts, David L. Mobley, and Scott P. Brown

INTRODUCTION

The ultimate goal of structure-based drug design is a simple, robust process that starts with a high-resolution crystal structure of a validated biological macromolecular target and reliably generates an easily synthesized, high-affinity small molecule with desirable pharmacological properties. Although pharmaceutical science has made significant gains in understanding how to generate, test, and validate small molecules for specific biochemical activity, such a complete process does not now exist. In any drug design project, enormous amounts of luck, intuition, and trial and error are still necessary.

For any small molecule to be considered a likely drug candidate, it must satisfy a number of different absorption/distribution/metabolism/excretion (ADME) properties and have a good toxicological profile. However, a small molecule must above all be active, which in most cases means that it must bind tightly and selectively to a specific location in the protein target before any of the other important characteristics are relevant. To design a drug, large regions of chemical space must be explored to find candidate molecules with the desired biological activity. High-throughput experimental screening methods have become the workhorse for finding such hits.[1,2] However, their results are limited by the quality and diversity of the preexisting chemical libraries, which may contain only molecules representative of a limited portion of the relevant chemical space for a given target. Combinatorial libraries can be produced to supplement these efforts, but their use requires careful design strategies and they are subject to a number of pitfalls.[3] More focused direct in vivo or in vitro measurements provide important information about the effect of prospective drugs in the complete biological system but provide relatively little information that can be directly used to further engineer new molecules. Given a small number of molecules, highly accurate assays of binding, such as surface plasmon resonance (SPR) or isothermal calorimetry (ITC), are relatively accessible though rather costly.

Ideally, small molecules with high potential biological activity could be accurately and reliably screened by computer before ever being synthesized. The degree of accuracy that is required of any computational method will depend greatly its speed. A number of rapid structure-based virtual screening methods, generally categorized as "docking," can help screen large molecular libraries for potential binders and locate a putative binding site (see Chapter 7 for more information on docking). However, recent studies have illustrated that although docking methods can be useful for identifying putative binding sites and identifying ligand poses, scoring methods are not reliable for predicting compound binding affinities and do not currently possess the accuracy necessary for lead optimization.[4–6]

Atomistic, physics-based computational methods are appealing because of their potential for high transferability and therefore greater reliability than methods based on informatics or extensive parameterization. Given a sufficiently accurate physical model of a protein/ligand complex and thorough sampling of the conformational states of this system, one can obtain accurate predictions of binding affinities that could then be robustly incorporated into research decisions. By using a fundamental physical description, such methods are likely to be valid for any given biological system under study, as long as sufficient physical detail is included. Yet another advantage of physics-based models is that the failures can be more easily recognized and understood in the context of the physical chemistry of the system, which cannot be easily done in informatics-based methods.

Despite this potential for reliable predictive power, few articles exist in the literature that report successful, prospective use of physics-based tools within industrial or academic pharmaceutical research. Some of the likely reasons for such failures are the very high computational costs of such methods, insufficiently accurate atomistic models, and software implementations that make it difficult for even experts to easily set up with each new project. Until these problems are resolved, there remain significant obstacles to the realization of more rigorous approaches in industrial drug research.

There have been a number of important technical advances in the computation of free energies since the late 1990s that, coupled with the rapid increase in computational power, have brought these calculations closer to