the goal of obtaining reliable and pharmaceutically useful binding energies. In this chapter, we briefly review these latest advances, with a focus on specific applications of these methods in the recent literature. Under "How Accurate Must Calculations of Affinity Be to Add Value" we first discuss the level of reliability and accuracy that binding calculations must have to add some degree of value to the pharmaceutical process. Under "Free Energy Methodologies" we give an overview of the methods currently used to calculate free energies, including recent advances that may eventually lead to sufficiently high throughput for effective pharmaceutical utility. Under "MM-PBSA Calculations" and "Alchemical Calculations" we review recent ligand binding calculations in the literature, beginning with relatively computationally efficient methods that are generally more approximate but still attempt to calculate a true affinity without system-dependent parameters and then address pharmaceutically relevant examples of most physically rigorous methods. We conclude with a discussion of the implications of recent progress in calculating ligand binding affinities on structure-based drug design.

HOW ACCURATE MUST CALCULATIONS OF AFFINITY BE TO ADD VALUE?

Physics-based binding calculations can be very computationally demanding. Given these time requirements, it is important to understand quantitatively what levels of precision, throughput, and turnaround time are required for any computational method to systematically effect the lead-optimization efforts of industrial medicinal chemists in a typical work flow. To be useful, a method does not necessarily need to deliver perfect results, as long as it can produce reliable results with some predictive capacity on time scales relevant to research decision-making processes. These issues are frequently addressed anecdotally, but rarely in a quantitative manner, and we will try to sketch out at least one illustration of what the requirements of a computational method might be.

A recent analysis of more than 50,000 small-molecule chemical transformations spanning over 30 protein targets at Abbott Laboratories found that approximately 80% of the resulting modified molecules had potencies lying within 1.4 kcal/mol (i.e., 1 pK_i log unit) of the starting compound.[7] Potency gains greater than 1.4 kcal/mol from the parent were found to occur approximately 8.5% of the time, whereas gains in potency greater than 2.8 kcal/mol were found with only 1% occurrence. Losses in binding affinity on modification were approximately equal in magnitude and probability to the gains for most types of modifications; presumably wholly random chemical changes would result in a distribution with losses in binding that are much more common than gains. We treat this distribution as typical of lead-optimization affinity gains obtained by skilled medicinal chemists and use this distribution to examine the ability of accurate and reliable computational methods to influence drug research productivity.

Suppose our chemist sits down each week and envisions a large number of modifications of a lead compound he or she would like to make and test. Instead of simply selecting only his or her best guess from that list, which would lead to a distribution in affinity gains similar to the one described above, this chemist selects N compounds to submit to an idealized computer screening program. The chemist then synthesizes the top-rated compound from the computer predictions. What is the expected distribution of affinities arising from this process for different levels of computational error?

To model this process, we assume the medicinal chemist's proposals are similar to the Abbott data and we approximate this distribution of binding affinity changes as a Gaussian distribution with mean zero and standard deviation of 1.02 kcal/mol, resulting in 8.5% of changes having a pK_i increase of 1.0. We assume the computational predictions of binding affinity have Gaussian noise with standard deviation ϵ. In our thought experiment, we generate N "true" binding affinity changes from the distribution. The computational screen adds Gaussian error with width ϵ to each measurement. We then rank the "noisy" computational estimates and look at distribution of "true" affinities that emerge from selecting the best of the corresponding "noisy" estimates. Repeating this process a number of times (for Figure 5.1, one million), we can generate a distribution of affinities from the screened process.

Shown in Figure 5.1 is the modeled distribution of experimental affinity changes from the chemist's predictions (blue) versus the distribution of the experimental affinity changes after computationally screening $N = 10$ compounds with noise $\epsilon = 0.5$ (pink), $\epsilon = 1.0$ (red), and $\epsilon = 2.0$ (purple). In other words, the blue distribution of affinities is what the medicinal chemist would obtain alone; the redder curves what the chemist would obtain synthesizing the computer's choice of his N proposed modification. The shaded area represents the total probability of a modification with affinity gain greater than 1.4 kcal/mol.

With 0.5 kcal/mol computational noise, screening just ten molecules results in an almost 50% chance of achieving 1 pK_i binding increase in a single round of synthesis, versus an 8.5% chance without screening. With 1 kcal/mol error, we still have 36% chance of achieving this binding goal with the first molecule synthesized. Surprisingly, even with 2 kcal/mol, computational noise almost triples the chance of obtaining a 1 pK_i binding increase. Similar computations can be done with large numbers of computer evaluations; unsurprisingly, the more computational evaluations can be done, the more computational noise can be tolerated and still yield useful time savings. For example, even with 2 kcal/mol error, screening 100 molecules results in the same chance of producing a 1 pK_i binding increase that is the same as if ten molecules are screened with 0.5 kcal/mol error.