

Specific Aims

Among the most fundamental molecular interactions in biology are those of small molecules with their macromolecular partners. Endogenous small molecules play the role of messengers in many signaling pathways of the cell, and small molecule drugs interact with proteins in signaling cascades to modulate their function. Understanding their interactions is vital to understanding many biological systems, and critical to drug development efforts. Despite having catalogued many of the physical driving forces behind small molecule recognition, there are enormous gaps in our knowledge preventing us from articulating a quantitative, predictive understanding of small molecule affinities and selectivities for biomolecules.

In principle, alchemical free energy calculations provide a framework for quantitatively describing all aspects of the thermodynamics of small molecule recognition. However, deficiencies in our quantitative understanding of binding create large challenges in the ability of these calculations to reproduce experimental data in many systems, holding back their use in probing function and aiding design. In this proposal, we address three of the most significant open challenges in the quantitative modeling of small molecule recognition.

Aim 1. The binding of charged ligands

The soluble form of many drugs and endogenous small molecules is charged. Yet, retrospective studies and predictive challenges show current free energy methodology fails for charged species. A number of corrections for calculations with charged ligand species have been proposed, but (1) consensus, and (2) a good model system to test and to confirm theory are lacking. We propose using a computationally and experimentally tractable model system—the association of small-molecule guests with high-affinity supramolecular hosts—as a way to test, validate, and refine both theory and algorithms applicable to charged ligands.

We will utilize isothermal titration calorimetry (ITC) experiments, which provide a gold standard biophysical assay for binding affinities. We will additionally develop Bayesian approaches to accurately quantify measurement error and allow for model-error propagation and improved automated experimental design. The experimental datasets collected here will allow us to perform free energy calculations, and to evaluate proposed corrections in order to refine theory and algorithms as necessary.

Aim 2. Protonation state effects on binding

Proteins and many small-molecule drugs contain titratable moieties that can change protonation state upon binding or sample mixtures of protonation states, often in a conformation-dependent manner. While detailed biophysical studies of a few specific model systems have demonstrated that these effects can contribute several kcal/mol in binding affinity, the true scope of the problem in ligand recognition in general is completely unknown.

We will use computational techniques to conduct a large scale survey to broadly assess the importance of protonation state issues in protein-small molecule interactions across a range of systems of pharmacological interest. Existing pKa prediction tools will be benchmarked against small molecule pKa data, and then used together with Monte Carlo titration codes (both extant and developed for this project) to provide an estimate of the scope of the problem. We will perform complementary experiments on a tractable but disease-relevant system—kinase catalytic domains that can be expressed in *E. coli*—using fluorescence binding experiments developed in our laboratory, as well as ITC experiments when possible.

Aim 3. Weak association and multiple binding

Weak binding and association of multiple ligands are ubiquitous interactions in biological and pharmaceutically relevant systems. In addition, drug discovery approaches such as fragment-based ligand design depend predominantly on a reliable method for integrating data from biophysical experiments with modeling for these situations. We will extend the framework of alchemical free energy calculations to include the potential for multiple (possibly weak) binding events using a semigrand canonical ensemble formalism. As a model system, we will use the pharmacologically relevant protein human serum albumin (HSA), known to bind many small molecules (sometimes multiple species at a time) to a variety of distinct binding sites. We will bridge experiment and theory by both developing new Bayesian ITC experimental analysis techniques to select among theoretical association models, as well as simulating ITC data directly from semigrand canonical ensemble simulations.

Completion of the work will provide modeling strategies for non-trivial challenges in protein-ligand association that currently have no working solution, as well as illustrate the scope of challenges that remain.