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 affinities 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 by alchemical free energy calculations**.

Aim 1. Establish a correct quantitative treatment of alchemical free energy calculations for binding of charged ligands. The predominant soluble form of many drugs and endogenous small molecules is charged. Yet, retrospective studies and predictive challenges show current free energy methods make substantial errors in the treatment of binding by charged species [1, 2]. A number of corrections for calculations with charged ligand species have been proposed [3–5], 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. The experimental datasets collected here will allow us to evaluate proposed corrections to free energy calculations in order to refine theory and algorithms as necessary. We will utilize isothermal titration calorimetry (ITC) experiments, which provide a "gold standard" biophysical assay for binding affinities. Because we require accurate experimental uncertainties to determine whether theory is discrepant from experiment in a statistically significant manner, we will additionally develop Bayesian approaches to accurately quantify measurement uncertainty and allow for model-error propagation to rectify deficiencies in current data analysis protocols.

Aim 2. Quantify the magnitude of 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 [6–10], **the true scope and magnitude of the protonation state effects in ligand recognition in general is completely unknown.**

We will use computational techniques to conduct a survey to assess the importance of protonation state issues for a tractable but disease-relevant system—kinase catalytic domains with known bacterial expression and their corresponding commercially-available small molecule inhibitors. Existing pK_a prediction tools will be benchmarked against experimental kinase inhibitor pK_a data, and then used as input for constant-pH alchemical free energy calculations. Using free energy calculations we can quantify the contribution of protonation state effects, which can be validated by performing complementary experiments. We will perform ITC experiments using buffers of different ionization enthalpies that allow for deconvolution of binding and protonation states effects.

Aim 3. Develop a framework for alchemical free energy calculations to describe weak association and cooperative ligand 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 disparate biophysical experiments with modeling [11]. We will extend the framework of alchemical free energy calculations to include the potential for multiple (possibly weak) binding events using a semi-grand canonical ensemble formalism.

As a model system, we will use the pharmacologically relevant protein human serum albumin(HSA), known to multivalently bind many different small molecules to a variety of distinct binding sites [12]. **Most theory and frameworks focus on simple 1:1 association models, which are insufficient to correctly describe association in these circumstances.** We will extend deficiencies by both developing new Bayesian ITC experimental analysis techniques to select among theoretical association models, as well as simulating ITC data directly from semi-grand canonical ensemble simulations to enable direct comparisons between experiment and theory.

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.

Research Strategy - Significance

Accurate treatment of alchemical free energy calculations.

The binding free energy of a ligand (related to the affinity) determines to what degree it will spontaneously associate with a protein. Access to information such as free energies can be of great utility in the design of ligands that can selectively and tightly interact with a biomolecular target. Alchemical free energy calculations are a powerful computational tool for computing binding free energies, as they allow for efficient sampling of the relevant states of protein-ligand complexes [13]. While these methods hold considerable promise in the accurate prediction of small molecule binding free energies [14–16], their domain of applicability is currently highly restricted; and the broad application of free energy calculations to numerous pharmacologically relevant targets requires a number of additional challenges to be overcome [17–19]. **In particular, the neglect of corrections required for the treatment of charged ligands [1, 2], dynamic protonation state effects [6–10], and the potential for weak interactions and multiple ligand association events [20] can have a great impact on the reliability of alchemical free energy calculations for numerous systems in which these effects are important.** There is a great need for validation and extension of available methods for treating these situations in order to apply alchemical binding free energy calculations to wider ranges of pharmacologically relevant systems.

Alchemical free energy calculations involving charged ligands in explicit solvent require charge corrections.

In order to apply alchemical free energy calculations to charged ligands, one needs to eliminate artifacts introduced into the calculation arising from the modeling of bulk solvent behavior using a small periodic system. In a physical system, when a ligand dissociates the interaction never actually disappears completely, though the interaction between a ligand and protein is weakened to the extent that it there is no measurable correlation between the interacting molecules. The elimination of a charge from a system is therefore not a physical process, and is merely an artifact of using artificial infinite systems to approximate a macroscopic system of bulk liquid. Bulk liquids are approximated in simulation, either by using periodic boundary conditions, or an implicit solvent. Often, to further reduce computation cost, we introduce truncated potentials and non-Coulombic electrostatics (such as particle mesh Ewald [PME] [21] and reaction field [RF] [22] potentials). These unphysical treatments bring with them several electrostatic interaction terms that are added on top of the physical binding free energy (Figure 1). These terms do not perturb the weights in a normal simulation, but cause substantial artifacts in alchemical simulations once charged species are eliminated or introduced as part of an alchemical perturbation. The estimated binding free energy as such needs to be corrected to eliminate these effects. Neglecting to correct for these contributions can severely undermine the ability of alchemical free energy calculations to reproduce experimental binding affinities [1, 2]. These corrections may also prove to be necessary when performing dynamical protonation state calculations, where both the number of ligand and protein protons (and hence the total system charge) may fluctuate during the simulation.
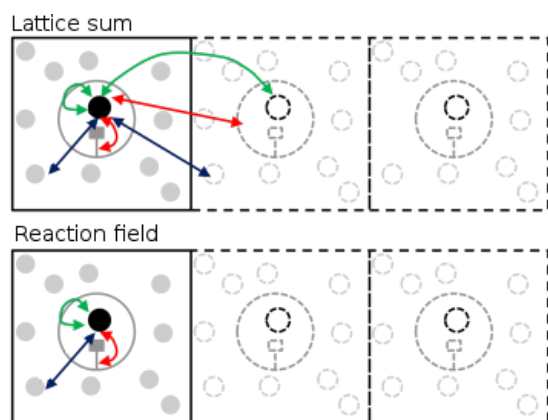
Figure 1: **Graphic illustration of contributions to binding free energy introduced when changing the net charge of the system.** It is common practice to truncate Coulomb interactions beyond a certain range to reduce calculation costs and use methods such as PME [21] or reaction field [22] solvation models to compensate for the loss of interaction beyond the cutoff. When there is a net charge change, this leads to errors in treatment of interaction between ligand and solvent molecules beyond cutoffs and interactions with periodic copies of the system (blue arrows). Likewise, interactions between macromolecule and ligand need to be corrected (red arrows). Lastly, self interacting terms may also introduce errors when charges are removed (green arrows). Figure adapted from Reif and Oostenbrink [3].

Protonation state effects can be important in kinase inhibitor binding, but the scope of this problem is unknown.

Kinases are important signaling proteins, and for a long time its been known that disregulation of some kinases can lead to cancer [23, 24]. They are a promising drug target, and many drugs have successfully been developed to inhibit kinases that are disregulated in cancer [25–27]. Unfortunately, intrinsic or induced mutations can lead to resistance, meaning there is a continued need for drugs to overcome resistance [26, 28–30].

  Imatinib (Gleevec) is a potent inhibitor of Abl kinase [31], and has shown high efficacy in patients with chronic myeloid leukemia [32]. At physiological pH , it has access to multiple protonation states (Figure 2b) [33]. Furthermore, from computation and experiment, there is evidence that there could be pH dependent effects to the binding affinity [34, 35]. The use of incorrect or inappropriate protonation states in modeling protein-ligand association can significantly degrade accuracy [36,

37]. But what if there is not a single protonation state that contributes to the binding affinity? With the exception of a few well-studied cases [6–10], the ubiquity of protonation state changes and the magnitude of their contribution to binding affinities in protein-ligand association in general is completely unknown. Preliminary data (Figure 2a) suggests that many kinase inhibitors have multiple accessible protonation states at physiological pH . We expect there are two separate effects that could be relevant:

1. A significant population of mixtures of protonation states may be present.

2. Changes in the dominant protonation state could occur upon binding.

Modeling the system with a single protonation state can in either case lead to large errors. As some kinase inhibitors bind with picomolar affinity to their target (corresponding to 28 $k_B T$) [38], a 500-fold affinity loss (6 $k_B T$, equivalent to 2.6 log unit change in the equilibrium constant) would still maintain sub-nanomolar potency. This suggests that a majority of these compounds could access multiple protonation states while still retaining a high binding affinity (Figure 2a). Neglecting the potential of sampling alternate protonation states would thus lead to large errors in predicted binding affinities. Such large contributions could play a significant role in the selectivity of kinase inhibitors for their targets. Therefore, it is essential to know whether these effects need to be incorporated for the prediction of kinase inhibitor binding affinities.
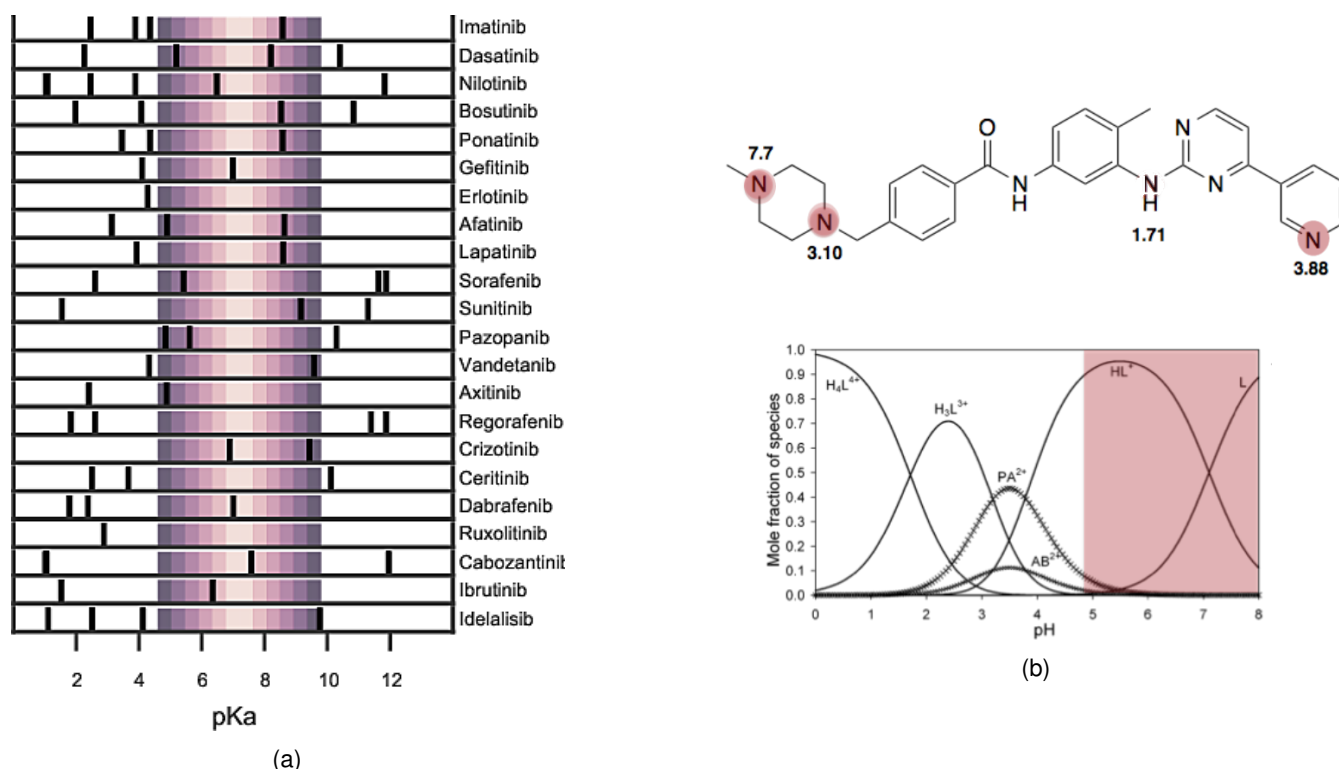

Figure 2: (a) **Protonation states accessible to FDA approved kinase inhibitors.** The intracellular pH of most tumors is around 7.2 [39, 40]. Colors indicate accessibility of states at pH 7.2 within blocks of $k_B T$ units of energy, where darker means more units of $k_B T$ away from the accessible region, up until 6 $k_B T$. Preliminary data obtained using Epik [41, 42]. (b) Top: The individual sites of protonation of imatinib (Gleevec) accessible with a penalty of 6 $k_B T$ under physiological conditions, highlighted in red [33]. Bottom: fraction of species of imatinib measured by titration and NMR [33], with the red region denoting states accessible with a penalty of 6 $k_B T$ under physiological conditions.

Multiple (weak) ligand binding models are required for studies of HSA and fragment based drug design.

Most experimental tools and theoretical frameworks focus on 1:1 association of proteins and ligands, but these tools must be extended if one wishes to accurately describe binding of multiple ligands to human serum albumin (HSA), or the weak association of multiple fragments with a protein. **This proposal addresses deficiencies in existing frameworks by dealing with the association of multiple ligands, as well as weak association, enabling accurate modeling of human serum albumin and protein-fragment binding.**

Because HSA controls the bioavailability of small molecules by absorbing them into the bloodstream [43, 44], it essential to understand how drugs interact with it if one wants to predict pharmacokinetics [12, 43, 45–47]. It binds multiple ligands simultaneously [44, 48–51], making it hard to model in existing frameworks. The (weak) association of multiple ligands is a significant problem in fragment based drug discovery (FBDD), that occurs because of the high concentrations used in experiments. FBDD is frequently applied and has impacted the way the pharmaceutical industry and academia alike

approach small molecule design [11]. The association of these fragments is hard to model computationally, because of the challenges associated with modeling weak interactions [20], and the absence of modeling frameworks for multiple association. Modeling deficiencies need to be addressed if free energy calculations are to be used to rationalize and predict fragment binding.

Research Strategy - Innovation

Reliable uncertainty estimates of isothermal titration calorimetry results by using Bayesian inference.

If experimental results are to be used to validate theoretical models, it is essential that one accurately estimates the uncertainty of the experiment. **Otherwise, it is impossible to tell whether experiment and theory differ from each other in a statistically significant way.**

The ABRF-MIRG'2 study revealed the disturbing fact that the uncertainty of ITC experiments is underestimated by several orders of magnitude by conventional analysis methods [52]. Panel D of Figure 3 indicates a statistical uncertainty of at least 10% in the estimation of extinction coefficients between all participants of the study, which is directly proportional to the concentration of reagents prepared by each lab. The error in reagent concentrations is not correctly propagated into the estimates of binding stoichiometry, affinity and enthalpy (Panels A, B and C of Figure 3), leading to inaccurate estimates of the uncertainty of these observables.

The underestimation of the error could have been prevented by performing adequate replicate experiments. By only performing technical replicates and no repeat experiments with newly prepared reagent solutions, the true uncertainty cannot be quantified. A hypothesis cannot be supported by a single experiment, and technical replicates perform but a single experiment, where only the measurement is carried out multiple times [53].

Alternatively, one can get an accurate estimate of the uncertainty by performing Bayesian inference on their experiments. We propose the usage of Bayesian statistics as a means of uncertainty estimation as opposed to errors from model fit based on single or replicate experiments. A Bayesian analysis of ITC experiments allows one to get more value out of single experiments, reducing resource costs. This is because Bayesian inference allows prior knowledge about possible errors of reagent concentrations to be propagated into the uncertainty estimates of model parameters such as binding affinity, enthalpy and stoichiometry. It thereby provides better uncertainty estimates from single experiments than standard analysis approaches, which do not propagate prior information on errors.
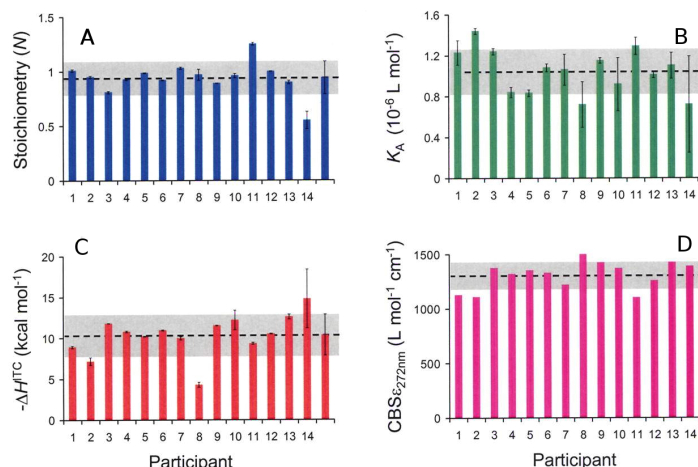

Figure 3: Binding measurements of CBS to bovine carbonic anhydrase II from the ABRF-MIRG'2 study. **Independent ITC as performed by 14 labs shows individual error estimates are orders of magnitude smaller than actual experimental variation.** A: Predicted stoichiometry of binding. B: Association constant K_A. C: enthalpic contribution to binding. D: Measured extinction coefficient of ligand CBS, as reported by the 14 participants [52]. Error bars indicate the error as reported by the individual labs that were part of the study. The gray region indicates the uncertainty when taking into account all of the measurements. As can be seen, some labs underestimate the overall error by over 2 orders of magnitude. The extinction coefficient, as displayed in panel D is directly proportional to the concentration, revealing that concentration errors of around 10 percent are likely not propagated correctly into the analysis.

Assessing proposed charge corrections in alchemical free energy calculations of charged ligands.

Alchemical free energy calculations involving charged ligands in explicit solvent suffer from artifacts due to the treatment of the system as an infinite periodic box, and approximations in the electrostatic models [1, 2]. Several means of correcting for these artifacts have been suggested [3–5]. **A quantitative comparison of these methodologies to each other and to experiment has not yet been performed, and the correctness of any of these available charge correction methods has not been established.** Our work will contribute substantially to the understanding of these charge corrections by providing an experimental dataset with small and quantified uncertainty, against which models can be validated. For the first time these models will be compared to each other by simulating the same system. We can determine whether these methods provide consistent corrections, and at the same time the experimental dataset allows the assessment of quantitative agreement with experiment.

Quantifying unknown contributions to kinase-inhibitor binding specificity by modeling protonation state effects.

Imatinib, a potent inhibitor of Abl kinase [31], has access to multiple protonation states at physiological pH (Figure 2b) [33], and there is evidence of pH dependent effects to the binding affinity [34, 35]. The magnitude of these contributions still remains to be elucidated, and preliminary data (Figure 2a) suggests that many other kinase inhibitors have multiple accessible protonation states at physiological pH. **This work will be the first attempt to comprehensively identify and quantify these effects for known kinase inhibitors, both experimentally and computationally.** Some well studied model systems have shown that protonation state changes can contribute significantly to ligand binding affinity [6–10], yet contribution of protonation state effects to binding affinities in protein-ligand association in general is completely unknown. Providing the evidence that this is important for kinase inhibitors may lead to further studies into how this contributes to inhibitor selectivity, a poorly understood characteristic of many kinase inhibitors [38, 54]. It may also warrant the investigation of these effects in other classes of proteins, while at the same time this proposal provides the tools to perform these investigations.

Estimating free energies of multiple ligands using semi-grand canonical ensemble free energy calculations.

Most experimental and modeling frameworks for treatment of protein-ligand binding focus on 1:1 association. Conversely, to model binding of ligands to serum albumin, or to model the weak binding of fragments to a protein, one needs to model the association of multiple ligands. Association of multiple ligands to a protein are not adequately treated by most analysis and modeling frameworks. **This work will provide novel computational and experimental frameworks that will allow for the simulation and analysis of multiple ligand association with a single target.** It also provides a solution for some of the challenges associated with modeling weak interactions [20], commonly observed in fragment binding [11]. The alchemical framework that we suggest handles the association of weak ligands by sampling the overall free energy of association as a function of the number of ligands bound, without focusing on individual binding sites. The model for multiple ligand association can be derived from semi-grand canonical ensemble theory [55]. The semi-grand canonical ensemble also allows for the direct, quantitative connection between alchemical free energy calculations and ITC experiments. This quantitative connection can be used both to suggest experiments that can validate the model, and to enable simulations of ITC experiments to optimize experimental design.

Research Strategy - Approach

Aim 1. Establish a correct quantitative treatment of alchemical free energy calculations for binding of charged ligands

Rationale

The soluble form of many drugs and endogenous small molecules is charged. Yet, retrospective studies and predictive challenges show current free energy methodology makes substantial errors in the treatment of binding by charged species [1, 2]. For explicit solvent simulations using particle mesh Ewald —the gold standard approach to simulating solvated biomolecular systems— a number of details of the calculation impact macroscopic observables when the total charge of the system or the overall charge density changes by removal of a charged ligand. A number of approaches for explicitly solvated PME free energy calculations with charged species have been proposed, but (1) consensus, and (2) a good model system to test and to confirm theory are lacking [3–5].

We propose using 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 use cucurbit[7]uril (CB7) as a host [56], which is known to bind a series of cations within a range of affinities spanning several orders of magnitude (Figure 4) [57]. In order to provide an experimental dataset with accurately quantified uncertainty, we will perform isothermal titration calorimetry (ITC) experiments of various guests binding to CB7. We will then perform alchemical free energy calculations to evaluate the available approaches in an attempt to support or refute them.

ITC experiments will benefit from the high solubility of CB7. CB7 is more soluble than many proteins, $20\,\text{millimolar}$ to $30\,\text{millimolar}$ [56] in water versus $< 1.0\,\text{millimolar}$ for a small soluble protein such as RNAase Sa [58]. This will allow us to reach higher concentrations, which will make for more tractable experiments.

The small system size provides an additional advantage for molecular dynamics simulation, since the limited degrees of freedom will make sampling more efficient and reduces computational cost. Unlike biomolecules, CB7 does not undergo large-scale conformational changes and has small correlation times [59, 60].
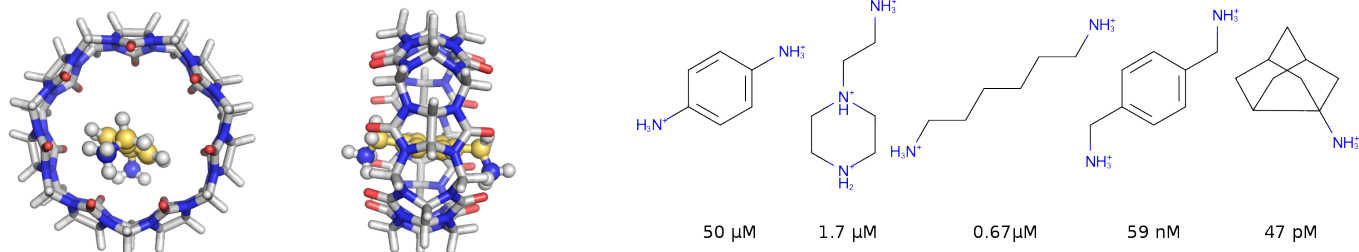
(a) A divalent guest bound to bound to the central cavity of CB7.

(b) A selection of guests and their K_d values from Cao and Isaacs [57].

Figure 4: **Cucurbit[7]uril (CB7) bound to several cationic guests from Cao and Isaacs [57]**. Different guests bind to the central cavity of CB7, their affinity determined by their geometry and size. Their affinities are in a range similar to many protein-ligand complexes, though the complexes are small, with few degrees of freedom, making them excellent candidates for our study. Figure prepared using Glide docking [61–64] into crystal structure pdb:QQ7 [65].

ITC has been shown to lead to serious underestimation of measurement error due to failure to propagate errors in the concentration of reactants [52, 66]. To provide an accurate assessment of error from single experiments, we will develop a novel analysis method using Bayesian inference. The framework of Bayesian inference allows for the propagation of errors throughout the experimental procedure, allowing us to accurately quantify the experimental uncertainty, thereby enabling us to clearly refute or support the various approaches to dealing with charged ligands. We will consider the approaches of Reif and Oostenbrink [3], Rocklin et al. [4], and Lin et al. [5] for eliminating the contribution of net charge. Each scheme mostly agrees on the source of error, but has a different way of solving the problem, as described in Subaim 1.2. We will also consider the elimination of neutral salt pairs, in lieu of corrections, to evaluate the necessity of these complex schemes.

Subaim 1.1: Develop an accurate approach to quantifying experimental uncertainty in ITC using Bayesian inference.

To test and confirm theoretical models, an accurate quantitative estimate is needed of the experimental error of the experiments that are used to validate. We will develop Bayesian approaches to accurately quantify measurement uncertainty and allow for model-error propagation. Errors in the concentration of prepared solutions are a major source of uncertainty in ITC data. Not taking these into account can lead to orders of magnitude underestimation of the uncertainty of the experiments [52, 66]. One way to take this into account is by doing repeat experiments in which all solutions are prepared anew for each replicate, instead of technical replicates [53]. This however is time and material consuming, and therefore not a popular option among most scientists.

In a Bayesian approach, we can assign prior distributions to the true concentrations of both species, allowing the uncertainty in these concentrations (tracked during solution preparation via standard propagation of error) to be used as input. Using *Markov Chain Monte Carlo* (MCMC) [67, 68], we can then sample from a joint posterior probability distribution of all thermodynamic parameters to estimate the actual concentration, using *Bayes' rule*

$$\mathcal{P}\left(\theta|\mathcal{D}\right) \propto \mathcal{P}(\mathcal{D}|\theta)\mathcal{P}\left(\theta\right) \quad ; \quad \theta = \left\{\Delta G_{\text{bind}}, \Delta H_{\text{bind}}, \Delta H_0, [\text{X}_{\text{syr}}], [\text{M}_{\text{cell}}], \sigma\right\} \tag{1}$$

where $\mathcal{P}(\mathcal{D}|\theta)$ is the likelihood of the data $\mathcal{D}$ (i.e., injections heats q_n^{obs}), given the individual thermodynamic parameters θ, and $\mathcal{P}(\theta)$ is a prior distribution of the set of independent thermodynamical parameters. Our posterior model contains parameters such as the binding affinity (ΔG_{bind}), the enthalpy of binding (ΔH_{bind}), an offset to incorporate mechanical and dilution effects (ΔH_0), the true reagent concentrations in the syringe ($[\text{X}_{\text{syr}}]$), and in the sample cell ($[\text{M}_{\text{cell}}]$), and σ, the square root of the variance of our measurements.

The simplest likelihood model of an ITC experiment consist of a series of n injections, resulting in injection observed heats q_n^{obs},

$$q_n^{\text{obs}} \sim \mathcal{N}(q_n^{\text{true}}, \sigma^2) \quad , \tag{2}$$

which are drawn from a normal distribution $\mathcal{N}$ with a variance of σ^2 and with the true injection heats q_n^{true} as the underlying mean. As stated by the central limit theorem (CLT), the sum of the individual power measurements, integrated to give the heat of single injections, will be normally distributed.

The methodology will allow incorporation of multiple experiments, such as calibrations, to accurately separate out heat effects due to e.g. dilution of the chemicals used or mechanical effects of the injections.

We will collect ITC data for the series of guests used by Cao and Isaacs [57] binding to CB7. Experiments will be performed using an Auto iTC-200 instrument, available at the Rockefeller University High-Throughput and Spectroscopy Resource Center (HTSRC). The experimental setup will be designed using a Python library [69] developed in house for the design of ITC experiments [70], and the setup will be carried out on our in house laboratory automation system.

Subaim 1.2: Perform a quantitative comparison of suggested correction models to experiments to establish a correct treatment of charged ligands in alchemical free energy calculations. We will compare several methods that correct for changes in the net charge of the system in alchemical free energy calculations [3–5], in order to estimate which methods can provide quantitatively correct estimates of the binding affinity. To do so, we will perform alchemical binding free energy calculations on the series of guests-CB7 complexes for which binding affinities were measured as part of Subaim 1.1 and compare the methods to the experimental results.

We will consider a total of four schemes that account for charged ligand binding. The approach of Reif and Oostenbrink [3] uses a thermodynamic cycle to define a raw charging contribution to binding free energy, and tries to deconvolve the various contributions so that they can be subtracted from the total binding free energy. Rocklin et al. [4] suggests a numerical, as well as an analytical scheme to calculate contributions, both using Poisson-Boltzmann (PB) calculations on the protein-ligand system. Then, there is the approach of Lin et al. [5], who use potential of mean force (PMF) calculations in a single simulation system with a macromolecule and ligand unbound, separating them by a large distance in explicit solvent. We will also test an additional model that does not apply corrections, in which a neutral ligand:counterion pair is alchemically eliminated.

We will first consider whether all methods agree with each other, and the extent of the difference between their estimates. Discrepancies between the results could indicate that some of the approaches are quantitatively incorrect. If the various approaches agree, we can compare them in terms of computational efficiency. Next, we will compare each of the approaches to the experimental data, to see whether any of them produce a quantitatively correct result. The accurately quantified uncertainty of our experimental data will provide us with a means to refute or buttress the proposed correction schemes.

Expected results

The ITC experiments performed on the CB7-guest complexes will be analyzed using our Bayesian ITC analysis method. This will provide us with estimates of their binding affinity with accurately quantified uncertainty. Then, we will perform alchemical binding free energy calculations, in an attempt to reproduce the experimental data. We will evaluate whether each of the individual schemes produces the same answer and will estimate the difference between their estimates. Next, we can compare the result obtained by each scheme to the experimental results. If their estimate falls within the range of the experimental estimate, then it supports the suggested approach. Each of the methods carries an increased computational cost with them, compared to not applying corrections. By applying them on the same system, we can also get an estimate on how efficient they are in costs versus improved accuracy. We will then be able to recommend a methodology to use when calculating the free energy of binding for charged species.

Pitfalls and alternatives

It is possible that all methods produce comparable results, in which case we can perform a comparison of efficiencies to determine which method is best. Another potential outcome is that none of the models are able to reproduce the experimental results. This could indicate problems with the force field used to simulate the systems. We can use a variety of different force fields to estimate errors that are due to the force field. For instance, we could compare the GAFF [71], CGENFF [72, 73], and OPLS [74] force field models to get an estimate of the force field error.

To further estimate force field error, we could perform the same experiments and alchemical free energy calculations using a neutral guest. This would leave out the contribution due to a net charge change, so that the accuracy of the force field can be assessed. We may find the need to try more complicated models for our Bayesian analysis of the ITC experiments. We can incorporate effects such as dilution heats, injection volume errors, and sophisticated baseline estimation to improve our estimates.

Aim 2. Quantify the magnitude of protonation state effects on binding

Rationale

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. Evidence exists that for the binding of imatinib to Abl kinase, pH dependent effects may contribute to the binding affinity[33–35]. Our preliminary investigations suggest that the same effects will be relevant to a large number of FDA approved kinase inhibitors (Figure 2a). Kinases are important signaling proteins, and for a long time its been known that disregulation of some kinases can lead to cancer [23, 24]. They are a promising drug target, and many drugs have succesfully been developed to inhibit kinases that are disregulated in cancer [25–27]. Kinase inhibitors typically have high potency [38], so protonation state effects may mostly go unnoticed in standard experimental assays. The usual approach in modeling ignores these effects and assumes a single protonation state for an entire system. However, they form a potential pitfall when modeling, since these unknown contributions potentially lead to large errors in the binding affinity.

While detailed biophysical studies of a few specific model systems have demonstrated that these effects can contribute several kcal/mol in binding affinity [6–10], the ubiquity and magnitude of protonation state effects in ligand recognition in

general is unknown. **Our aim is to identify kinase systems where protonation state effects influence binding, and to quantify the effect of protonation state effects on the binding affinity of kinase inhibitors.** We will first survey kinase-inhibitor complexes in the protein databank (PDB) [75] to identify complexes in which protonation state effects may play a major role. To identify whether protonation state changes may have an effect on binding in these systems, we will simulate them using a Monte Carlo titration framework called MCCE [76] which can rapidly sample the protonation states of both inhibitor and kinase subject to a constrained backbone and fixed binding mode. From the populations of protonation states sampled by MCCE, we can predict which kinase-inhibitor combinations are likely to sample multiple protonation states of importance or if large population shifts are observed upon binding. For systems where MCCE simulations suggest significant involvement of protonation state effects in binding, we will investigate these further using detailed alchemical free energy calculations into which we implement a dynamic protonation state sampling scheme [77–79]. For those systems where detailed free energy calculations show kinase and ligand experience a net change in charge and for which kinase catalytic domains can be expressed in bacteria, we will use isothermal titration calorimetry (ITC) experiments [80, 81] to experimentally verify and quantify the contributions of protonation state effects to binding.

Both MCCE and the dynamic protonation alchemical free energy calculations require that we know the pK_a of titratable sites in our small molecules. While a number of tools exist to predict these, their accuracy for kinase inhibitors is uncertain. We will first benchmark these tools using experimental kinase inhibitor pK_a data for FDA-approved kinase inhibitors. Once a suitable tool has been identified, we will use it to generate reference pK_as for the constant-pH free energy calculations, combining the two into a useful framework for calculating the free energy of binding.

Subaim 2.1: Benchmark small molecule pK_a prediction tools against experimental data for kinase inhibitors.

The dynamic protonation state simulations can only capture relevant protonation if we have accurate pK_a data for the inhibitors. To identify reliable tools, we will benchmark existing pK_a prediction tools using pK_a data obtained for a series of kinase inhibitors using a Sirius T3 instrument. The Sirius T3 instrument provides a gold standard for giving macroscopic pK_as using a combination of electrochemical and UV-metric titrations [82]. This will help us select a tool that can adequately predict pK_a values for kinase inhibitors with similar chemical components. For an efficient framework, we would like to incorporate a reliable computational pK_a prediction tool, so that we can estimate the pK_a and simulate a system using a single, straightforward procedure. Among the small molecule pK_a prediction tools considered will be **MoKa** [83], **Jaguar** [84], and **Epik** [41, 42]. MoKa generates pK_as based on atomistic descriptors, defined by the surrounding atoms. The descriptors are based on molecular interaction fields calculated using GRID [85] for a library of 3D fragments, but can successfully be applied on 2D structures. Schrodinger's Jaguar provides means of estimating pK_a values using quantum mechanical methods. Epik uses Hammett Taft linear free energy approaches [86] for predicting pK_a values.

Subaim 2.2: Survey the kinase:inhibitor cocrystal structures for possible protonation state effects in inhibitor binding.

We will systematically explore the effects of fixed and dynamic protonation states on kinase, inhibitor, and kinase:inhibitor complexes to determine the impact of assuming protonation states are fixed or dynamic, and to dissect the effect in terms of protein-dominated, ligand-dominated, or coupled protein-ligand effects. Preliminary data indicates that a number of FDA approved kinase inhibitors have pK_a values close to the physiological pH , including imatinib [33], (Figure 2a). We aim to quantify these contributions, in order to find out how important they are for correctly estimating the binding affinity of small molecule kinase inhibitors. We will first survey kinase:inhibitor crystal structures in the PDB interactions using MCCE [76], which will be extended to incorporate small molecules in its calculations using pK_a prediction tools to determine the appropriate populations of small molecule protomeric states in solution. MCCE calculations of kinase, inhibitor, and complex will provide populations of protonations states at intracellular pH (approximately 7.4), sampled using Monte Carlo with a restricted protein backbone but including rotamer sampling. We will then select those systems that exhibit large populations of multiple protonation states or significant population shifts of binding, as all of these effects could contribute to significant error in computed binding affinities when protonation states are assumed to be static. On these systems, we will quantify the effect of protonation state changes on the binding free energy and further dissect its origin.

Subaim 2.3: Dissect the determinants and impact of protonation state effects on binding affinity through free energy calculations and ITC experiments.

We expect that protonation state effects contribute significantly to binding affinity of the kinase-inhibitor combinations identified in the previous aim. To quantify the magnitude of these effects—as well as determine their origin—we will use alchemical free energy calculations, a reliable standard for calculating binding affinities [14–16], in which we additional implement a dynamic treatment of protonation states for a constant external pH [77–79]. By allowing the dynamic protonation states for kinase, inhibitor, both, and neither, we will be able to both quantify the magnitude of dynamic protonation state effects and determine the primary origin of these effects. The calculations will be performed using OpenMM [87] and Yank [88]. The dynamic protonation state sampling scheme, also known as constant-pH , allows the distribution of protons in the system to

change, at a constant chemical potential of protons μ_{H^+},

$$\text{pH} = -\log_{10}\left(a_{\text{H}^+}\right) = -\log_{10}\left(\exp\left(\frac{\mu_{\text{H}^+} - \mu_{\text{H}^+}^\ominus}{RT}\right)\right) \approx -\log_{10}\left([\text{H}^+]\right) \quad, \tag{3}$$

where a_{H^+} is the activity coefficient of the protons, the superscript $\ominus$ indicates a thermodynamic standard state, R is the molar gas constant and T is the absolute temperature. Protonation and deprotonation events are sampled using a Monte Carlo (MC) method, using the free energy of proton transfer, calculated from the electrostatic interactions, as an acceptance criterion. The free energy of proton transfer depends on the reference pK_a value, as obtained from our pK_a survey, which is why we need an accurate estimate for these reference values to quantify contributions from our simulations.

For systems in which the kinase can be expressed in bacteria and the alchemical free energy calculations with dynamic protonation states indicate there will be a net change in charge for kinase and inhibitor upon binding, we will use ITC experiments that are able to dissect out the binding affinities of different protonation species. Changes in protonation states will lead to an enthalpic contribution/penalty to the binding affinity. By performing ITC experiments at the same pH using buffers with different ionization enthalpies, one is able to dissect the contribution of each protonation state to the binding affinity [80, 81]. To analyze these experiments, we will extend our Bayesian ITC methods to model the different experiments together to provide correlated uncertainty estimates.

Expected results

Our initial survey with MCCE will allow us to estimate the prevalence of protonation state effects on binding of small molecule kinase inhibitors. We can then quantify the magnitude by which protonation state effects contribute to the binding affinity. Constant-pH alchemical binding free energy calculations should provide us with accurate estimates, as long as we can accurately predict small molecule pK_as . At the same time, we can validate these results by performing ITC experiments of these kinases-inhibitor combinations. This knowledge can be used to determine whether these effects play a significant role in the selectivity of kinase inhibitors for their targets. If the contribution appears to make no significant difference, we will have ruled out the need to use more expensive theoretical methods and more experimental resources. We will also have developed a framework that will enable future evaluation of these effects, in systems for which experiments have not been carried out to determine the presence of these effects.

Pitfalls and alternatives

In the case that our computational estimates of the protonation state contributions differ from our experimental estimates, this could implicate other effects have not been incorporated into our simulation. For instance, the conformational states of the protein considered for our free energy calculations might need to be expanded. We could use techniques such as Markov state models (MSM) [89] to determine the important metastable states of kinase systems and use this knowledge to enhance our free energy calculations. There could also be a problem with the predicted pK_a values. In case we cannot obtain reasonable pK_a estimates, we may attempt to use experimental pK_a data directly in our simulations.

In the case that there is disagreement with experiment because of force field errors, we can perform experiments and simulations on systems where protonation state changes are not expected to occur, in order to evaluate the magnitude of this error. We may also find that the charge corrections illuminated in Aim 1 are required for accuracy of dynamic protonation state calculations, so we could evaluate this if the results seem to require it. Additional verification using host-guest complexes where the host or guest experience protonation state effects can be employed as needed.

Aim 3. Develop a framework for alchemical free energy calculations to describe weak association and cooperative ligand binding.

Rationale

Weak binding and association of multiple ligands to proteins are ubiquitous interactions in biological and pharmaceutically relevant systems. It is particularly common when using drug discovery approaches such as fragment-based ligand design, where concentrations are scaled up to detect weak binding signals, often resulting in multiple associating fragments. Most available free energy calculation frameworks focus on 1:1 binding interactions, and are not suited for calculating free energies of an arbitrary number of ligands. **We will overcome deficiencies in current frameworks with a new framework that can provide predictions for ligands binding multiple times to a single macromolecule.** Our proposed framework will create an alchemical ladder from $0 \ldots N_{\max}$ ligands, and can provide a description of binding by using a semi-grand canonical ensemble methodology [90–92]. We can establish a quantitative connection between these semi-grand canonical alchemical free energy calculations and ITC experiments on these systems. The parameters obtained from our alchemical free energy calculations will allow us to simulate ITC experiments directly. Consequently, we can use ITC experiments as

a means to validate the alchemical binding free energy predictions. To do so, our Bayesian framework for ITC experiments will be extended to include models of multiple binding sites.

As a model system, we will study the binding of several nonsteroidal anti-inflammatory drugs (NSAIDs) known to bind the pharmacologically relevant protein human serum albumin (HSA) (Figure 5) [12]. It is a protein with multiple binding sites that bind a wide variety of ligands [46–48]. Because of its ability to bind many small molecules, it is of great importance to pharmacokinetics [93]. Its high solubility makes it easy to reach the concentrations required for ITC experiments [44]. Many of the compounds that bind to HSA are commercially available [12]. The NSAIDs selected by us for this study (Figure 5b) are commercially available in hydrate or soluble salt forms, making them suitable for ITC experiments as well.
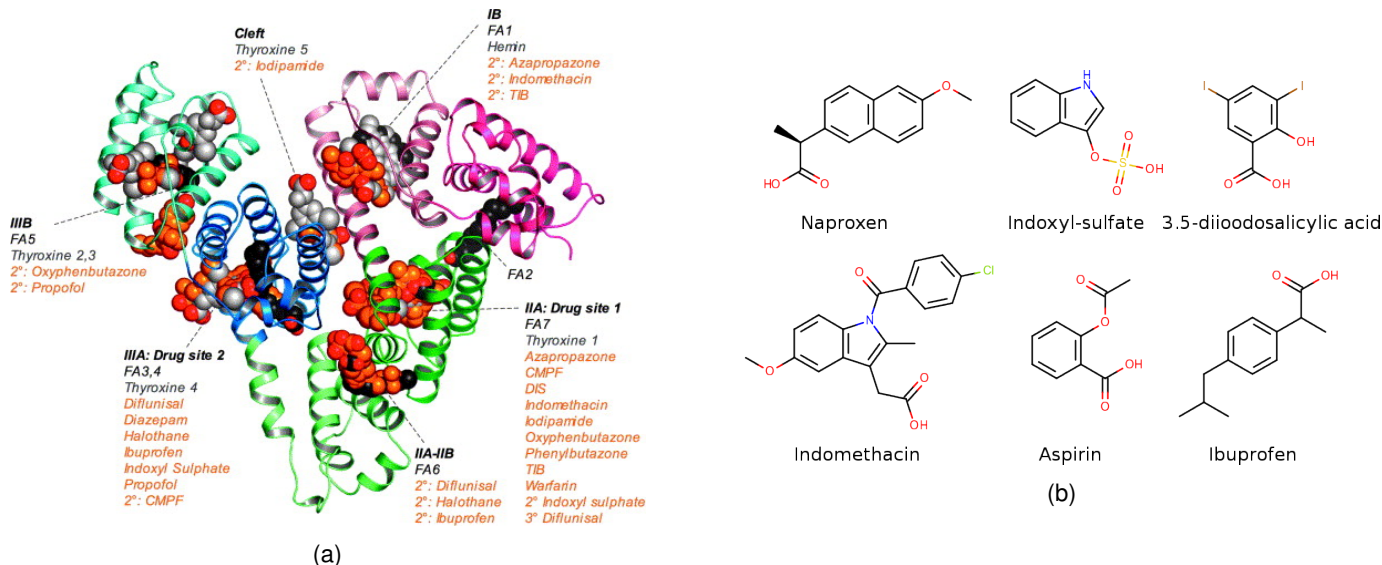
Figure 5: (a) **Summary of the ligand-binding capability of Human serum albumin structural studies.** Ligands are depicted in space-filling representation; oxygen atoms are colored red; all other atoms in fatty acids (myristic acid), other endogenous ligands (hemin, thyroxin) and drugs are coloured dark-gray, light gray and orange, respectively. Because of the many small molecule interactions of HSA, it is expected to be of great importance to the bioavailability of small molecules [93]. Its binding sites are asymmetric [48, 49] and one expects there to be a difference in binding affinity of a ligand for each site [94]. Figure 5a depicts the various binding sides of HSA, and shows how diverse the structure of each binding site is. It is known to bind several drugs [12, 45, 47, 50, 95, 96], as well as dietary supplements [51]. It is also essential for the bioavailability of endogenous small molecules such as hormones [97], and toxins such as cardiac glycosides [43]. From Figure 7 of [50]. (b) A series of commercially available NSAIDS that bind to human serum albumin, which will be used for this study [12].

Subaim 3.1: Extend alchemical free energy calculations to simulate multiple ligand binding.

Since current alchemical free energy calculation frameworks focus mostly on 1:1 association, we will need to extend available frameworks to incorporate multiple ligand binding. We will build upon the existing Yank framework [88] for alchemical free energy calculations to include the potential for multiple (possibly weak) binding events using a semi-grand canonical ensemble formalism. Our extension allows the calculation of the free energy of the different states of a protein-multiligand complex by using an alchemical ladder that allows for binding of of 0, 1, ..., $N_{\max}$ ligands. As a first application of the framework we will study the binding of NSAIDs (Figure 5b) to HSA. The compounds we have selected are known to bind to multiple sites on HSA [12], and are soluble enough to perform ITC experiments for validation of the simulation results.

To calculate the free energy of binding for multiple ligands, we employ semi-grand canonical ensemble theory. In the semi-grand canonical ensemble (constant μ, N_R, p, T), the number of ligand molecules is allowed to vary between 0 and N_{max}, and the number of receptor (R) molecules N_R is fixed. The pressure, p, and temperature, T, are also fixed. The free energy of observing n ligand molecules is,

$$g_n = -\ln \int \mathrm{d}x_1 \ldots \mathrm{d}x_n \, \mathrm{d}x_R \, e^{-u_n(x_1,\ldots,x_n,x_R)} \quad , \tag{4}$$

where $u_n(x_1,\ldots,x_n,x_R) \equiv \beta U_n(x_1,\ldots,x_n,x_R) + \beta p V(x)$ is the *reduced* potential energy, based on the coordinates of n ligand molecules and the receptor, β is the inverse temperature, defined as $1/k_B T$, and $V(x)$ is the instantaneous volume of the simulation cell. Similarly, we can also derive the enthalpy associated with binding of n ligands, h_n.

By knowing the free energies of each occupation state, we can estimate the bound concentrations for a given concentration of receptor [R] and ligand [L] from simulation. The term in the logarithm can be substituted by what is known as the

binding polynomial, Q

$$Q([L]) = 1 + K_1[L] + K_1 K_2 [L]^2 + \cdots + \left(\prod_{i=1}^{n} K_i\right)[L]^n = 1 + \sum_{l=1}^{n}\left(\prod_{i=1}^{l} K_l\right)[L]^l \quad , \tag{5}$$

which is also equivalent to the partition function of the system [55]. In this equation, K_n are the stoichiometric association constants, not site constants (see also Figure 6) meaning we do not define a specific site. They are obtained by taking the exponent of the free energy difference between binding of consecutive numbers of ligands

$$K_{n+1} = \frac{[\mathrm{RL}_{n+1}]}{[\mathrm{RL}_n][\mathrm{L}]} = e^{-(g_{n+1}-g_n)} \quad ; g_{n+1} - g_n = -\ln(K_{n+1}) \tag{6}$$

The form of the equation gives us an additional benefit of modeling weak association, where we can evaluate a free energy for a number of ligands associating, even without a strongly defined binding site.

From the estimated enthalpy and association constants we can simulate an ITC experiment. The heat of injection q_m is a function of the enthalpy change Δh_m, cell volume V_0 and the change in a complex concentrations $[\mathrm{PL_n}]$ of one receptor bound to n ligands

$$q_m = \sum_{n=0}^{N_{\max}} q_m^{(n)} = \sum_{n=0}^{N_{\max}} V_0 \Delta h_m^{(n)}\left([\mathrm{PL}_n]_m - [\mathrm{PL}_n]_{m-1}\right) \quad . \tag{7}$$

For a given receptor and ligand concentration, all of these properties are obtainable from free energy calculations. This allows us to establish a quantitative connection with ITC experiments.
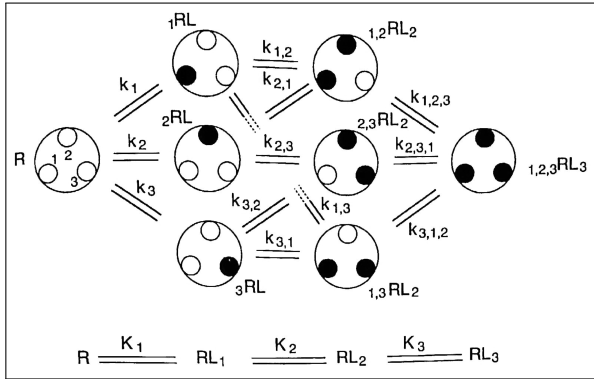
Figure 6: **In the semi-grand canonical ensemble, we consider the association of 0 up to $N_{\max}$ ligands to a protein.** By using alchemical perturbations, they can be inserted without pushing the system away from equilibrium. Rather than using site specific constants (k), for which specific binding sites need to be defined, we calculate the free energy of associating additional ligands to the protein by using binding polynomials of the stoichiometric binding constants K_i (Equation (5)). This will also allow us to calculate a free energy for weak associating ligands. Figure adapted from Klotz [98].

Subaim 3.2: Validate computational predictions by applying Bayesian model selection on ITC experiments of HSA and a series of NSAIDs.

In order to validate the results obtained from free energy calculations, we will perform complementary ITC experiments on human serum albumin (HSA). To design the ITC experiments, we can use the parameters obtained from the free energy calculations to simulate an ITC experiment. This will allow us to find an experimental setup that can provide us with maximal information on the different binding constants. We will be extending our Bayesian ITC framework, so that we can perform an adequate analysis. Current ITC analysis tools are too limited in terms of their model descriptions, nor do they provide accurate uncertainty estimates due to lack of error propagation. Tools like Origin [99] and NanoAnalyze [100], conventional ITC analysis software packages, uses cooperativity parameters n_i to describe multiple binding sites up to a maximum of two. This is too primitive to actually describe multiple binding events [98]. Furthermore, errors in the concentrations of the reagent solutions are not propagated, directly influencing the value of the estimated stoichiometry. More advanced tools such as SEDPHAT additionally uses terms describing "incompetent fractions" of experimental components [101, 102] and can deal with up to 3 sites. Unlike a Bayesian analysis, it does not propagate errors from the experimental procedure and just provides a "fudge factor" to improve the quality of the fit. We will extend our Bayesian analysis tools with the capacity to use binding polynomials (Equation (5)) to estimate the binding affinity and provide accurate uncertainty estimates. It will be able to propagate errors in the reagent concentrations of the prepared solutions, unlike the non-Bayesian models. Additionally, the binding polynomial will not be limited to necessarily 2, or 3 sites. It can be extended to incorporate as many ligands as needed. The number of parameters necessary can be selected by using Bayesian model selection techniques such as reversible jump Markov chain Monte Carlo (RJMCMC) [103].

Expected results

The newly developed free energy calculation framework will allow us to estimate the free energy of an arbitrary number of ligands binding to a protein. From these free energies, we can estimate stoichiometric binding constants, which will

allow us to calculate complex concentrations for a given receptor and ligand concentration. Our method will be applied on HSA, a pharmacologically relevant target, to predict the binding affinity of a series of NSAIDs. The information obtained information can then be used to simulate ITC experiments of multiple ligands binding to HSA at the same time, binding to multiple sites. If these estimates are correct, this will enable the design of experiments that provide more information on binding by reducing the uncertainty. We can also use our Bayesian ITC analysis methodology to experimentally measure binding affinities to HSA as a means of validation of the calculations. The experiments will yield credible intervals for each of the terms in the binding polynomial, with increasing uncertainty for the higher order terms. The number of terms that we can fit will be limited by this uncertainty, and we can use Bayesian model selection to estimate the number of parameters to fit. Using our experimental results, we can confirm whether our computational predictions fall within the experimental uncertainty.

Pitfalls and alternatives

It is possible that our simulations will not sample many different stoichiometries, if it is unfavorable to add more ligands. If that is the case, we will explore methods that drive the simulation to sample more states of the system. Otherwise, our estimates of the free energy of binding will be poorly converged. Higher order binding terms in the polynomial will have increased uncertainty, which will limit our ability to estimate binding free energies of accurately. In order to improve the information gain, we can explore various ways of performing Bayesian experimental design. This should allow us to decrease the uncertainty in the estimates. We can also explore the use of alternative model systems, such as host guest systems. Some hosts with larger cavities can bind multiple ligands, with their tractability making them attractive model systems. Also, some hosts themselves can associate with higher stoichiometry to a solute [104].

Conclusion

The aims of this proposal contribute to the limited set of tools available to deal with protein-ligand interactions in special, but not rare, cases where conventional tools provide no implementation of existing theoretical frameworks. Using Bayesian methodology will provide reliable uncertainty estimates on experimental data, which in turn will enable the distinction between theoretical model errors and experimental errors.

References

[1]   G. J. Rocklin et al. "Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site". In: J Mol Biol 425.22 (Nov. 2013), pp. 4569–4583. DOI: 10.1016/j.jmb.2013.07.030.

[2]   H. S. Muddana et al. "The SAMPL4 hostguest blind prediction challenge: an overview". In: J Comput -Aided Mol Des 28.4 (Mar. 2014), pp. 305–317. DOI: 10.1007/s10822-014-9735-1.

[3]   M. M. Reif and C. Oostenbrink. "Net charge changes in the calculation of relative ligand-binding free energies via classical atomistic molecular dynamics simulation". In: J Comput Chem 35.3 (Nov. 2013), pp. 227–243. DOI: 10.1002/jcc.23490.

[4]   G. J. Rocklin et al. "Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects". In: J Chem Phys 139.18 (2013), p. 184103. DOI: 10.1063/1.4826261.

[5]   Y.-L. Lin et al. "An Overview of Electrostatic Free Energy Computations for Solutions and Proteins". In: J Chem Theory Comput 10.7 (July 2014), pp. 2690–2709. DOI: 10.1021/ct500195p.

[6]   F. Dullweber et al. "Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition." eng. In: J Mol Biol 313.3 (Oct. 2001), pp. 593–614. DOI: 10.1006/jmbi.2001.5062.

[7]   A. Aleksandrov et al. "Protonation Patterns in Tetracycline:Tet Repressor Recognition: Simulations and Experiments". In: Chembiochem 8.6 (Apr. 2007), pp. 675–685. DOI: 10.1002/cbic.200600535.

[8]   P. Czodrowski, C. A. Sotriffer, and G. Klebe. "Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pKa Calculations and ITC Experiments". In: J Mol Biol 367.5 (Apr. 2007), pp. 1347–1356. DOI: 10.1016/j.jmb.2007.01.022.

[9]   H. Steuber et al. "Tracing Changes in Protonation: A Prerequisite to Factorize Thermodynamic Data of Inhibitor Binding to Aldose Reductase". In: J Mol Biol 373.5 (Nov. 2007), pp. 1305–1320. DOI: 10.1016/j.jmb.2007.08.063.

[10]   P. Czodrowski, C. A. Sotriffer, and G. Klebe. "Atypical Protonation States in the Active Site of HIV-1 Protease: A Computational Study". In: J Chem Inf Model 47.4 (July 2007), pp. 1590–1598. DOI: 10.1021/ci600522c.

[11]   P. J. Hajduk and J. Greer. "A decade of fragment-based drug design: strategic advances and lessons learned". In: Nat Rev Drug Discovery 6.3 (Feb. 2007), pp. 211–219. DOI: 10.1038/nrd2220.

[12]   F. Zsila et al. "Evaluation of drug-human serum albumin binding interactions with support vector machine aided online automated docking". In: Bioinformatics 27.13 (May 2011), pp. 1806–1813. DOI: 10.1093/bioinformatics/btr284.

[13]   M. R. Shirts, D. L. Mobley, and J. D. Chodera. "Alchemical free energy calculations: ready for prime time". In: Annu Rep Comput Chem 3 (2007), pp. 41–59.

[14]   H. Fujitani et al. "Direct calculation of the binding free energies of FKBP ligands". In: J Chem Phys 123.8 (2005), p. 084108. DOI: 10.1063/1.1999637.

[15]   Y. Deng and B. Roux. "Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant". In: J Chem Theory Comput 2.5 (Sept. 2006), pp. 1255–1273. DOI: 10.1021/ct060037v.

[16]   L. Wang et al. "Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field". In: J Am Chem Soc 137.7 (Feb. 2015), pp. 2695–2703. DOI: 10.1021/ja512751q.

[17]   J. D. Chodera et al. "Alchemical free energy methods for drug discovery: progress and challenges". In: Curr Opin Struct Biol 21.2 (2011), pp. 150–160. DOI: http://dx.doi.org/10.1016/j.sbi.2011.01.011.

[18]   N. Hansen and W. F. van Gunsteren. "Practical Aspects of Free-Energy Calculations: A Review". In: J Chem Theory Comput 10.7 (July 2014), pp. 2632–2647. DOI: 10.1021/ct500161f.

[19] V. Gapsys et al. "Calculation of Binding Free Energies". English. In: Molecular Modeling of Proteins. Ed. by A. Kukol. Vol. 1215. Methods in Molecular Biology. Springer New York, 2015, pp. 173–209. DOI: 10.1007/978-1-4939-1465-4_9.

[20] M. Gilson et al. "The statistical-thermodynamic basis for computation of binding affinities: a critical review". In: Biophys J 72.3 (Mar. 1997), pp. 1047–1069. DOI: 10.1016/s0006-3495(97)78756-3.

[21] U. Essmann et al. "A smooth particle mesh Ewald method". In: J Chem Phys 103.19 (1995), p. 8577. DOI: 10.1063/1.470117.

[22] I. G. Tironi et al. "A generalized reaction field method for molecular dynamics simulations". In: J Chem Phys 102.13 (1995), p. 5451. DOI: 10.1063/1.469273.

[23] A. D. Levinson et al. "Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein." eng. In: Cell 15.2 (Oct. 1978), pp. 561–572.

[24] I. Vivanco and C. L. Sawyers. "The phosphatidylinositol 3-kinase–AKT pathway in human cancer". In: Nat Rev Cancer 2.7 (2002), pp. 489–501.

[25] J. Baselga. "Targeting tyrosine kinases in cancer: the second wave." eng. In: Science 312.5777 (May 2006), pp. 1175–1178. DOI: 10.1126/science.1125951.

[26] K. Garber. "The second wave in kinase cancer drugs." eng. In: Nat Biotechnol 24.2 (Feb. 2006), pp. 127–130. DOI: 10.1038/nbt0206-127.

[27] T. O'Hare et al. "Targeting the BCR-ABL signaling pathway in therapy-resistant Philadelphia chromosome-positive leukemia." eng. In: Clin Cancer Res 17.2 (Jan. 2011), pp. 212–221. DOI: 10.1158/1078-0432.CCR-09-3314.

[28] N. P. Shah. "Overriding Imatinib Resistance with a Novel ABL Kinase Inhibitor". In: Science 305.5682 (July 2004), pp. 399–401. DOI: 10.1126/science.1099480.

[29] W. Pao et al. "Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain". In: PLoS Med 2.3 (2005), e73.

[30] J. Zhang, P. L. Yang, and N. S. Gray. "Targeting cancer with small molecule kinase inhibitors". In: Nat Rev Cancer 9.1 (2009), pp. 28–39.

[31] B. J. Druker et al. "Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia". In: N Engl J Med 344.14 (2001). PMID: 11287972, pp. 1031–1037. DOI: 10.1056/NEJM200104053441401. eprint: http://dx.doi.org/10.1056/NEJM200104053441401.

[32] S. G. O'Brien et al. "Imatinib Compared with Interferon and Low-Dose Cytarabine for Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia". In: N Engl J Med 348.11 (2003). PMID: 12637609, pp. 994–1004. DOI: 10.1056/NEJMoa022457. eprint: http://dx.doi.org/10.1056/NEJMoa022457.

[33] Z. Szakcs et al. "AcidBase Profiling of Imatinib (Gleevec) and Its Fragments". In: J Med Chem 48.1 (Jan. 2005), pp. 249–255. DOI: 10.1021/jm049546c.

[34] M. A. Seeliger et al. "c-Src Binds to the Cancer Drug Imatinib with an Inactive Abl/c-Kit Conformation and a Distributed Thermodynamic Penalty". In: Structure 15.3 (Mar. 2007), pp. 299–311. DOI: 10.1016/j.str.2007.01.015.

[35] Y.-L. Lin et al. "Explaining why Gleevec is a specific and potent inhibitor of Abl kinase". In: Proc Natl Acad Sci U S A 110.5 (Jan. 2013), pp. 1664–1669. DOI: 10.1073/pnas.1214330110.

[36] T. Polgr and G. M. Keser. "Virtual Screening for -Secretase (BACE1) Inhibitors Reveals the Importance of Protonation States at Asp32 and Asp228". In: J Med Chem 48.11 (June 2005), pp. 3749–3755. DOI: 10.1021/jm049133b.

[37] K. Wittayanarakul, S. Hannongbua, and M. Feig. "Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors". In: J Comput Chem 29.5 (2008), pp. 673–685. DOI: 10.1002/jcc.20821.

[38] Z. A. Knight and K. M. Shokat. "Features of Selective Kinase Inhibitors". In: Chemistry & Biology 12.6 (June 2005), pp. 621–637. DOI: 10.1016/j.chembiol.2005.04.011.

[39] J. R. Griffiths. "Are cancer cells acidic?" eng. In: Br J Cancer 64.3 (Sept. 1991), pp. 425–427.

[40] M. Stubbs et al. "Causes and consequences of tumour acidity and implications for treatment". In: Mol Med Today 6.1 (Jan. 2000), pp. 15–19. DOI: 10.1016/s1357-4310(99)01615-9.

[41] J. C. Shelley et al. "Epik: a software program for pK a prediction and protonation state generation for drug-like molecules". In: J Comput -Aided Mol Des 21.12 (Sept. 2007), pp. 681–691. DOI: 10.1007/s10822-007-9133-z.

[42] J. R. Greenwood et al. "Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution". In: J Comput -Aided Mol Des 24.6-7 (Mar. 2010), pp. 591–604. DOI: 10.1007/s10822-010-9349-1.

[43] T. W. Smith. "Pharmacokinetics, bioavailability and serum levels of cardiac glycosides." eng. In: J Am Coll Cardiol 5.5 Suppl A (May 1985), 43A–50A.

[44] T. Peters Jr. All About Albumin: Biochemistry, Genetics, and Medical Applications. Academic Press, 1995.

[45] B. Bannwarth et al. "Clinical pharmacokinetics of low-dose pulse methotrexate in rheumatoid arthritis." eng. In: Clin Pharmacokinet 30.3 (Mar. 1996), pp. 194–210.

[46] U. Kragh-Hansen, V. T. G. Chuang, and M. Otagiri. "Practical Aspects of the Ligand-Binding and Enzymatic Properties of Human Serum Albumin". In: Biol Pharm Bull 25.6 (2002), pp. 695–704. DOI: 10.1248/bpb.25.695.

[47] A. Sukowska. "Interaction of drugs with bovine and human serum albumin". In: J Mol Struct 614.1-3 (Sept. 2002), pp. 227–232. DOI: 10.1016/s0022-2860(02)00256-9.

[48] X. M. He and D. C. Carter. "Atomic structure and chemistry of human serum albumin". In: Nature 358.6383 (July 1992), pp. 209–215. DOI: 10.1038/358209a0.

[49] S. Curry et al. "Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites". In: Nat Struct Biol 5.9 (Sept. 1998), pp. 827–835. DOI: 10.1038/1869.

[50] J. Ghuman et al. "Structural Basis of the Drug-binding Specificity of Human Serum Albumin". In: J Mol Biol 353.1 (Oct. 2005), pp. 38–52. DOI: 10.1016/j.jmb.2005.07.075.

[51] S. Pal and C. Saha. "A review on structureaffinity relationship of dietary flavonoids with serum albumins". In: J Biomol Struct Dyn 32.7 (July 2013), pp. 1132–1147. DOI: 10.1080/07391102.2013.811700.

[52] D. G. Myszka et al. "The ABRF-MIRG'02 study: assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction." eng. In: J Biomol Tech 14.4 (Dec. 2003), pp. 247–269.

[53] D. L. Vaux, F. Fidler, and G. Cumming. "Replicates and repeats–what is the difference and is it significant? A brief discussion of statistics and experimental design." eng. In: EMBO Rep 13.4 (Apr. 2012), pp. 291–296. DOI: 10.1038/embor.2012.36.

[54] M. W. Karaman et al. "A quantitative analysis of kinase inhibitor selectivity". In: Nat Biotechnol 26.1 (Jan. 2008), pp. 127–132. DOI: 10.1038/nbt1358.

[55] K. A. Dill and S. Bromberg. Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience, 2nd Edition. Garland Science, 2010.

[56] J. Lagona et al. "The Cucurbit[n]uril Family". In: Angew Chem , Int Ed 44.31 (Aug. 2005), pp. 4844–4870. DOI: 10.1002/anie.200460675.

[57] L. Cao and L. Isaacs. "Absolute and relative binding affinity of cucurbit[7]uril towards a series of cationic guests". In: Supramol Chem 26.3-4 (Nov. 2013), pp. 251–258. DOI: 10.1080/10610278.2013.852674.

[58] C. N. Pace et al. "Protein structure, stability and solubility in water and other solvents." eng. In: Philos Trans R Soc Lond B Biol Sci 359.1448 (Aug. 2004), pp. 1225–34, 1225–34. DOI: 10.1098/rstb.2004.1500.

[59] W. L. Mock and N. Y. Shih. "Dynamics of molecular recognition involving cucurbituril". In: J Am Chem Soc 111.7 (Mar. 1989), pp. 2697–2699. DOI: 10.1021/ja00189a053.

[60] H. Tang et al. "Guest Binding Dynamics with Cucurbit[7]uril in the Presence of Cations". In: J Am Chem Soc 133.50 (Dec. 2011), pp. 20623–20633. DOI: 10.1021/ja209266x.

[61] T. A. Halgren et al. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening". In: J Med Chem 47.7 (Mar. 2004), pp. 1750–1759. DOI: 10.1021/jm030644s.

[62] R. A. Friesner et al. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy". In: J Med Chem 47.7 (Mar. 2004), pp. 1739–1749. DOI: 10.1021/jm0306430.

[63] R. A. Friesner et al. "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for ProteinLigand Complexes". In: J Med Chem 49.21 (Oct. 2006), pp. 6177–6196. DOI: 10.1021/jm051256o.

[64] Glide, version 6.4, Small-Molecule Drug Discovery Suite 2014-3. Schrdinger, LLC, New York, NY, 2014.

[65] Z. Feng et al. "Ligand Depot: a data warehouse for ligands bound to macromolecules". In: Bioinformatics 20.13 (Apr. 2004), pp. 2153–2155. DOI: 10.1093/bioinformatics/bth214.

[66] J. Tellinghuisen and J. D. Chodera. "Systematic errors in isothermal titration calorimetry: Concentrations and baselines". In: Anal Biochem 414.2 (July 2011), pp. 297–299. DOI: 10.1016/j.ab.2011.03.024.

[67] N. Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: J Chem Phys 21.6 (1953), p. 1087. DOI: 10.1063/1.1699114.

[68] W. K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". English. In: Biometrika 57.1 (1970), pp. 97–109.

[69] G. van Rossum. Python tutorial. Report CS-R9526. Amsterdam: Centrum voor Wiskunde en Informatica (CWI), Apr. 1995, pp. iii + 65.

[70] A. S. Rustenburg and J. D. Chodera. ITCtools. https://github.com/choderalab/itctools. Schrdinger, LLC, New York, NY, 2015.

[71] J. Wang et al. "Development and testing of a general amber force field." eng. In: J Comput Chem 25.9 (July 2004), pp. 1157–1174. DOI: 10.1002/jcc.20035.

[72] K. Vanommeslaeghe et al. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields." eng. In: J Comput Chem 31.4 (Mar. 2010), pp. 671–690. DOI: 10.1002/jcc.21367.

[73] W. Yu et al. "Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations." eng. In: J Comput Chem 33.31 (Dec. 2012), pp. 2451–2468. DOI: 10.1002/jcc.23067.

[74] OPLS 2.1. Schrdinger, LLC, New York, NY, 2014.

[75] H. M. Berman. "The Protein Data Bank". In: Nucleic Acids Res 28.1 (Jan. 2000), pp. 235–242. DOI: 10.1093/nar/28.1.235.

[76] Y. Song, J. Mao, and M. R. Gunner. "MCCE2: Improving protein pKa calculations with extensive side chain rotamer sampling". In: J Comput Chem (2009). DOI: 10.1002/jcc.21222.

[77] J. Mongan, D. A. Case, and J. A. McCammon. "Constant pH molecular dynamics in generalized Born implicit solvent". In: J Comput Chem 25.16 (Dec. 2004), pp. 2038–2048. DOI: 10.1002/jcc.20139.

[78] H. A. Stern. "Molecular simulation with variable protonation states at constant pH". In: The Journal of Chemical Physics 126.16 (2007), p. 164112. DOI: 10.1063/1.2731781.

[79] J. P. Nilmeier et al. "Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation". In: Proc Natl Acad Sci U S A 108.45 (Oct. 2011), E1009–E1018. DOI: 10.1073/pnas.1106094108.

[80] B. M. Baker and K. P. Murphy. "Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry." eng. In: Biophys J 71.4 (Oct. 1996), pp. 2049–2055. DOI: 10.1016/S0006-3495(96)79403-1.

[81] M. Neeb et al. "Chasing Protons: How Isothermal Titration Calorimetry, Mutagenesis, and p K a Calculations Trace the Locus of Charge in Ligand Binding to a tRNA-Binding Enzyme". In: J Med Chem 57.13 (July 2014), pp. 5554–5565. DOI: 10.1021/jm500401x.

[82] D. Schnherr et al. "Characterisation of selected active agents regarding pKa values, solubility concentrations and pH profiles by SiriusT3". In: Eur J Pharm Biopharm 92 (May 2015), pp. 155–170. DOI: 10.1016/j.ejpb.2015.02.028.

[83] F. Milletti et al. "New and Original p K a Prediction Method Using Grid Molecular Interaction Fields". In: J Chem Inf Model 47.6 (Nov. 2007), pp. 2172–2181. DOI: 10.1021/ci700018y.

[84] A. D. Bochevarov et al. "Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences". In: Int J Quantum Chem 113.18 (July 2013), pp. 2110–2142. DOI: 10.1002/qua.24481.

[85] P. J. Goodford. "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules". In: J Med Chem 28.7 (July 1985), pp. 849–857. DOI: 10.1021/jm00145a002.

[86] D. D. Perrin, B. Dempsey, and E. P. Serjeant. pKa Prediction for Organic Acids and Bases. Springer Science + Business Media, 1981. DOI: 10.1007/978-94-009-5883-8.

[87] P. Eastman et al. "OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation". In: J Chem Theory Comput 9.1 (Jan. 2013), pp. 461–469. DOI: 10.1021/ct300857j.

[88] J. D. Chodera et al. "Yank: An open, extensible platform for GPU-accelerated alchemical binding free energy calculations using the OpenMM toolkit." Unpublished. 2015.

[89] J.-H. Prinz et al. "Markov models of molecular kinetics: Generation and validation". In: J Chem Phys 134.17 (2011), p. 174105. DOI: 10.1063/1.3565032.

[90] D. A. Kofke and E. D. Glandt. "Monte Carlo simulation of multicomponent equilibria in a semigrand canonical ensemble". In: Mol Phys 64.6 (Aug. 1988), pp. 1105–1131. DOI: 10.1080/00268978800100743.

[91] D. A. Kofke. "Semigrand Canonical Monte Carlo Simulation; Integration Along Coexistence Lines". In: Adv Chem Phys (Jan. 1999), pp. 405–441. DOI: 10.1002/9780470141649.ch13.

[92] G. C. Lynch and B. Pettitt. "Semi-grand canonical molecular dynamics simulation of bovine pancreatic trypsin inhibitor". In: Chem Phys 258.2-3 (Aug. 2000), pp. 405–413. DOI: 10.1016/s0301-0104(00)00159-2.

[93] P. D. Metcalfe and S. Thomas. "Challenges in the prediction and modeling of oral absorption and bioavailability." eng. In: Curr Opin Drug Discov Devel 13.1 (Jan. 2010), pp. 104–110.

[94] G. Sudlow, D. Birkett, and D. Wade. "Further characterization of specific drug binding sites on human serum albumin". In: Mol Pharmacol 12.6 (1976), pp. 1052–1061.

[95] I. Sjholm et al. "Binding of drugs to human serum albumin: XI. The specificity of three binding sites as studied with albumin immobilized in microparticles". In: Mol Pharmacol 16.3 (1979), pp. 767–777.

[96] E. A. Perez. "Novel enhanced delivery taxanes: an update." eng. In: Semin Oncol 34.3 (June 2007).

[97] W. M. Pardridge. "4 Serum bioavailability of sex steroid hormones". In: Clin Endocrinol Metab 15.2 (May 1986), pp. 259–278. DOI: 10.1016/s0300-595x(86)80024-x.

[98] I. M. Klotz. Ligand-Receptor Energetics: A Guide for the Perplexed. Wiley, 1997.

[99] MicroCal, LLC. ITC Data Analysis in Origin: Tutorial Guide. 2004.

[100] NanoAnalyze Data Analysis, Version 3.3.0. TA Instruments, 2014.

[101] J. C. D. Houtman et al. "Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling." eng. In: Protein Sci 16.1 (Jan. 2007), pp. 30–42. DOI: 10.1110/ps.062558507.

[102] H. Zhao, G. Piszczek, and P. Schuck. "SEDPHAT - A platform for global ITC analysis and global multi-method analysis of molecular interactions." eng. In: Methods 76 (Apr. 2015), pp. 137–148. DOI: 10.1016/j.ymeth.2014.11.012.

[103] P. J. Green. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination". In: Biometrika 82.4 (1995), pp. 711–732. DOI: 10.1093/biomet/82.4.711.

[104] D. W. Armstrong et al. "Efficient detection and evaluation of cyclodextrin multiple complex formation". In: J Am Chem Soc 108.7 (Apr. 1986), pp. 1418–1421. DOI: 10.1021/ja00267a010.