

SPECIFIC AIMS

All proteins and many drugs can access a range of protomer and tautomer states, which differ in the number or location of protons. When a small molecule binds a biomolecular target, both partners may shift their equilibrium populations of protomers and tautomers due to changes in electrostatic environment and solvation. Current computational approaches to binding affinity prediction—used to drive the discovery and optimization of small molecules for use as chemical probes or potential pharmacological agents—almost universally assume that a *single* dominant protonation state for both binding partners persists in both complex and solution. Ignoring these *protonation state effects*—the fact that real molecules populate an equilibrium mixture of protomer and tautomer states, and that these can shift upon binding—can lead to large errors in predicted binding affinities. Surprisingly, the magnitude and pervasiveness of protonation effects in pharmacologically relevant systems is largely unknown.

Recently, a series of blind community challenges led by the NIH-sponsored Drug Design Data Resource (D3R) and Statistical Assessment of Modeling of Protein Ligand interactions (SAMPL) **have exposed the lack of rigorous approaches for treating protonation state effects as a critical roadblock to achieving quantitative predictive accuracy**. While there is tremendous enthusiasm that advanced modeling methods—especially alchemical free energy calculations—can now achieve useful quantitative accuracy in some systems, the domain of applicability of these methodologies is highly constrained without accurate treatment of protonation state effects. Since we currently lack computational tools to study protonation state effects in protein:ligand binding, we do not yet know which systems are most sensitive to their neglect, nor the magnitude of the resulting error.

Here, we combine two technologies to overcome these deficiencies and build an advanced modeling platform that explicitly treats protonation state effects: First, we develop a new hybrid nonequilibrium molecular dynamics / Monte Carlo approach to explicit solvent constant-pH dynamics that permits a rigorous treatment of protomer/tautomer equilibria for both small molecules and proteins. Second, we incorporate this methodology into our open source GPU-accelerated alchemical free energy platform to allow protonation state effects to be included in binding affinity calculations. To validate and optimize this approach, we adopt a joint computational-experimental approach using small molecule inhibitors binding to human kinase domains as a model system. In addition to being an excellent model system for quantitative validation of the methodology, lessons learned about the pervasiveness, magnitude, and biophysical origin of protonation state effects in kinase inhibition will provide new insight into a popular target on which billions of dollars are spent annually on pharmaceutical development.

AIM 1. Survey kinase:inhibitor complexes for protonation state effects to develop novel model systems. While there is ample evidence to suggest protonation state effects are ubiquitous in the binding of small molecule kinase inhibitors, only one case (Abl:imatinib) has been investigated in detail. To identify additional candidate systems for further study, we will incorporate small molecule protomer sampling into MultiConformation Continuum Electrostatics (MCCE) from the Gunner lab and survey of over 2000 kinase:inhibitor complexes from the PDB for protonation state effects. This tool which assumes a rigid backbone but rapidly samples protomer equilibria using discrete rotamer/protomer libraries. As solution protomer free energies are the foundation of our protomer sampling approach, we will investigate the accuracy of small molecule solution pK_a prediction for a set of small molecule kinase inhibitors, benchmarking a variety of pK_a prediction tools against new experimental titration data.

AIM 2. Incorporate efficient protomer sampling into molecular dynamics and alchemical free energy calculations and dissect the physical origin of protonation state effects in kinase:inhibitor systems. First, we will implement an efficient Monte Carlo protomer sampling scheme into molecular dynamics simulations in explicit solvent—making use of highly efficient nonequilibrium candidate Monte Carlo (NCMC) techniques developed in the Chodera lab—and optimize this to achieve high acceptance rates in kinase:inhibitor systems. Second, we will incorporate this approach into alchemical free energy calculations, integrating it into our GPU-accelerated open source codebase. We will then use these tools to computationally probe the magnitude and nature of protonation state effects for systems identified in Aim 1, examining the impact of fixed or dynamic protonation states on computed binding affinities, and investigating which functionalities (protein residues or ligand moieties) are primary contributors to protonation state effects across related kinases and inhibitors.

AIM 3. Experimentally probe kinase:inhibitor systems to quantify protonation state effects on binding. We will study experimentally tractable kinase:inhibitor systems identified in Aim 1 as likely to possess significant protonation state effects, and use insights from Aim 2 to inform the selection of experimental methodology applied to each system. We focus on bacterially-expressed human kinase domains, where conditions can be matched between experiments and computation. We will conduct three types of experimental investigations: (1) isothermal titration calorimetry to detect proton shuttling to solvent upon binding; (2) pH-dependent fluorescence binding assays to investigate pH sensitivity (relevant to acidified intracellular environments in cancer); (3) and the sensitivity of binding affinities to clinically observed mutations that may induce drug resistance via pK_a shifts.

SIGNIFICANCE

Computational techniques hold the potential to accelerate drug discovery, but can be hindered by neglect of protonation state effects. Computational modeling already plays an active role in current drug discovery programs, but neglecting protonation state effects can limit its effectiveness [28]. Recent blind predictive modeling challenges hosted by the Drug Design Data Resource (D3R) [29] and Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) [30, 31] have highlighted three major hurdles to achieving useful predictions on a wide variety of systems: Forcefield accuracy, adequate sampling of thermodynamically relevant conformations, and accurate modeling of chemical effects such as protomer and tautomer equilibria. This last issue forms the basis of this proposal. Progress requires improved methods for modeling how both inhibitor and protein change protomer and tautomer equilibria upon binding and how this change in equilibrium impacts overall binding affinity. Critically, the large number of accessible protonation states and tight coupling of many titratable sites—especially within the ligand, where titratable sites are often coupled through electronic effects—makes this problem challenging.

Current standard physical modeling approaches (and most drug discovery efforts) assume a single protomer dominates in solution and complex. In standard molecular dynamics or Monte Carlo simulations, both binding partners are locked in fixed protonation and tautomeric states (here referred to generically as a *protomer*) throughout the modeling process. Modelers must select protomers for both protein ionizable residues and ligand, be this the dominant solution protomer or a presumed dominant protomer in complex (based on local interactions). Thus, computed binding free energies and sampled conformations represent the affinity and interactions of a single subspecies which may be only a minor fraction of the true equilibrium protomer distribution. Several scenarios can lead to errors up to several kcal/mol in magnitude in computed binding affinities [32–35]: (1) ligand or receptor populate a mixture of protomers in solution or complex and therefore cannot be accurately modeled by a single protomer; (2) a single protomer is dominant in both solution and complex, but differs in free and bound states; and (3) a single protomer persists in solution and complex, but the incorrect protomer is selected by the modeler.

Neglect of correct protomer distribution leads to large errors in transfer free energies. While it is difficult to experimentally isolate protomer effects from forcefield and sampling challenges in experimental binding reactions, transfer free energies between aqueous and nonaqueous environments provide insight into the magnitude of protomer effects absent conformational sampling issues. For the recent SAMPL5 challenge, the Chodera lab measured cyclohexane-water distribution coefficients ($\log D$) of 53 druglike molecules at pH 7.4 [2], and multiple laboratories participated in blind prediction exercises using a variety of methods to account for protonation state effects. One group found that simple corrections for aqueous protomers resulted in a $1.1 \log_{10}$ unit (1.5 kcal/mol) improvement in average error [30], while a sophisticated QM approach incorporating nonaqueous protomer effects found corrections as large as 10 kcal/mol, reducing $\log D$ RMS errors by $2.5 \log_{10}$ units (3.5 kcal/mol) [36].

Limited work in model systems highlights to the importance of protomer/tautomer equilibria in binding. A large-scale computational survey of protein-ligand structures in the PDB using a continuum solvent methodology suggested that 60% of all protein-small molecule complexes undergo ionization state changes upon binding [37]. Despite this, there has been surprisingly little detailed experimental investigation into these protonation state effects. Joint experimental-computational studies by collaborator Czodrowski have conclusively demonstrated how protonation state effects can play a major role in protein:ligand association. Using an ITC technique to separate enthalpic changes due to proton exchange with buffer from enthalpic changes upon binding [38], they showed large shifts in protonation state from both protein and ligand in the binding of inhibitors by aldose reductase [39] or trypsin and thrombin [40–42], where pK_a shifts as large as $4.9 \log_{10}$ units (6.8 kcal/mol) were observed.

Recent work has suggested the importance of protonation state effects in kinase:inhibitor binding. Selective kinase inhibitors, which specifically target one or more kinases or their mutants, have already proven their enormous potential as powerful therapeutics in the treatment of kinase dysregulation diseases, which include many cancers. Imatinib (**Figure 1A**) specifically targets the Abl kinase dysregulated in chronic myelogenous leukemia (CML) patients to abate disease progression [43, 44]. At typical intracellular pH of 7.4, unbound imatinib populates a mixture of protomers, predominantly a neutral form ($0 k_B T$) and a higher free energy protonated form where the piperazine ring is protonated ($\sim 0.6 k_B T$) [45] (**Figure 1A**). The complex (which has an overall affinity of $17 k_B T$) prefers the charged form of imatinib, incurring an unfavorable energetic cost to populate the protonated form of imatinib [34, 45, 46]. Abl also populates a multiple protomers: Asp381 in the conserved DFG (Asp-Phe-Gly) loop must swing out to permit imatinib binding, and its protonation state is coupled to DFG loop conformation [26]. Both the cocrystal structure (pdbrid:2HYY) and free energy calculations suggest the dominant bound species is the protonated form of imatinib bound to the deprotonated Asp381, while in solution it is the deprotonated imatinib and mixture of Asp conformations and protonation states [32, 34], though exact protomer populations are still in debate [26, 33, 34, 46].

Most FDA-approved kinase inhibitors can access multiple protonation states at intracellular pH. Kinase inhibitors can bind to their targets with picomolar affinities ($\Delta G_{\text{bind}} = 28 k_B T$ or 16.5 kcal/mol). High-affinity inhibitors can easily tolerate affinity losses of 500-fold ($6 k_B T$, or 3.6 kcal/mol, or 2.6 pK_a units) while maintaining sub-nanomolar potency. As a result, kinase inhibitors could easily access protonation states up to 2.6 pK_a units (or more) away from intracellular pH, allowing groups with pK_a s in the range to 4.8–10 to modulate binding specificity and affinity (assuming intracellular pH of 7.4). A preliminary survey of Epik-computed pK_a s for 22 FDA-approved kinase inhibitors (Figure 2) shows *more than half* have at least one predicted pK_a in this range, suggesting multiple inhibitor protomers are likely important for numerous kinase inhibitors of clinical interest.

Modeling protomer equilibria in proteins is frustrated by the large number of possible protomer states. On average, ionizable residues comprise approximately 29% of proteins [47]. The number of potential protein protomers is combinatorially large, making the accurate modeling of protomer equilibria challenging. As a result, a number of approaches have relied on Monte Carlo methods to overcome the curse of dimensionality. Using reference pK_a data for amino acids in isolation, these methods incorporate electrostatic and solvation effects in a manner that predicts shifts the effective pK_a in the sampled protomer ensemble. Co-I Gunner developed the Monte Carlo continuum electrostatics (MCCE) approach [16–20], which uses a Poisson-Boltzmann solver in conjunction with sidechain rotamer, protomer, and tautomer libraries. While MCCE assumes a rigid backbone, newer Monte Carlo approaches have been introduced into implicit solvent molecular dynamics simulations to sample protomer equilibria in a fully flexible model [48, 49]. While applying standard instantaneous Monte Carlo to explicit solvent simulations results in vanishingly small acceptance rates, nonequilibrium protocols—which achieve increased efficiency by breaking the proposal into many small steps and substituting the nonequilibrium work for the energy in the acceptance criteria [9]—have been shown to achieve useful acceptance rates [50, 51]. An alternative approach propagates fractional protonation states as dynamical degrees of freedom for increased efficiency, but suffers from the inequivalence of equilibrium averages of fractional and discrete protonation states [52–55].

Rigorously incorporating protomer equilibria for both small molecules and biomolecular targets into free energy calculations has proven difficult. It is significantly more difficult to model protomer equilibria of small molecules. While reference pK_a data is readily available for naturally-occurring ionizable protein residues, titratable sites within small molecules are highly varied and may be have their pK_a s shifted significantly by electronic coupling effects to nearby groups, which may also be titratable. As a result, it may be inappropriate to assume the titratable sites are independent and interact only via electrostatic and solvation forces. Simple approaches to correcting results using relative protomer free energies of solvated ligands have been reported to enhance the performance of docking and virtual screening (e.g. [56]), but this scheme does not account for protein protonation state effects and is not readily scalable to more titratable species. Incorporating a titratable model into alchemical free energy calculations—where the ligand is alchemically decoupled from its environment in a series of stages to estimate the binding free energy [7, 57]—presents additional difficulties. One proposal—the enveloping distribution sampling (EDS) method of Brooks [58]—presents a clever approach in which a modified potential is used at the titratable site of a ligand to mimic both protonated and deprotonated states, with the physical system recovered by reweighting analysis, but becomes highly inefficient for more than a few titratable sites. By contrast, our approach—which makes use of efficient nonequilibrium Monte Carlo proposals within the context of alchemical free energy calculations—presents the first fully general and scalable method for treating both small molecule and protein protomer equilibrium in molecular dynamics simulations and alchemical free energy calculations.

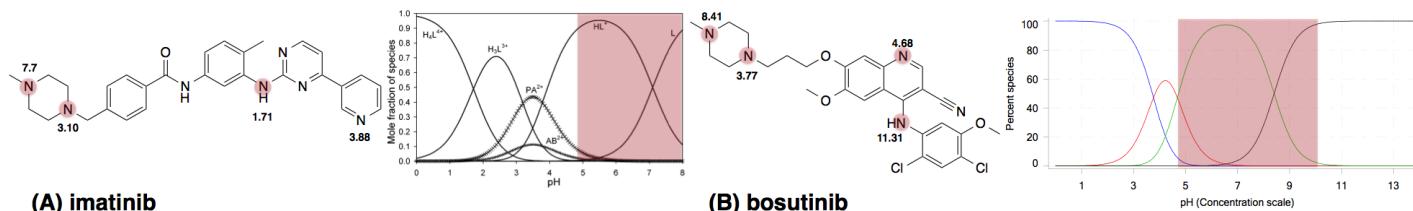


Figure 1. Experimentally-determined pK_a s of selective kinase inhibitors imatinib and bosutinib show multiple protonation states are relevant under physiological conditions. (A) The Abl-selective kinase inhibitor imatinib and its experimentally-determined macroscopic effective pK_a s are shown next to their corresponding titratable sites, along with microspecies populations in solution [45]. (B) The second-generation tyrosine kinase inhibitor bosutinib (nearly equipotent to Src and Abl) and its experimentally-determined macroscopic pK_a s and microspecies populations in solution (Sirius Analytical). The red shaded region in the microspecies population plots show the range of equilibrium protomer/tautomer populations that can be accessed if the electrostatics of binding induce a shift of only $6 k_B T$ (3.5 kcal/mol). Given that kinase inhibitors can bind with picomolar affinity ($29 k_B T$, 17 kcal/mol), there is ample energetic driving force available in binding to access these protomers or tautomers. Note that, due to potential coupling between titratable sites, our approach avoids decomposition into site pK_a s in favor of modeling the energetics of entire molecular protomer/tautomer states.

INNOVATION

This proposal integrates (1) extensive development of new, rigorous, fast methodologies that will allow full sampling of protein:ligand conformational and protomer space with (2) experimental validation of software packages to compute the protomer equilibria of unbound ligands; (3) experimental validation of predicted protonation state effects of ligand binding to specific kinases; combined with (4) a lower resolution broad computational survey of many kinase:inhibitor complexes to understand the global importance of protomer equilibria in selective kinase inhibition. We aim to implement a physically accurate treatment of small molecule and protein protomer equilibria within a rigorous statistical mechanical framework. The developed methods will move beyond protein-based constant-pH MD simulations [16, 18, 19, 48, 50, 50–54, 59], allowing a full treatment of ligand and protein protomer states with full conformational degrees of freedom in molecular dynamics and affinity calculations.

We will use the thermodynamically rigorous framework of modern alchemical free energy calculations. Alchemical methods can compute small molecule binding affinities at moderate computational cost [7, 11, 57, 60–62]. They employ molecular simulations—molecular dynamics (MD) or Monte Carlo (MC)—of unphysical, *alchemical* intermediate states that attenuate the interactions of the small molecule with its environment. These alchemical intermediate states include both the fully-interacting complex as well as replicas in which the ligand does not interact with its environment, and allow the total free energy of binding—including entropic and enthalpic contributions—to be efficiently computed [11, 61, 62]. Currently, alchemical methods can achieve useful predictive accuracy (\sim 1–2 kcal/mol) for well-behaved proteins in which there are no protonation state effects [7, 11, 63], which is sufficient to greatly accelerate lead optimization [64]. We make novel use of these calculations to investigate the nature of protomer effects and their influence on binding affinities by allowing specific titratable sites to be fixed or dynamic.

We will develop free energy calculations with dynamic treatment of protonation states and tautomers. This project will build on YANK [65], the first GPU-accelerated alchemical free energy tool [12, 65], developed by the Chodera lab. YANK uses the GPU-accelerated OpenMM library to achieve high performance on consumer-grade graphics cards (over 330 ns/day on the JAC DHFR PME benchmark on a GTX-1080) [66]; enhances sampling via Gibbs sampling Hamiltonian replica exchange [10]; extracts optimal estimates using reweighting [8]; includes anisotropic long-range dispersion corrections [5]; and efficiently samples ligand poses by mixing Monte Carlo and molecular dynamics [12]. We will build on work that incorporates MC protomer sampling [48, 50, 51] alongside MD propagation, extending this to (1) accurately treat small molecule protomers, (2) efficiently sample protomer equilibria in explicit solvent, and (3) dynamically sample protomers within alchemical free energy calculations. Our approach to small molecule protomer sampling extends a new jointly developed approach for MCCE recently reported by PI Chodera and co-I Gunner in *Nature Chemical Biology* [1]. Efficient MC sampling in explicit solvent will utilize techniques developed by the Chodera lab for incorporating nonequilibrium Monte Carlo proposals within MD simulations to increase acceptance rates by as much as 10^{27} [9], subsequently elaborated on by Roux [51]. Gunner lab expertise in highly efficient implicit solvent MC protomer sampling methods will help design highly efficient proposal schemes [16–21]. Efficiency will be further increased by optimizing alchemical nonequilibrium switching protocols using estimators we developed for this purpose [67, 68] and incorporating Metropolized Langevin integrators to suppress shadow work [69]. Alchemical free energy analysis will make new use of the multistate Bennett acceptance ratio estimator [8] extended to the semigrand ensemble.

Automated biophysical techniques will be used to validate and complement computational approaches to assessing protonation state effects. The Seeliger lab, which first demonstrated a significant role for protonation state effects in Abl:imatinib association [26], brings extensive expertise in kinase expression, biochemistry, and structural biology. Seeliger developed a widely cited bacterial expression system that allows rapid generation of large quantities of pure, active kinase for biophysical studies [27]; in collaboration with PI Chodera, this approach has been extended to automate the expression of a library of human kinase domains to be used here [70]. Automated biophysical experiments using a robotic wetlab capable of engineering many point mutations and assessing binding affinities via automated fluorescence or ITC experiments will be performed in the Chodera lab. Czodrowski contributes expertise in combining ITC techniques to dissect protonation state effects with computational methodologies in active use in drug discovery [39–42], while co-I Gunner brings extensive experience in organizing joint computational-experimental collaborative efforts to study protonation state effects in proteins [24]PMC2920746.

Our survey will provide the first estimate of the prevalence of protonation state effects in selective kinase inhibitor recognition. While protonation state effects have been studied in a few specific protein model systems by collaborator Czodrowski [39–42], little is known about the prevalence of protonation state effects in inhibitor binding in general [71], and less is known of the prevalence, nature, and magnitude of these effects in kinase inhibitor recognition outside of the well-studied Abl:imatinib system. The survey and subsequent detailed study we propose will greatly advance our understanding of the prevalence of protonation state effects in inhibitor binding.

APPROACH

This proposal has two primary goals:

1. Develop and validate a rigorous computational framework to incorporate the effects of protein and small molecule dynamic protonation states within the context of alchemical free energy calculations to compute accurate protein:ligand binding affinities
2. Utilize computational and experimental biophysical approaches—including this new framework—to assess the pervasiveness, nature, and magnitude of protonation state effects in kinase:inhibitor binding as a novel model system for exploring protonation state effects in drug binding

To achieve these goals, the work is organized into three main Aims that both develop new simulation methodologies that explicitly treat protomer equilibria and apply these approaches to successively filter kinase:inhibitor complexes into successively refined subsets that permit their detailed computational and experimental study. For tractability, we focus on kinase domains—the therapeutic target of kinase inhibitors—in isolation, matching conditions between experiment and computational modeling as precisely as possible.

AIM 1. Survey kinase:inhibitor complexes for protonation state effects to develop novel model systems.

Rationale: Aside from the well-studied Abl:imatinib system, little is known about the degree to which protonation state effects are relevant in other kinase:inhibitor systems. These effects could include both changes in the population mixtures of protomeric species and shifts in dominant protomeric species upon binding, either of which can cause large (several kcal/mol) errors in modeling. Using multiconformer continuum electrostatics methods in MCCE [17, 19] developed by the Gunner group for the study of protonation state effects, we will survey kinase:inhibitor complexes from the PDB for evidence of significant protonation state effects in ligand or protein. Systems exhibiting significant protonation state effects will be targeted for detailed study in subsequent Aims.

Our computational techniques for treating protomer/tautomer equilibria require as input the relative free energies of accessible protomers in solution for all titratable species (ligands, ionizable protein residues). While high-quality reference pK_a data is available for amino acids, we need a convenient manner to predict these aqueous protomer populations for small molecules—especially if the goal is to use these methods in molecular design. We therefore benchmark a variety of computational tools to assess their suitability for predicting small molecule aqueous protomer/tautomer energetics at physiologically relevant pH ranges. Because small molecules (such as kinase inhibitors) can possess multiple strongly-coupled titratable sites that interact through quantum chemical effects that cannot be captured solely in classical electrostatic coupling models, we treat the entire molecule as a single titratable entity possessing multiple protomer states, each with its own relative free energy in aqueous solution.

Subaim 1A. Benchmark small molecule pK_a calculation methodologies against experimental kinase inhibitor data

Approach:

The development of a quantitative approach to modeling protomer equilibria rests on the ability to predict protomer/tautomer populations for small druglike molecules in solution sufficiently accurately that this step is not accuracy-limiting. A large number of software tools are available for predicting equilibrium populations of protonation and tautomeric states for small molecules (**Table 1**), with some having reported pK_a errors as low as 0.45 log units [76]. For this benchmark, we will utilize the >30 FDA-approved small molecule kinase inhibitors, which are heavily represented in kinase:inhibitor structures in the PDB. When experimental data is unavailable, we will collect new data using electrochemical and UV-metric acid-base titrations using a Sirius T3 (via Sirius Analytical, example derived protomer populations shown in **Figure 1B**), resolving ambiguous site specificity by ^1H -NMR titrations (performed in the Seeliger laboratory). We will utilize a variety of available pK_a calculation methodologies (including those listed in **Table 1**) to assess their accuracy in predicting protomer/tautomer relative free energies and overall charge state over pH 5–10 (**Figure 2**). We will ask: (1) Which methodologies are most accurate for kinase inhibitors? (2) Are current pK_a -prediction methodologies sufficiently accurate to avoid being the accuracy-limiting step in incorporating protonation state effects into binding free energy calculations? While alchemical binding free energy calculations are believed to have accuracies of 1–2 kcal/mol under ideal conditions free of protonation state effects (0.7–1.4 log₁₀ units) [11, 29, 63, 77], neglect of protonation state effects has been reported to lead to significantly larger errors [32–35].

Tool	Supplier	Method
Epik [72]	Schrödinger	Hammett-Taft
pKa plugin	ChemAxon	machine learning
MoKa [73]	Molecular Discovery	machine learning
ACD/pKa	ACD/Labs	machine learning
Jaguar [74]	Schrödinger	pseudospectral QM
COSMO-RS [75]	SCM	DFT QM

Table 1. A subset of available software tools for predicting small molecule solution protomer populations/ pK_a s.

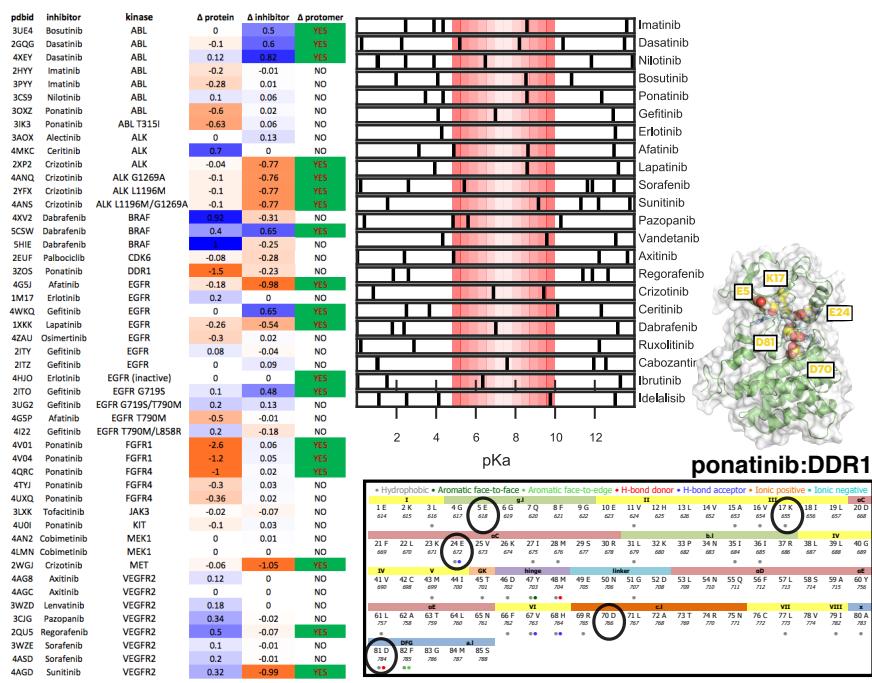


Figure 2. Preliminary MCCE survey of net proton change and tautomer shifts on binding at pH 7.4 for 49 kinase:inhibitor complexes. *Upper right:* pK_a s of 22 FDA-approved kinase inhibitors computed using Epik [72], with those accessible from pH 7.4 with up to 6 $k_B T$ energetic penalty shown in red shaded 1 $k_B T$ steps. *Left:* Shifts in total proton number in protein (Δ protein), total proton number of inhibitor (Δ inhibitor), or shift in dominant inhibitor protomer/tautomer upon binding (Δ protomer). Proton update is colored blue, while proton loss is colored red; major shifts in inhibitor protomer are colored green. Nearly all complexes surveyed appeared showed significant changes in protomer equilibria upon binding. AM1-BCC charges were assigned to each protomer [78, 79], and MCCE 2.3 was run in “default” mode to estimate equilibrium protomer populations of titratable protein residues and inhibitor. Statistical uncertainty in populations was $\sim 0.1\%$. Titratable protein residues across kinases found to change protonation state upon binding are highlighted in their corresponding positions on DDR1 (*right*) and in the kinase-ligand interaction fingerprints (KLIFS) [80] interaction fingerprint for DDR1:ponatinib (*lower right*).

Preliminary results: We used the Epik [72] tool to predict pK_a s for 22 FDA-approved kinase inhibitors (**Figure 2**, upper right), finding experimental pK_a s of imatinib, bosutinib, erlotinib, and gefitinib reproduced to < 1 log unit error (**Figure 1**; erlotinib and gefitinib measured by Sirius Analytical on a Sirius T3). More than half of the FDA-approved kinase inhibitors surveyed are predicted to possess pK_a s in a range that would require a free energy cost of 6 $k_B T$ or less to access an alternative protonation state at the typical intracellular pH of 7.4, indicating the potential for significant protonation state effects is widespread among selective kinase inhibitors (**Figure 2**).

Potential pitfalls and alternative approaches: Should pK_a predictions for kinase inhibitors be far less accurate than reported in large-scale benchmarks [76], we will focus subsequent aims on kinase inhibitors with available experimental pK_a data and utilize experimentally-derived protomer populations inferred from experimental data, allowing us to fully develop and validate our methodology. Our data would inform collaborator Schrödinger on how small molecule pK_a prediction methodologies (in Epik and Jaguar) could be further improved.

Aim 1B. Survey noncovalent kinase:inhibitor complexes for significant protonation state effects using MCCE.

Approach: We will extract all noncovalent kinase:inhibitor complexes in the PDB to perform a large-scale survey of potential protonation-state effects using MCCE 2.3 [17, 19]. MCCE uses Monte Carlo (MC) sampling of sidechain rotamers, protomers, and tautomers and a fast Poisson-Boltzmann implicit solvent model to efficiently sample the equilibrium distribution [17, 19]. All kinase inhibitors will have solution protonation and tautomer populations predicted as in Aim 1A and charges assigned using the AM1-BCC method [78, 79] to produce charges compatible with the MCCE titratable protein forcefield. For each complex, a complete model of the kinase catalytic domain (with missing heavy atoms added and internal loops modeled) will be generated using MODELLER [81, 82] and ROSETTA kinematic loop closure [83], and an explicit-hydrogen ligand constructed using the OpenEye toolkit. MCCE will be run on inhibitor alone, kinase alone, and complex to identify (1) whether mixtures of protonation states exist in significant populations in any of these states, and (2) if so, whether dominant protonation states of inhibitor or kinase are seen to shift upon binding. MCCE requires $\sim 6\text{--}12$ hr/complex on a single CPU, allowing all ~ 2100 complexes to be rapidly evaluated on our resources. Kinase:inhibitor complexes predicted to have large protonation state effects despite these assumptions will be flagged as candidates for subsequent detailed study.

We will ask many questions of the data from this first large-scale survey: Are protonation state effects mainly driven by ionizable protein residues (determined by comparing fixed or variable protomers), or by ligand moieties, and if so, which ones? Do these have commonalities across the kinome or in specific branches, or across inhibitors of similar scaffolds or substituents? How sensitive are these effects to conformation, phosphorylation, and point mutations? For cases where structures of antitargets (like imatinib:Src) are available, do inhibitors always bind both targets and antitargets in the same protonation state? Are there insights into exploiting protonation state penalties to achieve selectivity that could be exploited? A wealth of new insight into the role of protonation state effects in selective kinase inhibition is expected.

Preliminary data: The essential concept behind our methodology for incorporating small molecule protonation state effects into continuum electrostatics MC sampling was recently described by co-I Gunner and PI Chodera in *Nature Chemical Biology* [1]. We have performed a survey of 49 kinase:inhibitor complexes from the RCSB using MCCE in ‘default’ mode, where extensive sampling of sidechain rotamers and protomers is performed. These complexes were selected with the goal of identifying FDA-approved kinase inhibitors in complex with their targets of therapy. Protonation state penalties and charges were obtained as described above. **Figure 2** summarizes the results of this survey. The majority of complexes surveyed display some form of protonation state effect upon inhibitor binding. It appears that the kinase more frequently experiences protonation state changes than the inhibitor, and that both proton gain (blue) and loss (red) upon binding is possible. A surprising fraction demonstrate inhibitor protomer or tautomer shifts upon binding (green). Conserved protein residues for which protonation state effects are seen are highlighted in their corresponding positions on DDR1 (**Figure 2**, right), as well as their kinase-ligand interaction fingerprints (KLIFS) [80] (**Figure 2**, bottom) to better illustrate locations at which titratable residues can influence protonation state either through direct interactions or proximity to the ATP binding site.

We also used MCCE to assess the expected error in computed binding affinities if the *wrong protomer or tautomer* was selected in a binding free energy calculation. We used crizotinib, an FDA-approved inhibitor approved for treatment of non-small cell lung carcinoma (NSCLC), for which complexes with multiple proteins appear in the RCSB. The dominant solution protomer of crizotinib has a +2 charge, but can access 3 additional protomers with a maximum penalty of only 3.31 kcal/mol (as predicted by Epik). While crizotinib prefers the +2 state in solution, it prefers the +1 state (which requires only 0.3 kcal/mol to access) in complex. If, however, the modeler assumes the solution state is dominant, they make a ~7 kcal/mol error in binding affinity for ALK (4ANQ, 5AAA, 2YFX, 4ANS, 5AAB, 5AAC, 2XP2, all of which give comparable energetics) or a ~10 kcal/mol for binding to MEK (2WGJ).

Potential pitfalls and alternative approaches: If the list of kinase:inhibitor systems with significant protonation state effects is too numerous, we will prioritize kinases with facile bacterial expression and kinase inhibitors that can be easily purchased to facilitate detailed experimental investigations in Aim 3. Multiple simulation runs and statistical error analysis, as well as examination of different structures of the same kinase:inhibitor pair, will ensure reproducibility of calculations. It is possible that assuming the same kinase conformation persists for both bound and unbound states could be problematic; if so, we will test a model in which we use apo kinase structures (where available) or a single reference structure to model the unbound state.

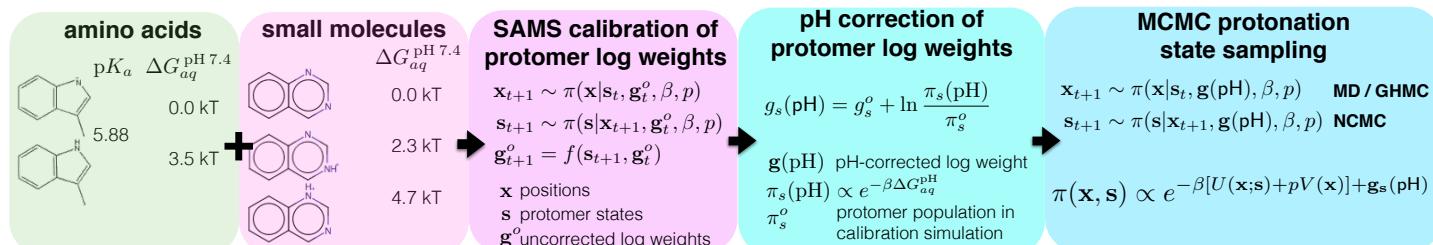


Figure 3. Overview of constant-pH molecular dynamics simulation methodology developed here for Aim 2. From left to right: For both **amino acids** and **small molecules**, we must first compute the aqueous relative free energies of each protomer species at the pH of interest, ΔG_{aq}^{pH} . For amino acids, we use reference pK_a s for amino acids in isolation, and compute ΔG_{aq}^{pH} at the pH of interest with the Henderson-Hasselbalch equation (or variants). For small molecules, we use a protomer aqueous population prediction tool (explored in Aim 1A; **Table 1**) to estimate ΔG_{aq}^{pH} , avoiding the uncoupling of titratable sites within the molecule that may be coupled by electronic effects, as in the quinazoline ring shown here. For each titratable entity (small molecule or ionizable residue in a reference state matching the reference pK_a), an explicit solvent **SAMS calibration of protomer log weights** follows, which uses a Gibbs sampling framework [10] in which positions x are updated by some form of (possibly Metropolized) molecular dynamics, protomer states s are updated by NCMC [9, 50, 51], and the optimal SAMS (self-adjusted mixture sampling) update scheme is used to update the log weights g (here using the function f) [84]. SAMS is an adaptive single-replica expanded ensemble [85] method similar in spirit to Wang-Landau [86] or simulated scaling [87], but has a provably asymptotically consistent and optimal convergence properties [84], a major advance over earlier methods. This calibration step uses equal SAMS target probabilities for each protomer species to ensure the relative molecular mechanics free energies of each species are estimated to $O(k_B T)$, so that these can be removed in the correction step. To preserve charge neutrality, water molecules are reversibly turned into monovalent counterions. This calibration step *must* be repeated whenever the forcefield, water model, electrostatics model (e.g. PME), temperature, or pressure are changed, or else the correct protomer populations will no longer be obtained if a constant-pH simulation is run for the amino acid or small molecule in isolation in solvent. Next, we perform **pH correction of log weights**, in which we remove the residual bias from the relative molecular mechanics free energies in explicit solvent (using the SAMS-derived log weights g_s^o and π_s^o , the empirical probability of observing protomer s in the SAMS simulation), to ensure the weight will generate the appropriate input protomer distribution in solvent constant-pH simulations. This can be used to compute the pH-corrected log weight for any pH of interest without additional SAMS calibration. Finally, in **MCMC protonation state sampling**, all titratable entities of interest (residues and small molecules) are included in the simulation, and we simulate use a modified version of the same Gibbs sampling updates we used in calibration, except that the pH-corrected log weights $g(pH)$ are fixed throughout the simulations. The overall distribution sampled is an expanded ensemble $\pi(x, s)$ that includes the protomer states s of all titratable entities in the system.

AIM 2. Incorporate efficient protomer sampling into molecular dynamics and alchemical free energy calculations and dissect the physical origin of protonation state effects in kinase:inhibitor systems.

Rationale: Current modeling techniques for protein-ligand interactions assume both protein and ligand retain a single, dominant protonation state in solution and complex—an assumption that can give rise to errors of multiple kcal/mol in systems with significant protonation state effects. We will remove this assumption by rigorously incorporating dynamic protomer sampling into quantitatively accurate explicit-solvent alchemical free energy calculations using a mixed MD/MC framework (**Figure 3**). We will build on existing work for explicit solvent constant-pH molecular dynamics simulations, optimizing the methodology to achieve high efficiency for solvated kinase:inhibitor systems. We will then integrate this methodology into our GPU-accelerated alchemical free energy code [12], extending our reweighting analysis [8] into the semigrand ensemble to extract pH-dependent binding affinities. Finally, we will apply both tools to computationally probe the magnitude and nature of protonation state effects for candidate systems identified in Aim 1, examining the error in computed binding affinities incurred when fixed protonation states are used for protein and/or inhibitor, as well as which species (protein or ligand, which residues/functions) are primary contributors. Quantitative data collected in Aim 3 will provide critical feedback on the relative importance of protonation state effects relative to other challenges to quantitative accuracy in kinase systems, such as conformational reorganization and forcefield limitations.

Aim 2A. Investigate kinase:inhibitor protonation state effects in detail with constant-pH molecular dynamics

Approach: To investigate kinase:inhibitor protonation state effects with a detailed method in explicit solvent, we first implement and then optimize a constant-pH simulation methodology that is applicable to both small molecules and titratable protein residues. We present a high-level overview of our approach to constant-pH simulations in explicit solvent allowing both small molecules and protein residues to sample protomer states in **Figure 3**. This approach builds on previous approaches to hybrid MD/MC protonation state sampling in implicit [16, 18, 19, 48] and explicit [50, 51] solvent to account for electrostatic interactions that influence protomer populations, and extends it to handle both small molecule and protein residue protomers in a scalable manner.

We will use our highly efficient nonequilibrium candidate Monte Carlo (NCMC) approach [9] to enormously boost Monte Carlo efficiencies. During the simulation, MC moves attempt to switch the current protonation (or proton tautomer) state of a residue or cluster of residues to another discrete, physical protonation state (**Figure 3**). MC moves will utilize NCMC [9], in which the initial configuration is stored, a short velocity Verlet [89] trajectory of n steps is generated in which the protomer incrementally changes to its new proposed protomer through some (possibly alchemical) nonequilibrium protocol in which the rest of the system is also allowed to respond. The final configuration and physical, discrete protomer is either accepted according to a Metropolis-like criteria involving the current potential energy, the solvent pH, and the solution population of the protomer. If rejected, the entire switching trajectory is rejected, allowing rigorous sampling the correct constant-pH ensemble—the intermediate nonphysical protonation states are never recorded [9]. If the NCMC switching move is long enough to keep the system close to equilibrium, astronomical increases in acceptance rates can be achieved (**Figure 4A**).

Our detailed investigation of kinase:inhibitor complexes will focus on those identified as likely candidates for protonation state effects from Aim 1, with an estimated 100–200 total complexes considered. (100 complexes \times 10 GPUs/complex \times 10 days/complex = 5000 GPU-days on the MSK 200-GPU cluster, or a total of 25 cluster-days over the course of this project.) For this detailed study, we will attempt to match experimental conditions

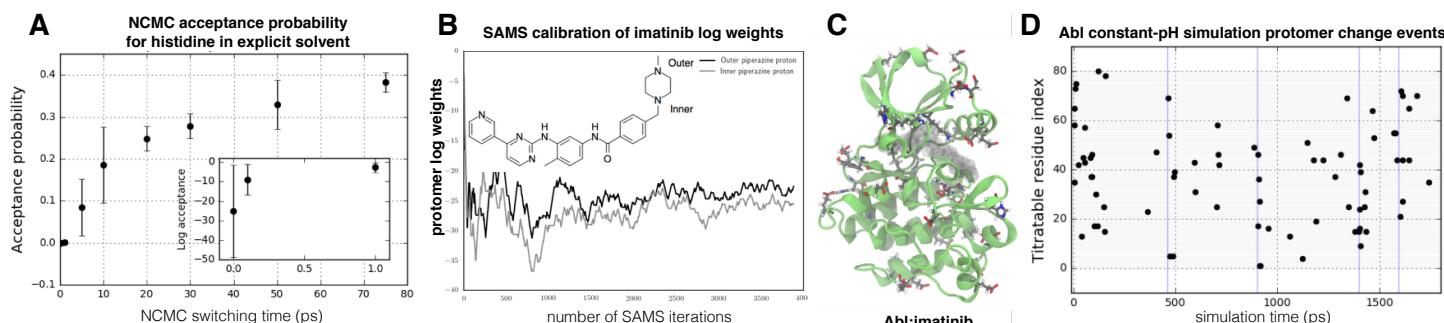


Figure 4. Protomer sampling in MD/MC simulation using GPU-accelerated OpenMM. The scheme described in **Figure 3** is used. (A) While explicit solvent acceptance rates for instantaneous MC protomer change proposals are vanishingly small ($\sim 10^{-24}$), NCMC proposals [9] with GHMC (Metropolized Langevin) integration to eliminate shadow work [69, 88] boost acceptance rates by a factor of $\sim 10^{23}$ for switching times of 10 ps or more. Error bars denote 95% confidence intervals. (B) Calibration of imatinib uncorrected protomer log weights in explicit solvent using SAMS [84] rapidly converge to $\sim 1 k_B T$ in 4000 iterations of 1 ps of GHMC propagation followed by a 10 ps NCMC protomer switch proposal. (C) Abl:imatinib complex showing ionizable residues observed to change protonation state in 2 ns of MD/MC simulation in explicit solvent at pH 7.4. (D) Accepted protonation state changes for ionizable residues in Abl kinase at pH 7.4.

(experimental protein construct, salt conditions, and pH) as closely as possible for feasible experimental assays in Aim 3. Our investigation will focus on (1) shifts in dominant protonation state in solution and complex, (2) shifts in protomer equilibria upon binding, (3) identification of protein residues and small molecule moieties predominantly responsible for protonation state effects (1) and (2), and (4) detailed biophysical interactions responsible for mediating protonation state effects (such as those indicated in the KLIFS fingerprints [80]). We ask similar questions to Aim 1B, but with the added detail of fully flexible explicit solvent equilibrium geometries.

Preliminary data: **Figure 4** shows a working prototype of the scheme described in **Figure 3** implemented using the GPU-accelerated OpenMM 7. As expected, NCMC greatly increases acceptance rates with switching time (**Figure 4A**) provided shadow work is suppressed with a Metropolized integrator. The essential small molecule self-adjusted mixture sampling (SAMS) [84] calibration step (described in **Figure 3**) indeed rapidly reaches convergence to $O(k_B T)$ as required (**Figure 4B**). Following calibration of ionizable sidechains, constant-pH simulations of Abl in explicit solvent (**Figure 4CD**) shows that many ionizable residues in the apo kinase can sample a variety of protomers at pH 7.4.

Potential pitfalls and alternative approaches: Even a GPU-accelerated of the constant-pH simulation methodology may be inefficient due to large number of states that must be sampled to thoroughly sample protomer equilibria. Numerous optimizations will be explored to further increase acceptance rates in kinase:inhibitor systems, including Metropolized Langevin integrators such as GHMC [88] to suppress shadow work [69], low-dissipation integrators that permit large timesteps like g-BAOAB [90], propagating only a local solvated region during NCMC switching to reduce work variance, biasing proposals by inexpensive MCCE-like MC moves, and optimized NCMC pathways that minimize thermodynamic length [91]. We will incorporate ideas from MCCE [16–20] to design clever proposal schemes that ensure that proposals are likely to be accepted, including rapidly evaluating potential protomer moves with implicit solvent models prior to proposal, and correcting for this bias in the Metropolis-Hastings criteria.

Aim 2B. Quantify protonation state effects on kinase:inhibitor binding with constant-pH free energy calculations

Approach: PI Chodera developed the first free, open-source, GPU-accelerated code YANK [12, 65] for computing protein-ligand binding affinities using rigorous alchemical free energy methods [12]. YANK combines numerous efficiency and accuracy improvements previously developed by Chodera and collaborators [3–6, 8–10, 92] into a single open source code that allows new algorithms to be easily developed and implemented. YANK uses a Gibbs sampling replica exchange molecular dynamics (MD) framework, where many exchange attempts among alchemical intermediates are periodically attempted and accepted with a Metropolis-like criterion to give correct equilibrium statistics [10, 12]. To propagate replicas between exchange attempts, YANK uses a general Markov chain Monte Carlo (MCMC) sampling scheme [93] that allows any composition of MCMC updates to be chained together. Currently, YANK uses a composition of (Metropolized) Langevin dynamics (GHMC [88]) and MC ligand rotation and displacement proposals [12].

We will integrate the constant-pH scheme developed in Aim 2B into the existing MD/MC replica propagation scheme within YANK. To complete our hybrid constant-pH alchemical methodology, we need two additional components: (1) an updated alchemical replica-exchange scheme for the constant-pH ensemble, and (2) an updated analysis method to extract pH-dependent free energy estimates. For (1), we first define an appropriate *reduced potential* $u(x, s; \lambda, \text{pH})$, where the equilibrium distribution sampled by a given replica is $\pi(x, s; \beta, p, \lambda, \text{pH}) \propto e^{-u(x, s; \beta, p, \lambda, \text{pH})}$, with x the current instantaneous configuration, $s \in \mathbb{Z}^N$ an integer vector specifying the current protonation state of all N titratable sites (across small molecules and protein), $\beta = 1/(k_B T)$ the inverse temperature, p pressure, $\lambda \in [0, 1]$ the alchemical parameter, and pH the current pH:

$$u(x, s; \beta, p, \lambda, \text{pH}) \equiv \beta [U(x, s, \lambda) + pV(x) + g_s(\text{pH})] \quad (1)$$

where $U(x, s, \lambda)$ is the potential energy function that depends on both protomers s and alchemical parameter λ , $V(x)$ the current volume, and $g_s(\text{pH})$ is the pH-corrected log weights. If velocities are regenerated from the Maxwell-Boltzmann distribution at the beginning of each Hamiltonian exchange iteration, the acceptance criteria is

$$P_{\text{exchange}}(\{x_i, s_i\}, \{x_j, s_j\}) = \min \left\{ 1, \frac{e^{-u(x_j, s_j; \beta_i, p_i, \lambda_i, \text{pH}_i)} - u(x_i, s_i; \beta_j, p_j, \lambda_j, \text{pH}_j)}{e^{-u(x_i, s_i; \beta_i, p_i, \lambda_i, \text{pH}_i)} - u(x_j, s_j; \beta_j, p_j, \lambda_j, \text{pH}_j)} \right\} \quad (2)$$

For (2), the free energy contribution for each alchemical leg is optimally estimated using the multistate Bennett acceptance ratio (MBAR) method we developed to analyze equilibrium data from arbitrary equilibrium thermodynamic

states [8], by solving the self-consistent equations for the dimensionless free energies \hat{f}_i , $i = 1, \dots, K$,

$$\hat{f}_i \equiv -\ln \sum_{k=1}^K \sum_{n=1}^{N_k} \left[\sum_{l=1}^K \exp \left\{ \hat{f}_l - [u(x_{kn}, s_{kn}; \beta_l, p_l, \lambda_l, \text{pH}_l) - u(x_{kn}, s_{kn}; \beta_i, p_i, \lambda_i, \text{pH}_i)] \right\} \right]^{-1} \quad (3)$$

where there are N_k samples x_{kn} from each of K alchemical states. This self-consistent set of equations can be efficiently solved in several ways [8]; we will use our `pymbar` code, which has been downloaded over 27,000 times. The free energy difference between the fully interacting state ($\lambda_1=1$, $\beta_1=\beta$, $p_1=p$, $\text{pH}_1=\text{pH}$) and noninteracting state ($\lambda_K=0$, $\beta_K=\beta$, $p_K=p$, $\text{pH}_K=\text{pH}$) at the inverse temperature, pressure, and pH of interest is simply given by $\Delta\hat{f}_{1K} \equiv f_K - \hat{f}_1$. Note that the Gibbs sampling replica exchange and MBAR framework is sufficiently general to permit the use of *multiple* pH values in a single simulation, allowing the estimation of binding affinities over a pH range of interest or pH sensitivities ($\partial\Delta f/\partial\text{pH}$). Indeed, including multiple pH values in a hybrid approximate implicit/explicit replica exchange method was found to improve mixing [94].

Initial validation of the treatment of small molecule protomers in alchemical free energy calculations will make use of cyclohexane-aqueous distribution coefficients collected by our laboratory for SAMPL5 [2, 30] and host-guest binding free energies [31], both of which proved challenging due to protonation state effects [30, 36]. Internal consistency will also be examined by ensuring agreement between transfer free energies with dynamic protomer sampling with transfer free energies with fixed protomers appropriately weighted by solution state protomer free energies used in calibration. Control experiments—such as the computation of equilibrium protomer populations and pK_a s for blocked amino acids and small molecules—will ensure the correctness of the implementation.

We will quantify the impact of protonation state effects on inhibitor binding affinities and selectivities. Many kinase:inhibitor binding affinities have been reported in the literature (e.g. [95–97]), or will be measured as part of Aim 3. A number of kinase:inhibitor complexes available in the PDB represent the same inhibitor complexed with different kinases—for example, structures of imatinib with both Src and Abl are available, despite the high selectivity for Abl over Src. For cases where experimental affinity data is available, we will assess the degree to which the inclusion of protonation state effects increases the accuracy in reproducing experimental inhibitor binding affinities. We will ask whether small molecule protonation state effects are crucial in discriminating a selective binder from a non-selective binder. We will assess whether neglect of accurate or dynamic protonation states in the protein and/or ligand leads to significant shifts in affinity, and assess their magnitude. We expect better understanding of when and where protonation effects influence drug binding will aid active drug discovery projects and inform the development of future accurate methodologies to guide drug discovery efforts.

Preliminary data: We used YANK to perform alchemical free energy calculations of Abl:imatinib binding (pdbid:2HYY) using 40 orientational restraints [98], AMBER14SB with GAFF [99] and AM1-BCC charges [78, 79], and PME. The binding free energy of the Epik-predicted solvent-dominant protomer (outer piperazine N protonated) is predicted to have an unfavorable affinity (-2 ± 1 kcal/mol). The second most stable protomer in solvent (inner piperazine N protonated, $\Delta G_{aq}^{\text{pH}7.4} = 0.21$ kcal/mol) has a computed binding affinity of -11.6 ± 0.8 kcal/mol, surprisingly close to the deprotonated form ($\Delta G_{aq}^{\text{pH}7.4} = 0.91$ kcal/mol), with computed affinity of -11.5 ± 0.7 kcal/mol.

Potential pitfalls and alternative approaches: Given the efficiency of the GPU-accelerated OpenMM library [66] on which YANK is based, simulations achieve 200–300 ns/day for DHFR in explicit solvent, which is expected to allow for 10^4 – 10^5 protonation state trials/day in a standard 10-GPU alchemical free energy calculation. Should convergence of dynamic protonation states prove problematic, we will pursue the optimization strategies described in Aim 2A alternative approaches. Slow conformational changes—especially those coupled to protonation state changes—can also impede convergence in constant-pH or alchemical free energy calculations. To mitigate this risk, we will compare free energies of binding computed with multiple crystal structures to see if these give rise to distinct binding free energy estimates. If so, we can use umbrella sampling techniques to compute the free energy of DFG flips (following [34]) to assess dependence of calculated affinities and protonation states on conformation. The accuracy of these calculations will also depend on the protein and small molecule forcefields. We will benchmark multiple forcefields, starting with the AMBER family of forcefields. If forcefield inaccuracies prove problematic at replicating experimental affinities, our calculations should provide qualitative insight into the impact of protonation state effects. Even if few significant protonation state effects are identified, this knowledge will aid in identifying other dominant sources of limiting accuracy in modeling tools.

AIM 3. Experimentally probe kinase:inhibitor systems to quantify protonation state effects on binding.

Rationale: In this Aim, we will experimentally investigate kinase:inhibitor systems predicted to have significant protonation state effects in Aim 1, using the results of Aim 2 to match determine which experimental measurements would be most appropriate for detecting observable protonation state effects. For experimental tractability, we restrict ourselves to kinases for which bacterial expression permits both large-scale production of milligram quantities of protein and the introduction of site-directed mutants, and inhibitors for which compounds can readily be obtained or purchased from vendors. The Chodera and Seeliger labs have applied the phosphatase coexpression strategy of pioneered by Seeliger [27] and demonstrated good bacterial expression (≥ 2 mg/L culture) for 52 human kinase catalytic domains with available crystal structures using an automated expression pipeline [70], providing ample opportunities for kinases to be carried forward for experimental investigation. We will primarily consider the dephosphorylated form of these kinases, though we will examine the effects of phosphorylation for some kinases for which the biochemically-accessible phosphorylation state is known and predicted to have impact on protonation state effects by the large-scale survey in Aim 1. Notably, these experiments can be run in parallel once candidate kinase:inhibitor complexes are identified in Aim 1 or from a preliminary survey (Figure 2).

Aim 3A. Observe protonation uptake/loss using ITC, and dissect its contribution to affinity

Isothermal titration calorimetry (ITC) provides a direct route to studying proton shuttling between solvent and complex upon binding. ITC measurements in buffers with identical pH and ionic strength but distinct ionization enthalpies (e.g. Tris, Hepes, and Tricine), allow the heat effects of proton shuttling between complex and solvent to be deconvoluted from heat effects of binding [38] (Figure 5B). With the assistance of the Seeliger lab in designing and optimizing these experiments, and Czodrowski in interpreting the data in a manner similar to his previous work in other systems [100, 101], these experiments will be conducted by the Chodera lab using, making use of the LabMinds Revo (which prepares solutions to ± 0.1 pH unit of specified pH, and quantifies pH to $\pm 0.5\%$) to automate buffer preparation, automated liquid handling to set up the experiments, and an automated calorimeter (Auto iTC-200) to execute them. From these experiments, we will obtain direct evidence of protonation state effects, extracting both effective pK_a and affinity for each protomer in the case of single relevant titration sites. We will use recommended techniques for optimizing experimental design and minimizing accuracy-limiting errors [14], and develop a Bayesian approach to propagate concentration and pH uncertainties into derived thermodynamic parameters for error [15].

Rational Selection criteria: For kinase:inhibitor complexes identified in Aims 1 or 2 as having proton shuttling events between complex and solvent upon binding (a net uptake/loss of protons for the kinase:inhibitor pair), and for which the inhibitor is sufficiently soluble in buffer to permit useful ITC experiments to be performed (ideally $\geq 250 \mu\text{M}$), we will use this technique to experimentally determine whether proton shuttling events can be observed, and to experimentally validate the predicted difference in affinity between protonated and deprotonated species.

Preliminary data: Co-I Seeliger demonstrated the feasibility of ITC experiments on kinase:inhibitor systems nearly a decade ago in published work with the Kuryian laboratory [27, 46] (Figure 5A).

Potential pitfalls and alternative approaches: Protein stability can be a challenge in ITC experiments. Should kinases prove unstable in obvious buffer choices, we will perform Thermofluor stability analysis for a range of buffer systems using the Chodera lab automation system [102, 103] to determine buffers with differing ionization enthalpies in which the kinases are maximally stable. Lack of knowledge of exact pH-dependent inhibitor solubilities may also complicate ITC experiments. To avoid this, for inhibitors for which we will collect Sirius T3 electrochemical titrations, we will also use the Sirius T3 to measure kinetic and thermodynamic solubilities to serve as a guide for optimal ITC experimental design. In complex cases with multiple relevant protonation states, it will be difficult to unambiguously interpret the data; in these cases, we will instead predict ITC titration data from the alchemical free energy calculations using both computed enthalpies upon binding and the experimentally-characterized buffer ionization enthalpies in a model of titration calorimetry. Finally, the limited dynamic range of affinities accessible by ITC may prove challenging, in which case competition experiments may be appropriate [104]. Fortunately, the

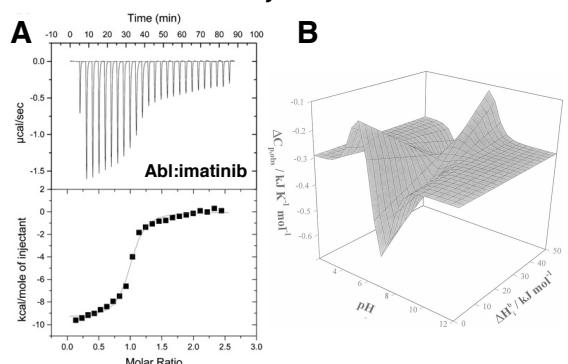


Figure 5. Isothermal titration calorimetry (ITC) allows protonation state effects to be separated from binding. (A) Co-I Seeliger has previously demonstrated the feasibility and utility of calorimetry on kinase:inhibitor binding for multiple kinases in multiple phosphorylation states [27, 46]. Data for titration of unphosphorylated Abl by imatinib from [46] is shown. (B) For a system with a single titratable site highly sensitive to the current pH, the observed calorimetric heat ($\Delta C_{p,\text{obs}}$) depends on both the buffer ionization enthalpy (ΔH_i^b) and the pH. Figure reproduced from [38].

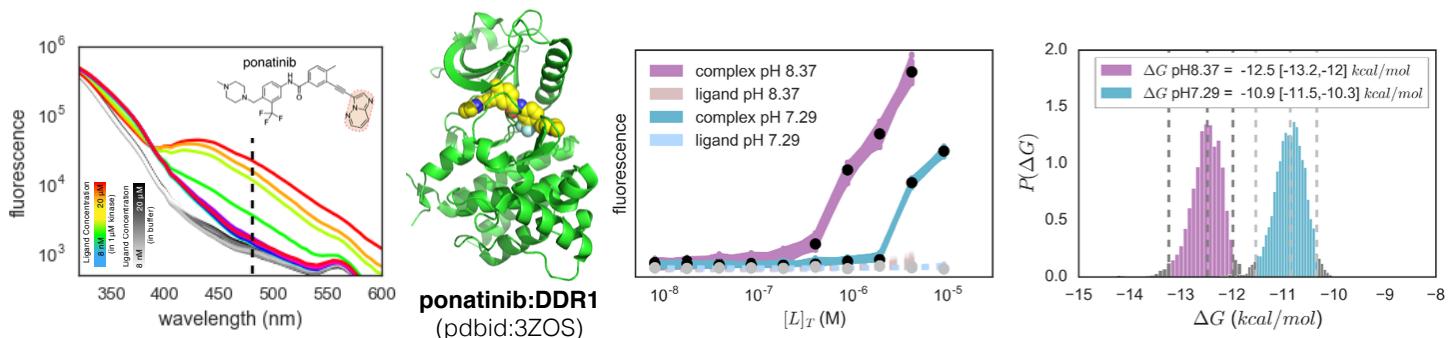


Figure 6. Dependence of DDR1:ponatinib affinity on pH measured by automated fluorescence assay for direct inhibitor binding.

By using ATP-competitive kinase inhibitor probes that greatly increase fluorescence upon binding (which include FDA-approved kinase inhibitors with quinoline, quinazoline, and pyrazolopyrimidine scaffolds, among others), we can rapidly and directly measure the affinity of the probes over a very large dynamic range (nM–mM) or measure other inhibitor affinities by competition assays. We have miniaturized a cuvette-based assay [105] into an automated high-throughput assay. *Left:* Titration of ponatinib (pyrazolopyrimidine scaffold) into $1 \pm 0.1 \mu\text{M}$ DDR1 at $\text{pH } 8.37 \pm 0.04$ causes large increases in fluorescence emission (color traces) over titrating identical concentrations into buffer alone (greyscale). Excitation at 280 nm (which excites Trp and Tyr residues in the kinase) was used. *Middle:* Fluorescence emission at 480 nm was fit with a Bayesian model that incorporates all known sources of uncertainty (protein concentration, ligand DMSO stock solution concentration, pipetting and dispensing errors, fluorescence detection errors) to generate an ensemble of model fits and posterior distributions of thermodynamic parameters. *Right:* Bayesian posterior distributions for free energy of DDR1:ponatinib binding at $\text{pH } 7.29 \pm 0.04$ and 8.37 ± 0.04 show a highly significant $\sim 1.6 \text{ kcal/mol}$ change in binding affinity. ΔG mean estimates and 95% confidence intervals are shown in brackets. Affinities for non-fluorescent inhibitors can be determined by competition versions of this assay (not shown).

large dynamic range of kinase:inhibitor affinities [96, 97] suggests many will be directly measurable by ITC.

Aim 3B. Use direct fluorescence binding affinity assays to examine binding events sensitive to pH

While pH sensitivity of binding affinities is a clear signature of protonation state effects, pH-dependent kinase inhibition may also be relevant to cancer therapy. Tumor environments are often acidic due to increased aerobic and anaerobic glycolysis [106]. While intracellular pH is more tightly regulated than the extracellular pH, significant shifts in intracellular pH have been observed to range from 6.4–7.4 [106], which can manipulate the energetics of protonation states by up to $2.3 k_B T$. For kinase:inhibitor complexes predicted to show significant pH-dependent binding affinities in Aims 1 or 2, we will measure pH-dependent binding affinities using a fluorescence assay, where FDA-approved kinase inhibitors featuring certain fluorescent scaffolds undergo a measurable large direct fluorescence increase upon binding (**Figure 6**, left), without requiring ATP or enzyme turnover [105]. The Chodera lab will use an automated plate-based version of this assay to measure the pH-dependent binding affinity of ATP-competitive inhibitors, either directly (for fluorescent inhibitors) or via competition with a fluorescent probe inhibitor. Bayesian analysis will be used to propagate concentration and measurement uncertainties into affinities.

Selection criteria: To assess the potential for significant changes in inhibitor effectiveness between unperturbed (~ 7.4) and acidified (~ 6.4) intracellular environments—and hence differences in kinase inhibitor efficacy between normal and tumor cells—we will use the tools developed in Aims 1 and 2 to compute the effect of acidified intracellular pH on binding affinities. Systems that show significant pH-dependence in binding affinities and for which inhibitor solubilities in aqueous assay buffer are at least $20 \mu\text{M}$ (to allow a sufficiently large dynamic range of affinities to be resolved) will be selected for this class of experiments.

Preliminary data: Our MCCE survey predicted ponatinib and DDR1 both undergo significant proton loss upon association (**Figure 2**, left). Since both compound and protein were available in our laboratories, we used this fluorescence assay to measure the binding affinity of ponatinib (which has a fluorescent pyrazolopyrimidine scaffold) to DDR1 in 20 mM Tris buffer under two different pH conditions separated by approximately one pH unit (**Figure 6**). Bayesian analysis (**Figure 6**, right) provides estimated free energies of binding of -12.5 kcal/mol (95% CI [-13.2, -12.0]) at $\text{pH } 7.29 \pm 0.04$ and -10.9 kcal/mol (95% CI [-11.5, -10.3]) at $\text{pH } 8.37 \pm 0.04$, giving a highly significant difference of $\sim 1.6 \text{ kcal/mol}$, validating that protonation state effects are clearly significant in this system.

Potential pitfalls and alternative approaches: Some inhibitors may exhibit pH-dependent fluorescence properties, limiting the useful pH range for these compounds. The pH-dependent solubility of kinase inhibitors may additionally limit range of affinities measurable by this assay; should signatures of small molecule aggregation appear (such as deviations in model fits, shown in **Figure 6**, middle), we will ensure that pH-dependent solubility data is collected for compounds for which Sirius T3 measurements are performed and limit our maximum concentrations to below the aqueous solubility limits. Protein stability can be compromised at low or high pH; if pH-dependent instability is observed, we will utilize Thermofluor stability analysis to bracket the range of pH for which the kinase is stable [102, 103].

Aim 3C. Engineer clinical mutations that modulate pK_a s and determine their impact on inhibitor affinity.

In systems where protonation state effects are significant, the introduction or elimination of charged residues in the vicinity of the binding site can cause electrostatically-induced shifts in pK_a that indirectly modulate inhibitor binding affinities, potentially manifesting a new mechanism of resistance. Such mutations would alter the electrostatic environments of key titratable groups directly or stabilize alternative conformations with shifted pK_a s (as has been observed for site-directed mutations [26]). Engineering such mutations into the kinase and utilizing the fluorescence inhibitor binding assay from Aim 3C for both wild-type and mutant kinases will reveal whether these mutations are capable of causing measurable alternations to inhibitor binding free energies.

Selection criteria: We will use both MCCE (Aim 1) and YANK (Aim 2) to computationally evaluate whether clinically-observed mutations (identified in the MSKCC cBioPortal) may cause measurable changes to inhibitor binding affinity via protonation state effects or induced pK_a shifts in other titratable sites. While both MCCE and YANK extensions developed in this project will be able to identify point mutants that appear to exhibit protonation state effects in ligand binding, MCCE can also compute whether these residues might cause significant pK_a s shifts in other titratable sites. We restrict ourselves to kinases identified to have significant protonation state effects and for which facile bacterial expression of the kinase permits site-directed mutagenesis, and for which direct changes in inhibitor fluorescence can be monitored due to the presence of a fluorescent scaffold (**Figure 6**). The Seeliger and Chodera labs will work together to engineer and express the mutants, and the Chodera lab will perform the fluorescence assays at pH 7.4 as described in Aim 3B.

Preliminary data: The Seeliger and Chodera laboratories have engineered a panel of 96 point mutants of Abl and Src harvested from the cBioPortal, using an automated approach to single-primer mutagenesis and bacterial expression screening similar to the one we used to engineer the panel of 52 human kinase catalytic domains that will be used here [27, 70]. While only 19/48 Abl mutants achieved expression levels sufficient for the work described here, 44/46 Src mutants achieved useful levels of expression, suggesting that many mutations will yield sufficient expression for experimental study in this subaim.

Potential pitfalls and alternative approaches: Clinically-observed mutations may simply be neutral in their impact on inhibitor binding, or may exert effects on inhibitor binding due to a variety of mechanisms. As a result, there is some concern that the yield of mutations that influence protonation state effects or pH sensitivity may be low if mutations are selected as only those that are predicted to modulate binding affinity. To mitigate this concern, we will first restrict our study to mutations at conserved sites already identified to exhibit protonation state effects in our large-scale MCCE screen (building on the analysis in **Figure 2**, right), later expand this set to include mutations involving charged residues, and finally expand this set to consider all observed clinical mutations.

COLLABORATION MANAGEMENT PLAN. The Chodera, Gunner, and Seeliger groups are all located in close proximity, have worked together over the past three years, and already have an excellent collaborative relationship. In particular, Chodera and Gunner have worked together to develop new methodologies for constant-pH simulation, while Chodera and Seeliger have worked to develop new bacterially-expressing kinase constructs and a novel fluorescence assay for measuring kinase inhibitor binding affinities. To facilitate the close collaboration this project requires, we will utilize modern communication technologies such as Slack, ensuring barrierless communication among members and PIs of all three groups. We will make use of shared public GitHub repositories to coordinate the development of simulation and analysis code and store simulation and experimental data. Frequent in-person visits between CCNY and MSKCC—the sites at which most software development will take place—are expected for the students heading software development. All project participants will receive accounts on the MSKCC High Performance Computing (HPC) resource, which provides a central location to carry out computation share large datasets. We will hold monthly virtual meetings for subgroups working on different Aims, quarterly virtual meeting of all PIs and personnel, and semiannual physical meetings at MSKCC to review progress and plans in depth. Schrödinger is also located in Manhattan, facilitating consultations, and reciprocal visits with collaborator Czodrowksi (Merck KGaA) occur several times a year for a collaboration focused on open source tools for automating alchemical binding affinity calculations.

The PI (Chodera) will be responsible for the oversight and coordination of the overall project. The Gunner and Chodera labs will work closely on the identification of candidate kinase:inhibitor systems (Aim 1), with Co-I Gunner directing work related to the MCCE code produced in her laboratory. The Gunner and Chodera labs will also work closely on developing protomer sampling methodologies and applying them to kinase:inhibitor systems (Aim 2), with Chodera directing this work. The Seeliger and Chodera labs will work closely on the experimental aspects of this project (Aim 3), with Co-I Seeliger directing experimental work in his laboratory (large-scale kinase expression, ITC optimization) and Chodera directing experimental work in his laboratory (automated ITC, automated screening of mutants, fluorescence binding assays).

Bibliography and References Cited

- [1] Intlekofer, A., Wang, B., Shah, H., Carmona-Fontaine, C., Rustenburg, A. S., Salah, S., Gunner, M. R., Chodera, J. D., Cross, J. R., and Thompson, C. B.: Acidification enhances production of L-2-hydroxyglutarate through alternative substrate use by dehydrogenase enzymes. *Nature Chem. Biol.* in press.
- [2] Rustenburg, A. S., Dancer, J., Lin, B., Feng, J. A., Ortwine, D. F., Mobley, D. L., and Chodera, J. D.: Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge. *J. Computer Aid. Mol. Des.* in press.
- [3] Mobley, D. L., Chodera, J. D., and Dill, K. A.: On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* 125: 084902, 2006. PMCID: PMC3583553.
- [4] Mobley, D. L., Chodera, J. D., and Dill, K. A.: Confine and release: Obtaining correct free energies in the presence of protein conformational change. *J. Chem. Theor. Comput.* 3: 1231–1235, 2007. PMCID: PMC2562444.
- [5] Shirts, M. R., Mobley, D. L., Chodera, J. D., and Pande, V. S.: Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B.* 111: 13052–13063, 2007.
- [6] Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., and Swope, W. C.: Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126: 155101, 2007.
- [7] Michael R. Shirts, D. L. M. and Chodera, J. D.: Alchemical free energy calculations: Ready for prime time? *Annu. Rep. Comput. Chem.* 3: 41–59, 2007.
- [8] Shirts, M. R. and Chodera, J. D.: Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 129: 124105, 2008. PMCID: PMC2671659.
- [9] Nilmeier, J. P., Crooks, G. E., Minh, D. D. L., and Chodera, J. D.: Nonequilibrium candidate Monte Carlo: A new tool for equilibrium simulation. *Proc. Natl. Acad. Sci. USA.* 108: E1009, 2011. PMCID: PMC3215031.
- [10] Chodera, J. D. and Shirts, M. R.: Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *J. Chem. Phys.* 135: 194110, 2011.
- [11] Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K. M., and Pande, V. S.: Alchemical free energy methods in drug discovery and design: Progress and challenges. *Curr. Opin. Struct. Biol.* 21: 150, 2011. PMCID: PMC3085996.
- [12] Wang, K., Chodera, J. D., Yang, Y., and Shirts, M. R.: Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput. Aided Mol. Des.* 27: 989, 2013. PMCID: PMC4154199.
- [13] Lee, T.-S., Chong, L. T., Chodera, J. D., and Kollman, P. A.: An alternative explanation for the catalytic proficiency of orotidine 5'-phosphate decarboxylase. *J. Am. Chem. Soc.* 123: 12837–12848, 2001.
- [14] Boyce, S. E., Tellinghuisen, J., and Chodera, J. D.: Avoiding accuracy-limiting pitfalls in the study of protein-ligand interactions with isothermal titration calorimetry. *bioRxiv.* <http://dx.doi.org/10.1101/023796> pp, 2015.
- [15] Tellinghuisen, J. T. and Chodera, J. D.: Systematic errors in isothermal titration calorimetry: Concentrations and baselines. *Anal. Biochem.* 414: 297, 2011.
- [16] Alexov, E. and Gunner, M. R.: Incorporating protein conformational flexibility into pH-titration calculations: Results on T4 lysozyme. *Biophys. J.* 74: 2075–2093, 1997. PMCID: PMC1184402.
- [17] Alexov, E. and Gunner, M. R.: A pragmatic approach to structure based calculation of coupled proton and electron transfer in proteins. *Biochim. Biophys. Acta.* 1458: 63–87, 2000.
- [18] Georgescu, R. E., Alexov, E. G., and Gunner, M. R.: Combining conformational flexibility and continuum electrostatics for calculating pK_a s in proteins. *Biophys. J.* 83: 1731–1748, 2002. PMCID: PMC1302268.
- [19] Song, Y., Mao, J., and Gunner, M. R.: MCCE2: Improving protein pK_a calculations with extensive side chain rotamer sampling. *J. Comput. Chem.* 30: 2231–2247, 2009. PMCID: PMC2735604.
- [20] Song, Y. and Gunner, M. R.: Using multiconformation continuum electrostatics to compare chloride binding motifs in α -amylase, human serum albumin, and Omp32. *J. Mol. Biol.* 387: 840–846, 2009.
- [21] Zheng, Z., Dutton, P. L., and Gunner, M. R.: The measured and calculated affinity of methyl and methoxy substituted benzoquinones for the QA site of bacterial reaction centers. *Proteins.* 78: 2638–2654, 2010. PMCID: PMC2920746.
- [22] Warncke, K., Gunner, M. R., Braun, B. S., Gu, L., Yu, C., Bruce, J. M., and Dutton, P. L.: Influence of hydrocarbon tail structure on quinone binding and electron-transfer performance at the QA and QB sites of the photosynthetic reaction center protein. *Biochem.* 33: 7830–7841, 1994.
- [23] Zhang, X. and Gunner, M. R.: Affinity and activity of non-native quinones at the Q(B) site of bacterial photosynthetic reaction centers. *Photosynth. Res.* 120: 181–196, 2013. PMCID: PMC4442677.

- [24] Nielsen, J. E., Gunner, M. R., and Garcia-Moreno, E. B.: The pK_a cooperative: a collaborative effort to advance structure-based calculations of pka values and electrostatic effects in proteins. *Proteins.* 79: 3249–3259, 2011. PMCID: PMC3375608.
- [25] Stanton, C. L. and Houk, K. N.: Benchmarking pK_a prediction methods for residues in proteins. *J. Chem. Theory Comput.* 4: 951–966, 2008.
- [26] Shan, Y., Seeliger, M. A., Eastwood, M. P., Frank, F., Xu, H., Jensen, M. O., Dror, R. O., Kuriyan, J., and Shaw, D. E.: A conserved protonation-dependent switch controls drug binding in Abl kinase. *Proc. Natl. Acad. Sci. USA.* 106: 139–144, 2009. PMCID: PMC2610013.
- [27] Seeliger, M. A., Young, M., Henderson, M. N., Pellicena, P., King, D. S., Falick, A. M., and Kuriyan, J.: High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci.* 14: 3135–3139, 2005. PMCID: PMC2253236.
- [28] Schneider, G.: Virtual screening: an endless staircase? *Nature Rev. Drug Discovery.* 9: 273–276, 2010.
- [29] Gathiaka, S., Liu, S., Chiu, M., Yang, H., Stuckey, J. A., Kang, Y. N., Delproposto, J., Kubish, G., Jr., J. B. D., Carlson, H. A., Burley, S. K., Walters, W. P., Amaro, R. E., Feher, V. A., and Gilson, M. K.: D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. *J. Computer Aid. Mol. Des.* 30: 651–668, 2016.
- [30] Bannan, C. C., Burley, K. H., Chiu, M., Shirts, M. R., Gilson, M. K., and Mobley, D. L.: Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge. *J. Computer Aid. Mol. Des.* in press.
- [31] Yin, J., Henrikson, N. M., Slochower, D. R., Shirts, M. R., Chiu, M. W., Mobley, D. L., and Gilson, M. K.: Overview of the SAMPL5 host-guest challenge: Are we doing better? *J. Computer Aid. Mol. Des.* in press.
- [32] Aleksandrov, A. and Simonson, T.: A molecular mechanics model for imatinib and imatinib:kinase binding. *J. Comput. Chem.* 31: 1550–1560, 2010.
- [33] Aleksandrov, A. and Simonson, T.: Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. *J. Biol. Chem.* 285: 13807–13815, 2010.
- [34] Lin, Y.-L., Meng, Y., Jiang, W., and Roux, B.: Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proc. Natl. Acad. Sci. USA.* 110: 1664–1669, 2013.
- [35] Hu, Y., Sherborne, B., Lee, T.-S., Case, D. A., York, D. M., and Guo, Z.: The importance of protonation and tautomerization in relative binding affinity prediction: A comparison of AMBER TI and Schrödinger FEP. *J. Computer Aid. Mol. Des.* 30: 533–539, 2016.
- [36] IV, F. C. P., König, G., Tofoleanu, F., Lee, J., Simmonett, A. C., Shao, Y., Ponder, J. W., and Brooks, B. R.: Blind prediction of distribution in the SAMPL5 challenge with QM based protomer and pK_a corrections. *J. Computer Aid. Mol. Des.* in press.
- [37] Aguilar, B., Anandakrishnan, R., Ruscio, J. Z., and Onufriev, A. V.: Statistics and physical origins of pK_a and ionization state changes upon protein-ligand binding. *Biophys. J.* 98: 872–880, 2010.
- [38] Baker, B. M. and Murphy, K. P.: Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophys. J.* 71: 2049–2055, 1996.
- [39] Steuber, H., Czdzrowski, P., Sotriffer, C. A., and Klebe, G.: Tracing changes in protonation: A prerequisite to factorize thermodynamic data of inhibitor binding to aldose reductase. *J. Mol. Biol.* 373: 1305–1320, 2007.
- [40] Dullweber, F., Stubbs, M. T., Musil, D., Stürzebecher, J., and Klebe, G.: Factorising ligand affinity: A combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J. Mol. Biol.* 313: 593–614, 2001.
- [41] Stubbs, M. T., Reyda, S., Dullweber, F., Möller, M., Klebe, G., Dorsch, D., Mederski, W. W. K. R., and Wurziger, H.: pH-dependent binding modes observed in trypsin crystals: Lessons for structure-based drug design. *ChemBioChem.* 02-03: 246, 2002.
- [42] Czdzrowski, P., Sotriffer, C. A., and Klebe, G.: Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pK_a calculations and ITC experiments. *J. Mol. Biol.* 367: 1347–1356, 2007.
- [43] Deininger, M., Buchdunger, E., and Druker, B. J.: The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood.* 105: 2640–2653, 2005.
- [44] Stegmeier, F., Warmuth, M., Sellers, W. R., and Dorsch, M.: Targeted cancer therapies in the twenty-first century: Lessons from imatinib. *Clin. Pharm. & Therap.* 87: 543–552, 2010.
- [45] Szakács, Z., Béni, S., Varga, Z., Örfi, L., Kéri, G., and Noszál, B.: Acid-base profiling of imatinib (Gleevec) and its fragments. *J. Med. Chem.* 48: 249, 2005.
- [46] Seeliger, M. A., Nagar, B., Frank, F., Cao, X., Henderson, M. N., and Kuriyan, J.: c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. *Structure.* 15:

- 299–311, 2007.
- [47] Jordan, I. K., Kondrashov, F. A., Adzhubei, I. A., Wolf, Y. I., Koonin, E. V., Kondrashov, A. S., and Sunyaev, S.: A universal trend of amino acid gain and loss in protein evolution. *Nature*. 433: 633–638, 2005.
- [48] Mongan, J., Case, D. A., and McCammon, J. A.: Constant pH molecular dynamics in a generalized born implicit solvent. *J. Comput. Chem.* 25: 2038–2048, 2004.
- [49] Mongan, J. and Case, D. A.: Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.* 15: 157–163, 2005.
- [50] Stern, H. A.: Molecular simulation with variable protonation states at constant pH. *J. Chem. Phys.* 126: 164112, 2007.
- [51] Chen, Y. and Roux, B.: Constant-pH hybrid nonequilibrium molecular dynamics-Monte Carlo simulation method. *J. Chem. Theor. Comput.* 11: 3919–3931, 2015.
- [52] Khandogin, J. and Brooks III, C. L.: Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* 89: 141–157, 2005.
- [53] Donnini, S., Tegeler, F., Groenhof, G., and Grubmüller, H.: Constant pH molecular dynamics in explicit solvent with λ -dynamics. *J. Chem. Theor. Comput.* 7: 1962–1978, 2011.
- [54] Goh, G. B., Hulbert, B. S., Zhou, H., and Brooks III, C. L.: Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins*. 82: 1319–1331, 2014.
- [55] Huang, Y., Chen, W., and Shen, J.: All-atom continuous constant pH molecular dynamics with particle mesh Ewald and titratable water. *J. Chem. Theor. Comput.* 12: 5411, 2016.
- [56] ten Brink, T. and Exner, T. E.: pK_a based protonation states and microspecies for protein-ligand docking. *J. Comput. Aided Mol. Des.* 24: 935–942, 2010.
- [57] Tembe, B. L. and McCammon, J. A.: Ligand-receptor interactions. *Computers & Chemistry*. 8: 281–283, 1984.
- [58] Lee, J., Miller, B. T., and Brooks, B. R.: Computational scheme for pH-dependent binding free energy calculation with explicit solvent. *Protein Sci.* 25: 231–243, 2015.
- [59] Dashi, D. S., Meng, Y., and Roitberg, A. E.: pH-replica exchange molecular dynamics in proteins using a discrete protonation method. *J. Phys. Chem. B.* 116: 8805–8811, 2012.
- [60] Michel, J. and Essex, J.: Hit identification and binding mode predictions by rigorous free energy simulations. *J. Med. Chem.* 51: 6654–6664, 2008.
- [61] Michel, J. and Essex, J. W.: Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aided Mol. Des.* 24: 639–658, 2010.
- [62] Gallicchio, E. and Levy, R. M.: Advances in all atom sampling methods for modeling protein–ligand binding affinities. *Curr. Opin. Struct. Biol.* 21: 161–166, 2011.
- [63] Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., Wester, R., Murcko, M., Frye, L., Farid, R., Lin, T., Mobley, D. L., Jorgensen, W. L., Berne, B. J., Friesner, R. A., and Abel, R.: 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. *J. Am. Chem. Soc.* 137: 2695–2703, 2015.
- [64] Shirts, M. R., Mobley, D. L., and Brown, S. P. Free energy calculations in structure-based drug design. In *Structure Based Drug Design*. Cambridge University Press, 2009.
- [65] Yank: A gpu-accelerated python framework for exploring algorithms for alchemical free energy calculations. <http://getyank.org>.
- [66] Eastman, P., Friedrichs, M., Chodera, J. D., Radmer, R., Bruns, C., Ku, J., Beauchamp, K., Lane, T. J., Wang, L.-P., Shukla, D., Tye, T., Houston, M., Stitch, T., and Klein, C.: OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J. Chem. Theor. Comput.* 9: 461, 2012. PMCID: PMC3539733.
- [67] Minh, D. D. L. and Chodera, J. D.: Optimal estimators and asymptotic variances for nonequilibrium path-ensemble averages. *J. Chem. Phys.* 131(13): 134110, 2009. PMCID: PMC2771048.
- [68] Minh, D. D. L. and Chodera, J. D.: Estimating equilibrium ensemble averages using multiple time slices from driven nonequilibrium processes: theory and application to free energies, moments, and thermodynamic length in single-molecule pulling experiments. *J. Chem. Phys.* 134: 024111, 2011.
- [69] Sivak, D. A., Chodera, J. D., and Crooks, G. E.: Using nonequilibrium fluctuation theorems to understand and correct errors in equilibrium and nonequilibrium simulations of discrete Langevin dynamics. *Phys. Rev. X.* 3: 011007, 2013.
- [70] Parton, D. L., Hanson, S. M., Rodríguez-Laureano, L., Albanese, S. K., Jeans, C., Gradia, S., Seeliger, M. A., and Chodera, J. D.: A panel of recombinant human kinase domain constructs for automated bacterial

- expression. *bioRxiv*. <http://dx.doi.org/10.1101/038711> pp, 2016.
- [71] Onufriev, A. V. and Alexov, E.: Protonation and pK_a changes in protein-ligand binding. *Quart. Rev. Biophys.* 46: 181–209, 2013.
- [72] Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchiyama, M.: Epik: A software program for pK_a prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 21: 681–691, 2007.
- [73] Milletti, F., Storchi, L., and Cruciani, G.: Predicting protein pK_a by environment similarity. *Proteins.* 50: 484–495, 2009.
- [74] Bochevarov, A. D., Watson, M. A., and Greenwood, J. R.: Multiconformation, density functional theory-based pK_a prediction in application to large, flexible organic molecules with diverse functional groups. *J. Chem. Theor. Comput.* in press.
- [75] Eckert, F., Diedenhofen, M., and Klamt, A.: Towards a first principles prediction of pK_a : COSMO-RS and the cluster-continuum approach. *Mol. Phys.* 108: 229–241, 2010.
- [76] Liao, C. and Nicklaus, M. C.: Comparison of nine programs predicting pK_a values of pharmaceutical substances. *J. Chem. Inf. Model.* 49: 2801, 2009.
- [77] Mobley, D. L. and Gilson, M. K.: Predicting binding free energies: Frontiers and benchmarks. *bioRxiv*. <http://dx.doi.org/10.1101/074625> pp, 2016.
- [78] Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I.: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. method. *J. Comput. Chem.* 21(2): 132–146, 2000.
- [79] Jakalian, A., Jack, D. B., and Bayly, C. I.: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J. Comput. Chem.* 23(16): 1623–1641, 2002.
- [80] van Linden, O. P. J., Kooistra, A. J., Leurs, R., de Esch, I. J. P., and de Graaf, C.: KLIFS: A knowledge-based structural database to navigate kinase-ligand interaction space. *J. Med. Chem.* 57: 249–277, 2014.
- [81] Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J., and Baker, D.: High-resolution structure prediction and the crystallographic phase problem. *Nature*. 450(7167): 259–264, November 2007.
- [82] Wang, C., Bradley, P., and Baker, D.: Protein-Protein Docking with Backbone Flexibility. *Journal of Molecular Biology.* 373(2): 503–519, October 2007.
- [83] Mandell, D. J., Coutsias, E. A., and Kortemme, T.: Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods.* 6: 551–552, 2009.
- [84] Tan, Z.: Optimally adjusted mixture sampling and locally weighted histogram analysis. *J. Comput. Graph. Stat.* in press.
- [85] Lyubartsev, A. P., Martsinovski, A. A., Shevkunov, S. V., and Vorontsov-Velyaminov, P. N.: New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.* 96: 1776–1783, 1992.
- [86] Wang, F. and Landau, D. P.: Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* 86: 2050–2053, 2001.
- [87] Li, H., Fajer, M., and Yang, W.: Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J. Chem. Phys.* 126: 024106, 2007.
- [88] Lelièvre, T., Stoltz, G., and Rousset, M.: *Free energy computations: A mathematical perspective*. Imperial College Press, 1st ed. edition, June 2002.
- [89] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R.: A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* 76(1): 637–649, 1982.
- [90] Leimkuhler, B. and Matthews, C.: Efficient molecular dynamics using geodesic integration and solvent-solute splitting. *Proc. Royal Soc. A.* 472, in press.
- [91] Sivak, D. A. and Crooks, G. E.: Thermodynamic metrics and optimal paths. *Phys. Rev. Lett.* 108: 190602, 2012.
- [92] Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Fischbach, M., Held, M., Chodera, J. D., Schütte, C., and Noé, F.: Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* 134: 174105, 2011.
- [93] Liu, J. S.: *Monte Carlo strategies in scientific computing*. New York, Springer-Verlag, 2nd ed. edition, October 2002.
- [94] Swails, J., York, D., and Roitberg, A.: Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *J. Chem. Theor. Comput.* 10: 1341–1352, 2014.
- [95] Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C. J., McLauchlan, H., Klevernic, I., Arthur, J. S. C., Aless, D. R., and Cohen, P.: The selectivity of protein kinase inhibitors: A further update. *Biochem. J.* 408: 297–315, 2007.

- [96] Fabian, M. A., Biggs, W. H., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M., Ford, J. M., Galvin, M., Gerlach, J. L., Grotfeld, R. M., Herrgard, S., Insko, D. E., Insko, M. A., Lai, A. G., Lélias, J.-M., Mehta, S. A., Milanov, Z. V., Velasco, A. M., Wodicka, L. M., Patel, H. K., Zarrinkar, P. P., and Lockhart, D. J.: A small molecule–kinase interaction map for clinical kinase inhibitors. *Nature Biotech.* 23: 329–336, 2005.
- [97] Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., Faraoni, R., Floyd, M., Hunt, J. P., Lockhart, D. J., Milanov, Z. V., Morrison, M. J., Pallares, G., Patel, H. K., Pritchard, S., Wodicka, L. M., and Zarrinkar, P. P.: A quantitative analysis of kinase inhibitor selectivity. *Nature Biotech.* 26: 127–132, 2008.
- [98] Boresch, S., Tettinger, F., Leitgeb, M., and Karplus, M.: Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. A.* 107(35): 9535–9551, 2003.
- [99] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A.: Development and testing of a general AMBER force field. *J. Comput. Chem.* 25: 1157–1174, 2004.
- [100] Czodrowski, P., Sotriffer, C. A., and Klebe, G.: Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pKa calculations and ITC experiments. *J. Mol. Biol.* 367: 1347–1356, 2007.
- [101] Steuber, H., Czodrowski, P., Sotriffer, C. A., and Klebe, G.: Tracing changes in protonation: A prerequisite to factorize thermodynamic data of inhibitor binding to aldose reductases. *J. Mol. Biol.* 373: 1305–1320, 2007.
- [102] Lavinder, J. L., Hari, S. B., Sullivan, B. J., and Magliery, T. J.: High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J. Am. Chem. Soc.* 131: 3794–3795, 2009.
- [103] Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J., and Weiss, M. S.: Optimization of protein buffer cocktails using Thermofluor. *Acta Cryst. F*96: 209–214, 2013.
- [104] Velazquez-Campoy, A., , and Freire, E.: Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nature Protocols.* 1: 186–191, 2006.
- [105] Levinson, N. M. and Boxer, S. G.: Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. *PLoS One.* 7: e29828, 2012.
- [106] Song, C. W., Griffin, R., and Park, H. J. *Cancer Drug Discovery and Development*, chapter Influence of Tumor pH on Therapeutic Response, pp 21–42. Humana Press, Totowa, New Jersey, 2006.