

Protocols Utilizing Constant pH Molecular Dynamics to Compute pH-Dependent Binding Free Energies

M. Olivia Kim,^{†,‡} Patrick G. Blachly,^{†,‡} Joseph W. Kaus,[†] and J. Andrew McCammon^{*,†,‡,§,||}

[†]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California 92093, United States

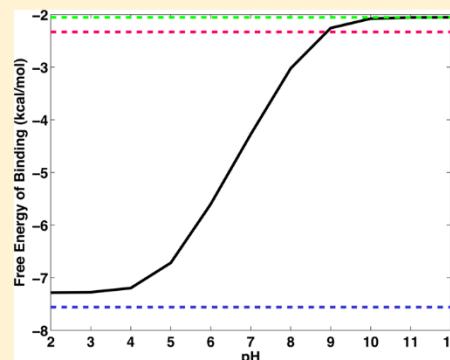
[‡]Department of Pharmacology, University of California San Diego, La Jolla, California 92093, United States

[§]Howard Hughes Medical Institute, University of California San Diego, La Jolla, California 92093, United States

^{||}Center for Theoretical Biological Physics, University of California San Diego, La Jolla, California 92093, United States

S Supporting Information

ABSTRACT: In protein–ligand binding, the electrostatic environments of the two binding partners may vary significantly in bound and unbound states, which may lead to protonation changes upon binding. In cases where ligand binding results in a net uptake or release of protons, the free energy of binding is pH-dependent. Nevertheless, conventional free energy calculations and molecular docking protocols typically do not rigorously account for changes in protonation that may occur upon ligand binding. To address these shortcomings, we present a simple methodology based on Wyman’s binding polynomial formalism to account for the pH dependence of binding free energies and demonstrate its use on cucurbit[7]uril (CB[7]) host–guest systems. Using constant pH molecular dynamics and a reference binding free energy that is taken either from experiment or from thermodynamic integration computations, the pH-dependent binding free energy is determined. This computational protocol accurately captures the large pK_a shifts observed experimentally upon CB[7]:guest association and reproduces experimental binding free energies at different levels of pH. We show that incorrect assignment of fixed protonation states in free energy computations can give errors of >2 kcal/mol in these host–guest systems. Use of the methods presented here avoids such errors, thus suggesting their utility in computing proton-linked binding free energies for protein–ligand complexes.



INTRODUCTION

The changes in the electrostatic environment that accompany binding of small molecules, nucleic acids, or other proteins may thus induce changes in the protonation states of titratable groups in the protein.^{1–8} Recently, Aguilar et al. conducted a computational survey of various protein–protein, protein–small molecule, and protein–nucleic acid complexes to ascertain the prevalence of protonation change in the protein receptor upon biomolecular association. Notably, in 60% of the protein–small molecule complexes considered, at least one titratable residue in the protein was found to assume different protonation states in its free and bound states.⁹ Furthermore, protonation changes that accompany small molecule binding to proteins are not limited to the protein partner: an estimated 60–80% of orally administered drugs are weak acids or bases, whose protonation states can also be tuned by the cellular pH and electrostatic environment of their protein binding partners.^{10–13} In cases where protein–ligand binding accompanies a net transfer of protons to either binding partner, the binding process is pH-dependent; i.e., the observed binding free energy is a function of pH.

Conventionally, both computational docking and more rigorous free energy computations, such as the thermodynamic integration (TI) and free energy perturbation (FEP) methods,

employ fixed protonation states that are identical for free and bound states in the computation of binding affinities. Clearly, in cases where ligand binding is linked to the (un)binding of protons, such approximations will lead to error. Improper assignment of protonation states in binding free energy computations may result in significant errors, making correct assignment of pK_a and protonation state essential to obtaining accurate free energies.

As stated, simulations of protein–ligand systems are typically preceded by the assignment of fixed protonation states to titratable groups on the two binding partners, often using programs such as H++^{14–16} and PROPKA^{17–20} to do so. Further, docking studies often employ empirical prediction algorithms, which sometimes use Hammett and Taft relations, to assign fixed protonation states to the free ligands being docked.^{21,22} These approaches, however, fail to account for changes in protonation that may follow from the altered electrostatic environment surrounding the two binding partners upon complex formation. Several computational methods,

Special Issue: William L. Jorgensen Festschrift

Received: June 10, 2014

Revised: August 8, 2014

Published: August 18, 2014

however, have been developed that permit the protonation of titratable residues to respond to changes in the electrostatic environment.^{2,23–28} For instance, various flavors of constant pH molecular dynamics (CpHMD) methodologies have emerged to incorporate pH as an added external thermodynamic parameter to conventional molecular dynamics (MD) simulations, allowing fluctuations in the protonation of titratable residues to accompany conformational sampling.^{29–34} To date, CpHMD simulations have been used to successfully predict pK_a values of titratable groups in proteins^{29–37} and nucleic acids,^{38–40} as well as to explain the mechanism behind the pH-dependent conformational changes critical to the function of proteins such as nitrophorin⁴¹ and rhodopsin.⁴²

The CpHMD method provides a framework through which the pH dependence of binding processes can be examined. To the best of our knowledge, there is currently no standard protocol available to rigorously account for proton-linked ligand binding. Multiple experimental and computational groups, however, have utilized the binding polynomial formalism devised by Wyman⁴³ to calculate the changes in binding free energy that accompany binding-induced protonation changes for both protein–protein^{34,44} and protein–nucleic acid binding.^{39,45,46} Motivated by Mason and Jensen's usage of this binding polynomial formalism to estimate the free energies of binding for protein–protein complexes using the PROPKA web server,⁴ we adopt a similar approach in conjunction with the CpHMD method by Mongan et al.³³ to obtain pH-dependent free energy profiles *in silico* for the binding of small molecules to the cucurbit[7]uril (CB[7]) host.

CB[7] is a synthetic molecule with seven repeating glycoluril units bridged by methylene groups (Figure 1).^{47,48} This 7-fold

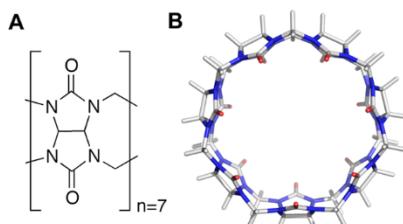


Figure 1. Structure of cucurbit[7]uril (CB[7]) host: (A) glycoluril unit; (B) top view of CB[7].

symmetric host has gained much attention due to its ability to encapsulate drug-like small molecules with high affinity as a stable host–guest complex.^{49–55} Benzimidazole (BZ) and a series of its derivatives (Figure 2) comprise a class of widely used fungicides and antihelmintic drugs^{56–58} that have been shown to bind to the CB[7] host and undergo the pK_a shifts as large as 4 pK units upon complex formation (Table 1).⁵⁹ At neutral pH, these weakly acidic guests are predominantly deprotonated when free in solution, but each binds a single proton upon encapsulation by CB[7]. Both the acid/base behaviors of BZ-derived guests and the small size and relative rigidity of CB[7] compared to a typical biomolecule make the CB[7]:BZ complexes ideal model systems to test theoretical methods for computing pH-dependent binding free energies.

In this work, we accurately reproduce the pK_a shifts of the various BZ derivatives upon binding to CB[7], using CpHMD simulations. Coupling these pK_a data with reference binding free energies taken either from experiment or from TI computations allows us to obtain a full description of CB[7]:guest binding free

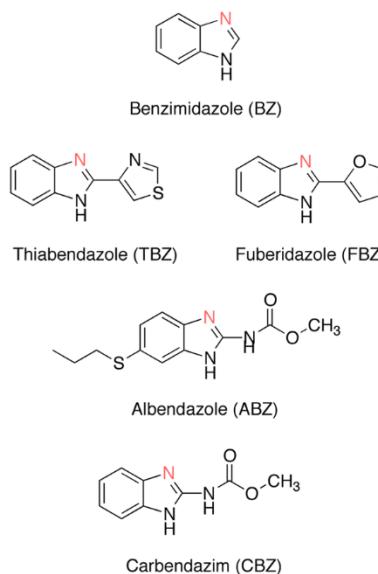


Figure 2. Chemical structures of benzimidazole and its derivatives.

Table 1. Experimental pK_a Shifts of Benzimidazole Guests upon Binding to CB[7]⁵⁹ ^a

guest	pK_a^F	$pK_a^{C,\text{exp}}$	ΔpK
BZ	5.5	9.0	3.5
TBZ	4.6	8.6	4.0
FBZ	4.8	8.6	3.8
ABZ	3.5	6.1	2.6
CBZ	4.5	7.0	2.5

^a pK_a^F denotes the pK_a of the free guest, and pK_a^C represents the pK_a of the guest in complex with CB[7].

energies as functions of pH. Additionally, we show that improper assignment of guest protonation states in binding free energy computations can produce errors in excess of 2 kcal/mol at neutral pH, highlighting the importance of accurately accounting for the pH effects in free energy calculations or docking.

THEORY

Binding Polynomial Formalism for Computing the pH Dependence of Binding Free Energies.

Mason and Jensen recently examined the pH dependence of protein–protein binding⁴ through an application of the binding polynomial formalism developed by Wyman⁴³ and used by Tanford to describe protein folding/unfolding.⁶⁰ Following the theoretical foundations of these groups, the binding of a titratable ligand (L) to a general macromolecular receptor (R) can be considered through a general equation for ligand association governed by the apparent equilibrium constant, K_{app} :

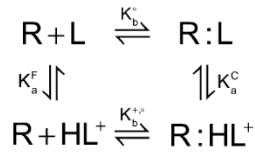


where the brackets indicate that the ligand (L) and complex (RL) ensembles may contain different protonated forms of the titratable ligand species. In the case of a ligand with a single titratable site binding to CB[7], which itself does not titrate in the biological range of pH levels, K_{app} can be written as

$$K_{\text{app}} = \frac{[\text{LR}] + [\text{HL}^+]}{[\text{R}]([\text{L}] + [\text{HL}^+])} \quad (2)$$

where the concentrations, rather than activities, of the given species are reported assuming ideal dilute solutions. Building from the thermodynamic cycle used to describe the proton-linked ligand binding to CB[7] (Scheme 1), K_{app} can be rewritten

Scheme 1. Thermodynamic Cycle for Complex Formation between a Receptor (R) and a Titratable Ligand (L)



according to eq 3, in which the concentrations of all species are presented in binding polynomials with respect to the concentrations of the deprotonated complex and ligand species:

$$K_{\text{app}} = \frac{[\text{LR}] \left(1 + \frac{[\text{HL}^+]}{[\text{LR}]} \right)}{[\text{R}][\text{L}] \left(1 + \frac{[\text{HL}^+]}{[\text{L}]} \right)} = K_b^* \frac{\left(1 + \frac{[\text{HL}^+]}{[\text{LR}]} \right)}{\left(1 + \frac{[\text{HL}^+]}{[\text{L}]} \right)} \quad (3)$$

Using the acid dissociation constants for the free ligand (K_a^F) and ligand–receptor complex (K_a^C), as illustrated by the vertical reactions in Scheme 1 (eqs 4 and 5),

$$K_a^F = \frac{[\text{HL}^+]}{[\text{L}]a_{\text{H}^+}} \quad (4)$$

$$K_a^C = \frac{[\text{HL}^+]}{[\text{LR}]a_{\text{H}^+}} \quad (5)$$

where the proton activity is denoted by a_{H^+} , eq 3 can be rewritten in terms of the overall free energy of binding for the ligand L to the receptor R (ΔG°):

$$\begin{aligned} \Delta G^\circ(\text{pH}) &= -k_B T \ln K_b^* - k_B T \ln \left(\frac{1 + 10^{pK_a^C - \text{pH}}}{1 + 10^{pK_a^F - \text{pH}}} \right) \\ &= \Delta G_{\text{ref}}^\circ - k_B T \ln \left(\frac{1 + 10^{pK_a^C - \text{pH}}}{1 + 10^{pK_a^F - \text{pH}}} \right) \end{aligned} \quad (6)$$

where the proton activity and acid dissociation constants have been converted to their respective logarithmic constants, pH and pK_a . The pH dependence of the binding free energy can thus be obtained having only the pK_a values of the ligand molecule free in solution (pK_a^F) and in complex with the receptor (pK_a^C), as well as the free energy of binding for a reference reaction shown in eq 6 (the top reaction in Scheme 1), $\Delta G_{\text{ref}}^\circ$, in which there is no net uptake or release of protons. This formalism for obtaining ΔG° as a function of pH can further be applied to cases where multiple ligand and receptor groups titrate in the pH range considered, assuming that proton binding occurs independently. In other words, eq 6 can only be applied when all titratable groups are uncoupled from each other.

As protein active sites often contain multiple titratable groups whose protonation states are coupled to perform a given function, it will sometimes be wrong to assume that all titratable groups remain uncoupled upon ligand binding. In such cases, Wyman^{43,61} derived a relation between K_{app} and pH such that

$$\frac{\partial \ln K_{\text{app}}}{\partial \ln [\text{H}^+]} = \Delta \nu_{\text{H}^+} = Z_{\text{LR}} - (Z_{\text{L}} + Z_{\text{R}}) \quad (7)$$

where, using the notation used by Tanford,⁶⁰ $\Delta \nu_{\text{H}^+}$ is the change in the number of bound protons in the receptor–ligand complex, relative to the number of protons bound to the ligand and receptor individually. Utilizing the unit charge of a proton, this relation is equivalent to the difference in total charge, Z , between reactants and product in eq 1. With $\Delta Z = Z_{\text{LR}} - (Z_{\text{L}} + Z_{\text{R}})$, integration of eq 7 provides a thermodynamic relation that holds for proton-linked ligand binding in cases where titratable sites may interact (eq 8):

$$\begin{aligned} \Delta G^\circ(\text{pH}) &= \Delta G_{\text{ref}, \text{pH}}^\circ - k_B T \ln(10) \\ &\quad \int_{\text{pH}_{\text{ref}}}^{\text{pH}} \{Z_{\text{LR}}(\text{pH}) - Z_{\text{L}}(\text{pH})\} \, d\text{pH} \end{aligned} \quad (8)$$

where Z_{R} is omitted, since the CB[7] receptor under consideration does not titrate in the pH range considered in this study. Since the integration is performed with respect to pH in the second term in eq 8, the reference binding free energy corresponds to the binding free energy at a specific pH.

Both eqs 6 and 8 thus provide frameworks for computing the pH-dependent binding free energy by adding a correction term to the reference free energy of binding. In the case of eq 6, the reference free energy, $\Delta G_{\text{ref}}^\circ$, is obtained for receptor–ligand binding with protonation states fixed, such that no net change of protonation occurs. Analogously, the reference free energy in eq 8 is required to be the free energy of binding at a given value of pH. These two reference free energies are not necessarily equivalent; however, the reference reaction can be chosen such that they have the same value.

Constant pH Molecular Dynamics. Baptista and co-workers developed constant pH molecular dynamics (CpHMD) with stochastic titration to enable concurrent sampling of both conformational and protonation spaces according to the semigrand canonical ensemble.³² Here, we use the simplified CpHMD formulation implemented in the standard release of AMBER 12⁶² that is similar to Baptista's formulation except that the simulation is performed in implicit solvent with generalized Born electrostatics.³³ In this method, an MD simulation is propagated from initial sets of coordinates and protonation states. After a chosen number of MD steps, the simulation is halted, at which point a Monte Carlo (MC) step evaluates whether a random titratable residue in the system should change protonation states. The acceptance of this new protonation state is contingent on the application of the Metropolis criterion to the computed transition free energy, ΔG_{trans} , obtained using eq 9, where pH enters as an external thermodynamic parameter and $k_B T$ is the Boltzmann constant multiplied by the temperature of the system.

$$\Delta G_{\text{trans}} = k_B T (\text{pH} - pK_{a,\text{ref}}) \ln 10 + \Delta G_{\text{elec}} - \Delta G_{\text{elec,ref}} \quad (9)$$

For the value of pH at which the simulation is conducted, the difference in electrostatic free energy that accompanies the change in protonation being considered, ΔG_{elec} , is computed with respect to the difference in electrostatic free energy that accompanies the analogous change in protonation for a model compound, $\Delta G_{\text{elec,ref}}$, which has a known pK_a value ($pK_{a,\text{ref}}$). In this manner, any nonclassical contributions to the transition free energy cancel. For a given CB[7]:guest system, the model compound that enters eq 9 is the guest molecule free in solution, its $pK_{a,\text{ref}}$ is the experimentally obtained pK_a value of the free guest (pK_a^F , Table 1), and $\Delta G_{\text{elec,ref}}$ is defined to be the electrostatic free energy that equally populates the protonated

and deprotonated forms of the free guest when the solution pH is equal to the experimental pK_a of the free guest. If the transition is accepted, then MD is continued with the new protonation state for the titratable residue; otherwise, MD continues without change in the protonation state. Repeated application of these steps builds an ensemble of protonation states along the MD trajectory.

Use of Constant pH Molecular Dynamics in the Binding Polynomial Scheme.

The CpHMD method is applied to obtain values for pK_a^C in eq 6 and ΔZ in eq 8 to provide pH-dependent correction terms to the reference binding free energies. In the case of eq 6, values of pK_a^C are obtained from simulating the CB[7]:guest system at a range of pH values. Each CpHMD simulation obtains a fractional protonation for the titratable guest being considered. By tabulating the fraction of deprotonated guest species (s) computed at each value of pH, application of the Hill equation can be used to predict pK_a^C as the midpoint of the titration, as well as the Hill coefficient, n (eq 10):

$$s = \frac{1}{1 + 10^{n(pK_a^C - \text{pH})}} \quad (10)$$

This method can reliably extract the pK_a when the titratable residue exhibits typical titration behavior.⁶³ In all fits, the Hill coefficient obtained is approximately 1, which is anticipated in the absence of cooperativity.

In the case of eq 8, the partial charges for the guest free in solution, Z_L , and the partial charges for the guest in complex with CB[7], Z_{LR} , can similarly be obtained from CpHMD simulations. BZ and its derivatives have charges of +1 when protonated and 0 when deprotonated. Consequently, Z_{LR} and Z_L are equivalent to the fraction of protonated species ($1 - s$) obtained from CpHMD simulations performed on the CB[7]:guest complex and the free guest, respectively.

METHODS

Parameterization of CB[7] and Benzimidazole Ligands for Molecular Dynamics Simulations. Partial charges for the CB[7] host have previously been derived⁶⁴ using the restrained electrostatic potential (RESP) procedure,^{65–67} conventionally used to parameterize nonstandard residues for molecular simulations performed with AMBER force fields. Analogously, the geometries of benzimidazole (BZ), albendazole (ABZ), carbendazim (CBZ), fuberidizole (FBZ), and thiabendazole (TBZ, Figure 2) are optimized at the B3LYP/6-31G(d) level of theory^{68–71} using the Gaussian 09 suite of programs.⁷² Subsequently, the electrostatic potentials (ESPs) associated with the optimized geometries of these guests are computed using MK radii⁷³ at the HF/6-31G(d) level of theory. The ESPs of the different guest molecules are submitted to the antechamber module⁶⁷ in the AmberTools 12 suite of programs,⁶² which applies the RESP procedure to extract atomic point charges for use in molecular dynamics (MD) simulations. All other CB[7] and guest ligand force field terms, including Lennard-Jones parameters, are taken from the general AMBER force field (GAFF).⁷⁴

Docking of Guest Molecules to CB[7]. To generate starting coordinates for MD simulations of different CB[7]:guest complexes, the various BZ derivatives are docked rigidly into the CB[7] cavity using the extra precision mode (XP) in Schrodinger's Glide program.^{75–77} Each CB[7]:guest docking experiment yields a single pose for the CB[7]:guest complex, and all guest molecules bind CB[7] similarly. For illustrative

purposes, the resulting CB[7]:FBZ complex obtained from Glide is shown in Figure 3. It is also worth noting that the docked

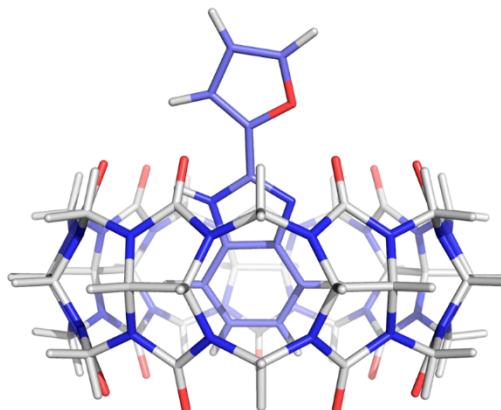


Figure 3. Structure of CB[7]:fuberidazole complex generated by docking.

poses of deprotonated and protonated guests in complex with CB[7] are similar. The hydrophobic core of the BZ guests is encapsulated by the CB[7] cavity, orienting the ligands similarly regardless of protonation, while additional furanyl, thiazole, amido, or thioether R-groups seen in the BZ derivatives protrude outside of the entrance to CB[7]. All poses show good agreement with experiments.⁵⁹

Constant pH Molecular Dynamics Simulation Details.

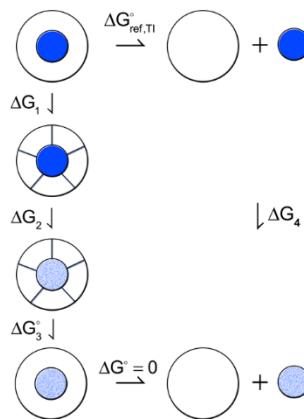
CpHMD simulations are performed using the AMBER 12 suite of programs for the range of pH values between 2 and 12 at increments of 0.5.^{33,62} All simulations employ the OBC generalized Born (GB) implicit solvent model (igb = 5)⁷⁸ with a salt concentration of 0.1 M. Starting from the docked CB[7]:guest structures, all systems are minimized for 5000 steps while applying positional constraints to all heavy atoms with a force constant of 20 kcal/mol Å². Following minimization, the system is heated to 300 K over the course of 500 ps using a Langevin thermostat⁷⁹ while maintaining the positional constraints applied to all heavy atoms with a force constant of 5 kcal/mol Å². After heating, a 1 ns equilibration simulation is performed at 300 K. Production simulations are then performed for 5 ns, with MC steps taken every 10 fs. In all equilibration and production steps, the bonds involving hydrogen are constrained using the SHAKE algorithm⁸⁰ and a cutoff of 30 Å for the computation of nonbonded interactions is enforced.

Computing Absolute Binding Free Energy with Thermodynamic Integration. The calculation of the pH-dependent binding free energy requires a reference binding energy obtained either in the absence of protonation change (eq 6) or at a specified pH value (eq 8). TI computations are performed to obtain the absolute binding free energy between CB[7] and guest molecules that are deprotonated both free in solution and in complex. In TI, the free energy change is evaluated as

$$\Delta G_{0 \rightarrow 1} = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle d\lambda \quad (11)$$

where U is the total potential energy of the system coupled to λ , which varies smoothly between the initial state of $\lambda = 0$ and the final state of $\lambda = 1$.⁸¹ The reference binding free energy is obtained from the thermodynamic cycle shown in Scheme 2 and is calculated using

Scheme 2. Thermodynamic Cycle for an Absolute Binding Free Energy Calculation^a



^aThe outer circle represents a CB[7] host, and the inner blue circle shows a guest molecule in the reference deprotonated state. ΔG_1 is the free energy for gradually turning on the restraints; ΔG_2 is for decoupling the guest from the host in the presence of the restraints; ΔG_3° is the analytical correction for removing the restraints; and ΔG_4 is the solvation free energy for the guest.

$$\Delta G_{\text{ref},\text{TI}}^\circ = -\Delta G_1 - \Delta G_2 - \Delta G_3^\circ + \Delta G_4 \quad (12)$$

where ΔG_1 is the free energy for gradually turning on restraints (see below), ΔG_2 is the free energy for decoupling the guest while bound to the host in the presence of the restraints, ΔG_3° is the free energy for turning off the restraint and correcting for the standard state, and ΔG_4 is the solvation free energy for the decoupled guest (Scheme 2).

The electrostatic and van der Waals (vdW) contributions to ΔG_2 and ΔG_4 are computed separately, the latter using the softcore potential algorithm.^{82–84} To improve the convergence for these computations, the virtual bond algorithm developed by Karplus and co-workers is applied, where a set of restraints are used to fix the position and orientation of the guest relative to CB[7]. The free energy for turning on the restraint, ΔG_1 , is computed using TI. The free energy for turning off the restraint, ΔG_3° , is calculated using an analytical expression, which corrects for the presence of restraints and also accounts for the standard state:⁸⁵

$$\Delta G_3^\circ = -k_B T \ln \left[\frac{8\pi^2 V^\circ}{r_{\text{A},\circ}^2 \sin \theta_{\text{A},\circ} \sin \theta_{\text{B},\circ}} \frac{(K_r K_{\theta_{\text{A}}} K_{\theta_{\text{B}}} K_{\phi_{\text{A}}} K_{\phi_{\text{B}}} K_{\phi_{\text{C}}})^{1/2}}{(2\pi k_B T)^3} \right] - k_B T \ln \frac{\sigma_{\text{R...L}}}{\sigma_{\text{R}} \sigma_{\text{L}}} \quad (13)$$

Here V° is the standard state volume of 1661 Å³ for ideal gas, $r_{\text{A},\circ}$ sin θ_{A} and sin θ_{B} are the distance and angle values used for each restraint, having corresponding harmonic force constants (K 's in eq 13), which are 5 kcal/mol Å² for the distance restraint and 20 kcal/mol rad² for the angle and dihedral restraints. The second term in eq 13 accounts for the symmetry in the system, where $\sigma_{\text{R...L}}$, σ_{R} and σ_{L} are the symmetry numbers for the host–guest complex, CB[7], and the guest molecule, respectively. For our system, $\sigma_{\text{R...L}}$ and σ_{L} are 1 and σ_{R} is 14.

The *pmemd* implementation of TI in AMBER 14 is used to calculate the reference binding free energy $\Delta G_{\text{ref},\text{TI}}^\circ$ for each guest to CB[7].^{86,87} The reference ionization state is chosen to be

deprotonated because all experimental values of $\Delta G_{\text{ref}}^\circ$ were measured with the guests deprotonated.⁵⁹ For the calculation of the electrostatic contribution to ΔG_2 and ΔG_4 , 11 equally spaced λ values are used (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0). For the calculation of the vdW contribution to ΔG_2 and ΔG_4 , 21 λ values are used (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.725, 0.75, 0.775, 0.8, 0.825, 0.85, 0.875, 0.9, 0.925, 0.95, 0.975, 1.0). For the computation of the free energy for turning on the restraints, ΔG_1 , 16 λ values are used (0.0, 0.01, 0.02, 0.03, 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0). The unequal spacing of λ windows is needed to capture a smoother transition of $\partial U(\lambda)/\partial \lambda$ along the λ parameter and reduce errors in integration. Integration is performed numerically using the trapezoidal rule, and uncertainties in the free energies are propagated as standard deviations.

Each CB[7]:guest complex was solvated with TIP3P water⁸⁸ with a region of 12 Å in any direction using the leap program.⁸⁹ The system was minimized for 5000 steps and heated to 300 K over 500 ps in the NVT ensemble using a Langevin thermostat,⁷⁹ followed by an equilibration for 500 ps in the NPT ensemble using a Berendsen barostat⁹⁰ with isotropic position scaling to bring the system to a stable density. All production simulations are performed in the NVT ensemble and are extended until the cumulative free energy computed for each individual transformation converges (changes in $\Delta G < 0.01$ kcal/mol).

RESULTS

Review of Experimental Results. Previously, Koner et al. observed enhancements in stabilities and solubilities of benzimidazole (BZ) derivatives upon encapsulation by the cucurbit[7]uril (CB[7]) host.⁵⁹ The authors obtained values of pK_a^F and pK_a^C by fitting the data from UV titrations and ¹H NMR spectroscopy.⁹¹ Henceforth, $pK_a^{\text{F,exp}}$ and $pK_a^{\text{C,exp}}$ will differentiate experimental pK_a values from their respective computed values, $pK_a^{\text{F,calc}}$ and $pK_a^{\text{C,calc}}$. The experimental data showed large shifts in pK_a ranging between 2.5 and 4 pK units upon complex formation with CB[7] (Table 1). Additionally, association constants of the complexes were obtained at basic pH where guests were presumably deprotonated in both bound and unbound states; association constants were also obtained for the binding of protonated guests through application of the thermodynamic cycle (see Scheme 2 in ref 59). In all cases, measurements of the binding free energies for different CB[7]:guest complexes indicated that the protonated guests are favored in the CB[7] cavity (discussed later; see Table 3).

pK_a Shifts upon CB[7]:Guest Complex Formation. To compute the pK_a values of various BZ derivatives in complex with CB[7], we perform CpHMD simulations on five CB[7]:guest complexes. In Figure 4, representative titration curves are shown for benzimidazole (BZ) and albendazole (ABZ), both in complex with CB[7] and free in solution. Similar curves corresponding to the other guests can be found in Figure S1 (Supporting Information). Titration of the free guest in solution offers preliminary examination of the ability of the CpHMD to reproduce the experimental pK_a ; for example, in the case of BZ, the $pK_a^{\text{F,calc}}$ value matches the $pK_a^{\text{F,exp}}$ value of 5.5, indicating proper calibration of the CpHMD method. From the titration curve of BZ free in solution (Figure 4A, green curve), it is apparent that free BZ is protonated at values of pH less than 4.5 and deprotonated at pH levels above 6.5. Between these pH levels, an ensemble of protonated and deprotonated states exists. Relative to the titration curve for free BZ, the titration curve for the CB[7]:BZ complex is shifted toward more basic values of pH

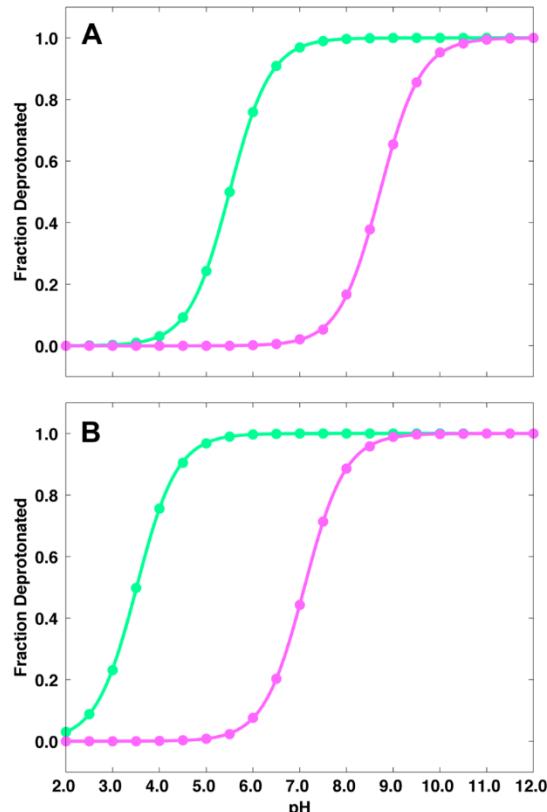


Figure 4. Titration curves from constant pH MD simulations of the guests free in solution (green) and in complex with CB[7] (purple): (A) benzimidazole; (B) albendazole.

(Figure 4A, purple curve). Indeed, the value of $pK_a^{C,\text{calc}}$ for BZ is found to be 8.7—a shift of more than 3 pK units above its pK_a^F (Table 2); consequently, complexed BZ is protonated at pH

Table 2. Comparison of pK_a^C Values Obtained from CpHMD Simulations ($pK_a^{C,\text{calc}}$) with Experimental Data ($pK_a^{C,\text{exp}}$)⁵⁹ ^a

guest	$pK_a^{C,\text{exp}}$	$pK_a^{C,\text{calc}}$	n
BZ	9.0	8.71 ± 0.01	0.99
TBZ	8.6	8.19 ± 0.01	1.01
FBZ	8.6	8.61 ± 0.01	1.01
ABZ	6.1	7.10 ± 0.01	0.99
CBZ	7.0	7.40 ± 0.01	1.03

^aThe corresponding Hill coefficients (n) are also shown.

below 7.5, indicating the preferred protonation state of BZ at neutral (typical physiological) pH differs depending on its bound state. The observed preference for the protonated guest in the cavity of CB[7] is due to the additional hydrogen bond between the titratable proton on BZ and one of the carbonyl oxygens at the entrance to the CB[7] cavity (Figure 5). It is worth noting that the Hill equation provides a reasonable estimate of the pK_a values for BZ both free and in complex with CB[7], with fitting errors of ~ 0.01 pK units. Furthermore, the $pK_a^{C,\text{calc}}$ value for BZ underestimates its $pK_a^{C,\text{exp}}$ by only 0.3 pK units (Table 2).

The chemical structure of ABZ differs from that of BZ by the presence of amido and thioether R-groups attached to the BZ core. Additionally, the experimentally determined pK_a for the CB[7]:ABZ complex remains acidic ($pK_a^{C,\text{exp}} = 6.1$). The titration curves obtained from CpHMD simulations of free and complexed ABZ are shown in Figure 4B. Qualitatively, the

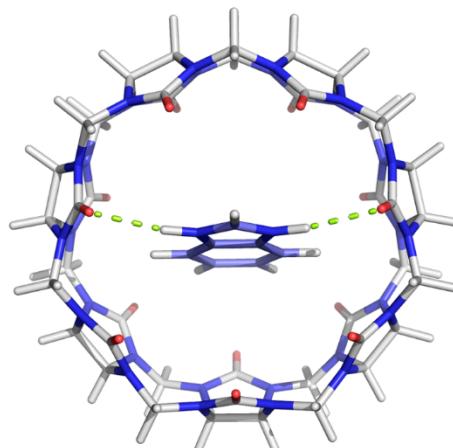


Figure 5. Hydrogen bonds formed between the protonated benzimidazole with the carbonyl oxygens of CB[7].

titration behavior of ABZ appears similar to that of BZ, as its $pK_a^{C,\text{calc}}$ of 7.1 is shifted toward a more basic value from its $pK_a^{F,\text{exp}}$ of 3.5. At neutral pH, these data suggest that ABZ is fully deprotonated when free in solution, whereas both its protonated and deprotonated forms are significantly populated when in complex with CB[7]. While the errors obtained for fitting the Hill equation to the titration data are minimal with errors observed for BZ and all other guests of less than 0.01 pK units, the value of $pK_a^{C,\text{calc}}$ for the CB[7]:ABZ complex has the greatest deviation from the experiment ($\Delta pK = 1.0$, Table 2).

The titration curves for the other guests follow a similar trend, where formation of the CB[7]:guest complex increases the pK_a of the guest (Figure S1, Supporting Information). These shifts are in line with the experimentally determined pK_a values (Table 2), with the largest deviation seen for ABZ as stated above. Overall, the CpHMD method provides accurate predictions of $pK_a^{C,\text{calc}}$ values, with a mean average error (MAE) of 0.42 pK units with respect to experiment ($pK_a^{C,\text{exp}}$, Table 2).

pH Dependence of the Binding Free Energy. As discussed above, the pK_a values of the BZ-derived guests differ when bound to CB[7] and when free in solution (Figure 4, Table 2). Since there are no other titratable groups in the CB[7]:guest complexes in the pH ranges studied here, the binding of the guests to CB[7] can have a net uptake of protons, which makes their binding free energies depend on the solution pH. In this section, we compute binding free energies as functions of pH using eqs 6 and 8. Both of these equations can be used to obtain the pH-dependent binding free energy by adding a pH-dependent correction term to a reference binding free energy. In eq 6, this reference free energy corresponds to the free energy of binding in the absence of proton binding. In contrast, eq 8 requires that the reference free energy be obtained at a specific pH. The reference binding free energies in these two equations can be identical if obtained at a specific value of pH where the protonation states do not change. Since experimental association constants were obtained at pH levels where both the free and bound guests are deprotonated,⁵⁹ we use reference binding free energies for the association of deprotonated guests with CB[7] in this work.

As a simple illustration of how pH-dependent binding free energies may be obtained, we first use the binding free energy measured experimentally for each of the different CB[7]:guest systems ($\Delta G_{\text{ref},\text{exp}}^\circ$) as the reference free energy term in eqs 6 and 8. We refer to this as a “hybrid” approach, as it obtains a pH-

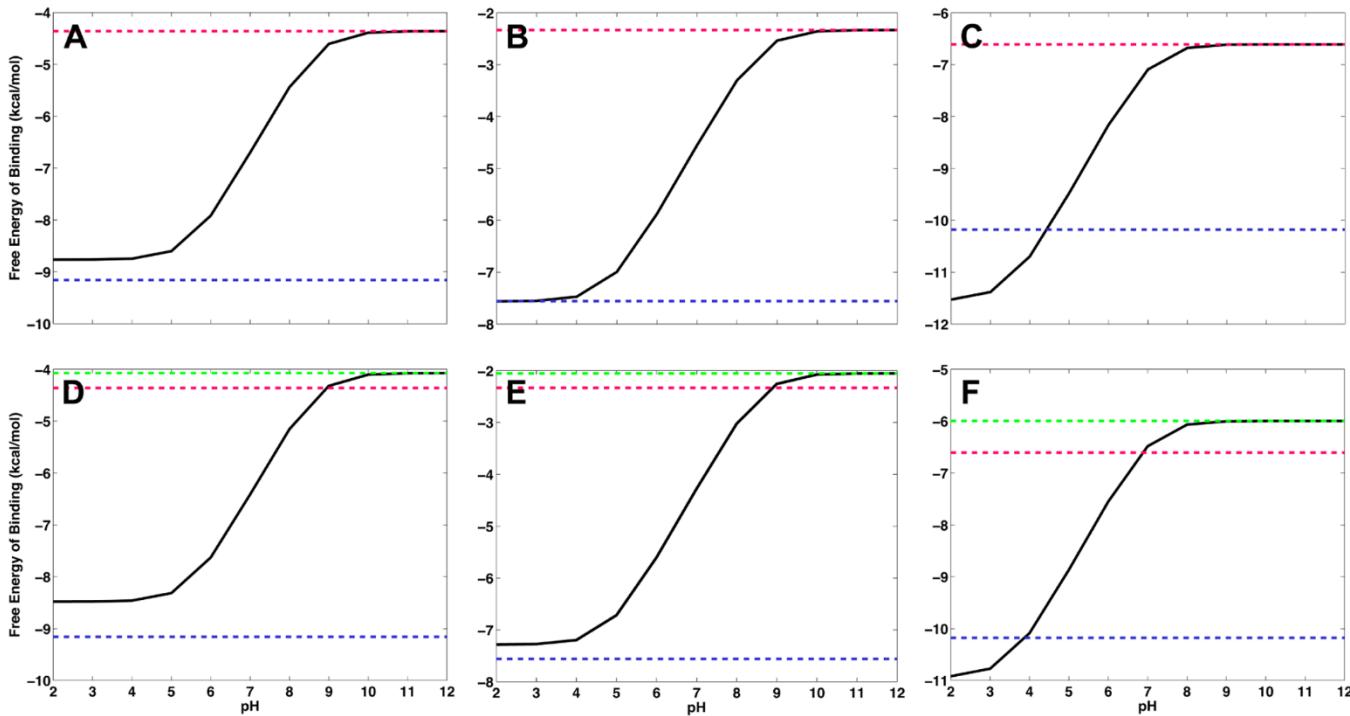


Figure 6. Binding free energies as functions of pH (black line). The top row is computed by the hybrid approach using the experimental reference binding energies ($\Delta G_{\text{ref,exp}}^{\circ}$, red line), and the bottom row uses the full computational approach with the reference binding energies computed by thermodynamic integration ($\Delta G_{\text{ref,TI}}^{\circ}$, green line). Experimentally derived binding free energies for the protonated guests are shown in blue. (A, D) Benzimidazole; (B, E) fuberidazole; (C, F) albendazole.

dependent binding free energy ($\Delta G_{\text{hybrid}}^{\circ}$) from the experimental reference binding free energy ($\Delta G_{\text{ref,exp}}^{\circ}$) and CpHMD-derived terms. These terms are either the pK_a for the complex, $pK_a^{C,\text{calc}}$, used in eq 6 or charge differences between the binding partners in complex and free in solution, ΔZ , in eq 8. While all results described here have been obtained using eq 6 with pH-dependent corrections requiring values of $pK_a^{C,\text{calc}}$, identical results have also been obtained using eq 8 (Figure S3 and Table S1, Supporting Information).

Plots of binding free energies as functions of pH for CB[7] complexes with BZ, FBZ, and ABZ are shown in Figure 6A–C. While these binding free energies are referenced to $\Delta G_{\text{ref,exp}}^{\circ}$ (Figure 6, red line), the use of $pK_a^{C,\text{calc}}$ in eq 6 and ΔZ in eq 8 to generate the full curve as a function of pH can be assessed by how well the computed binding free energy at acidic pH ($\Delta G_{\text{hybrid}}^{\circ+}$ in Table 3) matches the analogous value derived from experiment ($\Delta G_{\text{exp}}^{\circ+}$, blue line in Figure 6). For all CB[7]:guest complexes, the

values of $\Delta G_{\text{hybrid}}^{\circ+}$ deviate less than 1.35 kcal/mol from the respective experimental values (Table 3), with the greatest error observed for ABZ. These errors are entirely due to the errors in computing values of $pK_a^{C,\text{calc}}$, as the value for $\Delta G_{\text{exp}}^{\circ+}$ was derived using experimentally obtained values of pK_a^C , pK_a^F , and $\Delta G_{\text{ref}}^{\circ}$ ⁵⁹

From Figure 6A–C, it is evident that all guests bind more favorably when protonated. Indeed, the binding free energies observed for deprotonated guests (at extremely basic pH) are 3.4–5.6 kcal/mol more positive (less favorable) than those obtained at acidic pH when the guests are protonated. This tendency is most pronounced in the CB[7]:FBZ complex (Figure 6B), for which the binding free energy obtained when FBZ is predominantly protonated (−7.56 kcal/mol) is over 5 kcal/mol more favorable than its respective value when FBZ is deprotonated (−2.33 kcal/mol). This observation is consistent with experiment⁵⁹ and stems from the favorable hydrogen bond between the guest and CB[7] (Figure 5).

Taking a closer look at the pH-dependent binding free energies of the CB[7]:BZ complex, it is apparent that the binding free energy spans 4.8 kcal/mol between pH levels 4.5 and 10, a range that essentially encompasses the pH levels of most biological reactions (Figure 6A).⁹² At physiological pH (~7), free BZ is predominantly deprotonated, whereas BZ in complex with CB[7] is protonated (Figure 4A). In conventional free energy computations, ligand protonation states are typically assigned as the preferred protonation state for the free ligand. Consistent with this convention, BZ would be considered deprotonated in free energy computations performed at pH 7. In making this assumption, the binding free energy deviates from the pH-dependent binding free energy obtained here by ~2.3 kcal/mol for the CB[7]:BZ complex. Similar deviations are noted for the binding of other guests as well, with the magnitudes

Table 3. Binding Free Energies of the Guests upon Complex Formation with CB[7], Computed Using the Hybrid Approach with eq 6^a

guest	$\Delta G_{\text{ref,exp}}^{\circ}$	$\Delta G_{\text{exp}}^{\circ+}$	$\Delta G_{\text{hybrid}}^{\circ+}$	$\Delta G_{\text{ref,exp}}^{\circ}$ (pH 7)	$\Delta G_{\text{hybrid}}^{\circ}$ (pH 7)
BZ	−4.4	−9.2	−8.8	−7.1	−6.7
TBZ	−3.0	−8.6	−7.9	−5.2	−4.7
FBZ	−2.3	−7.6	−7.6	−4.5	−4.6
ABZ	−6.6	−10.2	−11.5	−6.7	−7.1
CBZ	−6.0	−9.5	−10.0	−6.4	−6.8

^aAll energies are reported in kcal/mol. $\Delta G_{\text{ref,exp}}^{\circ}$ is the experimental⁵⁹ binding free energy for the reference deprotonated guest; $\Delta G_{\text{exp}}^{\circ+}$ is the binding free energy for the protonated guest derived from $\Delta G_{\text{ref,exp}}^{\circ}$; and $\Delta G_{\text{hybrid}}^{\circ+}$ is the free energy obtained by using $pK_a^{C,\text{calc}}$ with $\Delta G_{\text{ref,exp}}^{\circ}$ in eq 6.

ranging between 0.4 and 2.3 kcal/mol (Table 3 and Figure S2, Supporting Information).

Full Prediction of the pH-Dependent Free Energy Profile. To demonstrate the utility of our method when experimental binding free energies are unavailable, we perform thermodynamic integration (TI) computations based on the thermodynamic cycle shown in Scheme 2 to obtain $\Delta G_{\text{ref,TI}}^{\circ}$, the reference binding free energies of the CB[7]:guest complexes with the guests deprotonated. The pH-dependent correction terms obtained either with eq 6 or 8 are then referenced to $\Delta G_{\text{ref,TI}}^{\circ}$ to obtain a full computational prediction (CpHMD/TI) of the pH-dependent free energy profiles.

The free energy profiles of CB[7]:guest complexes using $\Delta G_{\text{ref,TI}}^{\circ}$ are shown in Figure 6D–F, while the computed values of $\Delta G_{\text{ref,TI}}^{\circ}$ are reported in Table 4 for comparison with experiment

Table 4. Binding Free Energies of the Guests, Computed Using the Full Computational Approach (CpHMD/TI) and Compared to Experiment^{59 a}

guest	$\Delta G_{\text{ref,exp}}^{\circ}$	$\Delta G_{\text{ref,TI}}^{\circ}$	$\Delta G_{\text{exp}}^{\circ+}$	$\Delta G_{\text{TI}}^{\circ+}$	$\Delta G_{\text{ref,exp}}^{\circ}$ (pH 7)	$\Delta G_{\text{ref,TI}}^{\circ}$ (pH 7)
BZ	-4.4	-4.1 ± 2.0	-9.2	-8.5	-7.1	-6.4
TBZ	-3.0	-2.3 ± 2.6	-8.6	-7.3	-5.2	-4.0
FBZ	-2.3	-2.1 ± 2.6	-7.6	-7.3	-4.5	-4.3
ABZ	-6.6	-6.0 ± 3.0	-10.2	-11.0	-6.7	-6.5
CBZ	-6.0	-4.8 ± 2.7	-9.5	-8.7	-6.4	-5.5

^aAll energies are reported in kcal/mol. $\Delta G_{\text{ref,exp}}^{\circ}$ is the experimental⁵⁹ binding free energy for the reference deprotonated guest; $\Delta G_{\text{ref,TI}}^{\circ}$ is the absolute binding free energy obtained from TI computations for the reference state; $\Delta G_{\text{exp}}^{\circ+}$ is the binding free energy for the protonated guest derived from $\Delta G_{\text{ref,exp}}^{\circ}$; and $\Delta G_{\text{TI}}^{\circ+}$ is the binding free energy obtained by using $pK_a^{\text{C,calc}}$ with $\Delta G_{\text{ref,TI}}^{\circ}$ in eq 6.

($\Delta G_{\text{ref,exp}}^{\circ}$, Table 4). All values of $\Delta G_{\text{ref,TI}}^{\circ}$ agree well with experiment, showing absolute errors that are less than 1.3 kcal/mol. Further, the error with respect to experiment ($\Delta G_{\text{exp}}^{\circ+}$) in the predicted values for the binding free energy of protonated guests ($\Delta G_{\text{TI}}^{\circ+}$) are similarly low (<1.4 kcal/mol). Errors in $\Delta G_{\text{TI}}^{\circ+}$ arise from both the computation of $pK_a^{\text{C,calc}}$ (or ΔZ , when using eq 8; Table S2, Supporting Information) using the CpHMD method and the binding free energy computation with TI. These errors are not always of the same sign; for example, the deviation from $\Delta G_{\text{exp}}^{\circ+}$ obtained for ABZ decreased from 1.4 kcal/mol when using $\Delta G_{\text{ref,exp}}^{\circ}$ in the hybrid approach to 0.7 kcal/mol when using the $\Delta G_{\text{ref,TI}}^{\circ}$ reference, indicating some cancellation of error in the full computational approach. In contrast, the deviation for TBZ increased from 0.7 kcal/mol when using the experimental $\Delta G_{\text{ref,exp}}^{\circ}$ reference to 1.4 kcal/mol when using $\Delta G_{\text{ref,TI}}^{\circ}$. Regardless, results obtained using the full computational approach with $\Delta G_{\text{ref,TI}}^{\circ}$ show errors that are similar in magnitude to those observed using $\Delta G_{\text{ref,exp}}^{\circ}$ in the hybrid experimental/computational approach. Furthermore, the errors associated with the full computational protocol can be lower than the errors that arise from performing binding free energy computations with fixed protonation states assigned to the unbound CB[7] and guest molecules (Table 4).

DISCUSSION

Changes in the pK_a values and, consequently, the protonation states of ionizable species participating in biomolecular association processes are well documented. To address this phenomenon, we present a simple methodology for obtaining

the pH dependence of binding free energies for a series of cucurbit[7]uril (CB[7]):guest complexes. On the basis of Wyman's binding polynomial formalism,⁴³ binding free energies are computed as pH-dependent corrections to a reference binding free energy. Combining this formalism with constant pH molecular dynamics (CpHMD) simulations and free energy computations yields a reasonable protocol for evaluating the pH-dependent binding free energies of biomolecular systems.

Focusing on the application of CpHMD to eq 6 in order to obtain pH-dependent relative binding free energies, we assess how well CpHMD simulations can capture the pK_a of BZ guests in complex with CB[7] (Table 2). With the exception of albendazole (ABZ), the values of $pK_a^{\text{C,calc}}$ obtained for the different CB[7]:guest complexes deviate from experiment by less than 0.41 pK units. The $pK_a^{\text{C,calc}}$ value obtained for the CB[7]:ABZ complex, however, exhibits an error of 1.0 pK unit. Since the CpHMD simulations conducted in this study are only 5 ns long, we extended the simulation of CB[7]:ABZ to 25 ns to ascertain whether the value of $pK_a^{\text{C,calc}}$ had converged; however, the resulting $pK_a^{\text{C,calc}}$ remains unchanged. Further, the process of fitting titration data obtained from CpHMD simulations of CB[7]:guest complexes to the Hill equation is achieved with very little statistical error (<0.01 pK units, Table 2). Both of these findings indicate that the error in the $pK_a^{\text{C,calc}}$ values is not due to convergence problems in the CpHMD simulations. Instead, it is possible, though not explicitly demonstrated in this work, that inaccuracies in the computed pK_a^{C} values stem from problems with the force field due to the accuracies of similar magnitude to those seen in the previous CpHMD runs.^{93,94}

Since CpHMD simulations can reliably compute the values of $pK_a^{\text{C,calc}}$ (and, similarly, ΔZ in the case of eq 8) for the CB[7]:guest systems, we proceed to incorporate these $pK_a^{\text{C,calc}}$ values in eq 6 along with a reference experimental binding free energy ($\Delta G_{\text{ref,exp}}^{\circ}$) to obtain binding free energies as functions of pH. This hybrid experimental/computational approach is followed and shown for CB[7]:guest systems (Figure 6A–C). To evaluate the accuracy of this approach, we compare the computed value of $\Delta G_{\text{hybrid}}^{\circ+}$ to experiment ($\Delta G_{\text{exp}}^{\circ+}$) and observe good agreement, with errors of <1.4 kcal/mol arising from the computation of $pK_a^{\text{C,calc}}$.

Having established that eq 6 can successfully recapitulate pH-dependent binding free energies with an experimentally determined reference binding free energy, we consider the use of thermodynamic integration (TI) computations to remove this dependence on experiment. TI computations effectively reproduce the reference binding free energies observed from experiment ($\Delta G_{\text{ref,exp}}^{\circ}$) with absolute errors less than 1.3 kcal/mol ($\Delta G_{\text{ref,TI}}^{\circ}$, Table 4). The resulting pH-dependent free energy profiles using $\Delta G_{\text{ref,TI}}^{\circ}$ are similar to those computed with $\Delta G_{\text{ref,exp}}^{\circ}$, as shown in Figure 6. Furthermore, the absolute errors in predicting the free energies of the protonated guests, $\Delta G_{\text{TI}}^{\circ+}$, using the $\Delta G_{\text{ref,TI}}^{\circ}$ reference are less than 1.3 kcal/mol. These errors arise from both $\Delta G_{\text{ref,TI}}^{\circ}$ and the use of CpHMD simulations to obtain $pK_a^{\text{C,calc}}$ values. In regard to the computation of $\Delta G_{\text{ref,TI}}^{\circ}$, we do observe large statistical uncertainties for all CB[7]:guest complexes considered, which stem largely from the van der Waals decoupling simulations (Table S3, Supporting Information); however, the free energies computed for every transformation in the thermodynamic cycle shown in Scheme 2 have all converged, with the cumulative computed $\Delta G < 0.01$ kcal/mol.

The use of our CpHMD/TI approach to provide a full computational prediction of pH-dependent binding free energies

is particularly advantageous when experimental association constants are not available, as most experimental measurements face the limitations at extreme pH levels due to the highly possible destabilization or denaturation of the proteins under such severe conditions. Therefore, when combined with computational free energy calculations, our method is free from such concerns, eliminating the reliance on the availability of experimental data. While the CpHMD/TI computation of pH-dependent binding free energies is prone to greater error than the hybrid experimental/computational approach described previously, we find the absolute errors in the CpHMD-derived pK_a^C values and reference binding free energies obtained from TI computations are not necessarily of the same sign for the CB[7]:guest systems considered and may cancel out. Further, the observed errors with respect to the $\Delta G_{\text{exp}}^{o+}$ are relatively low (<1.4 kcal/mol). In contrast, the error in assigning incorrect protonation states in free energy computations without correcting for the pH dependence of the binding free energy can give errors in excess of 2 kcal/mol (Table 3). This observation underscores the importance of accounting for the linkage of proton binding or release to ligand binding in free energy computations and demonstrates the high utility of the CpHMD/TI approach.

Our results highlight the significant changes in pK_a and free energy of binding upon complex formation that accompanies a net proton uptake. Noting that the guests used here have a single titratable site, corresponding changes in free energy may sometimes be larger in protein–ligand binding where multiple titratable groups exist. Therefore, we believe that our method will have great utility in computer-aided drug discovery, where early stages of the structure-based drug design often focus on finding a high-affinity binder to a target protein. Extensions of our methodology to such more complex protein systems may require improvements to the computational protocols employed. The simple framework developed here allows for trivial incorporation of CpHMD methods that incorporate explicit solvent models³⁵ and/or enhanced sampling techniques, such as accelerated molecular dynamics³⁷ or replica exchange³⁶ to improve pK_a computations in systems where convergence is difficult.⁹⁴ Similarly, our protocol accommodates the use of alternative methods for obtaining the reference binding free energy required by eqs 6 and 8. Thus, the computational methodology for performing CpHMD/TI computations can be chosen to best address the system under consideration.

While we have focused on the results obtained using eq 6, which assumes all titratable groups are decoupled, the CpHMD/TI method is also compatible with the expression for obtaining the pH-dependent binding free energy given in eq 8, and these two expressions yield identical results in the case of the CB[7]:guest systems considered here. We intend to build on the computational protocol developed here, with a natural extension being the application of the CpHMD/TI method to obtain pH-dependent binding free energies of protein–ligand complexes. As protein–ligand systems are more complicated than the CB[7]:guest systems considered in this work, we believe the use of eq 8 will have high utility to address potential interactions between titratable groups.

Given the magnitude of errors in computed binding free energies obtained with fixed protonation states in the CB[7]:guest systems, our computational protocol represents a promising approach to remove these errors, thus implicating its utility in drug discovery workflows.⁹⁵ Though not specifically

addressed in this work, similar philosophies may also be applicable to the scoring functions in docking protocols.

CONCLUSIONS

In this work, we determined the pH-dependent changes in binding free energies for complex formation between cucurbit[7]uril (CB[7]) and a series of benzimidazole guests. Using constant pH molecular dynamics simulations combined with experimental data, we developed a hybrid protocol that could capture the significant changes in the CB[7]:guest binding free energies with high accuracy. Subsequently, we combined our method with thermodynamic integration (TI) to enable a full computational prediction of the pH-dependent free energy profiles. This protocol successfully accounted for the pH-dependent changes in the binding free energies during complex formation. Future work will include examination of pH-dependent binding free energies for protein–ligand complexes.

ASSOCIATED CONTENT

Supporting Information

Titration curves for thiabendazole, fuberidazole, and carbendazim; free energies of binding for thiabendazole and carbendazim; free energies of binding for the guests computed using eq 8; table for binding free energies using the hybrid approach with eq 8; table for binding free energies using the full computational approach with eq 8; and table for free energy for each segment in absolute binding free energy computations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jmccammon@ucsd.edu. Phone: 1-858-534-2905. Fax: 1-858-534-4974.

Author Contributions

[†]M.O.K., P.G.B.: These authors have equally contributed to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the NSF (MCB-1020765), NIH (NIH GM31749), Howard Hughes Medical Institute, Center for Theoretical Biological Physics (CTBP), National Biomedical Computation Resource (NCCR), and NSF supercomputer centers. P.G.B. and J.W.K. acknowledge support from the NIH Molecular Biophysical Training Grant (2T32GM008326-21).

REFERENCES

- (1) Onufriev, A. V.; Alexov, E. Protonation and pK_a Changes in Protein-Ligand Binding. *Q. Rev. Biophys.* **2013**, *46*, 181–209.
- (2) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. Prediction and Rationalization of Protein pK_a Values Using QM and QM/MM Methods. *J. Phys. Chem. A* **2005**, *109*, 6634–6643.
- (3) Jensen, J. H. Calculating pH and Salt Dependence of Protein-Protein Binding. *Curr. Pharm. Biotechnol.* **2008**, *9*, 96–102.
- (4) Mason, A. C.; Jensen, J. H. Protein-Protein Binding Is Often Associated with Changes in Protonation State. *Proteins* **2008**, *71*, 81–91.
- (5) Cassidy, C. S.; Lin, J.; Frey, P. A. A New Concept for the Mechanism of Action of Chymotrypsin: The Role of the Low-Barrier Hydrogen Bond. *Biochemistry* **1997**, *36*, 4576–4584.
- (6) Stranzl, G. R.; Gruber, K.; Steinkellner, G.; Zangerer, K.; Schwab, H.; Kratky, C. Observation of a Short, Strong Hydrogen Bond in the Active Site of Hydroxynitrile Lyase from Hevea Brasiliensis Explains a

- Large pKa Shift of the Catalytic Base Induced by He Reaction Intermediate. *J. Biol. Chem.* **2004**, *279*, 2699–3707.
- (7) Brandsdal, B. O.; Smalås, A. O.; Åqvist, J. Free Energy Calculations Show That Acidic P1 Variants Undergo Large pKa Shifts Upon Binding to Trypsin. *Proteins* **2006**, *64*, 740–748.
- (8) Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C. N.; Klabe, R. M.; Head, M. S.; Gilson, M. K. Thermodynamic Linkage between the Binding of Protons and Inhibitors to HIV-1 Protease. *Protein Sci.* **1999**, *8*, 180–195.
- (9) Aguilar, B.; Anandakrishnan, R.; Ruscio, J. Z.; Onufriev, A. V. Statistics and Physical Origins of pK and Ionization State Changes Upon Protein-Ligand Binding. *Biophys. J.* **2010**, *98*, 872–880.
- (10) Manallack, D. T. The Acid-Base Profile of a Contemporary Set of Drugs: Implications for Drug Discovery. *SAR QSAR Environ. Res.* **2009**, *20*, 611–655.
- (11) Leeson, P. D.; St-Gallay, S. A.; Wenlock, M. C. Impact of Ion Class and Time on Oral Drug Molecular Properties. *Med. Chem. Commun.* **2011**, *2*, 91–105.
- (12) Meanwell, N. A. Improving Drug Candidates by Design: A Focus on Physicochemical Properties as a Means of Improving Compound Disposition and Safety. *Chem. Res. Toxicol.* **2011**, *24*, 1420–1456.
- (13) Manallack, D. T.; Pranker, R. J.; Yuriev, E.; Oprea, T. I.; Chalmers, D. K. The Significance of Acid/Base Properties in Drug Discovery. *Chem. Soc. Rev.* **2013**, *42*, 485–496.
- (14) Gordon, J. C.; Myers, J. B.; Folta, T.; Soja, V.; Heath, L. S.; Onufriev, A. V. H++: A Server for Estimating pKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368–371.
- (15) Myers, J. B.; Grothaus, G.; Narayana, S.; Onufriev, A. V. A Simple Clustering Algorithm Can Be Accurate Enough for Use in Calculations of Pks in Macromolecules. *Proteins* **2006**, *63*, 928–938.
- (16) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulation. *Nucleic Acids Res.* **2012**, *40* (W1), W537–541.
- (17) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Interpretation of Protein pKa Values. *Proteins* **2005**, *61*, 704–721.
- (18) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes. *Proteins* **2008**, *73*, 765–783.
- (19) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. Propka3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (20) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Realization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
- (21) Perrin, D. D.; Dempsey, B.; Serjeant, E. P. *pKa Prediction for Organic Acids and Bases*; Chapman and Hall: London, 1981.
- (22) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for pKa Prediction and Protonation State Generation for Drug-Like Molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.
- (23) Alexov, E.; Gunner, M. R. Incorporating Protein Conformational Flexibility into pH Titration Calculations: Results on T4 Lysozyme. *Biophys. J.* **1997**, *74*, 2075–2093.
- (24) Georgescu, R. E.; Alexov, E.; Gunner, M. R. Combining Conformational Flexibility and Continuum Electrostatics for Calculating Pka's in Proteins. *Biophys. J.* **2002**, *83*, 1731–1748.
- (25) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (26) Warshel, A.; Russel, S. Calculations of Electrostatic Interactions in Biological Systems and in Solutions. *Q. Rev. Biophys.* **1984**, *17*, 283–422.
- (27) Sham, Y. Y.; Chu, Z. T.; Warshel, A. Consistent Calculations of Pka's of Ionizable Residues in Proteins: Semi-Microscopic and Microscopic Approaches. *J. Phys. Chem. B* **1997**, *101*, 4458–4472.
- (28) Li, H.; Robertson, A. D.; Jensen, J. H. The Determinants of Carboxyl pKa Values in Turkey Ovomucoid Third Domain. *Proteins* **2004**, *55*, 689–704.
- (29) Baptista, A. M.; Martel, P. J.; Petersen, S. B. Simulation of Protein Conformational Freedom as a Function of pH: Constant-pH Molecular Dynamics Using Implicit Titration. *Proteins* **1997**, *27*, 523–544.
- (30) Borjesson, U.; Hunenberger, P. H. Explicit-Solvent Molecular Dynamics Simulation at Constant pH: Methodology and Application to Small Amines. *J. Chem. Phys.* **2001**, *114*, 9706–9719.
- (31) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. Constant-pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins* **2004**, *56*, 738–752.
- (32) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH Molecular Dynamics Using Stochastic Titration. *J. Chem. Phys.* **2002**, *117*, 4184–4200.
- (33) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
- (34) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. Simulating Proteins at Constant pH: An Approach Combining Molecular Dynamics and Monte Carlo Simulation. *Proteins* **2002**, *47*, 469–480.
- (35) Swails, J.; York, D. M.; Roitberg, A. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352.
- (36) Meng, Y. L.; Roitberg, A. Constant pH Replica Exchange Molecular Dynamics in Biomolecules Using a Discrete Protonation Model. *J. Chem. Theory Comput.* **2010**, *6*, 1401–1412.
- (37) Williams, S. L.; de Oliveira, C. A. F.; McCammon, J. A. Coupling Constant pH Molecular Dynamics with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 560–568.
- (38) Goh, G. B.; Knight, J. L.; Brooks, C. L., III. Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36–46.
- (39) Goh, G. B.; Knight, J. L.; Brooks, C. L., III. pH-Dependent Dynamics of Complex RNA Macromolecules. *J. Chem. Theory Comput.* **2013**, *9*, 935–943.
- (40) Goh, G. B.; Knight, J. L.; Brooks, C. L., III. Towards Accurate Prediction of Protonation Equilibrium of Nucleic Acids. *J. Phys. Chem. Lett.* **2013**, *4*, 760–766.
- (41) Di Russo, N. V.; Estrin, D. A.; Marti, M. A.; Roitberg, A. pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pKas: The Case of Nitrophorin 4. *PLoS Comput. Biol.* **2012**, *8*, e1002761.
- (42) Laricheva, E. N.; Arora, K.; Knight, J. L.; Brooks, C. L., III. Deconstructing Activation Events in Rhodopsin. *J. Am. Chem. Soc.* **2013**, *135*, 10906–10909.
- (43) Wyman, J. Heme Proteins. *Adv. Protein Chem.* **1948**, *4*, 407–531.
- (44) Mitra, R. C.; Zhang, Z.; Alexov, E. In Silico Modeling of pH-Optimum of Protein-Protein Binding. *Proteins* **2011**, *79*, 925–936.
- (45) Misra, V. K.; Hecht, J. L.; Yang, A. S.; Honig, B. Electrostatic Contributions to the Binding Free Energy of the Lac Repressor to DNA. *Biophys. J.* **1998**, *75*, 2262–2273.
- (46) Moody, E. M.; Lecompte, J. T. J.; Belivacqua, P. C. Linkage between Proton Binding and Folding in RNA: A Thermodynamic Framework and Its Experimental Application for Investigating pKa Shifting. *RNA* **2005**, *11*, 157–172.
- (47) Freeman, W. A.; Mock, W. L.; Shih, N. Y. Cucurbituril. *J. Am. Chem. Soc.* **1981**, *103*, 7367–7368.
- (48) Kim, J.; Jung, I. S.; Kim, S. Y.; Lee, E.; Kang, J. K.; Sakamoto, S.; Yamaguchi, K.; Kim, K. New Cucurbituril Homologues: Syntheses, Isolation, Characterization, and X-Ray Crystal Structures of Cucurbit[N]Uril (N=5, 7, and 8). *J. Am. Chem. Soc.* **2000**, *122*, 540–541.
- (49) Liu, S.; Ruspis, C.; Mukhopadhyay, P.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L. The Cucurbit[N]Uril Family: Prime Components for Self-Sorting Systems. *J. Am. Chem. Soc.* **2005**, *127*, 15959–15967.
- (50) Rekharsky, M. V.; Mori, T.; Yang, C.; Ko, Y. H.; Selvapalam, N.; Kim, H.; Sobransingh, D.; Kaifer, A. E.; Liu, S.; Isaacs, L.; et al. A Synthetic Host-Guest System Achieves Avidin-Biotin Affinity by

- Overcoming Enthalpy-Entropy Compensation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *109*, 20006–20011.
- (51) Macartney, D. H. Encapsulation of Drug Molecules by Cucurbiturils: Effects on Their Chemical Properties in Aqueous Solution. *Isr. J. Chem.* **2011**, *51*, 600–615.
- (52) Ghosh, I.; Nau, W. M. The Strategic Use of Supramolecular pKa Shifts to Enhance the Bioavailability of Drugs. *Adv. Drug Delivery Rev.* **2012**, *64*, 764–783.
- (53) Ma, D.; Hettiarachchi, G.; Nguyen, D.; Zhang, B.; Wittenberg, J. B.; Zavalij, P. Y.; Briken, V.; Isaacs, L. Acyclic Cucurbit[N]Uril Molecular Containers Enhance the Solubility and Bioactivity of Poorly Soluble Pharmaceuticals. *Nat. Chem.* **2012**, *4*, 503–510.
- (54) Hettiarachchi, G.; Nguyen, D.; Wu, J.; Lucas, D.; Ma, D.; Isaacs, L.; Briken, V. Toxicology and Drug Delivery by Cucurbit[N]Uril Type Molecular Containers. *PLoS One* **2010**, *5*, e10514.
- (55) Saleh, N.; Meetani, M.; Al-Kaabi, L.; Ghosh, I.; Nau, W. M. Effects of Cucurbiturils on Tropicamide and Potential Applications in Ocular Drug Delivery. *Supramol. Chem.* **2011**, *23*, 654–661.
- (56) Horton, J. Albendazole: A Review of Anthelmintic Efficacy and Safety in Humans. *Parasitology* **2000**, *121*, S113–132.
- (57) Gould, M.; Nelson, L.; Waterere, D.; Hynes, R. Biocontrol of Fusarium Sambucinum, Dry Rot of Potato, by Serratia Plymuthica 5-6. *Biocontrol Sci. Technol.* **2008**, *18*, 1005–1016.
- (58) Gallidis, E.; Papadopoulos, E.; Ptochos, S.; Arsenos, G. The Use of Targeted Selective Treatments against Gastrointestinal Nematodes in Milking Sheep and Goats in Greece Based on Parasitological and Performance Criteria. *Vet. Parasitol.* **2009**, *164*, 53–58.
- (59) Koner, A. L.; Ghosh, I.; Saleh, N.; Nau, W. M. Supramolecular Encapsulation of Benzimidazole-Derived Drugs by Cucurbit[7]Uril. *Can. J. Chem.* **2011**, *89*, 139–147.
- (60) Tanford, C. Protein Denaturation. Part C. Theoretical Models for the Mechanism of Denaturation. *Adv. Protein Chem.* **1970**, *24*, 1–95.
- (61) Wyman, J. The Binding Potential, a Neglected Linkage Concept. *J. Mol. Biol.* **1965**, *11*, 631–644.
- (62) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Merz, K. M.; Roberts, B.; et al. *Amber 12*; University of California: San Francisco, CA, 2012.
- (63) Onufriev, A. V.; Case, D. A.; Ullmann, G. M. A Novel View of pH Titration in Biomolecules. *Biochemistry* **2001**, *40*, 3413–3419.
- (64) Rogers, K. E.; Ortiz-Sánchez, J. M.; Baron, R.; Fajer, M.; de Oliveira, C. A. F.; McCammon, J. A. On the Role of Dewetting Transitions in Host-Guest Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2013**, *9*, 46–53.
- (65) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Determining Atom-Centered Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (66) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
- (67) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (68) Vosko, S. H.; Nusair, L. W. M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (69) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (70) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (71) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (72) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.
- (73) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431–439.
- (74) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (75) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. I. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (76) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (77) *Glide*, Version 5.8; Schrödinger, LLC: New York, 2012.
- (78) Onufriev, A. V.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383–394.
- (79) Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of Trajectories in Canonical Molecular-Dynamics Simulations: Observation, Explanation, and Exploitation. *J. Chem. Phys.* **2004**, *120*, 6363–6374.
- (80) Ryckaert, J.-P.; Cicotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (81) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (82) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.
- (83) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations. *J. Chem. Phys.* **2007**, *127*, 214108.
- (84) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011**, *32*, 3253–3263.
- (85) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
- (86) Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A. Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *J. Chem. Theory Comput.* **2013**, *9*, 4131–4139.
- (87) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; et al. *Amber 14*; University of California: San Francisco, CA, 2014.
- (88) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (89) Schafmeister, C. E. A. F.; Ross, W. S.; Romanovski, V. *Leap*; University of California: San Francisco, CA, 1995.
- (90) Zhang, Y.; Feller, S. E.; Brooks, B. R.; Pastor, R. W. Computer Simulation of Liquid/Liquid Interfaces. I. Theory and Application to Octane/Water. *J. Chem. Phys.* **1995**, *103*, 10252–10266.
- (91) Zhang, X.; Gramlich, G.; Wang, X.; Nau, W. M. A Joint Structural, Kinetic, and Thermodynamic Investigation of Substituent Effects on Host-Guest Complexation of Bicyclic Azoalkanes by B-Cyclodextrin. *J. Am. Chem. Soc.* **2002**, *124*, 254–263.
- (92) Casey, J. R.; Grinstein, S.; Orlowski, J. Sensors and Regulators of Intracellular pH. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 50–61.
- (93) M, M.; Baptista, A. M. Is the Prediction of pKa Values by Constant-pH Molecular Dynamics Being Hindered by Inherited Problems? *Proteins* **2011**, *79*, 3437–3447.

- (94) Williams, S. L.; Blachly, P. G.; McCammon, J. A. Measuring the Successes and Deficiencies of Constant pH Molecular Dynamics: A Blind Prediction Study. *Proteins* **2011**, *79*, 3381–3388.
- (95) Jorgensen, W. L. Foundations of Biomolecular Modeling. *Cell* **2013**, *155*, 1199–1202.

