Received: 17 September 2023 | Revised: 11 January 2024 | Accepted: 16 January 2024

DOI: 10.1002/jcc.27318

Journal of COMPUTATIONAL CHEMISTRY WILEY

RESEARCH ARTICLE

Resolving coupled pH titrations using alchemical free energy calculations

Carter J. Wilson[1,2,3] | Bert L. de Groot[3] | Vytautas Gapsys[3,4]

[1]Department of Mathematics, The University of Western Ontario, London, Ontario, Canada

[2]Centre for Advanced Materials and Biomaterials Research (CAMBR), The University of Western Ontario, London, Ontario, Canada

[3]Computational Biomolecular Dynamics Group, Department of Theoretical and Computational Biophysics, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

[4]Computational Chemistry, Janssen Research & Development, Beerse, Belgium

Correspondence
Vytautas Gapsys, Computational Biomolecular Dynamics Group, Department of Theoretical and Computational Biophysics, Max Planck Institute for Multidisciplinary Sciences, 37077 Göttingen, Germany.
Email: vgapsys@gwdg.de

Abstract

In a protein, nearby titratable sites can be coupled: the (de)protonation of one may affect the other. The degree of this interaction depends on several factors and can influence the measured pK_a. Here, we derive a formalism based on double free energy differences (ΔΔG) for quantifying the individual site pK_a values of coupled residues. As ΔΔG values can be obtained by means of alchemical free energy calculations, the presented approach allows for a convenient estimation of coupled residue pK_as in practice. We demonstrate that our approach and a previously proposed microscopic pK_a formalism, can be combined with alchemical free energy calculations to resolve pH-dependent protein pK_a values. Toy models and both, regular and constant-pH molecular dynamics simulations, alongside experimental data, are used to validate this approach. Our results highlight the insights gleaned when coupling and microstate probabilities are analyzed and suggest extensions to more complex enzymatic contexts. Furthermore, we find that naïvely computed pK_a values that ignore coupling, can be significantly improved when coupling is accounted for, in some cases reducing the error by half. In short, alchemical free energy methods can resolve the pK_a values of both uncoupled and coupled residues.

KEYWORDS
computational alchemy, free energy calculations, molecular dynamics, pKa calculations, residue coupling

Protein function is known to depend on the acidity of the medium.[1–5] Such a pH dependence is caused by the (de)protonation of amino acid residues, whereby a proton is added or removed from an amino acid side chain. As this process is pH-dependent, at certain pH levels, the event will be more or less favorable and, by definition, at the pK_a, it will be equally probable (i.e., $\Delta G_{\text{deprot}} = 0$). Knowledge of the residue pK_a values in a protein is essential for understanding function. It not only allows for a rationalization of protein properties (e.g., stability,[6] solubility,[7] etc.) and interactions at a specific pH,[8,9] but in the context of enzymatic and redox reactions, pK_a values can provide insight into how favorable a proton transfer will be under certain conditions.[10,11]

As alluded to, the pK_a is fundamentally a free energy relationship; for an isolated, protonatable group, the value is proportional to the free energy of deprotonation:

$$\Delta G_{\text{deprot}} = RT\log(10)(\mathrm{p}K_a - \mathrm{pH}). \tag{1}$$

This relationship suggests that the free energy is linearly dependent on the solution pH; as the pH moves farther away from pK_a, the free energy required to (de)protonate also shifts. A purely linear relationship between pH and ΔG implies a joint relationship with the probability of finding a protonatable group i in a given state;

this follows from the rearranged Henderson–Hasselbalch (HH) equation:

$$\mathrm{p}K_\mathrm{a}^i = \mathrm{pH} + \log_{10}\left(\frac{\langle x_i\rangle}{1-\langle x_i\rangle}\right), \tag{2}$$

where $\langle x_i\rangle$ is the probability that residue i is protonated. However, such a curve, when computed from experiment, may be flatter or irregularly shaped,[12] often necessitating the application of specialized fitting procedures.[13,14] In such cases, not only does the curve suggest a non-linear dependence, an analysis of the complete pH-dependent behaviour is often more insightful than defining the residue by a single pK_a value.

In proteins and, in particular, enzymes, protonatable residues can, in only a few cases, be separated from their interactions with one another.[15–18] Although these associations will be more pronounced at an active site, even more distant residues can experience some degree of coupling,[19] interacting more or less strongly depending on their microenvironment, the pH of the solution, and their own protonation state. Indeed, this could result in a more challenging resolution of "the pK_a";[14] however, such interactions may provide insight into a reaction mechanism or suggest the functional importance of a residue pair.[20,21] In these scenarios, a modified HH-curve may still yield two clear inflection points; however, the assignment of pK_a values to specific residues could remain a challenge. Moreover, the protonation probability of a coupled residue, although potentially described by an HH-curve, is nonetheless a composite probability of microstates. As T. L. Hill,[22,23] J. T. Edsall,[24–26] and more recently G. Matthias Ullmann and co-workers[27–29] have helped formalize, these states are in a pH-dependent equilibrium with each other and collectively comprise the macroscopic probability observed experimentally. This knowledge gap between the measurable macrostates and the cryptic microstates suggests a potential role for theoretical and computational methods, which may help to resolve both the macroscopic pK_a and the microscopic pK_a values.

In this work we explore the effects of coupling on the shifts in pK_a between spatially neighbouring residues. We begin by illustrating the potential magnitude of such effects on a set of toy systems by systematically altering coupling strength. We then focus on two proteins where similar coupling behaviour between amino acid residues are observed. We demonstrate that explicitly accounting for the coupling between titratable sites significantly improves accuracy of the pK_a prediction and highlight the potential insights associated with coupling analysis (e.g., buffering between residues). Finally, we provide a practical guide for using alchemical free energy calculations to: (1) account for pK_a shifts in coupled residues based on the formalism derived in this work; and (2) compute the probabilities of individual protonation microstates based on an existing partition function framework.

THEORY

While from the perspective of statistical mechanics the behaviour of coupled titratable sites is well understood,[22,23,25,28,30,31] the subsequent challenge arises: What is the best way to extract information about changes in residue protonation using existing computational approaches? Molecular dynamics (MD) simulations offer an attractive solution to this question by providing efficient sampling of well defined thermodynamic ensembles. Over the years, several approaches have emerged to quantify pK_a shifts based on MD simulations, among the most popular are constant-pH simulations (CpHMD)[32–43] and alchemical free energy calculations.[44–46] While CpHMD explicitly models protonation changes dynamically over the course of a simulation, an alchemical approach considers discrete protonation states and computes the free energies between them (see Appendix A in the Supporting Information for additional details). Here, we focus on the latter approach and demonstrate how to connect double free energy differences with a statistical mechanics description of population changes in coupled residue protonations.

pK_a values and free energies

Consider the thermodynamic cycle given in Figure 1. To calculate the absolute protein pK_a value of a single residue (A), we must consider the free energy of proton transfer from the gas (g) phase into the solution (s) phase and then from the solution into the protein (p) phase. However, for many model compounds, the free energy associated with the proton transfer in solution is known. Using this reference pK_a value (pK_a°) allows us to only consider the free energies associated with the rightmost cycle, thus reducing our problem to solving $\Delta\Delta G_{s,p}(\mathrm{A^H}\rightarrow\mathrm{A^-})$ — the free energy difference between a deprotonation event in the solution phase and in the protein phase—which is related to the protein pK_a by

$$\begin{aligned}\mathrm{p}K_\mathrm{a} &= \mathrm{p}K_\mathrm{a}^\circ + \frac{\Delta G_{s,p}(\mathrm{A^-}) - \Delta G_{s,p}(\mathrm{A^H})}{RT\log(10)} \\ &= \mathrm{p}K_\mathrm{a}^\circ + \frac{\Delta\Delta G_{s,p}(\mathrm{A^H}\rightarrow\mathrm{A^-})}{RT\log(10)}.\end{aligned} \tag{3}$$

From here we will refer to $\Delta\Delta G_{s,p}(\mathrm{A^H}\rightarrow\mathrm{A^-})$ as $\Delta\Delta G$. Equation (3) implicitly contains two terms, which we here call $\Delta\Delta G^\mathrm{env}$ and ΔG^titr, following the notation of Sharp and Honig.[47] The first ($\Delta\Delta G^\mathrm{env}$) represents the free energy of dissociating a proton within a protein relative to the solvent environment, which we represent by a capped peptide. It is assumed that the protein is fixed in some protonation state and, based on Tanford and Kirkwood,[48] has often been taken to be the state in which all titratable sites in the protein are neutralized. Because these sites are fixed to some state, this free energy is pH-independent. The second free energy component (ΔG^titr) reintroduces pH dependence by capturing how the free energy of dissociating the proton within the protein will be more or less favorable, depending on the states of the other protonatable sites.

Once a reference protonation state is set, the $\Delta\Delta G^\mathrm{env}$ can be resolved and an intrinsic pK_a (pK_int) can be defined:

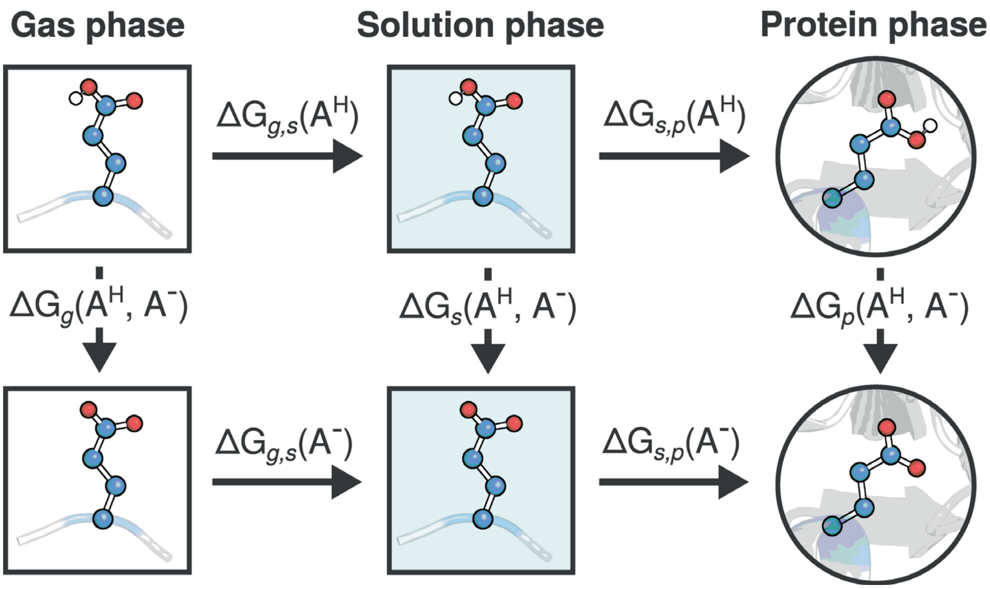

FIGURE 1 Complete pK_a thermodynamic cycle. The horizontal arrows mark the transfer of a titratable residue (A) between different environments: gas (g), solution (s), protein (p). The vertical arrows denote the free energy difference between the deprotonated and protonated form in a corresponding environment.

$$pK_{int} = pK_a^\circ + \frac{\Delta\Delta G^{env}}{RT\log(10)}. \tag{4}$$

As all other residues are in a fixed protonation state without the ability to titrate, this value is pH-independent. The true pK_a for a residue in a protein will depend on the dynamic protonation state of these other residues and will have a pH dependence:

$$pK_a = pK_{int} + \frac{\Delta G^{titr}(pH)}{RT\log(10)}. \tag{5}$$

Although the pK_{int} is pH-independent, it may still provide a strong estimate of the true pK_a depending on the reference protonation state assigned to the protein. Indeed, recent work[44,45] has demonstrated that alchemical free energy calculations can resolve protein pK_a values relying on the assumption: $pK_a \approx pK_{int}$. Nevertheless, it is inevitable that in some instances this will break down, and only by considering $\Delta G^{titr}(pH)$ can an accurate pK_a be resolved. Here, we investigate whether coupling can be meaningfully resolved within an alchemical free energy framework and assess the relative importance of this contribution for computing protein pK_a values.

In the following section, we outline two complementary approaches for extracting insights related to residue coupling from alchemical free energy calculations.

Partition function approach

We offer only a brief outline, additional details are provided in Appendix A in the Supporting Information.

As mentioned in the introduction, both the macroscopic and microscopic pK values are of interest. A partition function approach can account for coupling between residues and the microscopic pK values between individual states.[23,25,28,30]

$$\begin{tikzcd} \overset{[11]}{A_p^H + B_p^H} \arrow[r, "K_1"] \arrow[d, "K_2"'] & \overset{[01]}{A_p^- + B_p^H} \arrow[d, "K_3"] \\ \underset{[10]}{A_p^H + B_p^-} \arrow[r, "K_4"] & \underset{[00]}{A_p^- + B_p^-} \end{tikzcd}$$

FIGURE 2 Two-site protonation dyad. Equilibrium constants K describe the unbinding of a proton. Values in brackets indicate the microstate (e.g., [11]: doubly protonated, [10]: first residue protonated, etc.).

Note that the (de)protonation of the sidechain of an amino acid can be described using a standard equilibrium binding formalism. Specifically, the "binding" of protons can be fully described by the proton concentration (c) and the binding constant (K), from which it follows that the grand partition function is $\xi = 1 + Kc$.

This can be extended to a coupled residue system (Figure 2) in which the protonation of one site can influence the other. Given two residues, four protonation microstates are possible and the corresponding partition function is given by:

$$\xi = 1 + K_1c + K_2c + e^{-\beta w}K_1cK_2c. \tag{6}$$

Unlike in the single-site case, we now include an (un)cooperativity term that follows from the fact that: (1) the cycle is closed (i.e., $K_1 + K_3 = K_2 + K_4$), and (2) there is an "interaction free energy", w, associated with the second (de)protonation event given the first. Given that we can define the standard free energy of deprotonation as: $\Delta G^\circ(pH) = \Delta G(0) - \mu_{H^+}$, where μ_{H^+} is the chemical potential of the protons in solution: $\mu_{H^+} = RT\log(10)pH$, we can resolve the probabilities of the four protonation microstates:

$$\begin{tikzcd}
A_p^H + B_p^H \arrow[r, "\Delta\Delta G_0"] \arrow[d, "\Delta\Delta G_1"'] & A_p^- + B_p^H \arrow[d, "\Delta\Delta G_2"] \\
A_p^H + B_p^- \arrow[r, "\Delta\Delta G_3"] & A_p^- + B_p^-
\end{tikzcd}$$

FIGURE 3 Free energy cycle for a coupled residue scenario. The branches report the free energy change associated with the deprotonation of one residue, while the other is kept fixed with respect to the free energy change associated with the corresponding deprotonation in a capped peptide.

$$\begin{aligned}
\langle A^H B^H \rangle &= \frac{1}{\xi}, \\
\langle A^- B^H \rangle &= \frac{e^{-\beta\left(\Delta G^\circ_{01}(0) - \mu_{H^+}\right)}}{\xi}, \\
\langle A^H B^- \rangle &= \frac{e^{-\beta\left(\Delta G^\circ_{10}(0) - \mu_{H^+}\right)}}{\xi}, \text{ and} \\
\langle A^- B^- \rangle &= \frac{e^{-\beta\left(\Delta G^\circ_{00}(0) - 2\mu_{H^+}\right)}}{\xi},
\end{aligned}$$

and the macroscopic protonation probabilities:

$$\begin{aligned}
\langle A^H \rangle &= \langle A^H B^H \rangle + \langle A^H B^- \rangle \\
\langle B^H \rangle &= \langle A^H B^H \rangle + \langle A^- B^H \rangle.
\end{aligned}$$

Thermodynamic cycle approach

We offer only a brief outline, additional details are provided in Appendix A in the Supporting Information.

Consider a relabelled version of the same coupling scenario presented in Figure 2: residues A and B are close together, and their protonation free energies depend on the state of the other residue (Figure 3).

Here, $\Delta\Delta G_0$ corresponds to $\Delta\Delta G^{\mathrm{env}}$: the free energy of deprotonating residue A while in the presence of protonated B. Similarly, $\Delta\Delta G_3$ corresponds to the deprotonation of A in the presence of deprotonated B. For all branches in the cycle, the remaining protonatable sites in the protein are fixed to their model states at pH 7.4. Notice that $\Delta\Delta G_1$ and $\Delta\Delta G_2$ will shift the populations of "reactants" and "products" with respect to $\Delta\Delta G_0$ and $\Delta\Delta G_3$. Following a derivation provided in Appendix A in the Supporting Information, we resolve a pH-dependent $\Delta\Delta G(\mathrm{pH})$:

$$\begin{aligned}
\Delta\Delta G(\mathrm{pH}) = \ & \Delta\Delta G_0 + \frac{1}{\beta}\log\left(1 + e^{-\beta\Delta G_1(\mathrm{pH})}\right) \\
& - \frac{1}{\beta}\log\left(1 + e^{-\beta\Delta G_2(\mathrm{pH})}\right)
\end{aligned} \tag{7}$$

Considering Equation (7) and Equations (4) and (5) we note the equivalence:

$$\begin{aligned}
\Delta\Delta G^{\mathrm{env}} &= \Delta\Delta G_0, \text{ and} \\
\Delta G^{\mathrm{titr}}(\mathrm{pH}) &= \frac{1}{\beta}\log\left(1 + e^{-\beta\Delta G_1(\mathrm{pH})}\right) \\
&\quad - \frac{1}{\beta}\log\left(1 + e^{-\beta\Delta G_2(\mathrm{pH})}\right).
\end{aligned}$$

The free energy of deprotonating residue A in the protein with coupling accounted for can then be expressed as a function of pH:

$$\Delta G_{\mathrm{protein}}(\mathrm{pH}) = \Delta\Delta G(\mathrm{pH}) + \Delta G_{\mathrm{deprot}}(\mathrm{pH}), \tag{8}$$

where $\Delta G_{\mathrm{deprot}}(\mathrm{pH})$ corresponds to Equation (1) with a reference $\mathrm{p}K_a^\circ$ corresponding to residue A.

Equation (8) provides a family of solutions that depend on the pH value. To determine the $\mathrm{p}K_a$, we find the point where $\Delta G_{\mathrm{protein}}(\mathrm{pH}) = 0$. This pH corresponds to the $\mathrm{p}K_a$ that would be observed in a titration experiment and follows from the Henderson-Hasselbalch equation (Equation 2).

Computationally, we also have access to the whole set of $\mathrm{p}K_a$ solutions which are not necessarily limited by this Henderson-Hasselbalch relation. We can combine Equations (3) and (8) and compute these $\mathrm{p}K_a$ values at various pH:

$$\mathrm{p}K_a = \mathrm{p}K_a^\circ + \frac{\Delta G_{\mathrm{protein}}(\mathrm{pH}) - \Delta G_{\mathrm{deprot}}(\mathrm{pH})}{RT\log(10)} \tag{9}$$

METHODOLOGY

Non-equilibrium alchemy setup

pmx[49] was used for the system setup, hybrid structure and topology generation, and analysis. Initial structures: Δ+PHS Staphylococcal nuclease (SNase) variant[50] (PDB: 3BDC[50]) and protein deglycase DJ-1[51] (PDB: 1P5F[52]), were taken from the PDB database; in the case of 3BDC, the thymidine diphosphate molecule and calcium ion crystalized in the structure, but absent during the titration experiments,[50] were removed. A double system in a single box setup was used, with a 3 nm distance between the protein and peptide (ACE-AXA-NH_2); this ensured charge neutrality during the alchemical transition.[53] To prevent consequential protein–peptide interactions, a single $C\alpha$ in each molecule was positionally restrained. We used the CHARMM36m[54] (with CHARMM-modified TIP3P[55] water model) force field. A salt concentration consistent with the experimental setup was used. If no salt concentration was reported only K^+ or Cl^- counterions were added.

For all systems, an initial minimization using the steepest descent algorithm was performed. A constant temperature corresponding to the reference experimental setup was maintained implicitly using the leap-frog stochastic dynamics integrator[56,57] with an inverse friction constant of $\gamma = 0.5\ \mathrm{ps}^{-1}$. The pressure was maintained at 1 bar using the Parrinello-Rahman barostat[58] with a coupling time constant of

5 ps. The integration time step was set to 2 fs. Long-range electrostatic interactions were calculated using the Particle-mesh Ewald method[59] with a real-space cut-off of 1.2 nm and grid spacing of 0.12 nm. Lennard-Jones interactions were force-switched off between 1.0 and 1.2 nm. Bonds to hydrogen atoms were constrained using the Parallel LINear Constraint Solver.[60]

To improve sampling, systems were run for 50 ns in four independent replicas, and the first 10 ns of each simulation were discarded as equilibration. From the remaining 40 ns, 400 non-equilibrium transitions of 500 ps each were generated and work values from the forward and backward transitions were collected using thermodynamic integration. These values were then used to estimate the corresponding free energy difference with Bennett's acceptance ratio[61] as a maximum likelihood estimator relying on the Crooks fluctuation theorem.[62] Bootstrapping was used to estimate the uncertainties of the free energy estimates,[53,63] and these were propagated when calculating ΔΔG values.

Constant-pH setup

Constant-pH (CpHMD) simulations were performed using a GROMACS 2021 implementation with a modified CHARMM36m force field.[43] This approach is based on the λ-dynamics method developed by Brooks and co-corkers.[32,35,38] The system setup was similar to that discussed in the above section; however, to agree with the recommended CpHMD setup, both the leapfrog integrator and velocity rescaling with a 0.5 ps coupling time, as well as a PME Fourier spacing of 0.14 nm were used. Because only aspartate and glutamate were considered, a single-site representation (i.e., the proton can be bound to only one heavy atom in the residue) was employed; here, the A and B states represent the protonated and deprotonated forms of the titratable residue. The mass of the λ particle was set at 5 AU, and its temperature was maintained at 300 K using velocity rescaling with a 2 ps coupling time. The barrier height of the double-well potential was set at 5.0 kJ/mol. 50 buffer particles were used. For all systems, a minimization was performed using the steepest descent algorithm followed by a 100 ns simulation in the NpT ensemble.

We considered three SNase dyads and one DJ-1 dyad, resulting in four sets of simulations. SNase simulations were performed at pH values from −1 to 7 with 0.25 pH increments and from 7 to 14 with 0.5 pH increments, while DJ-1 simulations were performed from 0 to 14 with 0.5 pH increments. Regarding SNase, both residues in the dyad (i.e., D21–D19, D21–D40, and D21–D83) were allowed to (de)protonate as a function of pH, which in two cases resulted in non-sigmoidal curves. In the case of DJ-1, we allowed only E18 to titrate, while holding C106 fixed to a protonated or deprotonated state; this resulted in sigmoidal curves.

Block averaging was used to determine the protonation probabilities and standard deviations at each pH value. Following Ullmann[28] and similar to previous work,[64,65] constant-pH titration curves were fit to the pH-dependent protonation probabilities corresponding to the partition function in Equation (6). Specifically for each coupled pair we determine values for pK_1, pK_2, and w that best fit both titration probability curves $\langle x_1\rangle$ and $\langle x_2\rangle$:

$$\begin{tikzcd}
A_1^H + B_2^H \arrow[r, "0"] \arrow[d, "-8"'] & A_1^- + B_2^H \arrow[d, "-8"] \\
A_1^H + B_2^- \arrow[r, "0"] & A_1^- + B_2^-
\end{tikzcd}$$

FIGURE 4 Thermodynamic cycle for toy model 1. ΔΔG values (kJ/mol) are indicated along the branches.

$$\langle x_1\rangle = \frac{1+10^{pK_1-pH}}{Z} \qquad \langle x_2\rangle = \frac{1+10^{pK_2-pH}}{Z} \tag{10}$$

with $Z = 1+10^{pK_1-pH}+10^{pK_2-pH}+10^{pK_1+pK_2-w-2pH}$. These microscopic p$K$ values and the interaction energy w can then be used to resolve the macroscopic pK_a values (see Appendix A in the Supporting Information for additional details).

In the case of DJ-1, where only a single residue titrates, curves were fit to the sigmoidal pH equation:

$$\langle X\rangle = \frac{10^{pK_a-pH}}{1+10^{pK_a-pH}}. \tag{11}$$

RESULTS

Toy models

As has been considered previously[66] we illustrate the effects of coupling on several two residue (R_1, R_2) systems, of type A, B, or C, with corresponding reference pK_a° values of 3, 4, and 8, respectively. ΔΔG values can be related to the dissociation constants that link microstates via Equation (5).

Model 1: Similar reference pK_a° values: No coupling

Consider the system given in Figure 4.

Note that the interaction energy between the states is zero, as evidenced by the equality of opposite paths; the protonation of A_1 has no effect on the free energy required to protonate B_2 and vice-versa. It follows that $\Delta G^{titr}(pH)$ is zero for all pH and $\Delta G_{protein}(pH)$ is constant, implying pK_a = pK_{int} (Figure 6A). Computing the pK_a, we obtain values of approximately 3 and 2.6 for A_1 and B_2, respectively, unchanged and decreased from their respective reference values according to Equation (9).

Model 2: Similar reference pK_a° values: Weak coupling

Consider the system given in Figure 5.

In this example, unlike in the first, the interaction energy is nonzero. The free energies suggest that the first deprotonation event results in a less favorable second deprotonation. This is to be expected for nearby coupled residues, where the electrostatic repulsion associated with the introduction of a second negative charge would result in a less favorable free energy change. Instead of reporting a linear dependence, the $\Delta G_{\text{protein}}(\text{pH})$ curves each have an inflection point and two asymptotic values, corresponding to the protonation free energy of one residue, while the other residues remain in the same protonation state (Figure 6B, top). These asymptotic values correspond to the microscopic pK_a values (Figure 6B, middle), which put a bound on the range of computed pK_a values.

At low pH, both A_1 and B_2 are protonated, with large, unfavorable $\Delta G_{\text{protein}}$ values (Figure 6B, top). As the pH increases, $\Delta G_{\text{protein}}$ of deprotonating B_2 becomes more favorable, reaching $\Delta G_{\text{protein}} = 0$ near pH $\approx$ 3; at this point, B_2 begins to deprotonate. This deprotonation will result in a more unfavorable protonation free energy for A_1, as evidenced by the flattening in the ΔG curve and an increase in the apparent pK_a of A_1: the formation of B_2^- makes the formation of A_1^- less favorable. Because the reference pK_a° values of A_1 and B_2 are similar, this flattening occurs almost simultaneously with that of B_2. In this regime, the $\Delta G_{\text{protein}}$ for both residues changes slower and, in this example, remains relatively close to zero. We can think of this as the pH range over which the groups buffer each other, altering the favorability of protonation. As the pH continues to increase, a linear dependence is restored. Construction of the titration curves computed from the microscopic pK_a values reveals the coupling between residues (Figure 6B, bottom). The non-sigmoidal form of the curves follows from the fact that the singly protonated microstates for both residues occur with a similar probability.

$$\begin{tikzcd}
A_1^H + B_2^H \arrow[r, "0"] \arrow[d, "-8"'] & A_1^- + B_2^H \arrow[d, "1"] \\
A_1^H + B_2^- \arrow[r, "9"] & A_1^- + B_2^-
\end{tikzcd}$$

FIGURE 5 Thermodynamic cycle for toy model 2. $\Delta\Delta G$ values (kJ/mol) are indicated along the branches.

Model 3: Different reference pK_a° values: Weak coupling

Consider the system given in Figure 7.

In this example, the $\Delta\Delta G$ values along the branches are the same as in Example 2; however, the reference pK_a° values have changed. We now consider the coupling between a residue C that has a reference value 5 pK units higher than B. Although $\Delta G_{\text{protein}}(\text{pH})$ does report a non-linear dependence, the large difference in reference values means that the pH effects dominate; both inflection points of

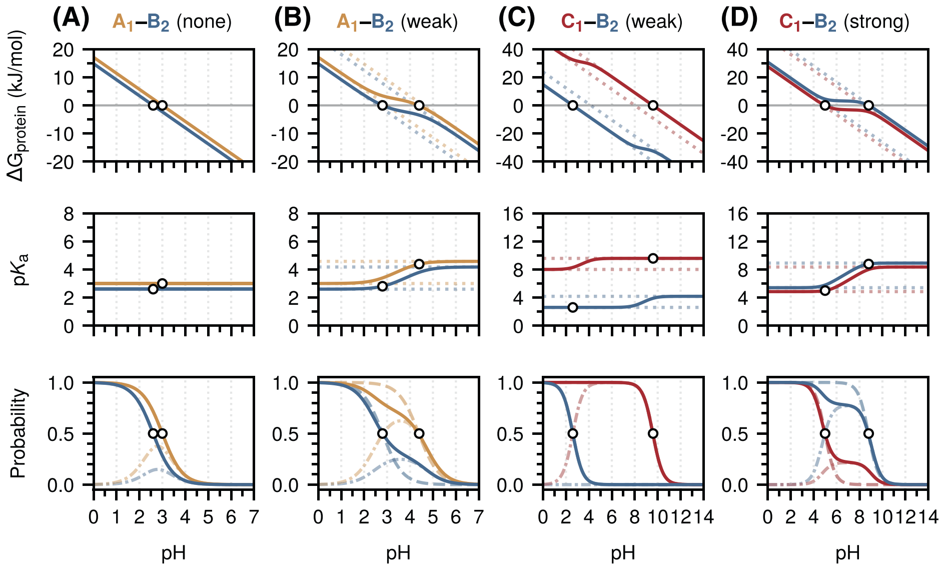

FIGURE 6 pH-dependent free energy, pK_a, and protonation probability curves: toy systems. The upper plots (solid lines) depict $\Delta G_{\text{protein}}(\text{pH})$ (Equation 8), which varies as a function of pH between its asymptotic values (dotted). The zero point of this curve (single dot) is used to resolve the corresponding pK_a value. The middle plots (solid) depict the pH-dependent pK_a value (Equation 9). Each residue has two limiting pK_a values (dotted) which correspond to the cases when the other coupled residue is protonated or deprotonated. As the pH changes, the probability that the other residue is deprotonated shifts, resulting in a pH-dependent pK_a. The lower plots depict several probabilities. The solid lines correspond to the protonation probabilities of the individual sites; these are a composite probability of the doubly protonated (i.e., $\langle A^H B^H \rangle$) and singly protonated (i.e., $\langle A^H B^- \rangle$) microstates (see Equation 6). Singly protonated probabilities are indicated with dashed-dotted lines. The dashed lines correspond to the standard Henderson-Hasselbalch curve computed for the pK_a determined by the point $\Delta G_{\text{protein}} = 0$. Because we resolve the pK_a at a single pH, these curves are sigmoidal. Observe that with no coupling (left column), the pK_a values are constant; however, when coupling is introduced, this is no longer the case. Columns (A)–(D) correspond to four different coupling scenarios.

$$\begin{tikzcd}
C_1^H + B_2^H \arrow[r, "0"] \arrow[d, "-8"'] & C_1^- + B_2^H \arrow[d, "1"] \\
C_1^H + B_2^- \arrow[r, "9"] & C_1^- + B_2^-
\end{tikzcd}$$

FIGURE 7 Thermodynamic cycle for toy model 3. ΔΔG values (kJ/mol) are indicated along the branches.

$$\begin{tikzcd}
C_1^H + B_2^H \arrow[r, "-18"] \arrow[d, "8"'] & C_1^- + B_2^H \arrow[d, "28"] \\
C_1^H + B_2^- \arrow[r, "2"] & C_1^- + B_2^-
\end{tikzcd}$$

FIGURE 8 Thermodynamic cycle for toy model 4. ΔΔG values (kJ/mol) are indicated along the branches.

the free energy curves and the inflection point of the pH-dependent pK_a values occur far from one another. As a result, the titration curves computed from the microscopic pK_a values suggest that there is no coupling between residues. The singly protonated microstates for each residue occur with a dramatically different probability, that is, residue B_2 is never protonated while C_1 deprotonates.

Model 4: Different reference pK_a° values: Strong coupling

Consider the system given in Figure 8.

Relative to Example 3 we have altered the ΔΔG values along the branches, while maintaining the reference pK_a° values. Here, the interaction between the residues is much stronger and a more favorable ΔΔG is assigned to the initial deprotonation of C_1. Unlike in Example 3, where the titration events occurred far from one another, both the $\Delta G_{\text{protein}}$ and pK_a curves resemble those in Example 2. However, unlike Example 2 the reference pK_a° values differ by 5 pK units; in this case, the reference pK_a gap is compensated by significant shifts in the ΔΔG values. Here, also note that the buffer region over which coupling occurs is larger than in Example 2 (Figure 6B, middle). In this region, the effect of pH on $\Delta G_{\text{protein}}$ is much less pronounced, and the $\Delta G_{\text{protein}}$ of protonation remains relatively constant (i.e., slope of $\Delta G_{\text{protein}}$ is zero), within 1 pK unit from zero.

As in Example 2, constructing the titration curves computed from the microscopic pK_a values reveals the coupling between residues (Figure 6C, bottom). Here, plateaus are evident for both residues over pH $\in [6,9]$, implying the existence of both singly protonated microstates. Moreover, the residue with the higher reference pK_a°, perhaps counterintuitively, titrates first.

Application to proteins

Alchemical free energies

We consider four potential dyads: three from the SNase variant[50] (PDB: 3BDC[50]) and one from protein deglycase DJ-1[51] (PDB: 1P5F[52]). These pairs are D19–D21, D21–D40, and D21–D83 in SNase and E18–C106 in DJ-1. In SNase, D19 and D21 are spatially adjacent to each other within the turn ($T1$) between $\beta 1$ and $\beta 2$, while D40 is located in the turn between $\beta 3$ and $\alpha 1$, which puts it close to D21 but farther away than D19 (Figure 9, top). D83 is located in the long linker between $\beta 4$ and $\beta 5$, and is the farthest from D21 of the dyads considered here. Although sequentially distant, E18 and C106 in DJ-1, located within $\alpha 1$ and near $\alpha 4$, respectively, are directly adjacent in space (Figure 9, bottom).

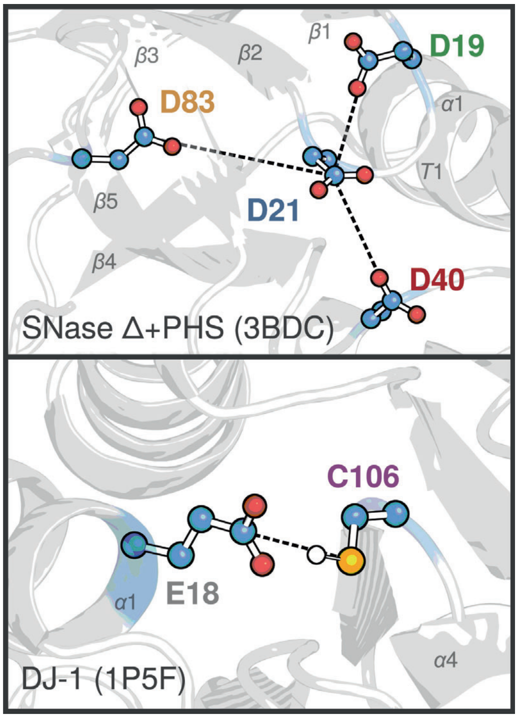

FIGURE 9 Coupled residues considered. Upper: Δ + PHS SNase (PDB: 3BDC); lower: monomeric DJ-1 (PDB: 1P5F). Carbon atoms are shown in blue, oxygen atoms in red, and sulfur atoms in yellow. Residue dyads are considered independently and are indicated with dashed lines.

The three SNase pairs exhibited different levels of coupling: the D21–D83 dyad showed almost no coupling (Figure 10A), as evidenced by the absence of inflection points in the $\Delta G_{\text{protein}}$ and pK_a curves, the D21–D40 showed moderate coupling (Figure 10B), and the D21–D19 dyad showed significant coupling (Figure 10C). In this latter case, as in the second toy example, the residues clearly acted to buffer one another, resulting in a flattening of $\Delta G_{\text{protein}}$ for both curves around pH $\in [3,5]$. Moreover, computing protonation probabilities for the individual sites revealed a non-sigmoidal form of the curves. We note that the coupling of these residues was

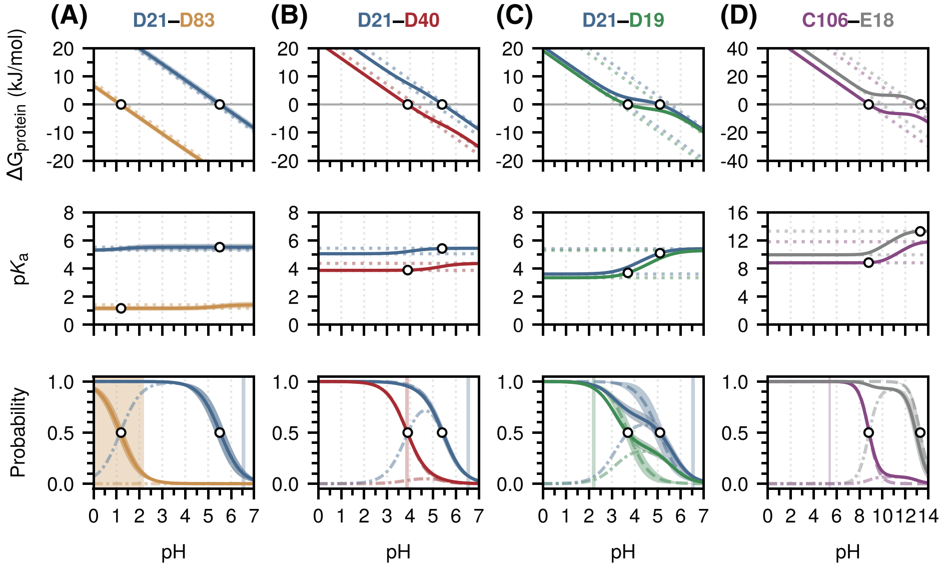

FIGURE 10 pH-dependent free energy, pK_a, and protonation probability curves: protein systems. The upper plots (solid lines) depict $\Delta G_{\text{protein}}(\text{pH})$ (Equation 8), which varies as a function of pH between its asymptotic values (dotted). The zero point of this curve (single dot) is used to resolve the corresponding pK_a value. The middle plots (solid) depict the pH-dependent pK_a value (Equation 9). Each residue has two limiting pK_a values (dotted) which correspond to the cases when the other coupled residue is protonated or deprotonated. As the pH changes, the probability that the other residue is deprotonated shifts, resulting in a pH-dependent pK_a. The lower plots depict several probabilities. Solid lines correspond to the protonation probabilities of the individual sites; these are a composite probability of the doubly protonated (i.e., $\langle A^H B^H \rangle$) and singly protonated (i.e., $\langle A^H B^- \rangle$) microstates (see Equation 6). The singly protonated probabilities are indicated with dashed-dotted lines. Dashed lines correspond to the standard Henderson-Hasselbalch curve computed for the pK_a determined by the point $\Delta G_{\text{protein}} = 0$. Because we resolve the pK_a at a single pH, these curves are sigmoidal. Observe that with no coupling (left column), the pK_a values are constant; however, when coupling is introduced, this is no longer the case. Vertical spans in the lower row indicate experimental pK_a values and uncertainties (note that D83 has a $pK_a < 2.2\,pK$). Error bands were bootstrapped. (A)–(C) correspond to the SNase + ΔPHS system, while (D) corresponds to the DJ-1 system.

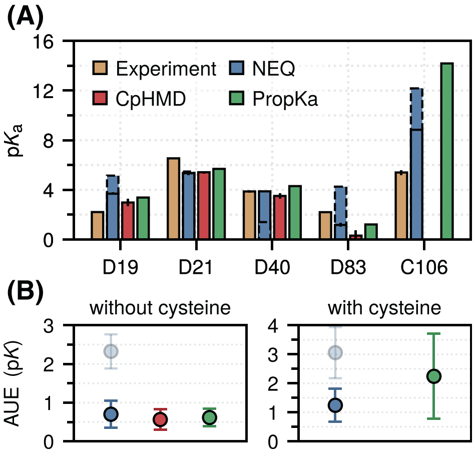

FIGURE 11 Performance of various methods for calculating protein pK_a values. Three methods are compared with experiment: NEQ with (solid) and without (dashed/transparent) coupling accounted for, constant-pH MD (CpHMD), and PropKa. Note that the pK_a of D83 is $< 2.2\,pK$. (A) Residue-wise performance. (B) Overall performance with or without cysteine included. Bootstrapped standard errors are depicted.

also observed in the experiment. The calculated pK_a values for SNase were: D19: $3.69 \pm 0.09\,pK$, D21: $5.34 \pm 0.16\,pK$, D40: $3.88 \pm 0.05\,pK$, D83: $1.17 \pm 0.16\,pK$, in good agreement with experiment (D19: $2.21 \pm 0.01\,pK$, D21: $6.54 \pm 0.02\,pK$, D40: $3.87 \pm 0.09\,pK$, and D83: $< 2.2\,pK$) (Figure 11A).

In the case of E18–C106, again a flattening in the $\Delta G_{\text{protein}}$ buffer region and, evidently, non-sigmoidal individual site pK titration curves suggested a coupling between the residues (Figure 10D). Here, computing the adjusted pK_a downshifts C106 from $12.18 \pm 0.07\,pK$ to $8.84 \pm 0.04\,pK$, bringing the estimate closer to the experimental value of $5.4 \pm 0.2\,pK$ (Figure 11A). However, this still leaves more than a 3 pK unit discrepancy between calculation and experiment. Previous work on homodimeric DJ-1 has revealed that two arginine residues (R48 and R28 from the other monomer) facilitate anion binding, which results in pK_a elevation[51] (Figure S1, right). In our simulations only positive counterions (i.e., no salt concentration) were present; however, through-space interactions as a result of a second arginine may also play a role in affecting the pK_a. To this end, we probed the pK_a of monomeric DJ-1. We found similar qualitative agreement in the curves between the dimeric and monomeric forms; however, rather than raising the pK_a, the elimination of the second arginine shifted the pK_a of C106 down to $6.78 \pm 0.19\,pK$ (Figure S1, left).

We note that in both cases an exceptionally high pK_a for E18 is predicted. While previous work on DJ-1 has suggested that E18 is protonated over the titration regime of C103[51] and glutamate residues have been reported with pK_a values greater than 9,[67] it would seem improbable that the pK_a is actually this high. Structurally, a second glutamate (E18) and a nearby histidine (H126) likely play roles within the lower pH regime (i.e., < 7). We preface the following section by noting that a high glutamate pK_a value is also suggested by the CpHMD simulations and PropKa as described further.

Constant-pH molecular dynamics and PropKa

To further investigate the reliability of our approach, we performed constant-pH molecular dynamics simulations of the same systems. Here, we found good agreement with the alchemical calculations.

We computed macroscopic pK_a values for D19/D21, D21/D40, and D21/D83 of: $2.98 \pm 0.29/5.53 \pm 0.16\,pK$, $5.20 \pm 0.10/3.51 \pm 0.21\,pK$, and $5.54 \pm 0.16/0.31 \pm 0.44\,pK$. In the case of D19/D21 we computed a non-zero interaction energy of $w = 1.83 \pm 0.23$. Pairwise comparison between the CpHMD and experimental values revealed a good agreement: 2.98 ± 0.29 versus 2.21 ± 0.07 (D19), 5.42 ± 0.09 versus 6.54 ± 0.02 (D21), 3.51 ± 0.21 versus 3.87 ± 0.09 (D40), and 0.31 ± 0.44 versus < 2.2 (D83).

Regarding DJ-1, the large pK_a value of E18 implied by free energy calculations is also suggested by the constant-pH simulations on monomeric DJ-1, where values of $8.75 \pm 0.27\,pK$ and $> 14 \pm 0.05\,pK$ were calculated in the cases of protonated C106 and deprotonated C106, respectively. C106 was not probed as cysteine residues are not yet supported by the current CpHMD implementation.[43]

In addition to the protonation probability we can also directly apply Equation (2) to each CpHMD pH simulation, this yields the pH-dependent pK_a. As discussed above, in the ideal case, uncoupled residues will have constant pK_a values while coupled pairs will exhibit sigmoidal behaviour. Computing these for the CpHMD simulations and the corresponding model curves revealed good agreement with those curves implied by the alchemical free energy calculations (Figure S2). In particular, for the coupled pairs (i.e., D19–D21 and D21–D40), we observed sigmoidal behavior.

We close this section by briefly comparing our results with the popular computational pK_a predictor PropKa (version 3.4).[68] Overall, for PropKa, we find good agreement for the SNase dyads, but a worse accuracy for C106 in DJ-1. We preface by noting that in the case of D19/D21, Propka identifies these as coupled and we can compute the pK_a values associated with the four underlying microstates: two values are computed by running default Propka and two are computed by running Propka with the `-d` flag, which probes an alternative protonation state. In default mode, D21 is protonated when D19 titrates (i.e., D19: $3.42\,pK$, D21: $5.65\,pK$) while in `-d` mode, D19 titrates after D21 (i.e., D19: $4.63\,pK$, D21: $4.44\,pK$). Given these microscopic values (i.e., $pK_1 = 4.63$, $pK_2 = 5.65$) we can employ the Hill/Ullmann formalism with $w = 4.63 - 3.42 = 5.65 - 4.44 = 1.21$ and resolve the macroscopic pK_as (see Appendix A in the Supporting Information for details). This yielded values of $3.38\,pK$ and $5.69\,pK$ for D19 and D21, respectively, and values of $4.30\,pK$, $1.21\,pK$, and $14.19\,pK$ for D40, D83, and C106, where coupling was not assumed by PropKa (Figure 11A). In the case of E18 in DJ-1, PropKa estimates a pK_a of $8.73\,pK$ and $7.38\,pK$ for the homodimeric and monomeric forms, respectively.

Here, our NEQ approach provided estimates of both D40 and C106 that were in closer agreement with the experimental values while exhibiting comparable performance on the other three residues. Considering the overall performance we found that the introduction of coupling dramatically improves agreement with experiment reducing our NEQ average unsigned error from $2.10 \pm 0.42\,pK$ to $0.69 \pm 0.34\,pK$ when cysteine is excluded and from $3.05 \pm 0.88\,pK$ to $1.23 \pm 0.57\,pK$ when it is included; with regard to the former, this performance was comparable to PropKa (AUE: $0.61 \pm 0.22\,pK$) (Figure 11B). We also found that CpHMD could accurately resolve the four aspartates with an unsigned error of $0.57 \pm 0.26\,pK$. It is probable that both a limited training set and a less frequent dyad (i.e., a large difference in reference pK_a° values) results in the markedly poorer estimate for C106 from PropKa.

Practical guide to using alchemical ΔΔG to account for residue coupling

We describe how to compute the pK_a corresponding to the upper branch in Figure 3. More specifically, we are interested in the pH at which $\Delta G_{\text{protein}}(\text{pH}) = 0$ (see Appendix A in the Supporting Information for details).

This function has a pH-independent component which captures the free energy of deprotonation that arises due to the local environment (e.g., $\Delta\Delta G_0$) without accounting for coupling, and a pH-dependent component that captures the coupling contribution arising due to nearby residues that shift the favorability of the underlying deprotonation reaction (e.g., $\Delta\Delta G_1$ and $\Delta\Delta G_2$). We resolve this function accordingly:

1. We run three variant protonation simulations using the double-system single-box setup:
   a. $A^H + B^H \rightarrow A^- + B^H$
   b. $A^H + B^H \rightarrow A^H + B^-$
   c. $A^- + B^H \rightarrow A^- + B^-$

   and compute the corresponding ΔΔG values in each case: $\Delta\Delta G_0$, $\Delta\Delta G_1$, and $\Delta\Delta G_2$. A fourth ΔΔG is implied by cycle closure; however, this can also be explicitly computed. Note that these are double free energy differences because of the simulation setup, where both the protein and peptide are present.
2. In the absence of any coupling, as the pH changes, the free energy of deprotonation in the protein shifts linearly according to

$$\Delta G_i(\text{pH}) = \Delta\Delta G_i + RT\log(10)\left(pK_a^{\circ} - \text{pH}\right)$$

   where pK_a° is the reference pK_a of the residue under consideration. We can use this relationship to calculate

$$\Delta G_1(\text{pH}) = \Delta\Delta G_1 + RT\log(10)\left(pK_a^{\circ} - \text{pH}\right)$$

and

$$\Delta G_2(\mathrm{pH}) = \Delta\Delta G_2 + RT\log(10)\left(\mathrm{p}K_a^\circ - \mathrm{pH}\right);$$

here, $\mathrm{p}K_a^\circ$ corresponds to the reference $\mathrm{p}K_a$ of residue B.

Note that $\Delta G_1(\mathrm{pH})$ and $\Delta G_2(\mathrm{pH})$ act on the "reactants" and "products" of the upper branch in Figure 3: a more favorable $\Delta G_1(\mathrm{pH})$ will increase $\Delta\Delta G(\mathrm{pH})$ while a more favorable $\Delta G_2(\mathrm{pH})$ will decrease $\Delta\Delta G(\mathrm{pH})$, and vice-versa.

3. We can resolve $\Delta G_{\mathrm{protein}}(\mathrm{pH})$ according to

$$\Delta G_{\mathrm{protein}}(\mathrm{pH}) = \Delta\Delta G(\mathrm{pH}) + \Delta G_{\mathrm{deprot}}(\mathrm{pH})$$

where

$$\begin{aligned}\Delta\Delta G(\mathrm{pH}) = \Delta\Delta G_0 \quad &+\frac{1}{\beta}\log\left(1+e^{-\beta\Delta G_1(\mathrm{pH})}\right)\\ &-\frac{1}{\beta}\log\left(1+e^{-\beta\Delta G_2(\mathrm{pH})}\right).\end{aligned}$$

and

$$\Delta G_{\mathrm{deprot}}(\mathrm{pH}) = RT\log(10)\left(\mathrm{p}K_a^\circ - \mathrm{pH}\right);$$

here, $\mathrm{p}K_a^\circ$ corresponds to the reference $\mathrm{p}K_a$ of residue A.

4. We can find the pH at which $\Delta G_{\mathrm{protein}}(\mathrm{pH}) = 0$; this pH will correspond to the apparent $\mathrm{p}K_a$.

We can also compute the $\mathrm{p}K_a$ at arbitrary pH and construct a pH-dependent curve via

$$\mathrm{p}K_a = \mathrm{p}K_a^\circ + \frac{\Delta G_{\mathrm{protein}}(\mathrm{pH}) - \Delta G_{\mathrm{deprot}}(\mathrm{pH})}{RT\log(10)}.$$

(Note that if we are only interested in the $\mathrm{p}K_a$, we can stop here. In order to resolve the individual site and microstate probabilities we need to follow the next three steps.)

5. Consider that the deprotonation free energy is related to the standard deprotonation free energy via

$$\Delta G_i = \Delta G_i(\mathrm{pH}=0) - \mu_{\mathrm{H}^+},$$

where $\Delta G_i(\mathrm{pH}=0)$ values are from Step 2 of this protocol (at standard conditions, $\mathrm{pH}=0$) and μ_{H^+} is the chemical potential of the protons in the solution: $\mu_{\mathrm{H}^+} = RT\log(10)\mathrm{pH}$.

6. We then have direct access to the partition function

$$\begin{aligned}\xi = 1 \quad &+ e^{-\beta\left(\Delta G_0(0)-\mu_{\mathrm{H}^+}\right)}\\ &+ e^{-\beta\left(\Delta G_1(0)-\mu_{\mathrm{H}^+}\right)}\\ &+ e^{-\beta\left(\Delta G_0(0)+\Delta G_2(0)-2\mu_{\mathrm{H}^+}\right)},\end{aligned}$$

and corresponding microstate probabilities

$$\begin{aligned}\langle \mathrm{A^H B^H}\rangle &= \frac{1}{\xi},\\ \langle \mathrm{A^- B^H}\rangle &= \frac{e^{-\beta\left(\Delta G_0(0)-\mu_{\mathrm{H}^+}\right)}}{\xi},\\ \langle \mathrm{A^H B^-}\rangle &= \frac{e^{-\beta\left(\Delta G_1(0)-\mu_{\mathrm{H}^+}\right)}}{\xi}, \text{ and}\\ \langle \mathrm{A^- B^-}\rangle &= \frac{e^{-\beta\left(\Delta G_0(0)+\Delta G_2(0)-2\mu_{\mathrm{H}^+}\right)}}{\xi}.\end{aligned}$$

7. The overall protonation probability for an individual site can be computed from:

$$\begin{aligned}\langle \mathrm{A^H}\rangle &= \langle \mathrm{A^H B^H}\rangle + \langle \mathrm{A^H B^-}\rangle\\ \langle \mathrm{B^H}\rangle &= \langle \mathrm{A^H B^H}\rangle + \langle \mathrm{A^- B^H}\rangle\end{aligned}$$

DISCUSSION

The importance of accounting for residue coupling is multifaceted and particularly relevant in the context of enzymatic active sites that are often enriched in protonatable residues. The theoretical formalism to describe such couplings in polyprotic acids has been detailed by Ullman.[28] Based on earlier work by Hill,[22,23] Edsall,[24–26] and coworkers, Ullman defines equilibrium protonation constants for all microstates and derives partition functions that fully describe the thermodynamics of these systems. This approach was then applied in a protein context using continuum electrostatics calculations to resolve the free energies between microstates.[29]

Here, we demonstrate that such a framework can be readily extended to double free energy difference calculations based on molecular dynamics simulations and introduce a thermodynamic cycle approach that allows one to resolve the apparent $\mathrm{p}K_a$s of coupled residues. This demonstration and extension is particularly relevant in the context of alchemical free energy calculations, which give access only to such ΔΔG values. Our overall aim was to demonstrate that insights into residue coupling can be resolved by means of alchemical free energy calculations. While such insights can be directly resolved via constant-pH molecular dynamics simulations—an exceedingly powerful method—we demonstrate with both toy examples and real protein systems that similar insights can be extracted via a different approach.

One such insight is the buffering of a residue dyad which maintained the free energy of protonation near zero (e.g., DJ-1: C106-E18). In various protein contexts, tuning the local residue environment surrounding pairs or groups of titratable residues to create large buffer regions over which the ΔG of protonation is close to zero could make the binding of a substrate more than sufficient to significantly alter the (de)protonation of a residue and ultimately the enzymatic activity.

Comparison with a recent GROMACS-based CpHMD implementation[43,69] also revealed a good $\mathrm{p}K_a$ prediction accuracy for coupled residues. This result suggests that both CpHMD and alchemical free energy methods can resolve pK_a values in both coupled and uncoupled contexts. The observation that CpHMD is capable of describing protonation coupling phenomena is not new and indeed such methods have been previously employed in this capacity.[38,70,71] Moreover, CpHMD has the distinct advantage that all residues can be simultaneously assessed for coupling, with deprotonation dynamically occurring over the course of a simulation due to changes in the protein or environment.[70,72] For instance, the ability to capture structural reorganizations due to (de)protonation has helped rationalize enzyme catalytic function: robustly identifying those residues that act as proton donors or nucelophiles and determining the basis for their pK_a shifts.[73] In addition to resolving coupled pK_a values,[71,73–76] CpHMD can accurately describe pH-dependent conformational changes,[77–81] protein interactions,[82,83] and ligand binding.[84–87] As CpHMD has developed, methodological refinements and extensions have improved performance and accuracy.[65,88–94] Nevertheless, the approach can still remain computationally expensive, necessitating multiple simulations across a pH spectrum, and requires residue-specific parameterizations (e.g., correction potentials,[35,69] empirical energy offsets[39,41]), which must be adjusted if the underlying residue parameters (e.g., partial charge) change. Here, we demonstrate that pK_a coupling insights can be extracted from "plain" molecular dynamics simulations paired with alchemical free energy methods; such an approach requires at most four $\Delta\Delta G$ evaluations per coupled pair.

A second comparison with PropKa suggested that although accurate estimates could be made for SNase, the pK_a of C106 in DJ-1 was significantly overestimated, possibly due to the limited cysteine data in the PropKa training set and the implicit assumption that E18 is deprotonated at the pH where C106 deprotonates.

The role of MD simulations and free energy calculation methods such as those employed here may provide insight not readily accessible to conventional prediction methods. One particular insight, namely the pK_a values of coupled residues, requires careful consideration of the role of nearby protonatable residues. Moreover, given that these residues are often found at the active site—frequently the target of engineered therapeutics—the relevance of this problem extends beyond basic research.

We underscore that while here we have employed non-equilibrium free energy calculations, the approaches outlined can be employed alongside any alchemical-based free energy method (e.g., FEP); however, the cost associated with converging such calculations should be further investigated. The ability to seamlessly and consistently integrate pH-dependent calculations into existing alchemical free energy workflows may prove useful for accurately resolving binding affinities or enzyme activities and thermostabilities.

To this end, we have elsewhere investigated the ability of alchemical free energy calculations to compute a large number of pK_a values in a variety of protein contexts. As with the results here, our approach showed strong performance, further suggesting the potential for a consistent integration of pH-dependent calculations into a broader alchemical free energy framework.[45]

ACKNOWLEDGMENTS

Carter J. Wilson thanks the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Government of Ontario for funding. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available at https://github.com/deGrootLab/pka_coupling_2024.

ORCID

Carter J. Wilson https://orcid.org/0000-0002-8992-6269
Vytautas Gapsys https://orcid.org/0000-0002-6761-7780

REFERENCES

[1] M. F. Perutz, *Science* **1978**, *201*, 1187.
[2] J. Srivastava, G. Barreiro, S. Groscurth, A. R. Gingras, B. T. Goult, D. R. Critchley, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14436.
[3] S. Kaptan, M. Assentoft, H. P. Schneider, R. A. Fenton, J. W. Deitmer, N. MacAulay, et al., *Structure* **2015**, *23*, 2309.
[4] S. E. Boyken, M. A. Benhaim, F. Busch, M. Jia, M. J. Bick, H. Choi, et al., *Science* **2019**, *364*, 658.
[5] Z. A. Ripstein, S. Vahidi, J. L. Rubinstein, L. E. Kay, *J. Am. Chem. Soc.* **2020**, *142*, 20519.
[6] M. Tollinger, K. A. Crowhurst, L. E. Kay, J. D. Forman-Kay, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 4545.
[7] C. N. Pace, G. R. Grimsley, J. M. Scholtz, *J Biol Chem* **2009**, *284*, 13285.
[8] H. Watanabe, C. Yoshida, A. Ooishi, Y. Nakai, M. Ueda, Y. Isobe, et al., *ACS Chem. Biol.* **2019**, *14*, 2729.
[9] X. Yao, C. Chen, Y. Wang, S. Dong, Y. J. Liu, Y. Li, et al., *Sci. Adv.* **2020**, *6*, eabd7182.
[10] A. Warshel, *Biochemistry* **1981**, *20*, 3167.
[11] Y. Y. Sham, Z. T. Chu, A. Warshel, *J Phys Chem B* **1997**, *101*, 4458.
[12] C. Tanford, R. Roxby, *Appl Lysozyme. Biochem* **1972**, *11*, 2192.
[13] R. I. Shrager, J. S. Cohen, S. R. Heller, D. H. Sachs, A. N. Schechter, *Biochemistry* **1972**, *11*, 541.
[14] H. Webb, B. M. Tynan-Connolly, G. M. Lee, D. Farrell, F. OMeara, C. R. Søndergaard, et al., *Proteins* **2010**, *79*, 685.
[15] Z. Y. Zhang, J. E. Dixon, *Biochemistry* **1993**, *32*, 9340.
[16] L. P. McIntosh, G. Hand, P. E. Johnson, M. D. Joshi, M. Körner, L. A. Plesniak, et al., *Biochemistry* **1996**, *35*, 9958.
[17] G. Dodson, *Trends Biochem. Sci.* **1998**, *23*, 347.
[18] Z. Du, Y. Zheng, M. Patterson, Y. Liu, C. Wang, *J. Am. Chem. Soc.* **2011**, *133*, 10275.
[19] K. Sakurai, Y. Goto, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15346.
[20] A. Warshel, S. T. Russell, *Q. Rev. Biophys.* **1984**, *17*, 283.
[21] K. M. Merz, *J. Am. Chem. Soc.* **1991**, *113*, 3572.
[22] T. L. Hill, *J. Phys. Chem.* **1944**, *48*, 101.
[23] T. L. Hill, *J. Am. Chem. Soc.* **1956**, *78*, 3330.
[24] J. T. Edsall, J. Wyman, *Biophysical Chemistry: Thermodynamics, Electrostatics, and the Biological Significance of the Properties of Matter*, Cambridge, MA, United States, Academic Press Inc, **1958**.
[25] J. T. Edsall, R. B. Martin, B. R. Hollingworth, *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 505.
[26] R. B. Martin, J. T. Edsall, D. B. Wetlaufer, B. R. Hollingworth, *J Biol Chem* **1958**, *233*, 1429.
[27] A. Onufriev, D. A. Case, G. M. Ullmann, *Biochemistry* **2001**, *40*, 3413.
[28] G. M. Ullmann, *J Phys Chem B* **2003**, *107*, 1263.

[29] E. Bombarda, G. M. Ullmann, *J Phys Chen B* **2010**, *114*, 1994.
[30] T. L. Hill, *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*, 1st ed. Springer Series in Molecular and Cell Biology, Springer, New York, NY **1985**.
[31] P. Kollman, *Chem. Rev.* **1993**, *93*, 2395.
[32] X. Kong, C. L. Brooks, *J Chem Phys* **1996**, *105*, 2414.
[33] A. M. Baptista, V. H. Teixeira, C. M. Soares, *J Chem Phys* **2002**, *117*, 4184.
[34] R. Bürgi, P. A. Kollman, W. F. van Gunsteren, *Proteins* **2002**, *47*, 469.
[35] M. S. Lee, F. R. Salsbury, C. L. Brooks, *Proteins* **2004**, *56*, 738.
[36] T. Simonson, J. Carlsson, D. A. Case, *J. Am. Chem. Soc.* **2004**, *126*, 4167.
[37] J. Mongan, D. A. Case, J. A. McCammon, *J. Comput. Chem.* **2004**, *25*, 2038.
[38] J. Khandogin, C. L. Brooks, *Biophys. J.* **2005**, *89*, 141.
[39] H. A. Stern, *J Chem Phys* **2007**, *126*, 164112-1. https://doi.org/10.1063/1.2731781
[40] S. Donnini, F. Tegeler, G. Groenhof, H. Grubmüller, *J. Chem. Theory Comput.* **2011**, *7*, 1962.
[41] Y. Chen, B. Roux, *J. Chem. Theory Comput.* **2015**, *11*, 3919.
[42] B. K. Radak, C. Chipot, D. Suh, S. Jo, W. Jiang, J. C. Phillips, et al., *J. Chem. Theory Comput.* **2017**, *13*, 5933.
[43] N. Aho, P. Buslaev, A. Jansen, P. Bauer, G. Groenhof, B. Hess, *J. Chem. Theory Comput.* **2022**, *18*, 6148.
[44] D. Coskun, W. Chen, A. J. Clark, C. Lu, E. D. Harder, L. Wang, et al., *J. Chem. Theory Comput.* **2022**, *18*, 7193.
[45] C. J. Wilson, M. Karttunen, B. L. de Groot, V. Gapsys, *J. Chem. Theory Comput.* **2023**, *19*, 7833.
[46] E. Awoonor-Williams, A. A. Golosov, V. Hornak, *J. Chem. Inf. Model.* **2023**, *63*, 2170.
[47] K. A. Sharp, B. Honig, *Annu Rev Biophys Biophys Chem* **1990**, *19*, 301.
[48] C. Tanford, J. G. Kirkwood, *J. Am. Chem. Soc.* **1957**, *79*, 5333.
[49] V. Gapsys, S. Michielssens, D. Seeliger, B. L. de Groot, *J. Comput. Chem.* **2014**, *36*, 348.
[50] C. A. Castañeda, C. A. Fitch, A. Majumdar, V. Khangulov, J. L. Schlessman, B. E. García-Moreno, *Proteins* **2009**, *77*, 570.
[51] A. C. Witt, M. Lakshminarasimhan, B. C. Remington, S. Hasim, E. Pozharski, M. A. Wilson, *Biochemistry* **2008**, *47*, 7430.
[52] M. A. Wilson, J. L. Collins, Y. Hod, D. Ringe, G. A. Petsko, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9256.
[53] V. Gapsys, S. Michielssens, J. H. Peters, B. L. Groot, H. Leonov, *Molecular Modeling of Proteins*, New York, NY, United States, Humana Press, **2015**, p. 173.
[54] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, et al., *Nat. Methods* **2016**, *14*, 71.
[55] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, et al., *J Phys Chem B* **1998**, *102*, 3586.
[56] W. F. V. Gunsteren, H. J. C. Berendsen, *Mol Sim* **1988**, *1*, 173.
[57] N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, H. J. C. Berendsen, *J. Chem. Theory Comput.* **2012**, *8*, 3637.
[58] M. Parrinello, A. Rahman, *J. Appl. Phys.* **1981**, *52*, 7182.
[59] T. Darden, D. York, L. Pedersen, *J Chem Phys* **1993**, *98*, 10089.
[60] B. Hess, *J. Chem. Theory Comput.* **2007**, *4*, 116.
[61] C. H. Bennett, *J. Comput. Phys.* **1976**, *22*, 245.
[62] G. E. Crooks, *Phys Rev E* **1999**, *60*, 2721.
[63] V. Gapsys, B. L. de Groot, *eLife* **2020**, *9*, e57589. https://doi.org/10.7554/elife.57589
[64] J. A. Henderson, Y. Huang, O. Beckstein, J. Shen, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 25517.
[65] J. A. Wallace, J. K. Shen, *J Chem Phys* **2012**, *137*, 184105-1. https://doi.org/10.1063/1.4766352
[66] A. R. Klingen, E. Bombarda, G. M. Ullmann, *Photochem. Photobiol. Sci.* **2006**, *5*, 588.
[67] G. R. Grimsley, J. M. Scholtz, C. N. Pace, *Prot Sci* **2008**, *18*, 247.
[68] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525.
[69] P. Buslaev, N. Aho, A. Jansen, P. Bauer, B. Hess, G. Groenhof, *J. Chem. Theory Comput.* **2022**, *18*, 6134.
[70] J. Mongan, D. A. Case, *Curr. Opin. Struct. Biol.* **2005**, *15*, 157.
[71] J. A. Wallace, J. K. Shen, *In: Predicting pKa Values with Continuous Constant pH Molecular Dynamics*, Cambridge, MA, United States, Academic Press, **2009**, p. 455. https://doi.org/10.1016/S0076-6879(09)66019-5
[72] V. Martins de Oliveira, R. Liu, J. Shen, *Curr. Opin. Struct. Biol.* **2022**, *77*, 102498.
[73] Y. Huang, Z. Yue, C. C. Tsai, J. A. Henderson, J. Shen, *J. Phys. Chem. Lett.* **2018**, *9*, 1179.
[74] J. Khandogin, C. L. Brooks, *Biochemistry* **2006**, *45*, 9363.
[75] A. Panahi, C. L. Brooks, *J Phys Chem B* **2015**, *119*, 4601.
[76] F. Hofer, J. Kraml, U. Kahler, A. S. Kamenik, K. R. Liedl, *J. Chem. Inf. Model.* **2020**, *60*, 3030.
[77] N. V. D. Russo, D. A. Estrin, M. A. Martí, A. E. Roitberg, *PLoS Comput. Biol.* **2012**, *8*, e1002761.
[78] M. Machuqueiro, A. M. Baptista, *Biophys. J.* **2007**, *92*, 1836.
[79] Z. Yue, W. Chen, H. I. Zgurskaya, J. Shen, *J. Chem. Theory Comput.* **2017**, *13*, 6405.
[80] A. Sarkar, A. E. Roitberg, *J Phys Chem B* **2020**, *124*, 11072.
[81] H. Torabifard, A. Panahi, C. L. Brooks, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 3583.
[82] X. Zeng, S. Mukhopadhyay, C. L. Brooks, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2034.
[83] S. M. Law, B. W. Zhang, C. L. Brooks, *Protein Sci.* **2013**, *22*, 595.
[84] M. O. Kim, P. G. Blachly, J. W. Kaus, J. A. McCammon, *J Phys Chem B* **2014**, *119*, 861.
[85] C. R. Ellis, C. C. Tsai, X. Hou, J. Shen, *J. Phys. Chem. Lett.* **2016**, *7*, 944.
[86] T. J. Paul, J. Z. Vilseck, R. L. Hayes, C. L. Brooks, *J Phys Chem B* **2020**, *124*, 6520.
[87] P. L. Gupta, J. S. Smith, A. E. Roitberg, *J Phys Chem B* **2021**, *125*, 9168.
[88] Y. Meng, A. E. Roitberg, *J. Chem. Theory Comput.* **2010**, *6*, 1401.
[89] J. A. Wallace, J. K. Shen, *J. Chem. Theory Comput.* **2011**, *7*, 2617.
[90] S. G. Itoh, A. Damjanović, B. R. Brooks, *Proteins* **2011**, *79*, 3420.
[91] W. Chen, J. A. Wallace, Z. Yue, J. K. Shen, *Biophys. J.* **2013**, *105*, L15.
[92] S. Donnini, R. T. Ullmann, G. Groenhof, H. Grubmüller, *J. Chem. Theory Comput.* **2016**, *12*, 1040.
[93] R. C. Harris, J. Shen, *J. Chem. Inf. Model.* **2019**, *59*, 4821.
[94] O. Bignucolo, C. Chipot, S. Kellenberger, B. Roux, *J Phys Chem B* **2022**, *126*, 6868.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: C. J. Wilson, B. L. de Groot, V. Gapsys, *J. Comput. Chem.* **2024**, 1. https://doi.org/10.1002/jcc.27318