Constant pH Molecular Dynamics with Proton Tautomerism

Jana Khandogin and Charles L. Brooks III

Department of Molecular Biology, The Scripps Research Institute, La Jolla, California

ABSTRACT   The current article describes a new two-dimensional λ-dynamics method to include proton tautomerism in continuous constant pH molecular dynamics (CPHMD) simulations. The two-dimensional λ-dynamics framework is used to devise a tautomeric state titration model for the CPHMD simulations involving carboxyl and histidine residues. Combined with the GBSW implicit solvent model, the new method is tested on titration simulations of blocked histidine and aspartic acid as well as two benchmark proteins, turkey ovomucoid third domain (OMTKY3) and ribonuclease A (RNase A). A detailed analysis of the errors inherent to the CPHMD methodology as well as those due to the underlying solvation model is given. The average absolute error for the computed pK_a values in OMTKY3 is 1.0 pK unit. In RNase A the average absolute errors for the carboxyl and histidine residues are 1.6 and 0.6 pK units, respectively. In contrast to the previous work, the new model predicts the correct sign for all the pK_a shifts, but one, in the benchmark proteins. The predictions of the tautomeric states of His[12] and His[48] and the conformational states of His[48] and His[119] are in agreement with experiment. Based on the simulations of OMTKY3 and RNase A, the current work has demonstrated the capability of the CPHMD technique in revealing pH-coupled conformational dynamics of protein side chains.

INTRODUCTION

The stability and function of proteins are dependent on the environmental pH (1). Well-known examples of pH-dependent biological processes include fibril formation from one of 16 normally soluble and functional human amyloidogenic proteins such as amyloid β-peptides, transthyretin, and prion proteins (2,3), membrane insertion of diphtheria toxin (4) and influenza virus hemagglutinin (5), and proton gradient-driven ATP synthesis (6), as well as the catalytic pathway of dihydrofolate reductase (7). In a conventional molecular dynamics (MD) simulation the protonation state of the protein is preset and kept fixed according to the known pK_a values of the ionizable side chains. Two problems arise from the assumption of static protonation. First, because the pK_a value of an ionizable side chain is typically unknown in its protein environment one has to resort to the value of an isolated amino acid (model pK_a), which sometimes deviates significantly from the value in the protein. Second, the protonation/deprotonation of an ionizable side chain is a process that is coupled to the dynamics of the local environment. By artificially fixing the protonation state in a MD simulation, the dynamics of the ionization equilibrium is neglected, thereby precluding detailed understanding of the specific interactions that give rise to the pH-coupled biological phenomena.

The history of pH-coupled molecular dynamics is relatively short. A decade ago, Mertz and Pettitt (8) demonstrated a titration simulation of acetic acid using an open system Hamiltonian. Sham et al. (9) showed that the pK_a values of lysozyme can be accurately modeled using the protein dipoles Langevin dipoles model, which treats the protein relaxation in the microscopic framework of the linear response approximation. In more recent years, several simulation techniques that allow pH-coupled molecular dynamics have emerged. Most of these techniques are based on the combination of molecular dynamics (MD) and Monte Carlo (MC) sampling of protonation states, with the major differences being the solvent representation and the way the free energy change accompanying the switch of the protonation state is computed. In the methods of Baptista et al. (10,11) and Bürgi et al. (12), explicit solvent is used to generate the MD trajectory. While the method of Bürgi et al. (12) utilizes thermodynamic integration to determine the protonation free energy change of a single residue, the method of Baptista et al. (11) makes use of static pK_a calculations based on the implicit solvent Poisson-Boltzmann (PB) model. More recently, these hybrid MD/MC schemes were extended to implicit solvent simulations: Dlugosz and Antosiewicz (13) make use of the analytic continuum electrostatic and the PB models for generating the MD trajectory and evaluating the protonation free energy change respectively; Mongan et al. (14) use the Generalized-Born (GB) model throughout the simulation.

There are several issues that have to be dealt with in the hybrid MD/MC constant pH simulation schemes. First, the periodic abrupt switch in the protonation state introduces a discontinuity in energy and force, which may result in conformational and energetic instabilities. Second, since only one protonation site at an instantaneous conformation of the protein is randomly chosen during a MC step, sampling convergence with regard to protonation and conformational states may be slow for proteins containing many ionizable sites. Third, some issues remain relating to the determination of the free energy change accompanying the protonation

state switch. In the methods of Baptista et al. (11) and Dlugosz and Antosiewicz (13), which use the instantaneous protein conformation for the evaluation of the free energy difference, a somewhat ad hoc protein dielectric constant has to be chosen to mediate the electrostatic interactions in the otherwise static protein configuration. On the other hand, in the method of Bürgi et al. (12), which employs a short run (tens of picoseconds) of free energy simulation such as thermodynamic integration, the increased computational cost associated with the large of number of MC steps has to be considered.

In contrast to the discrete methods, the methods of Börjesson and Hünenberger (15), and of Lee et al. (16) are based on simultaneous propagation of continuous titration and conformational degrees of freedom. Both approaches use linear combinations of the deprotonated and protonated states to model nonbonded interactions. Although the method of Börjesson and Hünenberger relies completely upon explicit solvent simulation and uses the λ-variable to represent the extent of protonation for each ionizable site (15), the method of Lee et al., which has its roots in the λ-dynamics method (17), utilizes the implicit GB solvent model (18) and attaches physical meaning only to the end points. To minimize the possible distortion due to the unphysical mixed states that are nonetheless necessary for the transition between the end-point states, an artificial titration barrier is utilized in the continuous pH molecular dynamics method (CPHMD) of Lee et al. (16), which serves to prolong the residency time of the end-point states. The combined advantage of a stable dynamics trajectory and low computational cost makes the CPHMD approach of Lee et al. particularly suitable and attractive as a means of exploring pH-dependent conformational processes in proteins.

One prerequisite for the application of the constant pH molecular dynamics method to pH-coupled protein dynamics is reasonable accuracy in the prediction of protonation states or pK_a values for ionizable side chains. Among the aforementioned methods, so far, only three have been tested on proteins. While two of them, the hybrid MD/MC methods of Bürgi et al. (12) and Mongan et al. (14) were tested only on hen egg white lysozyme, the CPHMD method of Lee et al. (16) was tested on four different proteins, including turkey ovomucoid third domain (OMTKY3), bovine pancreatic ribonuclease A (RNase A), bovine trypsin inhibitor, and hen egg white lysozyme. In this pilot study, the total average absolute error over all proteins was 1.6 pK units (16).

A quick survey of the computed pK_a values in the work of Lee et al. (16) reveals that the average absolute error is largest for the histidine groups (3.2 pK units), followed by the carboxyl (1.6 pK units) and amino groups (1.2 pK units). One obvious reason for the increased errors in predicting the pK_a of histidine is the simplistic treatment of proton isomerism, in which case the imidazole ring is split into two titration halves involving the ND1 and NE2 sites with an added penalty potential to prevent double deprotonation

(16). This model will be referred to as the split model in the rest of the article. In a constant pH MD simulation, the split model gives rise to two problems. First, since the two titration halves have common atoms, proper atomic charge states cannot be assigned to the neutral histidine tautomers, HSE (protonated on NE2) and HSD (protonated on ND1). Second, the penalty function to eliminate double deprotonated histidine, $K\lambda_1\lambda_2$, introduces an undesired biasing force.

In the present work, a new tautomeric state titration model, also referred to as the double-site model, is developed. This model allows simultaneous titration at two competing protonation sites, such as the NE2 and ND1 atoms in histidine. The new model utilizes an additional dynamic variable that is treated on an equal footing to the titration degrees of freedom to control the tautomer interconversion process. Besides histidine, which exhibits the true proton tautomerism in solution, carboxyl groups such as Asp, Glu, and the C-terminus can be considered as having two tautomeric protonation sites, since the exchange of two titrating oxygens due to rotation around the single bond is slow on the routine MD simulation timescale. In essence, our tautomeric state titration model is a two-dimensional λ-dynamics approach that can, in general, be extended to additional dimensions. For example, the titration of amino groups can be viewed as a tautomeric equilibrium among three states and can be dealt with by constructing a three-dimensional model. However, since most amino groups do not exhibit large pK_a shifts, as seen for example in hen egg white lysozyme, the necessity for the development of a more sophisticated model is questionable. In fact, as mentioned previously, the errors in the computed pK_a values for amino groups are the smallest in the existing CPHMD method of Lee et al. (16) Finally, it is worthwhile noting that although the two-dimensional λ-dynamics approach is applied to the coupled protonation and tautomeric interconversion equilibria in the present work, the approach is general and can be applied to other problems.

The CPHMD method is further extended in the present study to incorporate the newly developed GBSW implicit solvent model (19) because of the added capability of this method for simulations in a membrane environment (20) and for the greater computational efficiency (21) relative to the GBMV method (18) that was implemented in the original CPHMD code. Since the accuracy of the CPHMD simulations is intimately linked to conformational sampling and the underlying implicit solvent model, the applications of CPHMD will benefit from the improved sampling schemes as well as the solvent model.

The purpose of the present work is to extend and improve the existing CPHMD method developed by Lee et al. (16). In their work (16) it was noted that CPHMD simulations using the GBMV solvent model gave an overestimation of 0.3 pK units for the blocked aspartic acid. This has prompted us to examine the details of the titration simulations of model

compounds aimed at understanding the convergence behavior and errors inherent to the CPHMD methodology. Other errors in the prediction of protein ionization equilibrium arising from the force field and implicit solvent model will be explored by revisiting two proteins: OMTKY3 and RNase A. Since the predictions of pK_a shifts in the carboxyl and histidine residues in these proteins proved particularly problematic in the previous work (16), they provide excellent test beds for the tautomeric state model as well as the GBSW solvent model.

The rest of the article is organized as follows. In the Theory section, following a brief review of the CPHMD methodology, the formalism of the two-dimensional λ-dynamics approach is outlined and discussed for two special cases where tautomerism manifests in either the protonated or the deprotonated state. In Methods, the derivations of model potential functions as well as the macroscopic pK_a values for the titration simulations involving histidine and carboxyl side chains are given. Force-field issues related to the construction of the protonation state models are discussed. The last subsection in Methods details the simulation protocol. In Results and Discussion, the convergence behavior and errors in the titration simulations of blocked aspartic acid and histidine are discussed. The pK_a values of OMTKY3 and RNase A obtained from the CPHMD simulations using the new model are analyzed and compared to those from the previous work of Lee et al. (16), from the static PB calculations and experimental data. The issues related to the force field and implicit solvation, as well as the dynamic coupling between protonation and the local conformation. The last section of the article draws conclusions and makes suggestions for future directions of research.

THEORY

Continuous constant pH molecular dynamics

To facilitate derivations and discussions of the new model, we briefly review the methodology of the continuous implicit solvent constant pH molecular dynamics (CPHMD) (16). In the CPHMD method, the protonation/deprotonation (titration) process at a protein titrating residue i is described by a set of continuous coordinates,

$$\lambda_i = \sin^2(\theta_i), i = 1, n, \quad (1)$$

the end points of which define the deprotonated ($\lambda_i = 1$) and protonated states ($\lambda_i = 0$). Thus, the CPHMD simulation is a special case of λ-dynamics (17), where the λ-variables define the titration coordinates.

In classical simulations, the deprotonation free energy of a protein side chain is obtained through modeling the difference between the free energy of deprotonating a side chain in the protein environment and that in solvent. Since the latter energy, namely, the deprotonation energy of a blocked amino

acid (model compound) in solvent, is experimentally accessible, the deprotonation energy in the protein environment can be obtained as

$$\Delta G(protein) = \Delta G_{class}(protein) - \Delta G_{class}(mod) + \Delta G_{exp}(mod), \quad (2)$$

where $\Delta G_{exp}(mod)$ is given as

$$\Delta G_{exp}(mod) = ln\,10 \times k_B T(pK_a - pH), \quad (3)$$

and the pK_a value reflects the pH condition at which the populations of protonated and deprotonated states are equal.

Based on the above considerations, the total potential energy of the protein in the CPHMD simulation can be written as

$$U^{tot}(\lambda_i; i = 1, n) = U^{int} + U^{npolar} + U^{vdW}(\lambda_i; i = 1, n)$$
$$+ U^{elec}(\lambda_i; i = 1, n) + U^{GB}(\lambda_i; i = 1, n) - U^{mod}(\lambda_i; i = 1, n)$$
$$+ \sum_{i=1}^{n} U^{pH}(\lambda_i) - \sum_{i=1}^{n} U^{barr}(\lambda_i), \quad (4)$$

where $\{\lambda_i; i = 1, n\}$ is a set of titration coordinates for the n titrating residues as defined in Eq. 1. In Eq. 4, the two terms are the protonation-independent internal (bonded) energy and nonpolar solvation energy, and the rest are the protonation-dependent nonbonded and biasing potential energies.

The van der Waals energy for the interaction between the titrating proton i and another atom j is given by

$$\bar{U}^{vdW}(i,j) = \begin{cases} (1 - \lambda_i)U^{vdW}(i,j) & j : \text{other} \\ (1 - \lambda_i)(1 - \lambda_j)U^{vdW}(i,j) & j : \text{titrating} \end{cases}, \quad (5)$$

where $U^{vdW}(i,j)$ represents the protonation independent (full-strength) van der Waals interaction energy. The protonation dependence of the Coulomb and GB electrostatic energies is incorporated through the atomic partial charges on the titrating residue, which are linearly interpolated between the values in the deprotonated and protonated states,

$$q_{i,\alpha} = \lambda_i q_{i,\alpha}^u + (1 - \lambda_i)q_{i,\alpha}^p \quad (6)$$

where $q_{i,\alpha}$ represents the partial charge of atom α on residue i; and $q_{i,\alpha}^u$ and $q_{i,\alpha}^p$ are the charges in the deprotonated and protonated states, respectively.

The first biasing potential $-U^{mod}$ in Eq. 4 originates from the term $-\Delta G_{class}(mod)$ in Eq. 2 and is a potential of mean force (PMF) along the λ-coordinate. Due to the nature of the nonbonded terms (Eq. 5 and Eq. 6), it is a quadratic function of the titration coordinate,

$$U^{mod}(\lambda_i) = A_i(\lambda_i - B_i)^2, \quad (7)$$

and can be obtained via thermodynamic integration. The second biasing potential in Eq. 4 originates from the term

$\Delta G_{exp}(mod)$ in Eq. 2 and models the pH-dependence of the deprotonation free energy. Thus,

$$U^{pH}(\lambda_i) = ln\,10 \times k_B T \times \lambda_i(pK_a(i) - pH),\qquad(8)$$

where $pK_a(i)$ is the pK$_a$ value of the titrating group i.

The last term in Eq. 4 is called a barrier potential,

$$U^{barr}(\lambda_i) = 4\beta_i\left(\lambda_i - \frac{1}{2}\right)^2,\qquad(9)$$

where β_i is the barrier at the center of the titration coordinate. The barrier potential does not alter the energy difference between the end-point states ($\lambda = 0$ or $\lambda = 1$) but serves as an umbrella to prolong sampling time in the region of the end-point coordinates. This is necessary because the electrostatic and van der Waals interactions of the titrating residue associated with a mixed state ($0 < \lambda < 1$) are unphysical.

Two-dimensional λ-dynamics method

Tautomeric state model

As in most pK$_a$ calculations, the existing CPHMD method assumes a particular protonating atom for residues that contain multiple titration sites, such as histidine and carboxyl groups. In what follows, a new model is introduced that removes this bias and treats the titration of the proton tautomers on an equal footing.

In the new model we assign an additional continuous coordinate x, which offers control of interconversion between

The acid-base and proton tautomeric equilibria of histidine and aspartic acid are two special cases of two-dimensional λ-dynamics, where the two λ-driven processes are degenerate at one end point (either $\lambda = 0$ or $\lambda = 1$) and are coupled to each other via the x-coordinate at the other end point (either $\lambda = 1$ or $\lambda = 0$). Despite the formal identity of the two cases, somewhat different expressions have to be derived for the λ- and x-dependent potential energy functions (the second and third lines in Eq. 4). In our remaining discussions we will refer to both cases as the deprotonated (histidine) and protonated (aspartic acid) tautomerism.

Nonbonded interactions

Following the addition of the tautomeric state coordinate x, the atomic charge on a titrating group can be written as a linear combination of four charge states,

$$q_{i,\alpha} = \lambda_i[x_i q_{i,\alpha}(11) + (1 - x_i)q_{i,\alpha}(01)] \\ + (1 - \lambda_i)[x_i q_{i,\alpha}(10) + (1 - x_i)q_{i,\alpha}(00)],\quad(10)$$

where the values for the λ- and x-coordinates are given in parentheses. Since histidine has only one protonated state, $q_{i,\alpha}(01) = q_{i,\alpha}(00)$. Analogously, for aspartic acid, $q_{i,\alpha}(11) = q_{i,\alpha}(10)$. Notice that Eq. 6 can be recovered from Eq. 10 if the charge states of the tautomers are identical.

In the case of the deprotonated tautomerism (Fig. 1 A), the van der Waals energy for the interaction of titrating atom i and another atom j is given by

$$\tilde{U}^{vdW}(i,j) = \begin{cases} (1 - \lambda_i f^m(x_i))U^{vdW}(i,j) & j : \text{other} \\ (1 - \lambda_i f^m(x_i))(1 - \lambda_j f^n(x_j))U^{vdW}(i,j) & j : \text{titrating} \end{cases},\qquad(11)$$

the tautomeric states. This coordinate will be treated in the same manner as the titration coordinate λ. Thus, the titration of histidine involves interconversion among three states, which can be defined in terms of the titration and tautomeric coordinates as (0) for the doubly protonated state HSP; (1,1) for the NE2-protonated state HSE; and (1,0) for the ND1-protonated state HSD, where the first number in the parenthesis refers to the λ-value and the second the x value (Fig. 1 A). Notice that HSP can take on any x-value since its energy is degenerate with respect to x. Analogously, the three states in the titration of aspartic acid are defined as (1) for the deprotonated ASP; (0,1) for the OD1-protonated state ASPP1; and (0,0) for the OD2-protonated state ASPP2 (Fig. 1 B). Notice that ASP can take on any value of x.

where the superscripts m and n indicate the nature of the titrating atoms i and j, respectively. For example, in the titration of histidine, a value of 1 is associated with the titrating proton HE2 and 0 is associated with HD1. For a nontautomeric titrating proton, a value of -1 is assigned. Thus, the function $f^m(x_i)$ can be written as

$$f^m(x_i) = \begin{cases} x_i & m = 1 \\ 1 - x_i & m = 0 \\ 1 & m = -1 \end{cases},\qquad(12)$$

where the function $f^n(x_j)$ can be defined analogously. Notice that Eq. 5 can be recovered from Eq. 11 using the last definition of $f^m(x_i)$, or $f^n(x_i)$, in the absence of tautomerism.

In the case of the protonated tautomerism (Fig. 1 B), the van der Waals energy is given by

$$\tilde{U}^{vdW}(i,j) = \begin{cases} (1 - \lambda_i)f^m(x_i)U^{vdW}(i,j) & j : \text{other} \\ (1 - \lambda_i)(1 - \lambda_j)f^m(x_i)f^n(x_j)U^{vdW}(i,j) & j : \text{titrating} \end{cases},\qquad(13)$$
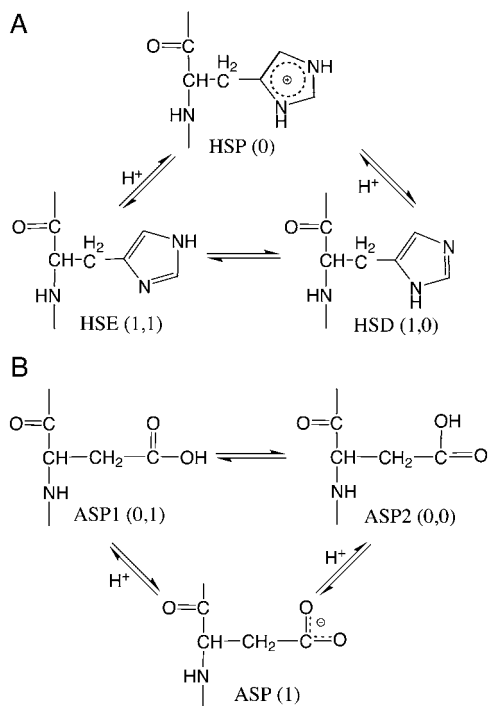
FIGURE 1 Acid-base and proton tautomeric equilibria of histidine (A) and aspartic acid (B).

where the function $f^m(x_i)$ (or $f^n(x_i)$) m (or n) takes on the value of 1 for the titrating proton HD1, 0 for HD2, and -1 for a nontautomeric titrating proton. Eq. 5 can again be recovered from Eq. 13 in the absence of tautomerism.

Biasing potentials

In the presence of proton tautomerism, the pH-dependent potential function becomes

$$U^{pH} = ln\,10 \times k_B T \times \lambda_i [x_i(pK_a^1 - pH) + (1 - x_i)(pK_a^2 - pH)],$$ (14)

where pK_a^1 and pK_a^2 refer to the microscopic pK_a values for the two titration sites. To suppress the population of mixed states in terms of x, a tautomer interconversion barrier potential in analogy to the titration barrier potential (Eq. 9) is applied to each coordinate x_i.

The functional form of the model potential in terms of the titration and tautomeric interconversion coordinates is determined by the functional forms of the Coulomb, GB electrostatic and van der Waals interaction energies. In the absence of tautomerism this leads to a quadratic function as mentioned earlier (Eq. 7). In the presence of proton tautomerism the coupling between λ and x results in a complex functional form for the model potential. However, since linear interpolation is used in the expressions for both the atomic partial charges and van der Waals interaction energy, the model potential function is expected to remain quadratic in either the λ- or x-variable.

Before attempting the derivation of the model potential function in the two-dimensional coordinate space (λ, x), it is instructive to consider the scenarios where either the titration or tautomeric interconversion coordinate is fixed at one of its end points. In the case of the deprotonated tautomerism, each process (or equilibrium) displayed in Fig. 1 A, is associated with a quadratic function. Since the energy of the protonated form, HSP, is independent of x, the potential function at HSP must be a constant. The following equations present the boundary conditions for the model potential function,

$$U^{mod}(\lambda_i, 1) = A_1(\lambda_i - B_1)^2$$
$$U^{mod}(\lambda_i, 0) = A_0(\lambda_i - B_0)^2$$
$$U^{mod}(1, x_i) = A_{10}(x_i - B_{10})^2$$
$$U^{mod}(0, x_i) = 0,$$ (15)

where A_1, B_1, A_0, B_0, A_{10}, and B_{10} are the parameters in the quadratic functions that describe the one-dimensional processes and are related to each other through the relationship

$$\Delta U^{mod}(\lambda_i = 0 \rightarrow 1, 1) - \Delta U^{mod}(\lambda_i = 0 \rightarrow 1, 0)$$
$$= \Delta U^{mod}(1, x_i = 1 \rightarrow 0).$$ (16)

In other words, the difference in the free energies of deprotonation for HSP$\rightarrow$HSE and HSP$\rightarrow$HSD is the same as the free energy of tautomer interconversion HSE$\rightarrow$HSD. We will show later that the model potential parameters can be derived analytically by making use of the boundary conditions given in Eq. 15. It should be noted that the model potential function contains an arbitrary constant since the difference deprotonation energy (see Eq. 2) is the interesting quantity in this work. For the case of the protonated tautomerism, there exists an analogous set of equations that describes the one-dimensional conditions.

Since the model potential function is quadratic in either the λ- or x-variable, a general expression in the form of a bivariate polynomial containing eight parameters can be found:

$$U^{mod}(\lambda_i, x_i) = a_0\lambda_i^2 x_i^2 + a_1\lambda_i^2 x_i + a_2\lambda_i x_i^2 + a_3\lambda_i x_i$$
$$+ a_4\lambda_i^2 + a_5 x_i^2 + a_6\lambda_i + a_7 x_i.$$ (17)

In the case where tautomerism exists for one protonation form (either $U^{mod}(0, x_i) = 0$ or $U^{mod}(1, x_i) = 0$) there are, at most, six nonvanishing parameters in Eq. 17. One way to identify the nonzero terms is to follow the procedure below,

Step 1, at fixed values of x, obtain parameters for the quadratic function of λ: $f(\lambda;x) = A(x)[\lambda - B(x)]^2$;

Step 2, fit the parameters $A(x)$ and $B(x)$ to the second-order polynomial: $b_1 x^2 + b_2 x + b_3$;

Step 3, at fixed values of λ, obtain parameters for the quadratic function of x: $f(x;\lambda) = A(\lambda)[x - B(\lambda)]^2$; and

Step 4, fit the parameters $A(\lambda)$ and $B(\lambda)$ to the second-order polynomial: $c_1\lambda^2 + c_2\lambda + c_3$,

where some of the coefficients in the polynomial obtained in Steps 2 and 3 vanish. By matching the same order terms in $f(\lambda;x)$ and $f(x;\lambda)$ and utilizing the difference in the λ- and x-dependent polynomial forms (Steps 2 and 3), the nonzero terms in the general form of the model potential (Eq. 17) can be determined. A detailed derivation of the parameters in the model potential functions for the deprotonated (histidine) and protonated tautomeric (carboxylic acid) cases is given in the Supplementary Material.

METHODS

Macroscopic pK$_a$ values from the tautomeric state model

In the CPHMD simulation, the fractional population of unprotonated states (unprotonated fraction) for a titrating residue is given as

$$S^{unprot} = \frac{\rho^{unprot}}{\rho^{unprot} + \rho^{prot}} \approx \frac{N^{unprot}}{N^{unprot} + N^{prot}}, \quad (18)$$

where ρ^{unprot} and ρ^{prot} are the probabilities for the unprotonated and protonated states, which can be related to the number of times unprotonated (N^{unprot}) and protonated (N^{prot}) states are observed in the simulation, respectively.

Since bounded continuous coordinates are utilized for the propagation of protonation and tautomeric states, some cutoff values have to be used to define the pure protonation and tautomeric states. In this work we extend the convention employed previously by Lee et al. (16) to include the tautomeric states. Thus, the number of unprotonated and protonated states are defined as

$$N^{unprot} = N(\lambda > 0.9; x < 0.1 \quad \text{or} \quad x > 0.9)$$
$$N^{prot} = N(\lambda > 0.1; x < 0.1 \quad \text{or} \quad x > 0.9). \quad (19)$$

Notice that the mixed tautomeric states ($0.1 < x < 0.9$) are discarded from the above definition, although the value of x is irrelevant for the doubly protonated state of histidine (Fig. 1 A) or the deprotonated carboxylate residue (Fig. 1 B). However, since the protonation states in the simulation are defined through some cutoff values, their energies do depend on x. Therefore, the mixed tautomeric states are not included in the statistics.

To obtain the pK$_a$ value for a titrating residue, the unprotonated fractions at different pH were used to fit to the generalized form of the Henderson-Hasselbach equation,

$$S^{unprot} = \frac{1}{1 + 10^{n(pK_a - pH)}}, \quad (20)$$

where n is the Hill coefficient. The deviation of the value of n from unity reflects the degree of cooperativity (coupling) between groups that interact and ionize over the same pH range. In this work the Hill coefficient is included into the fitting procedure to account for the couplings between ionizable sites. It is noted, however, that the computed pK$_a$ is only marginally affected by the value of n.

Given the equilibrium constants k_1 and k_2 for the microscopic deprotonation processes HSP → HSE, and HSP → HSD, respectively (Fig. 1 A), the equilibrium constant k for the macroscopic deprotonation of histidine is given as

$$k = k_1 + k_2. \quad (21)$$

Thus, by applying the definition of pK$_a$ (pK$_a = -\log_{10} k$, the macroscopic pK$_a$ for the histidine titration can be related to the microscopic counterparts, pK$_1$ and pK$_2$ by

$$pK_a = -\log_{10}(10^{-pK_1} + 10^{-pK_2}). \quad (22)$$

Provided the reference microscopic value of 6.6 (pK$_1$) and 7.0 (pK$_2$) for the ND1- and NE2-titration in the blocked histidine, respectively, Eq. 22 leads to a macroscopic pK$_a$ of 6.45.

Given the equilibrium constants k_1 and k_2 for the microscopic deprotonation processes of aspartic acid (Fig. 1 B), the equilibrium constant k for the macroscopic deprotonation can be written as

$$k = 1 / \left(\frac{1}{k_1} + \frac{1}{k_2} \right), \quad (23)$$

which leads to the relation

$$pK_a = \log_{10}(10^{pK_1} + 10^{pK_2}). \quad (24)$$

Assuming a reference microscopic value 4.0 for both pK$_1$ and pK$_2$, Eq. 24 leads to a macroscopic pK$_a$ of 4.3. Since the experimental macroscopic pK$_a$ for the blocked aspartic acid is 4.0, a postcorrection is made (see later discussions).

Protonation state models

In the CPHMD method, the transition between protonation states of a titrating residue is modeled by linear attenuation of nonbonded energies while keeping the topology, bonded, and van der Waals parameters of the residue intact. Consequently, in a deprotonated state, the titrating hydrogen atom becomes a dummy atom with zero charge and has no van der Waals interaction with other atoms. This strategy is therefore similar to the single-topology scheme in free energy simulations.

The titration of histidine in the tautomeric state model is based on the topology and associated bonded as well as van der Waals parameters of the doubly protonated histidine (residue type HSP in CHARMM). The atomic charges of the three protonation states (Fig. 1 A) are identical to those in CHARMM type HSP, HSE, and HSD, respectively. In the intermediate or mixed states, the nonbonded interactions are computed by linear interpolation via the λ- and x-variables as described in the Theory section.

For the titrations of the carboxyl residue, a new CHARMM residue type is constructed based on carboxylic acid such that a second hydrogen atom is placed on the unprotonated oxygen. Let us consider aspartic acid. The titration of aspartic acid in the tautomeric state model is based on the topology, bonded and van der Waals parameters of the new hybrid residue, which will be referenced as ASDP. The force-field parameters of ASDP differ from those of aspartic acid (CHARMM type ASPP, protonated on OD2) in that the CG, OD1, and OD2 atoms are assigned with the corresponding bonded parameters of the deprotonated aspartate (CHARMM type ASP) whereas the titrating (dummy) hydrogens HD1 and HD2 are assigned with the bonded parameters of the aspartic acid (CHARMM type ASPP). From the hybrid residue ASDP, three protonation states are constructed: the deprotonated state ASP, with both dummy hydrogens uncharged and having no van der Waals interactions; two protonated states (ASP1 and ASP2), with either HD1 or HD2 charged and having van der Waals interactions. Notice that there is no interaction between the two dummy hydrogens due to the unphysical nature of the doubly protonated state, although both hydrogens are allowed to interact with other atoms in a mixed state.

One drawback of the dummy atom representation of the protonation states for the carboxyl residues is that the uncharged dummy hydrogen can lose the ability to gain charge once it is rotated to the *anti* position (16,14). A remedy used previously by Lee et al. (16) was to significantly lower the barrier to the rotation around the carboxylate bond such that the rotation back to the *syn* position is facilitated. A caveat is, however, that this can lead to nonnative hydrogen bonds as a result of stabilization of the unphysical out-of-plane conformations (16). The *syn* conformation is thermodynamically more favorable relative to the *anti* one by ~2.8 kcal/mol, according to a solid-state NMR experiment (22). Quantum calculations and molecular dynamics simulations have shown a free energy difference of ~1.6–1.9

kcal/mol (22,14). On the basis of these considerations, our solution is to kinetically disfavor the *anti* position by raising the C-O bond rotation barrier from 4.1 to 6.0 kcal/mol and placing the dummy carboxyl hydrogens at the *syn* positions at the beginning of the simulation.

Simulation protocol

All of the simulations described in this work were performed using the PHMD module with the all-atom CHARMM22 force field (24) for proteins including the dihedral cross term corrections (CMAP) (25) and the GBSW solvent model (20,19) in the CHARMM molecular dynamics program (26). The original PHMD module (16), in which the CPHMD method was implemented, was extended to allow tautomeric state titrations and to include pH-dependent solvation energy and force with the GBSW solvent model. The SHAKE algorithm was applied to hydrogen bonds and angles, such that a timestep of 2 fs could be used. A 22 Å distance truncation was applied to both the nonbonded and GB energy/force evaluations. All simulations utilized a Nosé-Hoover thermostat to ensure a canonical distribution of atomic velocities at a constant temperature of 298 K. For the GB calculations, a smoothing length of 0.6 Å at the dielectric boundary with 24 radial integration points up to 20 Å and 38 angular integration points were used. The nonpolar solvation energy was computed using the surface tension coefficient of 0.03 kcal mol^{-1} Å^{-2}.

For the propagation of titration and tautomer interconversion coordinates, the velocities of λ and x were coupled to a three-mass (30, 50, and 70 amu) Nosé-Hoover chain (27,28) to ensure a canonical distribution at the target temperature of 298 K. Unless otherwise specified, the default initial velocity seed was used for the simulation. Model compounds used in this work were the blocked amino acids obtained by acetylating the N-terminus and amidating the C-terminus. To obtain the one-dimensional model potential functions, thermodynamic integrations based on 1-ns MD simulations (100 ps equilibration and 1 ns production) were performed at several fixed points along the titration or tautomer interconversion coordinates, as described previously (16). In single-site titration simulations, a titration barrier of 1.25 kcal mol^{-1} was used (16). For double-site titrations, a titration barrier of 1.75 and a tautomer interconversion barrier of 2.75 kcal mol^{-1} for the carboxyl groups and 2.5 kcal mol^{-1} for histidines were applied, which in general keeps the fractional population of mixed titration and tautomeric states below 25% and 15%, respectively.

The microscopic reference pK$_a$ values of 4.0, 4.4, and 3.8 were used for the Asp, Glu, and C-terminus carboxyl residues, respectively. This introduced a positive error of 0.3 units when both the carboxylate oxygens participated in the titration process (see Methods). Thus, we included a correction that accounts for this and other errors that occur in the titration simulations of model compounds in the reported pK$_a$ values for the test proteins, OMTKY3 and RNase A (see the legends of Tables 3 and 5). For histidine, the microscopic reference values of 6.6 and 7.0 were used for the ND1- and NE2-titration, respectively. A correction that accounted for the error in the titration of blocked histidine was included in the reported pK$_a$ values for the histidines in RNase A (see Table 5 legend). The reference pK$_a$ value of 7.5 was used for the N-terminus amino group.

Since the GBSW model was parameterized to closely mimic the solvation energies from the finite-difference Poisson method, the optimized radii for the latter method (29,30) were adopted to define the solvent-solute dielectric boundary in the present work with some small modifications to better match the experimental solvation free energies. Specifically, the solvation radii for the oxygen atom in carboxylate, Tyr, Ser, and Thr residues were adjusted by 0.1, −0.1, −0.14, and −0.14 Å, respectively.

Development of atomic radii for use with the GB model is a continuous effort. Previous radii parameterizations for the PB methods were targeted toward the experimental solvation energies of small molecules as well as the calculated interaction energy of small molecules containing charged side chains with explicit waters. However, this strategy is not ideal, for at least two reasons. First, the solvation treatment of small molecules is not entirely transferable to a protein environment. Second, reproducing the explicit

solvent behavior using the TIP3P (32), or any other explicit water model carries along inherent problems of these models (33,34). Thus, current effort in the group is directed toward reoptimization of the existing PB radii by performing control GB simulations of model peptides and mini-proteins aimed at finding the balance among stability, foldability, and the underlying physical principles (31). Since the strength of the electrostatic interactions in a microscopic approach such as the current CPHMD method is sensitive to changes in solvation or desolvation energies, the accuracy of the constant pH simulations will greatly benefit from the improvement of the GB model.

In the current method, a fixed set of solvation radii was used throughout the titration simulations. This is a drawback, because the radii of ND1 and NE2 on histidine in the GBSW model vary according to the charge state. Although for the neutral state (HSD and HSE), the radius is 1.8 Å, it is 0.5 Å larger in the charged state (HSP). A solution would be to linearly interpolate the radius between these two states, which would, however, result in a deviation from the quadratic behavior in the model potential function. Thus, in the present work, we assumed a radius of 2.0 Å for ND1 and NE2 in the pH range, expanding one unit below and one unit above the reference pK$_a$ value, and the default radii for the neutral or charged states under other pH conditions. Since the pK$_a$ values of carboxyl groups in OMTKY3 and RNase A are a few units below that of the histidine, the error due to the fixed histidine radii approach is minor.

RESULTS AND DISCUSSION

Model compounds

Fig. 2 shows the two-dimensional potential of mean force (PMF) maps for the blocked histidine and aspartic acid along the titration and tautomer interconversion coordinates. Notice that the PMF of the doubly protonated histidine HSP (Fig. 2, A $\lambda = 0$) and the deprotonated aspartate (Fig. 2, B, $\lambda = 1$) is constant. From the PMF map for histidine (Fig. 2, A), it is seen that the ND1-protonated form HSD (*lower-right corner* where $\lambda = 0$, $x = 1$) is higher in energy than the NE2-protonated form HSE (*upper-right corner* where $\lambda = 0$, $x = 0$). The barrier for HSP → HSD is higher and occurs later relative to that for HSP → HSE. From the PMF map for the aspartic acid (Fig. 2, B), it is seen that the OD1-protonated form ASP1 and OD2-protonated form ASP2 have almost identical energy. The barriers for ASP1 → ASP and ASP2 → ASP are very small (<5 kcal/mol, not visible in Fig. 2) and occur very early in the deprotonation process.
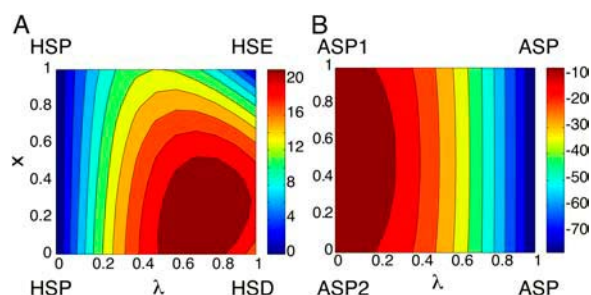

FIGURE 2 Two-dimensional potential of mean force (PMF) map along the titration coordinate λ and tautomeric interconversion coordinate x for the blocked His (*left*) and Asp (*right*) residues. The color bars shown on the right indicate that the PMF increases from dark blue to dark red.

Blocked histidine

Table 1 summarizes the running pK_a values for the blocked histidine using the double- and single-site titration models. The pK_a values were obtained from the data collected from 2-ns time windows along five 10-ns trajectories that were initiated with random initial velocity seeds. Given the reference pK_a of 6.6 for the ND1 and 7.0 for the NE2 titration, the average running pK_a values in the first 2-ns simulations have a deviation of 0.4 and −0.2 pK units for the ND1- and NE2-titration, respectively. The associated standard deviation due to different initial velocity seeds is 0.45 and 0.27 pK units for the ND1- and NE2-titration, respectively. It is interesting to see that the error and standard deviation in the single-site titrations remain approximately constant over the entire 10 ns, suggesting the convergence of the protonation state sampling. The greater variation from the standard values and standard deviation in the ND1-titration simulations can be attributed to an inherent problem in the implicit solvent simulation of the charged histidine as discussed further below.

Given the macroscopic pK_a value of 6.45 (see Methods), the error from the double-site titrations is ~0.35 pK units in the 0–8-ns simulations and decreases to 0.15 in the last 2-ns simulations. The standard deviation from the double-site titrations is closer to that from the ND1-titrations in the first 4-ns simulations and decreases almost steadily to 0.15 pK units in the last 2-ns simulations. It is also interesting to observe that the average pK_a from the last 2-ns double-site titrations (6.6) coincides with that obtained by combining the pK_a values from the two tautomer titrations (7.0 and 6.8) using Eq. 22 from Methods, above. This suggests that with sufficient sampling of tautomeric states, the errors in the double-site titrations mainly originate from those in the single-site titrations, as will be seen in the titration simulations of blocked aspartic acid.

As mentioned in Methods, the difference between the solvation radii for ND1 and NE2 in the neutral and charged states of histidine is 0.5 Å. Since our approach uses a fixed set of radii for the λ-dependent GB energy and force, the protonation state whose assigned solvation radii are closer to the "true" values (as given in the PB radii set; see Refs. 28,29) may be artificially favored. For example, in the extreme case, if the true radius for the charged state HSP is used in the single-site titration simulations, the titrating nitrogen can become undersolvated resulting in a strong electrostatic attraction between the titrating proton and the partially negatively charged backbone carbonyl oxygen, which prevents the deprotonation to occur. Curiously, this has only occurred in titration simulations at the ND1 site. To better understand this bias, we performed a standard GBSW simulation of the blocked doubly protonated histidine. The top plot of Fig. 3 shows that up until 1 ns of MD simulation, the distance from the backbone oxygen to ND1 (solid) is ~3.4 Å, which is >1 Å smaller than the distance to NE2 (shaded). At slightly after 1 ns, a conformational switch occurs, leading to the merge of the ND1-O and NE2-O distances. The middle plot of Fig. 3 shows a switch of the χ_1 dihedral angle from 50° to 59° at ~1 ns. The bottom plot of Fig. 3 shows that the χ_2 dihedral angle remains approximately constant during the entire 2-ns simulation. Thus, the difference in the electrostatic environment between the ND1 and NE2 provides an explanation for the failure of titration at ND1. As the nitrogen radius for solvation is decreased to 2.0 Å, both the ND1-O and NE2-O distances increase but the former remains ~1 Å smaller even after a conformational switch occurs at 1 ns (data not shown).

In this work, we fix the solvation radii at 2.0 Å for both the neutral and charged states of histidine in the titration simulation. Also, we restrict the pH range for the titration to no more than 3 pK units, around the pH at which the

TABLE 1 Computed running pK_a values for the blocked histidine from the double- and single-site titration simulations

	0–2 ns	2–4 ns	4–6 ns	6–8 ns	8–10 ns
1	6.8	6.4	7.1	6.8	6.6
2	6.4	7.0	6.6	6.7	6.4
3	6.8	6.5	6.8	6.7	6.8
4	6.9	6.7	6.8	6.6	6.7
5	7.3	7.3	7.3	7.1	6.6
Mean (std.)*	6.8 (0.32)	6.8 (0.37)	6.9 (0.28)	6.8 (0.19)	6.6 (0.15)
ND1[†]	7.0 (0.45)	7.0 (0.39)	7.0 (0.40)	7.0 (0.48)	7.0 (0.53)
NE2[†]	6.9 (0.27)	6.8 (0.18)	6.8 (0.19)	6.8 (0.21)	6.8 (0.21)

The pK_a values were obtained from 2-ns time windows with a total simulation time of 10 ns at pH values 6, 7, and 8. An additional 100 ps was allotted for equilibration. A random initial velocity seed was used for each simulation run.

*Average and standard deviation of pK_a values obtained from five double-site titration simulations.

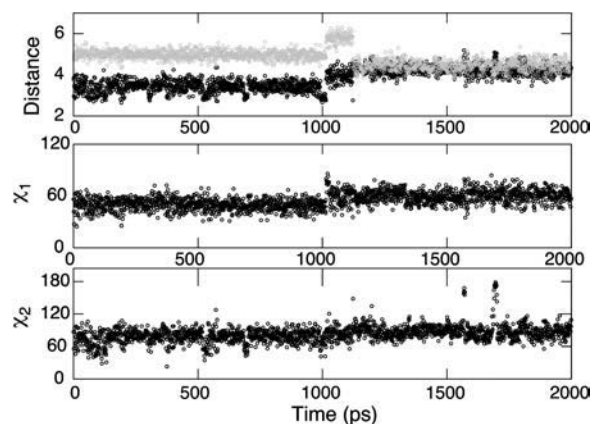[†]Average and standard deviations of pK_a values obtained from five single-site titration simulations.

FIGURE 3 2-ns standard GBSW simulation of the blocked doubly protonated histidine. The top plot shows the time series of the distance from the backbone carbonyl oxygen to ND1 (solid) and that to NE2 (shaded). The middle and bottom plots show the time series of the χ_1 and χ_2 angles, respectively.

protonation state is expected to change. A rigorous way to deal with the charge-state dependence of solvation radii is to introduce some mixing scheme, for example, a linear interpolation of radii with respect to λ. This is not implemented in the current version of the CPHMD module, because it would significantly complicate the derivation of an analytic form for the model potential function. As the CPHMD method becomes more elaborated, one can envision the inclusion of bonded parameters and solvation radii in the mixing scheme, in which case the derivation of the model potential function has to rely upon some numerical fitting scheme.

Blocked aspartic acid

Table 2 summarizes the running pK_a values for blocked aspartic acid in the double-site and single-site titration simulations. Given the target pK_a of 4.3 (see Methods), the double-site titrations lead to an underestimation of 0.5–0.6 pK units, with a standard deviation of 0.33 pK units in the first 2-ns simulation time. The magnitude of error and standard deviation decreases as the simulation continues. The last 2-ns gives an average error of -0.2 pK units with a standard deviation of 0.25.

Table 2 also shows results from two types of single-site simulations. First, simulations were conducted by keeping the topology and bonded parameters (of the ionic aspartate, see Methods, above) intact, while allowing only one carboxylate oxygen (OD1 or OD2) to titrate. For titrations at OD2 an average error of -0.3 pK units is observed for the first 2-ns simulations, the magnitude of which is reduced to

0.2 in the last 2-ns simulations. For titrations at OD1 there is an average error of -0.1 pK units in the first 2-ns simulations. The standard deviations in the single-site simulations are $\sim$0.1 pK units in the first 2-ns simulations and are smaller as the simulation progresses. A possible reason for the small underestimation of the pK_a value in the single-site titrations is a slight overstabilization of the ionic aspartate relative to the neutral aspartic acid due to insufficient relaxation of the peptide conformation in the latter state. In the titration simulation, the residency time of the end-point states is on the order of picoseconds (16), which is much smaller relative to the length of the thermodynamic integration simulations at discrete λ-values (typically 1 ns) that were used to determine the model potential function. Consequently, the employment of the bonded parameters of the ionic aspartate that have two C-O bonds with a formal bond order of 1.5 for both the ionic and neutral states in the case of insufficient sampling, can slightly favor the ionic state leading to a negative error in the predicted pK_a value.

If the above explanation were true, it would predict a positive error for single-site titrations using the bonded parameters of the neutral aspartic acid that has one single and one double C-O bond, as in the work of Lee et al. (16). This is indeed the case. The last two rows of Table 2 show a persistent overestimation by 0.2 pK units in the single-site titrations at OD1 or OD2 using the force-field parameters that are designed for aspartic acid. The same single-site model when used with the GBMV solvation method also gave a overestimation of 0.3 pK units in the predicted pK_a (16). Notice that the single-site titrations using the force-field parameters for the neutral species exhibit standard deviations twice those relative to using the parameters for the ionic species. Based on the above considerations and given the limitation of the current CPHMD methodology that allows only single topology, titration simulations of carboxyl group with the bonded parameters designed for the ionic species seems to be a better choice. It should be noted that the problem related to the force-field parameters does not exist in the titration simulations of histidine, since the latter has identical bonded parameters for both charge states in the CHARMM22 force field.

OMTKY3

Table 3 summarizes the computed pK_a values for turkey ovomucoid third domain (OMTKY3) from the double-site and single-site simulations and compares them with the results from experiment and other simulations. The double-site titration simulations with the crystal structure (column *Cryst.*) yield an average absolute error of 1.0 pK unit based on the first 2-ns simulations and 1.2 pK units based on the second 2-ns simulations. The double-site simulations with the NMR structure (column *NMR*) yield pK_a shifts (relative to the model compound reference values) of the same sign and the average absolute errors of similar magnitude. Thus

TABLE 2 Computed running pK_a values for the blocked aspartic acid from the double- and single-site titration simulations (see legend below)

	0–2 ns	2–4 ns	4–6 ns	6–8 ns	8–10 ns
1	3.9	3.5	4.2	4.3	4.3
2	3.4	3.5	4.1	4.0	4.2
3	3.7	3.7	3.7	3.9	3.7
4	4.3	4.5	4.7	4.2	4.3
5	3.8	3.1	3.4	3.8	4.0
Mean (std.)*	3.8 (0.33)	3.7 (0.52)	4.0 (0.50)	4.0 (0.21)	4.1 (0.25)
OD1[†]	3.9 (0.10)	4.0 (0.089)	4.0 (0.041)	4.0 (0.041)	4.0 (0.041)
OD2[†]	3.7 (0.12)	3.9 (0.055)	3.7 (0.045)	3.9 (0.083)	3.8 (0.083)
OD1[‡]	4.2 (0.31)	4.2 (0.28)	4.2 (0.25)	4.2 (0.23)	4.2 (0.13)
OD2[‡]	4.2 (0.34)	4.2 (0.28)	4.2 (0.31)	4.2 (0.23)	4.3 (0.23)

The pK_a values were obtained from 2-ns time windows with a total simulation length of 10 ns at pH values 3, 4, and 5. An additional 100 ps is allotted for equilibration. A random initial velocity seed was used for each run.

*Average and standard deviation of pK_a values from five double-site titration simulations.

[†]Average and standard deviation of pK_a values from five single-site titration simulations using the bonded parameters for the ionic aspartate (see text).

[‡]Average and standard deviation of pK_a values from five single-site titration simulations using the bonded parameters for the neutral aspartic acid (see text).

TABLE 3 Computed pK$_a$ values for OMTKY3 (see legend below)

Residue	Cryst.	NMR	OD1 (OD2)	GBMV	Static	Expt.
Asp[7]	2.8 (2.4)	3.2 (3.2)	4.0 (1.9)	4.5	2.9 (2.9)	<2.7
Glu[10]	3.1 (2.3)	2.6 (2.6)	2.5 (2.7)	5.1	3.4 (3.4)	4.2
Glu[19]	2.3 (2.1)	2.6 (2.5)	<0.0 (0.7)	2.7	3.2 (2.6)	3.2
Asp[27]	4.7 (4.7)	4.0 (3.9)	5.0 (4.6)	6.1	4.0 (3.6)	<2.3
Glu[43]	5.3 (5.5)	4.3 (4.6)	4.1 (4.5)	6.0	4.3 (4.4)	4.8
CT-Cys	1.3 (1.8)	1.0 (1.3)	0.9 (1.0)	4.6	2.7 (2.5)	<2.5
Avg. abs. error*	1.0 (1.2)	1.1 (1.0)	1.9 (1.5)	1.7	0.6 (0.6)	—

The pK$_a$ values were obtained from 4-ns simulations of OMTKY3 at pH values 0, 1, 2, 3, 4, 5, and 6. An additional 200 ps was allotted for equilibration. Unless otherwise noted, the pK$_a$ values obtained from 2-4 ns simulation time are listed in parentheses next to those from 0 to 2 ns. A value of 0.3 was added to all the pK$_a$ values from the double-site simulations to correct the errors found in the model compound simulations (see Table 2). For the same reason, the computed pK$_a$ values from the single-site titrations were corrected by adding 0.1 and 0.3, respectively (see Table 2). The double-site titrations were based on the crystal structure PDB:1PPF (44) (column *Cryst.*) or the first entry of the NMR structure ensemble PDB:1OMU (45) (column *NMR*). Column *OD1 (OD2)* refers to the 2-ns single-site titration at either OD1 or OD2, with the bonded parameters of the ionic aspartate. Column *GBMV* refers to the 1-ns single-site (OD2) simulations with the GBMV solvation model (16). Column *Static* refers to the static PB calculations of Forsyth et al. (46) and Antosiewicz et al. (in parentheses) (38) using the same crystal structure. Protein dielectric constant was set to 20 in these calculations. Column *Expt.* refers to the experimental data taken from Schaller and Robertson (47).
*Average absolute deviations from experimental data. The upper limit is used when a range is given.

we do not observe a strong initial structure dependence to the overall pK$_a$ comparisons. The single-site OD1 and OD2 titrations with the crystal structure (column *OD1 (OD2)*) result in pK$_a$ shifts of the same sign as from the double-site simulations, but much larger average absolute errors (1.9 and 1.5, respectively). As expected, the magnitude of these errors is similar to that from the single-site simulations with the GBMV solvent model (column *GBMV*). Compared to experimental data, the current simulations predict correct signs of pK$_a$ shifts for all but one residue (Asp[27], see later discussions), whereas the GBMV simulations predict only one correct sign of the pK$_a$ shift (Glu[19]). Finally, compared to the CPHMD simulations, the two sets of PB calculations employing a protein dielectric constant of 20 (column *Static*) based on the static crystal structure yield smaller average absolute errors (0.6 pK units).

To characterize the local electrostatic environment of titrating residues in OMTKY3, Table 4 summarizes the distances among titrating residues from the double-site simulations at three pH values and compares them with the distances measured in the crystal structure. These three pH values represent the conditions in which the specified residue is fully protonated (acidic), partially protonated, or fully deprotonated (basic). Below we will discuss the correlation between the pK$_a$ value and the dominant electrostatic interactions involving the titrating residue as revealed by simulations and the crystal structure data. In a related study,

Li et al. (35) showed that pK$_a$ values in OMTK3 could be reproduced using QM methods with minimal models including key interacting residues. It should be noted that since the crystal structure was obtained under basic (pH = 10) conditions, where all carboxyl groups were deprotonated, the measured distances between oppositely charged residues represent lower bounds to those measured at lower pH.

Asp[7] is a solvent-exposed residue that electrostatically interacts mainly with Glu[10] and Lys[34]. The positively charged Lys[34] provides stabilization for the ionic form of Asp[7]. The crystal structure (Table 4) shows that Asp[7] is closer to Lys[34] (5 Å) than to Glu[10], which explains the downfield pK$_a$ shift of the experimental pK$_a$ of Asp[7] relative to the model compound value. Under acidic conditions (pH = 0, Asp[7] is mainly neutral), Lys[34] is further away from Asp[7] with the average distance of 6.3 Å and 7.1 Å in the first (0–2 ns) and second (2–4 ns) half of the simulation, respectively. With increasing pH, the Asp[7]-Lys[34] distance becomes smaller. Under neutral conditions (pH = 2), the Asp[7]-Lys[34] distance is 6.6 Å and 5.3 Å in the first and second half of the simulation, respectively. Under more basic conditions (pH = 4, Asp[7] is mainly ionized), a salt bridge is formed between Asp[7] and Lys[34]: the average distance is 4.0 and 3.8 Å in the first and second half of the simulation, respectively. The lower pK$_a$ value (pK$_a$ = 2.4) of Asp[7] in the second half of the simulation relative to that in the first half (pK$_a$ = 2.8) is mainly due to the additional electrostatic attractions from Arg[21], which moves on average ~5 Å closer to Asp[7] in the second half of the simulation. In contrast to the Asp[7]-Lys[34] distance, the Asp7-Arg[21] distance becomes larger at pH 4.

The pK$_a$ (2.8) for Asp[7] obtained from the first 2-ns simulations using the double-site model is between the values computed using the OD1 (pK$_a$ = 3.9) and OD2 (pK$_a$ = 1.9) single-site model, and agrees well with the experimental value of 2.7 (Table 3). This is consistent with the reasonable agreement of the simulated major electrostatic interactions (Asp[7] with Glu[10] and Lys[34]) under basic conditions with those from experiment (crystal structure).

The electrostatic environment of Glu10 is more complex, and includes the titratable groups Asp[7], Lys[13], Arg[21], and Lys[34] (Table 4). In the crystal structure, Lys[13] is the closest residue with the Lys[13]-Glu[10] distance of 4.8 Å, followed by Lys[34], Asp[7], and Tyr[11], that are >6 Å away. Arg[21] is ~18 Å from Glu[10] in the crystal structure. In contrast, MD simulations under neutral to basic conditions show that Arg[21] moves to be <7 Å away from Glu[10] (Table 4).

The top plot of Fig. 4 shows how the motion of Glu10 is correlated with that of Asp[7], Lys[34], and Arg[21] at pH 3, where Glu[10] is partially deprotonated. At the beginning of the simulation, Lys[34] is in salt-bridge contact with Glu[10] (*dashed curve*) and is ~6.5 Å away from Asp[7]. Due to the attractive electrostatic force and the diminished mobility of Lys[34], Asp[7] is driven closer to form a salt bridge with Lys[34] at ~320 ps (*shaded dashed curve*), which also results in a sharp

TABLE 4 Distances between titrating residues and other charged residues in OMTKY3 (see legend below)

							Cryst.
Asp^7	pH = 0		pH = 2		pH = 4		
Glu^{10}	11.0 (1.9)	6.4 (1.0)	7.2 (1.8)	6.6 (1.1)	7.6 (1.6)	6.8 (0.8)	6.7
Arg^{21}	10.7 (1.3)	12.1 (0.9)	11.9 (2.1)	6.8 (3.0)	8.8 (1.5)	9.7 (1.3)	17.4
Lys^{34}	6.3 (0.7)	7.1 (1.1)	6.6 (1.4)	5.3 (1.7)	4.0 (0.9)	3.8 (0.6)	5.0
Glu^{10}	pH = 1		pH = 3		pH = 5		
Asp^7	7.3 (1.8)	6.8 (1.6)	8.3 (1.4)	6.7 (1.1)	8.5 (2.5)	6.5 (0.6)	6.7
Lys^{13}	4.6 (0.9)	4.7 (0.9)	7.4 (2.2)	7.2 (2.6)	7.8 (2.4)	9.6 (1.2)	4.8
Arg^{21}	12.5 (2.2)	11.1 (2.6)	7.3 (2.0)	5.1 (1.5)	6.8 (1.6)	5.8 (0.9)	18.2
Lys^{34}	6.8 (1.5)	6.5 (1.8)	6.0 (2.7)	4.6 (1.6)	4.4 (1.9)	3.4 (0.3)	6.3
Glu^{19}	pH = 1		pH = 2		pH = 4		
Lys^{13}	6.7 (1.0)	6.4 (1.4)	8.2 (1.9)	7.2 (2.0)	12.2 (1.6)	13.0 (1.4)	10.3
Arg^{21}	9.7 (1.8)	6.9 (2.3)	9.2 (2.0)	7.9 (1.1)	4.5 (1.1)	4.2 (0.5)	8.6
Lys^{34}	7.4 (1.7)	7.6 (1.7)	6.8 (1.8)	4.4 (1.3)	8.3 (1.6)	9.6 (1.2)	10.7
Asp^{27}	pH = 2		pH = 4		pH = 6		
Lys^{29}	6.6 (0.7)	6.5 (1.1)	4.6 (1.1)	6.7 (1.1)	6.8 (1.7)	5.7 (1.7)	6.1
Tyr^{31}	5.0 (0.7)	5.2 (0.9)	4.9 (0.8)	4.7 (0.6)	5.3 (1.0)	5.4 (1.1)	3.5
CT-Cys	pH = 0		pH = 2		pH = 3		
His^{52}	4.8 (0.6)	4.6 (0.6)	4.6 (0.6)	4.6 (0.6)	4.3 (0.6)	4.3 (0.6)	4.9
Lys^{55}	9.0 (0.5)	9.1 (0.2)	8.9 (0.8)	8.5 (1.5)	8.8 (1.3)	9.2 (0.4)	7.9

Average and RMS fluctuations of distances (in Å) between the titrating and other charged residues in the first (0–2 ns, *left column*) and second half (2–4 ns, *right column*) of the simulations of OMTKY3 starting from the crystal structure PDB:1PPF (44), under three pH conditions. The pH values represent the conditions in which the specified residue is fully protonated (*left*), partially protonated (*middle*), and fully deprotonated (*right*). Only those distances smaller than 10 Å are shown. For Asp^{27}, Tyr^{31} is listed since it forms a hydrogen bond with Asp^{27} in the crystal structure (see text). The residue-residue distance is defined as that to the carboxyl carbons, NE2 of histidine, NZ of lysines, and CZ of arginines. The column *Cryst.* refers to the distances measured in the crystal structure at pH 10.

drop in the Asp^7-Glu^{10} distance from 9 Å down to below 6 Å (*solid curve*). However, this local environment is not stable, due to the electrostatic repulsion between Asp^7 and Glu^{10}. At the same time, the electrostatic attraction between Arg^{21} and Glu^{10} facilitates the departure of Glu^{10} from Lys^{34} and the formation of a loose salt-bridge contact between Glu^{10} and Arg^{21} at 450 ps (*dotted curve*), which then breaks off at

FIGURE 4 Coupling between Asp^7 and Glu^{10} in a 2-ns simulation of OMTKY3 at pH 3. The top plot shows the time series of the running average distances between Asp^7 and Glu^{10} (*solid*), Asp^7 and Lys^{34} (*shaded dashed*), Glu^{10} and Lys^{34} (*dashed*), and Glu^{10} and Arg^{21} (*dotted*) computed from 100-ps time windows. The definitions of the residue-residue distances are given under Table 4. The bottom plot shows the running unprotonated fraction for Asp^7 (*solid*) and Glu^{10} (*dashed*) computed from 100-ps time windows.

520 ps due to the motion of Glu^{10}. The dynamics in the following period of time (until the end of the first half of the simulation), is dominated by the motion of Glu^{10}, driven by the attraction from Lys^{34} and Arg^{21}. This continues at the beginning of the second half of the simulation and starting at $\sim$1.1 ns, Glu^{10} arrives at a position that allows it to form salt-bridge contacts with both Lys^{34} and Arg^{21}, although not as tight as the Lys^{34}-Asp^7 salt bridge. The dynamics of the last part of the simulation (between 1.6 and 2 ns) is again characterized by the breaking and making of the electrostatic contacts of Glu^{10}-Lys^{34} and Glu^{10}-Arg^{21}.

A comparison of the bottom plot of Fig. 4 with the top one shows a direct correlation between the unprotonated fraction of Asp^7 (*solid curve*), that of Glu^{10} (*dashed curve*), and the existence of the salt bridge. Accordingly, Asp^7 is protonated at the beginning of the simulation and becomes unprotonated once it forms and maintains the salt bridge with Lys^{34}. On the other hand, Glu^{10} is unprotonated at the beginning and becomes protonated upon its departure from both Lys^{34} and Arg^{21}. In the second half of the simulation it looses a proton again due to salt-bridge formation with Lys^{34} and Arg^{21}. Finally, the simultaneous loss of salt-bridge contacts between Glu^{10} and Lys^{34} and between Glu^{10} and Arg^{21} is reflected in the plot as a dip in the curve of the unprotonated fraction. Fig. 4 also reveals that the salt-bridge formation in the second half of the simulation is the cause for the lowering of Glu^{10}'s pK_a relative to the first half of the simulation, consistent with the observation for Asp^7.
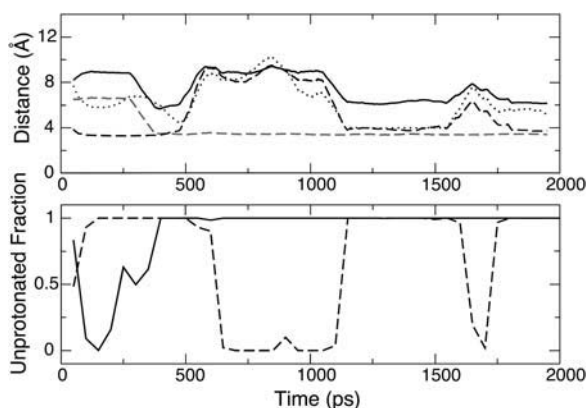
In contrast to the simulation, which reveals two *on* and *off* salt bridges (Glu[10]···Lys[34] and Glu[10]···Arg[21]), the crystal structure shows only one salt bridge (Glu[10]···Lys[13]). This is the reason for the negative difference (-1.1 pK units) in the computed pK_a for Glu[10] relative to experiment. In the last 2-ns simulations, the simultaneous formation of the two salt bridges gives rise to a larger difference (-1.9 pK units) relative to the first 2-ns simulations (-1.1 pK units). The salt-bridge effects seem to be more pronounced in single-site titrations. As a result, the differences are larger: -1.8 for the OD1 and -1.5 pK units for the OD2 titrations. Interestingly, the PB calculation also gives an underestimation of 0.8 pK units.

In the crystal structure under basic conditions, Glu[19] interacts with three positively charged groups (Lys[13], Arg[21], and Lys[34]) at a distance over 8 Å (Table 4). However, in the simulation at pH >2, Glu[19] forms a persistent salt bridge with Arg[21], thereby preventing the protonation of Glu[19]. This is the reason for the negative difference of Glu[19]'s pK_a (-0.9 pK units) using the double-site simulations (Table 3) relative to experiment. The magnitude of the underestimation using the OD1 and OD2 titrations is ~2 pK units larger.

The prediction of the pK_a for Asp[27] is most problematic, since it is partially sequestered from solvent and yet it is the most acidic residue in OMTKY3 according to experiment (Table 4). In the simulation at pH 4, Asp[27] is a hydrogen-bond acceptor to its backbone amide nitrogen and to that of Lys[29] with an occupancy of 20%. In addition, Asp[27] interacts with the hydroxyl oxygen of Tyr[31] as both a hydrogen-bond acceptor and donor with an occupancy of 30%. In the first 2-ns, Asp[27] is >40% deprotonated due to the salt-bridge interaction with Lys[29] (Table 4). Although the mobility of Asp[27] is low, the fluctuation of the side-chain position of Lys[29] is large as seen from the >1 Å root-mean square (RMS) deviation of the Asp[27]-Lys[29] distance. In the last 2-ns, Asp[27] loses the salt-bridge contact with Lys[29] and dominantly occupies the protonated state (Table 4). The convergence of the protonation state sampling for Asp[27] is incomplete, which can be seen from a large percentage of mixed states (~36% in the first and 42% in the last 2-ns of simulation time).

The double-site titration model overestimates the pK_a for Asp[27] by 2.4 pK units, the largest error among all the residues in OMTKY3 (Table 3). Interestingly, Asp[27] shows the largest computational error in the static PB calculations as well, although to a lesser degree (overestimated by 1.7 and 1.3 pK units). As compared to the single-site model, the double-site model does not offer any advantage for the titration of Asp[27]. One obvious reason for the large deviation from experiment is the aforementioned issue of incomplete sampling of protonation states. Conformational relaxation of buried residues in response to ionization requires much longer simulation time as demonstrated in a recent work by Simonson et al. (36). Another possible reason for the difference may be the overestimation of the desolvation effect due

to undersolvation for the protonated aspartic acid relative to the charged aspartate, since the current GB model employs one atomic solvation radius for the both the neutral and charged forms.

Glu[43] does not have strong electrostatic interactions with other charged groups and has therefore a small experimental pK_a upshift of 0.4 units (Table 3). Although the OD1 and OD2 titrations give a downshift of 0.4 and a upshift of 0.1 pK units, respectively, the double-site model predicts an upshift of 0.9 units, consistent with the experimental values. The deviations from experiment are within statistical errors, as shown in the titration simulations for model compounds. Interestingly, the static PB calculation is unable to capture the sign of the experimental pK_a shift of Glu[43].

The local electrostatic environment around the α-carboxyl group CT-Cys includes two positively charged residues His[52] and Lys[55] (Table 4). The pK_a downshift of CT-Cys is dominated by the interaction with His[52], since the simulation data reveals the formation of a salt bridge between CT-Cys and His[52] that does not exist in the crystal structure (Table 4). This explains why the double-site simulations result in a pK_a downshift that is 1.2-pK-units greater relative to experiment and the single-site titrations yield even larger errors.

RNase A

Bovine pancreatic ribonuclease A (RNase A) is an enzyme that catalyzes the transphosphorylation and hydrolysis reactions of RNAs (37). Table 5 summarizes the computed pK_a values from the double-site simulations with the GBSW solvation model (current model), in comparison with the split-model simulations (imidazole ring is split into two titration halves involving ND1 and NE2) with the GBMV solvation model (previous work) (16), and the PB static calculations (38), as well as experimental data. The current model gives an average absolute error of 0.6 pK units for the histidine groups, representing a significant improvement over the previous work using the split-model (2.5 pK units). In fact, the error from the current model is even smaller than that from the static PB calculations (0.8 pK units). It is noteworthy that results from the split-model simulations with the GBSW solvation model yield even larger differences (data not shown). The current model also gives smaller average absolute error (1.6 pK units) for the carboxyl groups relative to the previous work (2.0 pK units). For the N-terminus group, the current model yields a difference of 0.6 pK units, as compared to 1.0 pK unit from the previous work. Again, the current model predicts the correct sign for all pK_a shifts in RNase A, in contrast to the previous work that predicts the correct sign for only seven pK_a values (with a total of 17).

The protonation equilibria of active-site residues His[12] and His[119] are responsible for the pH-dependent catalytic mechanism of RNase A. A traditional view is that His[12] and His[119] act as general acid and base in RNA hydrolysis, respectively,

TABLE 5 Calculated pK_a values for RNase A (see legend below)

Residue	This work	Lee	Static	Expt.
NT-Lys	7.2	6.6	7.0	7.6
Glu[2]	0.0	<−1.0	2.5	2.8
Glu[9]	2.6	5.8	4.1	4.0
His[12]	5.1	2.8	4.3	6.2
Asp[14]	3.1	3.5	2.0	2.0
Asp[38]	1.5	2.4	2.8	3.1
His[48]	6.3	7.7	6.5	6.0
Glu[49]	5.0	6.4	4.6	4.7
Asp[53]	3.8	4.5	3.6	3.9
Asp[83]	−0.9	7.4	2.1	3.5
Glu[86]	3.9	5.9	3.8	4.1
His[105]	7.3	10.8	6.2	6.7
Glu[111]	3.7	5.8	3.9	3.5
His[119]	5.7	7.5	6.5	6.1
Asp[121]	0.4	3.3	1.3	3.1
CT-Val	0.0	<−1.0	2.3	2.4
Avg. abs. error*	1.6	2.0	0.5	
Avg. abs. error†	0.6	2.7	0.8	
Abs. error‡	0.4	1.0	0.6	

The pK_a values were obtained from 2-ns simulations of RNase A at a pH range between 0 and 8 with an increment of 1 using the crystal structure PDB:7RSA (39).To correct the errors found in the model compound simulations a value of 0.4 was subtracted from the computed pK_a values for His[105], and His[119] whereas 0.1 was added to the computed pK_a values for His[12] and His[48] since they titrate via NE2 exclusively (see Table 1). For the same reason, the computed pK_a values were corrected by adding 0.3 units for all the carboxyl residues except for Asp[53], in which case 0.1 units was added to the computed value since it titrates via OD1 exclusively (see Table 2). Column *Lee* refers to the work of Lee et al. (16). Column *Static* refers to the static PB calculations of Antosiewicz et al. (38) using a protein dielectric constant of 20. Cited are the values averaged over the calculations which allow the protonation of histidine to occur at either NE2 or ND1 and with His[119] being in the crystal conformation A. Column *Expt.* refers to the experimental data cited in the work of Antosiewicz et al. (38).
*Average absolute error of the computed pK_a values for the carboxyl residues.
†Average absolute error of the computed pK_a values for the histidine residues.
‡Absolute error of the computed pK_a for NT-Lys.

TABLE 6 Distances between histidine residues and other nearby residues in RNase A (see legend below)

				Cryst.
His[12]	pH = 4	pH = 5	pH = 7	
r(NE2···ND1[119])	4.3 (1.0)	4.8 (1.2)	3.9 (0.9)	6.3
His[48]	pH = 4	pH = 6	pH = 8	
r(ND1···CG[14])	3.8 (0.3)	4.1 (0.3)	7.1 (0.8)	4.3
r(ND1···O[14])	3.3 (0.5)	3.7 (1.2)	4.1 (1.1)	3.3
His[105]	pH = 5	pH = 7	pH = 8	
r(ND1···O[75])	4.8 (0.7)	4.1 (0.9)	4.1 (1.0)	3.0
r(NE2···O[76])	3.6 (0.6)	3.8 (0.6)	4.0 (0.6)	4.6
His[119]	pH = 4	pH = 5	pH = 8	
r(NE2···OD1[121])	4.0 (1.1)	5.3 (2.0)	7.1 (1.0)	1.9

Average and RMS fluctuations of the distances (in Å) between the histidine and other strongly interacting residues in the 2-ns simulations of RNase A starting from the crystal structure PDB:7RSA (38) under three pH conditions. The pH conditions are arranged such that the left and right ones give the fully protonated and deprotonated states, respectively, whereas the middle one gives partial protonation/deprotonation. The atoms used in the distance measurements are specified in parentheses, where the first atom refers to the underlined residue and the second one is indicated by the superscript. The column *Cryst.* refers to the distances measured in the crystal structure at pH 5.3.

is correlated with the strength of the electrostatic interaction between them. As His[12] and His[119] are >7 Å away from each other between 400 and 600 ps, the interaction is small, allowing both residues to be in the charged state at pH 5. As His[119] moves closer to His[12] between 750 and 1000 ps, His[119] starts to lose its proton more often. As the distance continues to shorten, the coupling between His[12] and His[119]) is so strong that only one residue (His[119]) is allowed to be charged to minimize the electrostatic repulsion. Thus, the strong coupling raises the pK_a of His[119] and lowers that of His[12]. The upper plot of Fig. 5 reveals that the fluctuation in the His[12]-His[119] distance is linked to the change in the χ_2

whereas their roles are reversed in the transphosphorylation step (37). The crystal structure of the phosphate-free RNase A (PDB:7RSA, see Ref. 39; pH = 5.3), shows that His[12] is completely buried whereas His[119] is partially buried in the protein interior. Thus, desolvation effects are expected to result in their pK_a downshifts. Distance measurements in the crystal structure reveal an electrostatic interaction between His[12] and His[119] and a hydrogen bond between the NE2 of His[119] and OD1 of Asp[121] (Table 6).

The lower plot of Fig. 5 shows that His[12] is predominantly charged in the first half of the simulation (0–1 ns), and switches to be in the neutral state in the second half of the simulation (1–2 ns) under pH 5 conditions. On the other hand, the degree of deprotonation of His[119] fluctuates at ~0.5 in the first half of the simulation but becomes much smaller in the second half of the simulation. Fig. 5 also reveals that the degree of deprotonation on His[12] and His[119]
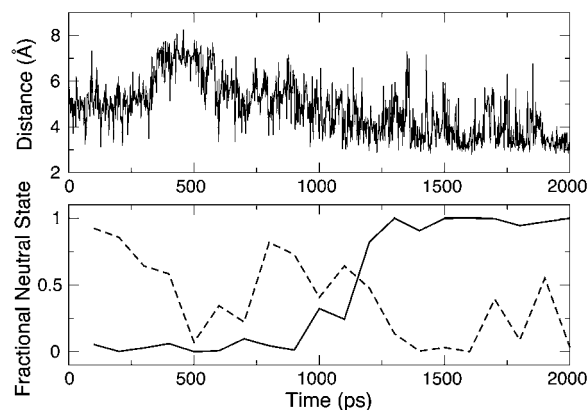

FIGURE 5   Coupling between His[12] and His[119] in a 2-ns simulation of RNase A at pH 5. The top plot shows the time series of the distance between NE2 of His[12] and ND1 of His[119]. The bottom plots shows the running fraction of neutral states for His[12] (*solid*) and His[119] (*dashed*) computed from 100-ps time windows.

dihedral angle of His[119]. In contrast, His[12] has a much lower mobility since it is buried and subject to steric hindrance from neighboring groups.

Another factor that affects the ionization equilibrium of His[119] is its interaction with Asp[121], which is a hydrogen-bond acceptor according to the crystal structure. In the simulation the hydrogen bond NE2-HE2[119]···OD1[121] only exists under pH conditions below 5. At pH 5, His[119] becomes partially deprotonated and the average distance between NE2[119] and OD1[121] is 5.3 Å (Table 6). The loss of the hydrogen-bonding interaction with Asp[121] contributes to the stronger interaction between His[119] and His[12], relative to that observed in the static crystal structure. This could arise from the fixed atomic solvation radii used in the current simulation. At pH 5, His[119] in the charged state experiences less attraction from Asp[121] due to the enhanced solvent screening as a result of the smaller solvation radius. On the other hand, the enhanced solvent screening decreases the unfavorable interaction between His[119] and His[12]. It should be noted, however, that the degree of the enhanced solvent screening is small, since both the NE2 and ND1 atoms are partially sequestered from solvent.

The interaction between His[119] and His[12] is one reason for our underestimation of the pK_a for His[12]. It may also be linked to the exaggerated desolvation effect, which can be again partially attributed to the fixed atomic solvation radii approximation. In this case, the neutral state is preferentially stabilized, as a result of a larger desolvation penalty for the charged state. It is interesting to see that the static PB calculation also gives a large underestimation for the pK_a of His[12].

The current simulations reveal that whereas the deprotonation process of His[119] occurs via both ND1 and NE2 atoms, the deprotonation of His[12] occurs predominantly (>90%) at NE2 (data not shown). This is in agreement with a direct observation via the high-resolution x-ray diffraction of RNase A (sulfate bound) under different pH conditions (39) and is in support of the proposed mechanism, which suggests that the NE2 atom of His[12] exchanges a proton with the substrate (40). It is worthwhile emphasizing the importance of predicting the correct tautomeric state for the histidine titration. The PB calculation and the split-model simulation gave a significant underestimation of His[12]'s pK_a by 2.0 and 3.4 units, respectively.

His[119] has two conformations due to the variation in the χ_1 dihedral, which are referred to as *trans* and *gauche*. In the above experimental study (40) both conformations were observed at the acid pH (pH = 5.2), whereas only the *trans* conformation was observed at the basic pH (pH = 8.8). The current simulations, however, show that the *trans* conformation exists exclusively in the pH range between 5 and 8. This is in line with the argument put forward by the authors that the coexistence of both conformations is made possible through hydrogen bonding with the sulfate ion whereas the *trans* conformation is favored in the ion-free environment due to its strong interaction with Asp[121] (40).

His[48] is completely embedded in the protein interior and interacts with a nearby negatively charged embedded residue Asp[14]. Thus, the protonation equilibrium of His[48] is a result of the balance between the desolvation effect that favors the neutral state and the electrostatic interaction that favors the charged state. In fact, the experimental pK_a reveals only a small downshift. The simulation at pH 6 (partial deprotonation) shows that the charged state of His[48] is stabilized by a salt-bridge interaction with Asp[14], as well as the hydrogen bond between HD1 of His[48] and the backbone carbonyl oxygen of Asp[14] (Table 6). Thus, it appears that the ND1 atom is protected from deprotonation through its hydrogen bonding with Asp[14]. This is consistent with the observation made in the aforementioned experiment that the deprotonation of His[48] occurs via its NE2 atom (40). As His[48] becomes deprotonated, the current simulations show that the χ_2 dihedral angle changes from an average of $-61°$ (pH = 6) to an average of $-53°$ (pH = 8), in line with the experimental observation (40).

It is curious that His[105] exhibits an error of 0.6 units for its computed pK_a, although it is fully immersed in solvent and free of strong electrostatic interactions with other charged residues. A close examination of the MD trajectory at pH 7 (partial deprotonation) reveals occasional hydrogen-bonding contacts between its HD1 atom and the backbone carbonyl oxygen of Ser[75] with an occupancy of 24%, as well as between the HE2 atom and the backbone carbonyl oxygen of Tyr[76] with an occupancy of 13%, which do not exist in the crystal structure (Table 6). Consequently, the doubly protonated form of His[105] is stabilized resulting in an over-estimation of its pK_a unit relative to experiment (Table 5).

From Table 5 it can be seen that the average absolute error of the computed pK_a values for RNase A is dominated by four groups, Glu[2], Asp[83], Asp[121], and CT-Val. The pK_a downshifts for these groups are calculated to be too large by 2.8, 4.4, 2.7, and 2.4 units relative to the corresponding experimental values, respectively. Table 7 summarizes the computed distances between these and other nearby charged residues in comparison to the measured values. It should be noted that since the crystal structure was obtained at pH 5.3, where all carboxyl groups are deprotonated, the measured distances between the oppositely charged residues represent lower bounds to the distances measured at lower pH.

The protonation equilibria of Glu[2] and Asp[38] are coupled. Both the crystal structure and simulation at pH 0 show that the ionic form of Glu2 is stabilized through two charged hydrogen bonds to Arg[10]: OE1 with NH2 and OE2 with NE (Table 7). The simulation also reveals a slightly closer distance of Glu[2]-Lys[7] relative to experiment. As pH is increased to 2, the simulation shows that whereas the double hydrogen bonds between Glu[2] and Arg[10] are maintained, Asp[38] moves closer to Arg[10] and forms a charged hydrogen bond to Arg[10]'s NH2 via its carboxylate oxygens alternately (Fig. 6 A). In contrast to the strong interaction between Asp[38] and Arg[10] as revealed by simulation results, the crystal structure shows

TABLE 7 Distance between the carboxyl residues and other nearby charged residues in RNase A (see legend below)

		Cryst.
Glu^2	pH = 0	
r(CD···NZ^7)	6.8 (1.3)	7.3
r(OE1···NH2^{10})	3.5 (1.5)	2.8
r(OE2···NE^{10})	3.9 (1.6)	2.8
Asp^{38}	pH = 2	
r(CG···CZ^{10})	6.2 (2.8)	10.3
r(CG···NZ^{37})	7.9 (1.9)	4.1
r(CG···CZ^{39})	6.8 (1.3)	8.2
Asp^{83}*	pH = 0	
r(OD1···NE^{85})	2.9 (0.5)	6.6
r(OD2···NH2^{85})	2.9 (0.2)	3.1
Asp^{121}	pH = 0	
r(OD1···N^{66})	3.8 (1.0)	3.7
r(OD2···N^{66})	3.8 (1.0)	2.8
r(OD1···NE2^{119})	4.5 (1.7)	2.9
CT-Val	pH = 0	
r(OT2···NZ^{104})	4.3 (0.9)	4.7
r(OT1···N^{105})	2.9 (0.3)	5.0
r(N···O^{105})	2.8 (0.1)	2.9

Average and RMS fluctuations (in Å) between the carboxyl residues that exhibit large errors in the calculated pK$_a$ values (>1.5 pK units) and other nearby charged residues in the 2-ns simulations of RNase A starting from the crystal structure PDB:7RSA (39) under a pH condition that gives partial protonation/deprotonation. The atoms used in the distance measurements are specified in parentheses, where the first atom refers to the underlined residue and the second one is indicated by the superscript. Column *Cryst.* refers to the distances measured in the crystal structure at pH 5.3.
*Taken from the crystallographic conformation where Arg^{85} is pointing toward Asp^{83} (see text).

a salt bridge between Asp38 and Lys^{37} (Table 7). Thus, it appears that the simulation overestimates the strength of the electrostatic interaction between a negatively charged residue and arginine.

The prediction of the pK$_a$ for Asp^{83} seems to be most problematic due to its interaction with Arg^{85}. The crystal structure shows that Arg^{85} occupies two conformations with its amino groups either pointing toward or swinging away from the carboxyl group of Asp^{83}. In the former case, a salt bridge is formed between OD2 of Asp^{83} and NH2 of Arg^{85}. However, independent of the starting conformation, the simulation leads to two charged hydrogen bonds between the carboxylate oxygens of Asp^{83} and the hydrogens that are attached to the NE and NH2 atoms of Arg^{85} (Fig. 6 *B*) within just a few hundreds of picoseconds.

The charged hydrogen bonding triad plays a major role in the protonation equilibria of Asp^{121} and CT-Val. The simulation at pH 0 shows that Asp^{121} is a hydrogen-bond acceptor to the backbone amide of Lys^{66} via OD1 and to the NE2 atom of His^{119} via OD1 (Fig. 6 *C*). In contrast, the crystal structure shows only salt-bridge interactions between Asp^{121} and Lys^{66} and between Asp^{121} and His^{119} (Table 7), although it has been suggested in other experimental work

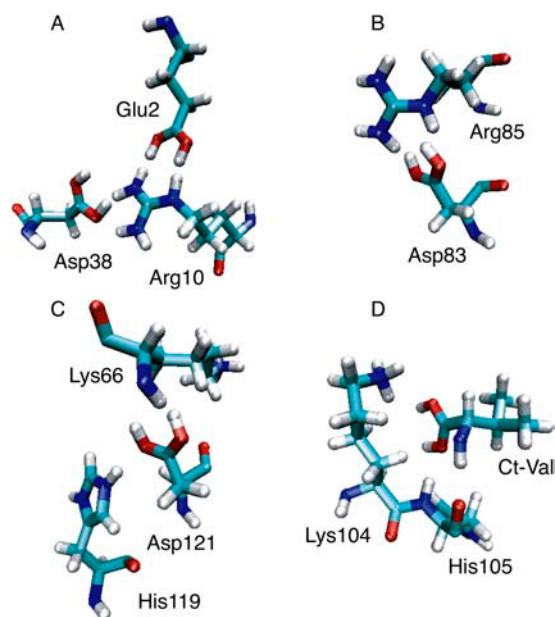
FIGURE 6 Snapshots from the 2-ns simulation of RNase A showing the salt-bridge and hydrogen-bond interactions involving Glu^2, Asp^{83}, Asp^{121}, and CT-Val under pH conditions of 2, 0, 0, and 0, respectively.

that the hydrogen bond between Asp^{121} and the backbone of Lys^{66} was used to restrict the conformational entropy of the protein (41). Similarly, the simulation at pH 0 shows that CT-Val is a hydrogen-bond acceptor to the backbone amide of His^{105} via OT1 that occasionally accepts a hydrogen bond from the NZ atom of Lys^{104} (Fig. 6 *D*). Interestingly, the backbone amide of CT-Val is hydrogen bonded with the backbone carbonyl of His^{105} as well (Fig. 6 *D*).

CONCLUDING REMARKS

In this work, we have presented a two-dimensional λ-dynamics approach to include proton tautomerism in continuous constant pH molecular dynamics (CPHMD) simulations. Consequently, a new tautomeric state model, also called the double-site titration model, is constructed for titration simulations involving histidine and carboxyl residues.

The double-site titration model combined with the GBSW solvation method was tested on blocked histidine and aspartic acid. Given sufficient sampling of tautomeric and protonation states, our calculations give good quantitative estimates of site-specific pK$_a$ values. Sources of deviation in the titration simulations of blocked histidine and aspartic acid are the fixed-radii approximation, the use of the same bonded parameters for both the neutral and charged states, and sampling convergence of conformational variables. These issues represent areas of continuing improvement of such models.

The results of the CPHMD simulations using the double-site titration model and the GBSW solvation method are very encouraging and provide a significant improvement over the previous work. In contrast to the single-site model with the

GBMV solvation, the new model predicts the right sign for all the pK_a shifts but one in the benchmark proteins. The new model offers a striking accuracy in the prediction of the protonation and tautomeric states of the histidine residues in RNase A.

However, our detailed analyses of the pH-dependent dynamics of the local electrostatic environment in OMTKY3 and RNase A point to potential areas for improvement of the underlying solvation model to ameliorate the undersolvation of salt bridges. This factor may account for the systematic underestimation of the pK_a values for carboxyl residues that interact with positively charged residues in solvent. The overstabilization of salt bridges in the context of continuum electrostatics is a known problem in the community of protein simulations. One remedy, as used in the work of Georgescu et al. (42), is to introduce some empirical function to dampen the ion-pair interactions. A straightforward but perhaps less pragmatic approach is to reoptimize the atomic solvation radii aiming for a balance among protein stability, foldability, and the underlying physical principles.

Conformational and protonation state sampling also influence the quantitative accuracy of calculated pK_a values, especially for coupled titrating residues such as His^{12} and His^{119} in RNase A and buried residues such as Asp^{27} in OMTKY3. The CPHMD approach will clearly benefit from the introduction of more efficient sampling techniques, such as replica-exchange MD (43). Our findings, nevertheless, clearly indicate that environmentally dependent factors, and their fluctuating character, mediate the protonation/deprotonation processes for such systems coupled to a pH bath.

One of our main goals in the development of the CPHMD technique is to model important pH-triggered conformational phenomena of proteins, such as the pH-dependent folding/unfolding, membrane insertion, ligand binding, and catalysis. Using OMTKY3 and RNase A, the current work has demonstrated in detail the capability of the CPHMD technique in revealing the pH-coupled conformational dynamics of protein side chains. Future work will include the development of more elaborate mixing schemes, which improve the representation of the λ-dependence in computing solvation energy, covalent interactions, and the addition of more complex sampling schemes.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

REFERENCES

1. Matthew, J. B., F. R. Gurd, B. Garcia-Moreno, M. A. Flanagan, K. L. March, and S. J. Shire. 1985. pH-dependent processes in proteins. *CRC Crit. Rev. Biochem.* 18:91–197.

2. Kelly, J. W. 1997. Amyloid fibril formation and protein misassembly: a structural quest for insights into amyloid and prion diseases. *Structure.* 5:595–600.

3. Clippingdale, A. B., J. D. Wade, and C. J. Barrow. 2001. The amyloid-β peptide and its role in Alzheimer's disease. *J. Pept. Sci.* 7:227–249.

4. O'Keefe, D., V. Cabiaux, S. Choe, D. Eisenberg, and R. J. Collier. 1992. pH-dependent insertion of proteins into membranes: B-chain mutation of diphtheria toxin that inhibits membrane translocation, Glu-349→Lys. *Proc. Natl. Acad. Sci. USA.* 89:6202–6206.

5. Bullough, P. A., F. M. Hughson, J. J. Skehel, and D. C. Wiley. 1994. Structure of Influenza haemagglutinin at the pH of membrane fusion. *Nature.* 371:37–43.

6. Rastogi, V. K., and M. E. Girvin. 1999. Structural changes linked to proton translocation by subunit c of the ATP synthase. *Nature.* 402: 263–268.

7. Howell, E. E., J. E. Villafranca, M. S. Warren, S. J. Oatley, and J. Kraut. 1986. Functional role of aspartic acid-27 in dihydrofolate reductase revealed by mutagenesis. *Science.* 231:1123–1128.

8. Mertz, J. E., and B. M. Pettitt. 1994. Molecular dynamics at a constant pH. *Int. J. Supercomput. Appl. High Perform. Comput.* 8:47–53.

9. Sham, Y. Y., Z. T. Chu, and A. Warshel. 1997. Consistent calculations of pK_as of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem. B.* 101:4458–4472.

10. Baptista, A. M., P. J. Martel, and S. B. Petersen. 1997. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins.* 27:523–544.

11. Baptista, A. M., V. H. Teixeira, and C. M. Soares. 2002. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* 117: 4184–4200.

12. Bürgi, R., P. A. Kollman, and W. F. van Gunsteren. 2002. Simulating proteins at constant pH: an approach combining molecular dynamics and Monte Carlo simulation. *Proteins.* 47:469–480.

13. Dlugosz, M., and J. M. Antosiewicz. 2004. Constant-pH molecular dynamics simulations: a test case of succinic acid. *Chem. Phys.* 302: 161–170.

14. Mongan, J., D. A. Case, and J. A. McCammon. 2004. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* 25:2038–2048.

15. Börjesson, U., and P. H. Hünenberger. 2001. Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines. *J. Chem. Phys.* 114:9706–9719.

16. Lee, M. S., F. R. Salsbury Jr., and C. L. Brooks III. 2004. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins.* 56:738–752.

17. Kong, X., and C. L. Brooks III. 1996. λ-dynamics: a new approach to free energy calculations. *J. Chem. Phys.* 105:2414–2423.

18. Lee, M. S., F. R. Salsbury Jr., and C. L. Brooks III. 2002. Novel generalized Born methods. *J. Chem. Phys.* 116:10606–10614.

19. Im, W., M. S. Lee, and C. L. Brooks III. 2003. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* 24:1691–1702.

20. Im, W., M. Feig, and C. L. Brooks III. 2003. An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* 85:2900–2918.

21. Feig, M., A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III. 2004. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* 25:265–284.

22. Gu, Z., C. F. Ridenour, C. E. Bronnimann, T. Iwashita, and A. McDermott. 1996. Hydrogen bonding and distance studies of amino acids and peptides using solid state 2D ^1H-^{13}C heteronuclear correlation spectra. *J. Am. Chem. Soc.* 118:822–829.

23. Chen, J. L., L. Noodleman, D. A. Case, and D. Bashford. 1994. Incorporating solvation effects into density functional electronic structure calculations. *J. Phys. Chem.* 98:11059–11068.

24. MacKerell Jr., A. D., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

25. Feig, M., A. D. MacKerell Jr., and C. L. Brooks III. 2003. Force field influence on the observation of π-helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B.* 107:2831–2836.

26. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.

27. Nosé, S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81:511–519.

28. Hoover, W. G. 1985. Canonical dynamics: equilibration phase-space distributions. *Phys. Rev. A.* 31:1695–1697.

29. Nina, M., D. Beglov, and B. Roux. 1997. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B.* 101:5239–5248.

30. Nina, M., W. Im, and B. Roux. 1999. Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophys. Chem.* 78:89–96.

31. Im, W., J. Chen, and C. L. Brooks III. 2005. Peptide and protein folding and conformational equilibria: theoretical treatment of electrostatics and hydrogen bonding with implicit solvent models. *In* Advances in Protein Chemistry. R. Baldwin and D. Baker, editors. Elsevier, San Diego, CA.

32. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

33. Young, S. W., and C. L. Brooks III. 1997. A reexamination of the hydrophobic effect: exploring the role of the solvent model in computing the methane-methane potential of mean force. *J. Chem. Phys.* 106:9265–9269.

34. van der Spoel, D., P. J. van Maaren, and H. J. C. Berendsen. 1998. A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* 108:10220–10230.

35. Li, H., D. A. Robertson, and H. J. Jensen. 2004. The determinants of carboxyl pK_a values in turkey ovomucoid third domain. *Proteins.* 55:689–704.

36. Simonson, T., J. Carlsson, and D. A. Case. 2004. Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* 126:4167–4180.

37. Raines, R. T. 1998. Ribonuclease A. *Chem. Rev.* 98:1045–1065.

38. Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pK_a values in proteins. *Biochemistry.* 35:7819–7833.

39. Wlodawer, A., L. A. Svensson, L. Sjölin, and G. L. Gilliland. 1988. Structure of phosphate-free ribonuclease A refined at 1.26 Ångstrom. *Biochemistry.* 27:2705–2717.

40. Berisio, R., V. S. Lamzin, and F. A. Sica, K. S. W. Zagari, and L. Mazzarella. 1999. Protein titration in the crystal state. *J. Mol. Biol.* 292:845–854.

41. Quirk, D. J., C. Park, J. E. Thompson, and R. T. Raines. 1998. His···Asp catalytic dyad of ribonuclease A: conformational stability of the wild-type, D121N, D121A, and H119A enzymes. *Biochemistry.* 37:17958–17964.

42. Georgescu, R. E., E. G. Alexov, and M. R. Gunner. 2002. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophys. J.* 83:1731–1748.

43. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.

44. Bode, W., A. Z. Wei, R. Huber, E. Meyer, J. Travis, and S. Neumann. 1986. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J.* 5:2453–2458.

45. Hoogstraten, C. G., S. Choe, W. M. Westler, and J. L. Markley. 1995. Comparison of the accuracy of protein solution structures derived from conventional and network-edited NOESY data. *Protein Sci.* 4:2289–2299.

46. Forsyth, W. R., M. K. Gilson, J. Antosiewicz, O. R. Jaren, and A. D. Robertson. 1998. Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. *Biochemistry.* 37:8643–8652.

47. Schaller, W., and A. D. Robertson. 1995. pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. *Biochemistry.* 34:4714–4723.