

# Constant pH Molecular Dynamics in Explicit Solvent with $\lambda$ -Dynamics

Serena Donnini, Florian Tegeler,<sup>†</sup> Gerrit Groenhof,\* and Helmut Grubmüller\*

Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

 Supporting Information

**ABSTRACT:** pH is an important parameter in condensed-phase systems, because it determines the protonation state of titratable groups and thus influences the structure, dynamics, and function of molecules in solution. In most force field simulation protocols, however, the protonation state of a system (rather than its pH) is kept fixed and cannot adapt to changes of the local environment. Here, we present a method, implemented within the MD package GROMACS, for constant pH molecular dynamics simulations in explicit solvent that is based on the  $\lambda$ -dynamics approach. In the latter, the dynamics of the titration coordinate  $\lambda$ , which interpolates between the protonated and deprotonated states, is driven by generalized forces between the protonated and deprotonated states. The hydration free energy, as a function of pH, is included to facilitate constant pH simulations. The protonation states of titratable groups are allowed to change dynamically during a simulation, thus reproducing average protonation probabilities at a certain pH. The accuracy of the method is tested against titration curves of single amino acids and a dipeptide in explicit solvent.

## 1. INTRODUCTION

Together with temperature, pressure, and ionic strength, pH is one of the key parameters that determine the structure and dynamics of proteins in solution. Most notably, many proteins denature at low pH values,<sup>1</sup> and aggregation, such as formation of amyloid fibrils in Alzheimer's disease<sup>2</sup> and insulin aggregation,<sup>3</sup> is pH-dependent. Because the function of a protein depends on its structure, pH is critical for protein function. Examples of pH-dependent regulation of protein function are the pH-controlled gating of membrane channels,<sup>4–6</sup> or activation of influenza virus in host cells.<sup>7</sup>

pH affects protein structure, because the protonation state of the ionizable groups of a protein depends on pH, in particular histidine amino acids for which the proton affinity ( $pK_a$ ) is very close to the physiological pH. Mainly via its charge, the protonation state of each ionizable group influences, in turn, the physicochemical properties of proteins, their structure, and their function.

Despite its relevance to biomolecular structure and function, pH and changes of protonation state of titratable groups of a protein are usually not described in computer simulations. Typically, a structure with fixed protonation states is used, chosen according to the most probable protonation arrangement at a given pH. This choice is often not straightforward, because hydrogens are usually not resolved in X-ray crystallography and the acid dissociation constant ( $K_a$ ) values of the ionizable groups, in most cases, are not known. Therefore, the protonation state must be inferred from NMR<sup>8</sup> or spectroscopic data,<sup>9</sup> or from electrostatic calculations (e.g., Poisson–Boltzmann (PB)<sup>10,11</sup> or Generalized Born<sup>12</sup> approaches). Furthermore, changes in the protonation state, either due to a change in the environment pH or in the protein conformation, as well as equilibrium protonation fluctuations leading to fractional protonation probabilities, are not taken into account by conventional simulations. As a consequence, the understanding of many biological phenomena, which involve a redistribution of charge, such as ligand binding reactions

inducing a proton redistribution,<sup>13,14</sup> peptide insertion in membranes (e.g., fusion peptides),<sup>15,16</sup> and pH-dependent conformational changes,<sup>2,6</sup> would greatly benefit from a dynamic description of the protonation states.

Several attempts have been made to overcome these limitations. The most-accurate way of modeling (de)protonation events is to describe the system at a quantum mechanics level, where the electronic structure responds to changes in the local environment. However, these calculations are very expensive, in terms of computational cost. This drawback has been partly overcome in mixed quantum mechanics/molecular mechanics approaches,<sup>17</sup> where only the ionizable groups of the protein are treated at the quantum level.

Computationally more affordable approaches to describe proton transfer events are EVB<sup>18–21</sup> and QHOP<sup>22</sup> methods. Here, the potential energy surface on which protons move is parametrized by *ab initio* calculations, whereas the rest of the system is described by a molecular mechanics force field.

A complication common to these approaches is that the equilibrium state is generally reached at time scales that are much slower than those accessible to molecular dynamics (MD) simulations. This is particularly true for protein systems, where typical deprotonation times of ionizable groups in the interior of a protein are microseconds or slower.<sup>23</sup> As a consequence, enhanced sampling of the transitions between the protonated and deprotonated state is particularly relevant for simulations of protein systems. For the aforementioned approaches, however, there is no obvious way how to enhance proton transfer rates.

A further problem concerns the proper description of the pH of a solution. The average hydronium concentration in a typical simulation box can be described by a time average, as well as via an ensemble average. In the case of a time average, because of the fact that the concentrations of hydronium considered are low,

Received: January 26, 2011

Published: April 25, 2011

typically pH 7, it might require very long simulation times to sample the hydronium distribution in the solution. In the case of the ensemble average, however, unpractically large simulation boxes would have to be considered, as, for example, for a typical simulation box of  $\sim 30\,000$  water molecules, one hydronium ion already corresponds to a pH of  $\sim 2-3$ , thus increasing the computational cost of the calculation.

To address these issues in the context of force field simulations, several approaches have been proposed, all of which use a titration coordinate  $\lambda$ , which describes the protonation state of a certain ionizable group. For example, values of  $\lambda = 0$  and  $\lambda = 1$  correspond to the protonated and deprotonated states of a titratable group, respectively, as will be used in this work. Two main categories of approaches can be distinguished depending on the nature—discrete or continuous—of this titration coordinate.<sup>24</sup>

A discrete titration coordinate is typically used by methods combining MD and Monte Carlo (MC) simulations for the sampling of the protonation reaction coordinate. At intervals during the MD simulation, a MC step is performed, in which the protonation state of a residue is changed. The acceptance criterion to keep the new protonation state is based on the protonation free energy of the titratable group, which is computed at every MC step. The major differences between the approaches in this category concern the way that this free energy is computed. In the approaches of Baptista and co-workers,<sup>25</sup> Dlugosz and Antosiewicz<sup>26</sup> and Mongan and Case,<sup>12</sup> the contribution of each protonation state to the protonation partition function is evaluated, and the protonation free energy (and  $\text{p}K_{\text{a}}$ ) is then obtained from the partition function. Because all possible protonation states of the system have to be considered, the computational effort formally scales exponentially ( $2^N$ ) with the number of titratable sites in the system ( $N$ ). In practice, however, MC sampling and cutoffs are applied to reduce computational effort. To estimate the free energy of each state, implicit solvent Poisson–Boltzmann (PB)<sup>25,26</sup> or Generalized Born<sup>12</sup> approaches are used. The use of continuum approximations in the estimation of protonation free energies has the advantage of reducing degrees of freedom of the system. However, to describe more-complex systems, such as membrane proteins, or systems such as channels for which explicit water molecules are crucial, continuum solvent models are of limited use.

In contrast, Bürgi et al. suggested to evaluate the protonation free energy at the MC step by a short thermodynamic integration (TI) simulation.<sup>27</sup> However, the cost of the free-energy calculation step can become significant, because it has to be evaluated at each trial. Also, inclusion of interactions between titratable sites is difficult.

In contrast to MD/MC simulations, in the second category of approaches, the titration coordinate  $\lambda$  is allowed to change continuously between the protonated and deprotonated states. Börjesson and Hünenberger<sup>28,29</sup> developed the “acidostat” method, where the extent of deprotonation is relaxed to equilibrium by weak coupling to a proton bath in a way similar to methods for constant temperature and pressure.<sup>30</sup> Equilibrium fluctuations of the protonation states are not described, and each site thus experiences the average effect of the others.

In a different approach, introduced by Merz and Pettitt,<sup>31</sup> the continuous  $\lambda$  coordinate is treated as an additional particle of the system, which is propagated in time, according to the equations of motion. The potential of the system is coupled to the chemical potential, which is a function of pH, of the reactants and of the

products. Along the same lines, the successive  $\lambda$ -dynamics approach<sup>32</sup> and  $\lambda$ -adiabatic free-energy dynamics<sup>33</sup> treat  $\lambda$  as a dynamical variable in the Hamiltonian. In particular, the  $\lambda$ -dynamics approach was applied to constant pH simulations in implicit solvent by Lee et al.<sup>34</sup> and Khandogin and Brooks.<sup>35,36</sup> In their approach, the potential energy landscape, which drives continuous changes of  $\lambda$ , is modulated by the potentials of isolated model titratable groups, and by the pH. Protons are not transferred explicitly to bulk water, forming  $\text{H}_3\text{O}^+$ ; rather, similar to the acidostat of Börjesson and Hünenberger,<sup>28,29</sup> the proton-solvation contribution to the force acting on  $\lambda$  is implicitly taken into account. Because this contribution depends on pH, by setting the pH parameter in the simulation, the effect of the proton concentration is included. Coupling between titratable sites, described by multiple  $\lambda$  particles, is implicitly taken into account via the potential energy landscape. In principle, linear scaling of the calculation with the number of protonatable sites is achieved. Because of the continuous character of the titration coordinate, fractional  $\lambda$  values can occur, which correspond to a partially protonated state. To decrease the population of these *unphysical* states, a barrier potential is used.<sup>34</sup> This is introduced as a separate parabolic function centered at  $\lambda = 0.5$ .<sup>34</sup> Alternatively, *ad hoc* nonlinear interpolation schemes between the potentials of the end states sampled by  $\lambda$  have been used to decrease the population of intermediate  $\lambda$  values, and thus obtain minima at  $\lambda = 0$  and  $\lambda = 1$ .<sup>33</sup>

As seen, most of the approaches for constant pH simulations both in the first and second category rest on an implicit description of the solvent. We are not aware of any fully atomistic description that (i) achieves sampling of the relevant space of the titration coordinate (i.e., the physically meaningful end states) and (ii) allows one to control the protonation/deprotonation rate.

In this work, we develop and test a framework to describe changes in protonation states at constant pH that meets all of these requirements. Our method extends the  $\lambda$ -dynamics approach of Brooks and co-workers<sup>32,34,35</sup> by introducing a new coupling scheme to describe chemically coupled titratable sites, such as those on the side chain of histidine. Both pH and, via the height of the barrier potential, the protonation rates can be controlled to either reflect experimental proton transfer rates, if available, or to enhance sampling of the protonation space. The method has been implemented within the MD package GROMACS.<sup>37–39</sup>

To test our method, the titration behavior of simple systems in an explicit solvent was analyzed. First, we considered glutamic acid with neutral termini. To provide a simple example of interactions that can occur in a protein environment, a small dipeptide of sequence Glu-Ala was simulated. Because of its importance in protein systems, imidazole and histidine were chosen as a test case for chemically coupled titratable sites. Finally, the effect of different temperature coupling schemes and different barrier potential heights on deprotonation/protonation rates was assessed.

## 2. THEORY

To clarify the notation, we will first summarize the thermodynamic integration and  $\lambda$ -dynamics approaches. Subsequently, we will describe and develop the building blocks of our approach. First, we will describe how the interval sampled by the titration coordinate  $\lambda$  is constrained, to describe the protonated and deprotonated states of the system during the constant pH

simulation. We will then specify how  $\lambda$  is coupled to a temperature bath. After introducing the thermodynamic cycle that is used to couple the protonated and deprotonated states to the appropriate reference states, we will develop the constant pH MD method. Finally, we will generalize the  $\lambda$ -dynamics approach for multiple protonation sites in a protein.

**2.1. Thermodynamic Integration.** Thermodynamic integration (TI)<sup>40</sup> is used to calculate the free-energy difference ( $\Delta G$ ) between a reactant state R and a product state P:

$$\Delta G_{\text{PR}} = \int_{\lambda=0}^{\lambda=1} d\lambda \left\langle \frac{\partial H_{\text{TI}}(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (1)$$

Here,  $H_{\text{TI}}$  is the Hamiltonian of the system, and  $\lambda$  is a coupling parameter that interpolates between the R ( $\lambda = 0$ ) and P ( $\lambda = 1$ ) states, e.g.,

$$H_{\text{TI}}(\lambda) = (1 - \lambda)H_0 + \lambda H_1 \quad (2)$$

To calculate  $\Delta G$  via eq 1,  $\lambda$  is changed from 0 to 1 during the simulation, thus forcing the system from its reactant to its product state. The ensemble average in eq 1 is then taken from the MD ensemble generated from the Hamiltonian  $H_{\text{TI}}(\lambda)$ .

For later use, and following the notation of Kong and Brooks,<sup>32</sup> we split the Hamiltonians of the reactant and product in  $\lambda$ -dependent ( $\tilde{H}_0$  and  $\tilde{H}_1$ ) and  $\lambda$ -independent ( $H_{\text{Env}}$ ) parts:

$$H_{\text{TI}}(\lambda) = (1 - \lambda)\tilde{H}_0 + \lambda\tilde{H}_1 + H_{\text{Env}} \quad (3)$$

**2.2.  $\lambda$ -Dynamics.** In the  $\lambda$ -dynamics approach,<sup>32</sup> a Hamiltonian similar to eq 3 is used. In contrast to TI,  $\lambda$  is defined as an additional dynamic degree of freedom of the system with mass  $m$ , coordinate  $\lambda$ , and velocity  $\dot{\lambda}$ . Accordingly, the Hamiltonian of the system is now expressed by<sup>32</sup>

$$H(\lambda) = (1 - \lambda)\tilde{H}_0 + \lambda\tilde{H}_1 + H_{\text{Env}} + \frac{m}{2}\dot{\lambda}^2 + U^*(\lambda) \quad (4)$$

with a force acting on  $\lambda$ ,

$$F_{\lambda} = -\frac{\partial V(\lambda)}{\partial \lambda} \quad (5)$$

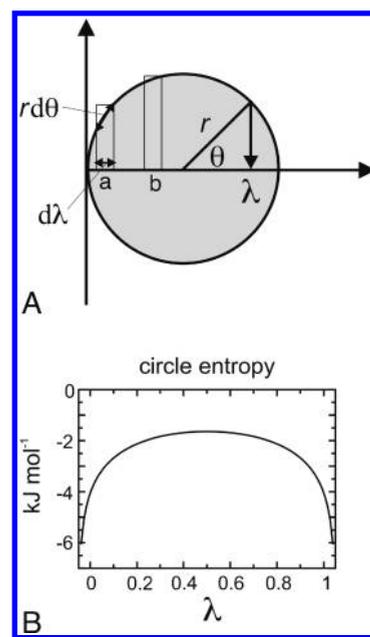
where  $V(\lambda)$  is the potential energy part of the Hamiltonian in eq 4:

$$V(\lambda) = (1 - \lambda)\tilde{V}_0 + \lambda\tilde{V}_1 + V_{\text{Env}} + U^*(\lambda) \quad (6)$$

In eq 4,  $(m/2)\dot{\lambda}^2$  is the kinetic energy term associated with the  $\lambda$  "particle". The  $\lambda$ -dependent potential term  $U^*(\lambda)$  will serve as a biasing potential to limit the range of  $\lambda$ ; this will be defined further below.

**2.3. Constraining the Interval of  $\lambda$ .** Because only  $\lambda = 0$  and  $\lambda = 1$  represent physical states of the system—the protonated and deprotonated states—we require  $\lambda$  to be close to these values for most of the simulation time. More specifically, we require that:

- (1) the  $\lambda$  space is limited to the interval between the two physical states;
- (2) the average values of  $\lambda$  in the protonated and deprotonated states are close to 0 and 1, respectively;
- (3) the time spent at intermediate states by the system is short, i.e., the transitions between the protonated and deprotonated states are fast;
- (4) the residence time at the physical states is sufficiently long to allow conformational sampling of each state; and
- (5) the frequency of transitions can be controlled.



**Figure 1.** (A) Schematic describing the angular coordinate.  $\lambda$  is defined as a function of the angular  $\theta$  coordinate,  $\lambda = r \cos(\theta) + (1/2)$ , with the radius of the circle being defined as  $r = (1/2) + \sigma$ , and  $\sigma$  a fluctuation size (see main text). The segments of circumference corresponding to the intervals  $a$  and  $b$  close to the end and center of the  $\lambda$  interval, respectively, are indicated. (B) Entropic free-energy term introduced by the use of an angular coordinate  $\theta$ .

To address condition 1, a projection of an angular coordinate on the  $\lambda$  space has been proposed in previous applications.<sup>33,34,41</sup> Here, we will extend this approach to meet also condition 2. Following Lee et al.,<sup>34</sup> we will address condition 3 by using a suitably chosen biasing potential. Finally, we will meet conditions 4 and 5 by adjusting the height of the biasing potential, taking into account the entropic part introduced by the use of the angular coordinate.

Note that a similar shape of the  $\lambda$  free-energy profile, which meets conditions 3 and 4, can be achieved also by designing *ad hoc* interpolation schemes between the potentials of the protonated and deprotonated states of  $\lambda$ , as previously proposed in the  $\lambda$ -adiabatic free-energy dynamics approach by Tuckerman and co-workers.<sup>33</sup> By adjusting the temperature of the  $\lambda$  particle, Tuckerman and co-workers,<sup>33</sup> ensured efficient barrier crossing, also meeting the last condition.

**2.3.1. Projection of the Angular  $\theta$  Coordinate on the  $\lambda$  Space.** In order to constrain the space sampled by  $\lambda$ , we switch to a new dynamic angular coordinate  $\theta$ , as shown in Figure 1. By this modification, the actual dynamics takes place in  $\theta$  space, and  $\lambda$  is redefined as the projection of  $\theta$  on the abscissa (see Figure 1),

$$\lambda = r \cos(\theta) + \frac{1}{2} \quad (7)$$

The force acting on  $\theta$  is

$$F_{\theta} = -\frac{\partial V(\lambda(\theta))}{\partial \theta} = r \sin(\theta) \frac{\partial V(\lambda(\theta))}{\partial \lambda} \quad (8)$$

with  $V$  being the potential energy of the system, as defined in eq 6.

In contrast to previous approaches,<sup>33,34,41</sup> where  $r = 1/2$ , and to meet condition 2, we chose  $r = (1/2) + \sigma$ , with an appropriate

fluctuation size  $\sigma$ . Several values of  $\sigma$  were tested. We have used a value of  $\sigma = 0.05$ , because, with this value, the average  $\lambda$  at the physical states was  $\sim 0$  (protonated state) and  $\sim 1$  (deprotonated state).

**2.3.2. Biasing Potential.** To meet condition 3, a parabolic biasing potential of the form<sup>34</sup>

$$U^*(\lambda) = 4h\lambda(1 - \lambda) \quad (9)$$

is used. By adjusting its height  $h$ , the frequency of the protonation transitions can be controlled, as required by condition 5.

Note that the choice of the above angular coordinate implies an entropic contribution to the effective free energy governing the  $\lambda$ -dynamics. This contribution originates from the higher density of  $\lambda$  states at the end points of the  $\lambda$  interval, with respect to the center of the interval, as indicated by the mapping of the intervals  $a$  and  $b$  in Figure 1A onto the circumference. The segment length for a given value of  $\lambda$  is

$$d\lambda = d\theta \sqrt{r^2 - \left(\lambda - \frac{1}{2}\right)^2} \quad (10)$$

resulting in a free-energy contribution of

$$\begin{aligned} A(\lambda) = -TS(\lambda) &= RT \ln \left| \frac{d\lambda}{d\theta} \right| \\ &= \frac{1}{2} RT \ln \left[ r^2 - \left(\lambda - \frac{1}{2}\right)^2 \right] \end{aligned} \quad (11)$$

where  $R$  is the gas constant and  $T$  is the temperature. The  $A(\lambda)$  term in eq 11 stabilizes the end parts of the  $\lambda$  interval by a barrier of a few  $\text{kJ mol}^{-1}$ , as shown in Figure 1B. This barrier needs to be taken into account when adjusting the height  $h$  of the biasing potential.

Note that the free energy  $A(\lambda)$  in Figure 1B, as well as the corresponding probability distribution, diverges for  $\lambda = 0$  and  $\lambda = 1$ . This is, however, not a problem, because, for any finite interval  $[\lambda_1 \cdots \lambda_2]$ , there is a finite probability for the system to be within this interval. Similarly, the partition function integral

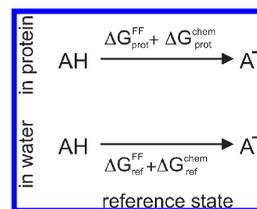
$$Z = \int_{\lambda=0}^{\lambda=1} \exp(-\beta A(\lambda)) d\lambda = \pi \quad (12)$$

with  $\beta = 1/(\kappa_B T)$ , over every finite  $\lambda$  interval of the free-energy curve is also finite.

**2.4.  $\lambda$ -Dynamics Thermostat.** The temperature of the  $\lambda$  particle is kept constant by coupling the particle to an external heat bath. We have considered two coupling schemes, the Berendsen,<sup>30</sup> or weak coupling thermostat, and the Andersen thermostat.<sup>42</sup>

It is not clear *a priori* whether to couple the  $\lambda$  particle to (i) the same heat bath as the *real* atoms of the system, or (ii) a separate heat bath. In the first situation, the temperature is computed from the total kinetic energy of the real atoms and the  $\lambda$  particle. In the second, different heat baths are used to couple the  $\lambda$  particle and real atoms separately, and the kinetic energy of the  $\lambda$  particle is used to calculate the temperature of the  $\lambda$  subsystem.

Therefore, we have tested the two coupling schemes. For variant (i), we used the Berendsen thermostat (with a coupling time of 0.10 ps), whereas, for variant (ii), the  $\lambda$  particle was coupled to the Andersen thermostat (with a coupling time of 0.15 ps), and the rest of the system to the Berendsen thermostat.



**Figure 2.** Equilibria between the protonated (AH) and deprotonated ( $A^-$ ) forms of a titratable site in a protein and in the reference state in water.  $\Delta G_{\text{prot}}^{\text{FF}}$  and  $\Delta G_{\text{ref}}^{\text{FF}}$  are obtained from a molecular dynamics (MD) simulation, whereas  $\Delta G_{\text{prot}}^{\text{chem}}$  and  $\Delta G_{\text{ref}}^{\text{chem}}$  include contributions from the environmental pH and from bonded terms, which are missing in the force field. We assume  $\Delta G_{\text{prot}}^{\text{chem}} \approx \Delta G_{\text{ref}}^{\text{chem}}$ .

We have used the Andersen thermostat, because the Berendsen thermostat is not suitable for low-dimensional systems, such as the  $\lambda$  subsystem.<sup>30</sup> At 300 K and with a  $\lambda$  particle mass of 20 u, the latter coupling scheme generated  $\lambda$ -trajectories that were more suitable (i.e., sufficiently long residence time at the physical states, fast transitions) to simulate biomolecular systems (see the Results section).

**2.5. Constant pH MD Simulations with  $\lambda$ -Dynamics.** To describe protonation and deprotonation events of a titratable site at a given pH, we included, within the Hamiltonian in eq 4, (i) the effect of the external pH bath on protonation and (ii) contributions to the free energy of protonation due to breakage and formation of chemical bonds, which are not described by the force field. These two free-energy contributions will be described by an additional term  $V^{\text{chem}}(\lambda)$ , which will shift the protonation equilibria by a certain free energy ( $\Delta G^{\text{chem}}$ ).

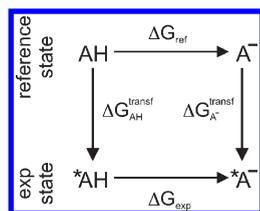
To determine  $\Delta G^{\text{chem}}$ , we considered the equilibrium between a protonated (AH) and a deprotonated acid ( $A^-$ ), in a (solvated) protein (see Figure 2, top) and in water (see Figure 2, bottom). We will use the latter as a *reference state*. This state is chosen such that a measured deprotonation free energy is available, and the reference compound AH is chemically similar for the reference and protein states, generally a solvated amino acid. Note that no  $\text{H}^+$  or  $\text{H}_3\text{O}^+$  species appears on the right side of the equilibria in Figure 2, since, here, we consider the free-energy difference between the protonated and deprotonated forms of the titratable site. Below, we will describe how the pH dependency of this free energy is taken into account.

The free energies for the top (prot) and bottom (ref) reactions of Figure 2 are split into a contribution  $\Delta G^{\text{FF}}$  (obtained via a force field calculation) and  $\Delta G^{\text{chem}}$  (contributions (i) and (ii) from pH bath and bond breakage and formation, respectively). Because of the choice of the reference state,  $\Delta G^{\text{chem}}$  is not expected to differ significantly between the top and bottom reactions in Figure 2.<sup>18,43,44</sup> Thus, the dominant contribution to the difference in the free energies of these two reactions is due to the different environment of the titratable site in the protein and in water. This contribution essentially depends on the long-range interactions of the titratable group, which are described by the force-field free-energy terms  $\Delta G_{\text{prot}}^{\text{FF}}$  and  $\Delta G_{\text{ref}}^{\text{FF}}$ .

Accordingly,

$$\Delta G_{\text{prot}}^{\text{chem}} \approx \Delta G_{\text{ref}}^{\text{chem}} = (\ln 10)RT(\text{p}K_{\text{a,ref}} - \text{pH}) - \Delta G_{\text{ref}}^{\text{FF}} \quad (13)$$

where  $\text{p}K_{\text{a,ref}}$  is the measured  $\text{p}K_{\text{a}}$  of the reference titratable site in the reference state. The pH term describes the pH dependency of



**Figure 3.** Thermodynamic cycle for the calculation of the reference free-energy difference  $\Delta G_{\text{ref}}$ . AH and  $A^-$  are transformed to chemically similar compounds  $*AH$  and  $*A^-$ , respectively, for which the free-energy difference has been experimentally measured ( $\Delta G_{\text{exp}}$ ).

the equilibria in Figure 2, thus accounting for the missing proton in Figure 2.

The last contribution in eq 13,  $\Delta G_{\text{ref}}^{\text{FF}}$  is obtained from a thermodynamic integration calculation (reference free-energy simulation), which is performed prior to starting the constant pH simulation,

$$\begin{aligned} \Delta G_{\text{ref}}^{\text{FF}} &= G_{\text{ref}}^{\text{FF}}(\lambda = 1) - G_{\text{ref}}^{\text{FF}}(\lambda = 0) \\ &= \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H_{\text{ref}}(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \end{aligned} \quad (14)$$

where  $H_{\text{ref}}(\lambda)$  is the Hamiltonian of the reference system.

Having determined  $\Delta G^{\text{chem}}$  for the protein state, the following potential  $V^{\text{chem}}(\lambda)$  serves to implement the desired free-energy difference in the  $\lambda$ -dynamics calculations:

$$V^{\text{chem}}(\lambda) = \lambda(\ln 10)RT(\text{p}K_{\text{a,ref}} - \text{pH}) - \tilde{\Delta G}_{\text{ref}}^{\text{FF}}(\lambda) \quad (15)$$

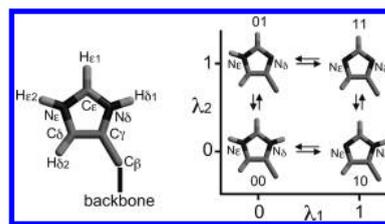
with  $\tilde{\Delta G}_{\text{ref}}^{\text{FF}}(\lambda)$  as a polynomial fit to  $G_{\text{ref}}^{\text{FF}}(\lambda)$ , which is typically close to a parabola.<sup>45,46</sup>

Note the use of  $\tilde{\Delta G}_{\text{ref}}^{\text{FF}}(\lambda)$  to describe the  $\Delta G_{\text{ref}}^{\text{FF}}(\lambda)$  contribution, instead of a linear function of  $\lambda$  (analogously to the first term in eq 15). By this choice, the free-energy profile of the reference state ( $G_{\text{ref}}^{\text{FF}}(\lambda)$ ) is effectively subtracted (except for fluctuations) from the one of the protein state ( $G_{\text{prot}}^{\text{FF}}(\lambda)$ ). In the simplest case of a constant pH simulation of an amino acid in water, in which case  $G_{\text{ref}}^{\text{FF}}(\lambda)$  and  $G_{\text{prot}}^{\text{FF}}(\lambda)$  are the same,  $\tilde{\Delta G}_{\text{ref}}^{\text{FF}}(\lambda)$  will, therefore, remove the barrier in the energy landscape between the protonated and deprotonated states of the titratable site. Therefore, the barrier is given and controlled directly by the height of the biasing potential, which thus can be adjusted to achieve the desired transition rates. In the less trivial case of a protein simulation,  $\tilde{\Delta G}_{\text{ref}}^{\text{FF}}(\lambda)$  will not remove the barrier completely, but still the remaining perturbation can be assumed to be small also in the general case.

**2.5.1. Reference Thermodynamic Cycle.** If a measured  $\text{p}K_{\text{a}}$  is available only for a compound that is similar, but not identical, to that considered in the reference state, a thermodynamic cycle can be used to calculate and correct for the free-energy difference due to this modification. In Figure 3, the free-energy difference of the reference state ( $\Delta G_{\text{ref}}$ ) is given by

$$\Delta G_{\text{ref}} = \Delta G_{\text{exp}} + (\Delta G_{\text{AH}}^{\text{transf}} - \Delta G_{\text{A}^-}^{\text{transf}}) \quad (16)$$

where  $\Delta G_{\text{AH}}^{\text{transf}}$  and  $\Delta G_{\text{A}^-}^{\text{transf}}$  indicate the free-energy differences for the transformation of the protonated and deprotonated forms of the reference state into the corresponding compounds of the experimentally known state (*exp*), respectively. The terms  $*AH$  and  $*A^-$  in Figure 3 denote compounds chemically similar to those in the reference state.



**Figure 4.** Four protonation states of the histidine side chain:  $\lambda_1$  and  $\lambda_2$  are the titration coordinates of the  $N_{\epsilon}$  and  $N_{\delta}$  deprotonation sites, respectively.  $\lambda_1 = \lambda_2 = 0$  (00) corresponds to the fully protonated and positively charged histidine;  $\lambda_1 = 0, \lambda_2 = 1$  (01) and  $\lambda_1 = 1, \lambda_2 = 0$  (10) correspond to the neutral histidine; and  $\lambda_1 = \lambda_2 = 1$  (11) corresponds to the negatively charged fully deprotonated histidine.

After calculation of  $\Delta G_{\text{AH}}^{\text{transf}}$  and  $\Delta G_{\text{A}^-}^{\text{transf}}$  by conventional TI, these two free-energy differences are included in eq 4, similar to  $V^{\text{chem}}(\lambda)$  in eq 15:

$$V^{\text{transf}}(\lambda) = \lambda(\Delta G_{\text{AH}}^{\text{transf}} - \Delta G_{\text{A}^-}^{\text{transf}}) \quad (17)$$

This approach will be used further below to parametrize the  $\lambda$ -dynamics simulation of histidine.

**2.6. Generalization to Multiple Titratable Groups.** The above formulation of the  $\lambda$ -dynamics approach for constant pH simulations is extended to multiple titratable groups by assigning a separate  $\lambda$ -coordinate to each titratable group in the protein.<sup>34,35</sup> In order to illustrate the approach, we first will consider the case of two titratable sites on a protein and derive the Hamiltonian for this system. We will then distinguish the case of two sites, which are (i) chemically *uncoupled* and (ii) chemically *coupled*. In the first case of uncoupled sites, interactions between titratable sites are mainly governed by electrostatics. In terms of the force field, these sites interact only via nonbonded interactions, which are described by the Coulomb and Lennard-Jones potential energies. For this reason, the Hamiltonian for uncoupled sites can be extended in a straightforward manner to any number  $N$  of uncoupled titratable sites in a protein,<sup>34,35</sup> and formally linear scaling with the number of sites is achieved. As this approach will be used later on, we will review it below. For chemically coupled sites, this straightforward approach is not applicable. In this case, the chemical character, which is described in the force field by a set of parameters, such as atomic charges, bonds, and angles, of the titratable sites depends on the protonation states of the respective other coupled sites. Because of this dependency, cross terms occur in the expression for the potential energies, which have to be taken into account explicitly, and the contributions of interacting atoms cannot be rearranged as conveniently as those for uncoupled sites. Therefore, unavoidably, in this case, the number of calculations scales exponentially with the number of sites, rather than linearly. Here, we will discuss the example of histidine, where the two deprotonation sites on the side chain are chemically coupled. Note that, in this case, since only two sites are coupled, the calculations still scale linearly. We will also discuss how this description of histidine differs from the treatment of Khandogin and Brooks.<sup>35</sup>

**2.6.1. Constant pH  $\lambda$ -Dynamics of Two Titratable Sites on a Protein.** We start by considering the case of two titratable sites on a protein. Each of the two sites  $i$  and  $j$  is described by a  $\lambda$ -coordinate,  $\lambda_i$  and  $\lambda_j$ , respectively. At  $\lambda = 0$ , the site is protonated; at  $\lambda = 1$ , the site is deprotonated. Independent of whether the two titratable groups are uncoupled or coupled, four protonation states are relevant. In Figure 4, these four states for histidine are denoted as 00 (both sites  $i$  and  $j$  protonated), 10

(site  $i$  deprotonated and site  $j$  protonated), 01 (site  $i$  protonated and site  $j$  deprotonated), and 11 (both sites  $i$  and  $j$  deprotonated).

Applying eq 4 in a first step to each group  $i$  and  $j$  separately, and combining the two resulting Hamiltonians again, according to eq 4 in a second step, yields

$$H(\lambda_i, \lambda_j) = (1 - \lambda_i)[(1 - \lambda_j)\tilde{H}_{00} + \lambda_j\tilde{H}_{01}] + \lambda_i[(1 - \lambda_j)\tilde{H}_{10} + \lambda_j\tilde{H}_{11}] + H_{\text{Env}} + \left(\frac{m_i}{2}\right)\dot{\lambda}_i^2 + \left(\frac{m_j}{2}\right)\dot{\lambda}_j^2 + U^*(\lambda_i) + U^*(\lambda_j) \quad (18)$$

where the first four Hamiltonians on the right side of the equation describe the titratable sites of the protein in the four protonation states in Figure 4, and  $U^*(\lambda)$  is the biasing potential discussed earlier in section 2.3.2.

Similarly, the potential energy of the system described by the Hamiltonian  $H(\lambda_i, \lambda_j)$  in eq 18 is given as

$$V(\lambda_i, \lambda_j) = (1 - \lambda_i)[(1 - \lambda_j)V_{00} + \lambda_jV_{01}] + \lambda_i[(1 - \lambda_j)V_{10} + \lambda_jV_{11}] + V_{\text{Env}} + U^*(\lambda_i) + U^*(\lambda_j) \quad (19)$$

where the first four potential energies  $V$  on the right side of the equation describe the interactions of the titratable sites in their respective states (see Figure 4), with forces acting on  $\lambda_i$  and  $\lambda_j$ , respectively:

$$-\frac{\partial V(\lambda_i, \lambda_j)}{\partial \lambda_i} = -[(1 - \lambda_j)(V_{10} - V_{00}) + \lambda_j(V_{11} - V_{01})] - \frac{dU^*(\lambda_i)}{d\lambda_i} \quad (20)$$

and

$$-\frac{\partial V(\lambda_i, \lambda_j)}{\partial \lambda_j} = -[(1 - \lambda_i)(V_{01} - V_{00}) + \lambda_i(V_{11} - V_{10})] - \frac{dU^*(\lambda_j)}{d\lambda_j} \quad (21)$$

As can be seen for the case of two interacting titratable sites, the force acting on each site depends on the protonation state of the respective other site, which also holds true for the general case of  $N$  interacting sites. This interdependence entails an exponential scaling.

**2.6.2. Chemically Uncoupled Titratable Sites.** If the two titratable sites are chemically uncoupled, however, the computational complexity is dramatically reduced. Uncoupled sites interact only via long-range (nonbonded) interactions. Below, we will show how these interactions (Coulombic and van der Waals) are efficiently described, achieving linear scaling of the calculations.

**Coulombic Interactions.** For two uncoupled titratable sites  $i$  and  $j$ , the Coulombic potential energy ( $V^c$ ) for two interacting atoms simplifies (from eq 19) to

$$V^c(\lambda_i, \lambda_j) = \frac{1}{4\pi\epsilon}[(1 - \lambda_i)q_0^i + \lambda_iq_1^i][(1 - \lambda_j)q_0^j + \lambda_jq_1^j] \quad (22)$$

where  $q_0$  and  $q_1$  are the atomic charges in the protonated ( $\lambda = 0$ ) and deprotonated ( $\lambda = 1$ ) states, respectively, of the corresponding atoms,  $r$  is the distance between the two atoms, and  $\epsilon$  is the permittivity. Note that eq 22 involves only two states, compared to the four states of eq 19.

Accordingly, the force acting on  $\lambda_i$  is

$$-\frac{\partial V^c(\lambda_i, \lambda_j)}{\partial \lambda_i} = -[V^c(\lambda_i = 1, \lambda_j) - V^c(\lambda_i = 0, \lambda_j)] \quad (23)$$

where the Coulombic energies  $V^c(\lambda_i = 0, \lambda_j) = [1/(4\pi\epsilon)]q_0^i[(1 - \lambda_j)q_0^j + \lambda_jq_1^j]$  and  $V^c(\lambda_i = 1, \lambda_j) = [1/(4\pi\epsilon)]q_1^i[(1 - \lambda_j)q_0^j + \lambda_jq_1^j]$  are evaluated at  $\lambda_j$ .

Equation 23 is extended in a straightforward manner to  $N$  uncoupled interacting sites:

$$-\frac{\partial V^c(\lambda_1, \dots, \lambda_{i-1}, \lambda_i, \lambda_{i+1}, \dots, \lambda_N)}{\partial \lambda_i} = -[V^c(\lambda_1, \dots, \lambda_{i-1}, \lambda_i = 1, \lambda_{i+1}, \dots, \lambda_N) - V^c(\lambda_1, \dots, \lambda_{i-1}, \lambda_i = 0, \lambda_{i+1}, \dots, \lambda_N)] \quad (24)$$

and linear scaling of the calculation with the number of interacting uncoupled sites is achieved.

**van der Waals Interactions.** The remaining long-range interactions are somewhat less straightforward. We consider the usual case where the van der Waals energies, together with the Pauli repulsion, are described by a Lennard-Jones potential  $V_{\text{LJ}}$ :

$$V_{\text{LJ}} = \frac{A}{r^{12}} - \frac{B}{r^6} \quad (25)$$

where  $r$  is the distance between the two atoms, and  $A$  and  $B$  are two parameters, which depend on the pairs of interacting atoms  $i$  and  $j$ ,

$$A = (A^i A^j)^{1/2} \quad (26)$$

and similarly for  $B$ .

For two uncoupled titratable sites  $i$  and  $j$ , the Lennard-Jones potential energy for two interacting atoms is (here, we treat only the  $r^{12}$  part; the  $r^6$  part is very similar)

$$V_{\text{LJ}}^{12}(\lambda_i, \lambda_j) = \frac{(1 - \lambda_i)[(1 - \lambda_j)A_{00} + \lambda_jA_{01}] + \lambda_i[(1 - \lambda_j)A_{10} + \lambda_jA_{11}]}{r^{12}} \quad (27)$$

where the indices of the  $A$  parameter indicate the protonation states of the two titratable sites (see Figure 4).

Similar to the Coulombic energy, eq 27 is rearranged in terms of the protonated ( $\lambda = 0$ ) and deprotonated ( $\lambda = 1$ ) values of the  $A^i$  and  $A^j$  Lennard-Jones parameters,

$$V_{\text{LJ}}^{12}(\lambda_i, \lambda_j) = \frac{[(1 - \lambda_i)(A_0^i)^{1/2} + \lambda_i(A_1^i)^{1/2}][(1 - \lambda_j)(A_0^j)^{1/2} + \lambda_j(A_1^j)^{1/2}]}{r^{12}} \quad (28)$$

with force acting on  $\lambda_i$

$$-\frac{\partial V_{\text{LJ}}^{12}(\lambda_i, \lambda_j)}{\partial \lambda_i} = -[V_{\text{LJ}}^{12}(\lambda_i = 1, \lambda_j) - V_{\text{LJ}}^{12}(\lambda_i = 0, \lambda_j)] \quad (29)$$

The potentials  $V_{\text{LJ}}^{12}(\lambda_i = 1, \lambda_j)$  and  $V_{\text{LJ}}^{12}(\lambda_i = 0, \lambda_j)$  are obtained by evaluating the second term in square brackets on the right side of eq 28 prior to starting the force calculation, analogous to the calculation of the Coulombic forces. As a more technical remark, note that, in GROMACS,<sup>39</sup> the Lennard-Jones parameters are not accessible in a straightforward manner in the MD source code. Therefore, instead of interpolating linearly between  $(A_0^j)^{1/2}$  and  $(A_1^j)^{1/2}$ , we define the atom type ( $a$ ) of the  $j$  atom, which is

used to determine  $A_0^j$  and  $A_1^j$ , prior to calculating the force, by

$$a^j(\lambda_j) = \begin{cases} a_0 & \lambda_j \leq 0.5 \\ a_1 & \lambda_j > 0.5 \end{cases} \quad (30)$$

This yields, effectively, an approximation to the second term in square brackets on the right side of eq 28. Note that, in the GROMOS96 force field,<sup>47</sup> only the  $A$  term of the atoms of the carboxylic group changes upon deprotonation. Since, in the Lennard-Jones potential (eq 25), the  $A$  (the repulsion) term decays with  $1/r^{12}$ , the approximation in eq 30 is not expected to introduce significant artifacts.

**2.6.3. Chemically Coupled Titratable Sites.** We move now to the situation of chemically coupled sites. To illustrate this case, Figure 4 shows the four protonation states of histidine, where  $\lambda_1$  and  $\lambda_2$  denote the titration coordinates of the  $N_\epsilon$  and  $N_\delta$  sites, respectively. In contrast to the chemically uncoupled situation, here, the protonation state of one site (e.g.,  $N_\delta$ ) does affect the charge of the other site ( $N_\epsilon$ ). Depending on the chemistry, other force-field parameters also may be affected. This prevents further simplification of eq 18, which leaves us with four Hamiltonians ( $\tilde{H}_{00}$ ,  $\tilde{H}_{01}$ ,  $\tilde{H}_{10}$ , and  $\tilde{H}_{11}$ ) and four states for the atomic charges ( $q_{00}$ ,  $q_{01}$ ,  $q_{10}$ , and  $q_{11}$ ). Therefore, the calculations will scale exponentially with the number of coupled sites, as each combination of the protonation states of the sites must be evaluated.

We note that this description of histidine differs from that of Khandogin and Brooks,<sup>35</sup> in that each of the two titratable sites on the side chain is described by a titration coordinate, and the coupling between the two sites is taken into account explicitly. Accordingly, our treatment also describes the doubly deprotonated, negatively charged form of histidine, which is not included in the model of Khandogin and Brooks,<sup>35</sup> where only three states are considered. Furthermore, our treatment is readily generalized to more than two chemically coupled titratable sites.

**Chemically Coupled Reference States.** The chemical coupling between titratable sites also must be taken into account for the reference states in a constant pH simulation. For example, when  $\lambda_2$  changes from 0 to 1 in histidine, the reference deprotonation reaction of the titratable site described by  $\lambda_1$  changes from the bottom ( $00 \rightleftharpoons 10$ ) to the top ( $01 \rightleftharpoons 11$ ) deprotonation equilibrium in Figure 4.

To account for this dependency, we define  $V^{\text{chem}}(\lambda_1, \lambda_2)$  (see for comparison  $V^{\text{chem}}(\lambda)$  in eq 15), e.g., for group  $\lambda_1$ , as

$$V^{\text{chem}}(\lambda_1, \lambda_2) = \lambda_1(\ln 10)RT(\text{p}K_{a,\text{ref}}^*(\lambda_2) - \text{pH}) - \Delta\tilde{G}_{\text{ref}}^{\text{FF}}(\lambda_1, \lambda_2) \quad (31)$$

where

$$\text{p}K_{a,\text{ref}}^*(\lambda_2) = (1 - \lambda_2)\text{p}K_{a,\text{ref}}(00 \rightleftharpoons 10) + \lambda_2\text{p}K_{a,\text{ref}}(01 \rightleftharpoons 11) \quad (32)$$

and  $\Delta\tilde{G}_{\text{ref}}^{\text{FF}}(\lambda_1, \lambda_2)$  is a polynomial fit to  $G_{\text{ref}}^{\text{FF}}(\lambda_1, \lambda_2)$ , which is the force-field free-energy profile for the reference deprotonations. To determine  $\Delta\tilde{G}_{\text{ref}}^{\text{FF}}(\lambda_1, \lambda_2)$ , several reference free-energy simulations at different values of  $\lambda_2$  are performed (see the Methods section).

Similarly to the reference state, the reference thermodynamic cycle (in section 2.5.1) of chemically coupled titratable sites will depend on the protonation state of the respective other sites. For

the example of histidine, eq 17 becomes, e.g., for group  $\lambda_1$ ,

$$V^{\text{transf}}(\lambda_1, \lambda_2) = \lambda_1[(1 - \lambda_2)(\Delta G_{\text{AHH}^+} - \Delta G_{\text{AH}}) + \lambda_2(\Delta G_{\text{AH}} - \Delta G_{\text{A}^-})] \quad (33)$$

with  $\Delta G_{\text{AHH}^+}$ ,  $\Delta G_{\text{AH}}$ , and  $\Delta G_{\text{A}^-}$  being the transfer free energies of the double protonated (00), singly protonated (10 or 01), and fully deprotonated (11) forms of histidine (see Figure 4).

### 3. METHODS

**3.1.  $\text{p}K_a$  Calculations.** To estimate the  $\text{p}K_a$  of a titratable compound, constant pH simulations of the compound at different pH values were performed, similar to a titration experiment. From each simulation, the fraction ( $S$ ) of deprotonated acid was calculated, and the Henderson–Hasselbalch equation was fitted to the obtained titration curve,

$$S^{\text{deprot}} = \frac{1}{10^{(\text{p}K_a - \text{pH})} + 1} \quad (34)$$

which, for  $N$  noninteracting titratable sites, takes the form

$$S^{\text{deprot}} = \sum_i^N \frac{1}{10^{(\text{p}K_{a,i} - \text{pH})} + 1} \quad (35)$$

In one case, where the fit was not satisfactory, the Hill equation has been used,

$$S^{\text{deprot}} = \frac{1}{10^{n(\text{p}K_a - \text{pH})} + 1} \quad (36)$$

where  $n$  is the Hill coefficient, which accounts for the degree of cooperativity ( $n > 1$ ) or anticooperativity ( $n < 1$ ) of the system.<sup>48,49</sup>

The fraction of deprotonated acid  $S$  in a constant pH simulation was calculated from the titration coordinate  $\lambda$  during the simulation, where all steps with  $\lambda < 0.1$  were recorded as protonated and those with  $\lambda > 0.9$  as deprotonated. The error in the calculated  $S$  was estimated via a Bayesian approach from the number of transitions observed during the simulations between the protonated and deprotonated states (see the Supporting Information).

In contrast to a conventional titration experiment, in a constant pH simulation, the titration coordinates of each titratable site in the compound are accessible. Therefore, both the *macroscopic* (or *apparent*)  $\text{p}K_a$  values of the entire compound, and the *microscopic*  $\text{p}K_a$  values of each site, can be estimated.

For a compound with two titratable sites, such as histidine, the equilibrium constant for the deprotonation of the first proton ( $K_{a,I}$ ) is related to the equilibrium constants for the deprotonations at sites  $N_\epsilon$  and  $N_\delta$  ( $K'_{a,1}$  and  $K'_{a,2}$ , respectively) by

$$K_{a,I} = K'_{a,1} + K'_{a,2} \quad (37)$$

from which follows

$$\text{p}K_{a,I} = -\log_{10}(10^{-\text{p}K'_{a,1}} + 10^{-\text{p}K'_{a,2}}) \quad (38)$$

with  $\text{p}K_{a,I}$  the (macroscopic)  $\text{p}K_a$  value for the deprotonation of the first proton of histidine, and  $\text{p}K'_{a,1}$  and  $\text{p}K'_{a,2}$  the (microscopic)  $\text{p}K_a$  value for the deprotonation of the first proton of histidine at sites  $N_\epsilon$  and  $N_\delta$ , respectively.

Similarly, the equilibrium constant for the deprotonation of the second proton of histidine ( $K_{a,II}$ ) is related to the equilibrium constants for the deprotonations at sites  $N_\epsilon$  and  $N_\delta$  ( $K''_{a,1}$  and

$K''_{a,2}$ , respectively) by

$$K_{a,\text{II}} = \frac{1}{\frac{1}{K''_{a,1}} + \frac{1}{K''_{a,2}}} \quad (39)$$

from which follows

$$\text{p}K_{a,\text{II}} = \log_{10}(10^{\text{p}K''_{a,1}} + 10^{\text{p}K''_{a,2}}) \quad (40)$$

where  $\text{p}K_{a,\text{II}}$  is the second macroscopic  $\text{p}K_a$  value of histidine, and  $\text{p}K''_{a,1}$  and  $\text{p}K''_{a,2}$  the microscopic  $\text{p}K_a$  values for the deprotonation of the second proton at sites  $N_\epsilon$  and  $N_\delta$ , respectively.

In all cases, the error in the calculated  $\text{p}K_a$  has been determined from the standard deviation of a set of four or five  $\text{p}K_a$  obtained from different fragments of the simulations.

**3.2. Constant pH MD Simulations.** The constant pH MD simulation method, as described above, was implemented in the GROMACS MD package (version 3.3).<sup>37–39</sup>

As test cases, constant pH simulations were carried out for four compounds: glutamic acid (Glu) with neutral termini, a dipeptide of sequence glutamic acid-alanine (Glu-Ala), imidazole, and a capped histidine (acetyl-NH-CHR-CO-methylamide with R the side chain of histidine). Glu, Glu-Ala, and histidine (His) were described with the GROMOS96 53A6 force field.<sup>50</sup> Force-field parameters of imidazole were adapted from histidine (atomic charges are listed in Table s1 in the Supporting Information). For the fully deprotonated form of histidine, no force-field parameters are available in GROMOS96.<sup>50</sup> Charges for this protonation state were thus taken from imidazole and, therefore, are not very accurate. However, in the pH interval considered here (pH 4–10), the doubly deprotonated state should never be visited, because the  $\text{p}K_a$  value for the second deprotonation of histidine is far beyond than the pH interval.<sup>1</sup> Thus, we do not expect a large influence of the charges on the protonation populations.

Each compound was placed in a dodecahedral box, which was subsequently filled with  $\sim 4200$ – $5200$  SPC (simple point charge) water molecules.<sup>51</sup> Interactions between atoms within 1.0 nm were evaluated at every step of the simulation, while interactions with atoms beyond 1.0 nm were evaluated every five steps. The Lennard-Jones long-range cutoff was set to 1.6 nm. The Particle Mesh Ewald (PME)<sup>52,53</sup> was used for the long-range electrostatic interactions, with a grid spacing of 0.12 nm and an interpolation order of 4. Constant pressure and temperature were maintained by weakly coupling the system to an external bath at 1 bar and 300 K, using the Berendsen barostat and thermostat<sup>30</sup> with coupling times of 1.0 and 0.1 ps, respectively. A leapfrog integrator was used with an integration time step of 2 fs. The bond distances and bond angles of water were constrained using the SETTLE algorithm.<sup>54</sup> All other bond distances were constrained using the LINCS algorithm.<sup>55</sup> Prior to the simulations, the potential energy of each system was minimized using a steepest descent approach. A 50-ps MD simulation with position restraints (with a force constant of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ ) on the amino acid/peptide atoms was then performed to relax the water molecules. Finally, a 5-ns simulation was performed to equilibrate each system before starting the constant pH MD simulations.

Deprotonation of a site was achieved by transforming the titratable hydrogen into a dummy atom, which is topologically bound to the acid, but has no interactions with the rest of the system. Charges and atom types of the ionizable groups were changed accordingly, from their force-field values in the

protonated state ( $\lambda = 0$ ) to the deprotonated state ( $\lambda = 1$ ). Bonded terms (bonds, angles, and torsions) were maintained in the protonated state. For glutamic acid and C-terminal, this effectively yields an approximate description of the deprotonated state. For N-terminal, imidazole, and histidine, instead, the bonded terms do not differ in the protonated and deprotonated states of the GROMOS96<sup>50</sup> force field. For glutamic acid in explicit solvent, the free energy of deprotonation was calculated, as described in the next section for the reference free-energy simulations, with and without change in the bonded terms. The difference was less than  $2 \text{ kJ mol}^{-1}$  (see Table s2 in the Supporting Information).

To compare constant pH simulations performed with two different force fields, the titration curve of Glu with neutral termini was calculated also with OPLSA<sup>56</sup> and TIP4P<sup>57</sup> water molecules, and the titration curves for a tripeptide Ala-Glu-Ala were calculated with GROMOS96<sup>50</sup> and OPLSA<sup>57</sup> in SPC<sup>51</sup> water. When OPLSA<sup>57</sup> was used to describe the system, in addition to the bonded terms, atom types also were maintained in their protonated state.

The temperature of the  $\lambda$  degree of freedom was set to 300 K. Unless indicated otherwise, each  $\lambda$  particle was coupled to a separate heat bath via the Andersen thermostat<sup>42</sup> with a coupling parameter of  $6 \text{ ps}^{-1}$ . A fixed barrier height of  $3.0 \text{ kJ mol}^{-1}$  was used for the biasing potential.

The mass of  $\lambda$  was set to 20 u. With this value of the mass, the calculations yielded suitable  $\lambda$ -trajectories (i.e., small ratio between transition time and residence time) for the simulated systems (see the Results section). At the same time, the mass of  $\lambda$  is in the same range as that for the other atoms in the system.

Finally, we note that during the change of the protonation state in the constant pH simulations, the overall charge of the system is (eventually) changed. In this situation, artifacts can arise due to the use of Ewald and related methods to describe electrostatic interactions. In particular, these artifacts are related to the periodic boundary conditions and the background charge that is used to neutralize the system.<sup>58,59</sup> However, for small compounds in a high dielectric medium (water), such as those investigated here, these effects are expected to be negligible.<sup>28,58</sup>

**3.3. Reference States and Reference Free-Energy Simulations.** Constant pH simulations require a reference state for each of the simulated titratable sites. The measured and calculated (force field) deprotonation free energies of this reference state were used to include the effect of the pH bath, and the effect of the breakage and formation of chemical bonds in the simulation (see eq 13).

Table 1 lists the titratable sites and their reference states, as well as the measured  $\text{p}K_a$  values obtained from the literature<sup>1,60–62</sup> and force-field deprotonation free energies ( $\Delta G_{\text{ref}}^{\text{FF}}$ ). Note that two measured  $\text{p}K_a$  values and  $\Delta G_{\text{ref}}^{\text{FF}}$  are reported for imidazole. These correspond to the *microscopic*  $\text{p}K_a$  values for the first and second deprotonation reaction of imidazole, respectively (the second microscopic  $\text{p}K_a$  value is obtained using eq 40, with the second macroscopic  $\text{p}K_a$  value being approximated from histidine, for which there are experimental data<sup>1</sup>). The first and second microscopic  $\text{p}K_a$  values of the  $N_\delta$  and  $N_\epsilon$  sites are identical, because of the symmetry of the imidazole molecule.

For the Ala-Glu-Ala tripeptide, which was added to the compounds set to compare the GROMOS96<sup>50</sup> and OPLSA<sup>57</sup> force fields, the reference states were chosen as follows: acetyl-Glu-methylamide ( $\text{p}K_{a,\text{ref}} = 4.25$ ,<sup>60</sup>  $\Delta G_{\text{ref}}^{\text{FF(GROMOS)}} = -225.6 \text{ kJ mol}^{-1}$ ,

**Table 1.** Reference States, Reference  $pK_a$  Values, and  $\Delta G_{\text{ref}}^{\text{FF}}$  Values

| titratable site                       | reference state             | reference $pK_a$                      | $(\ln 10)RT$ ( $pK_{a,\text{ref}}$ ) ( $\text{kJ mol}^{-1}$ ) | $\Delta G_{\text{ref}}^{\text{FF}}$ ( $\text{kJ mol}^{-1}$ ) |
|---------------------------------------|-----------------------------|---------------------------------------|---|--|
| Glu                                   | Glu (neutral termini)       | 4.25 <sup>60</sup>                    | 24.4  | −220.8   |
| N-terminus                            | di-Ala (neutral C-terminus) | 8.0 <sup>61</sup>                     | 45.9  | 332.8  |
| C-terminus                            | di-Ala (neutral N-terminus) | 3.5 <sup>61</sup>                     | 20.1  | −231.3   |
| imidazole ( $N_\delta$ ) <sup>a</sup> | imidazole ( $N_\delta$ )    | 7.28, <sup>62</sup> 14.4 <sup>1</sup> | 41.8, 82.7  | 155.4, −211.7  |

<sup>a</sup> For imidazole, only  $N_\delta$  is reported; values for  $N_\epsilon$  are the same.

$\Delta G_{\text{ref}}^{\text{FF(OPLSA)}} = -370.5 \text{ kJ mol}^{-1}$ , di-Ala-methylamide ( $pK_{a,\text{ref}} = 8.0$ ,<sup>61</sup>  $\Delta G_{\text{ref}}^{\text{FF(GROMOS)}} = 331.7 \text{ kJ mol}^{-1}$ ,  $\Delta G_{\text{ref}}^{\text{FF(OPLSA)}} = 219.0 \text{ J mol}^{-1}$ ), and acetyl-di-Ala ( $pK_{a,\text{ref}} = 3.5$ ,<sup>61</sup>  $\Delta G_{\text{ref}}^{\text{FF(GROMOS)}} = -230.7 \text{ kJ mol}^{-1}$ ,  $\Delta G_{\text{ref}}^{\text{FF(OPLSA)}} = -338.2 \text{ kJ mol}^{-1}$ ), for titratable sites Glu, N-terminus, and C-terminus, respectively.

The force-field deprotonation free energies for the reference states  $\Delta G_{\text{ref}}^{\text{FF}}$  were determined via conventional thermodynamic integration (see eq 14) as follows. Each reference compound was placed in a dodecahedral box filled with SPC<sup>51</sup> water molecules. The reference free-energy simulations consisted of 5-ns MD, during which  $\lambda$  was continuously increased from 0 to 1, thus deprotonating the reference compound, as described above for the constant pH simulations. The size and shape of the box in the reference and constant pH simulations was identical. Using the same simulation conditions in the reference and constant pH simulations, differences due to approximations of the force field and of the interaction potentials are minimized.<sup>28</sup>

$\Delta \tilde{G}_{\text{ref}}^{\text{FF}}(\lambda)$  (eq 15) was derived from a least-squares fit to  $\partial H/\partial \lambda$  obtained from the reference free-energy simulation. Since the deprotonation reaction in explicit water showed a nonlinear  $\partial H/\partial \lambda$  profile,<sup>28</sup> a third-order polynomial was used. Coefficients of these polynomials are given in Table s3 in the Supporting Information.

The two titration coordinates  $\lambda_1$  and  $\lambda_2$  of imidazole (Figure 4) are chemically coupled and, therefore, deserved particular attention. Here, the reference state changes as a function of the protonation state of the respective other site. Thus,  $\Delta G_{\text{ref}}^{\text{FF}}(\lambda_1, \lambda_2)$ , and, accordingly,  $\Delta \tilde{G}_{\text{ref}}^{\text{FF}}(\lambda_1, \lambda_2)$ , are a function of both,  $\lambda_1$  and  $\lambda_2$  (see eq 31). For this reason, reference free-energy simulations of one titratable site (e.g., the site described by  $\lambda_1$ ) were carried out for  $\lambda_2 = 0, 0.1, \dots, 0.9, 1$ . For each of these 11 simulations, a third-order polynomial in  $\lambda_1$  was fitted to its  $\partial H/\partial \lambda_1$  profile, in a manner similar to the case of chemically uncoupled sites. To describe the dependency from  $\lambda_2$ , third-order polynomials in  $\lambda_2$  were subsequently fitted to the coefficients of these polynomials, and vice versa for the titratable site described by  $\lambda_2$ . These two sets of polynomials served to calculate continuous forces for the two degrees of freedom  $\lambda_1$  and  $\lambda_2$ .

**3.3.1. Histidine Reference State.** As the reference state for the constant pH simulations of histidine, we chose imidazole, such that contributions from the backbone to the proton affinities of the side chain  $N_\epsilon$  and  $N_\delta$  titratable sites were present in the constant pH simulations, but not in the reference free-energy simulations. Because the force-field parameters of imidazole and histidine differed, imidazole was transformed to a modified imidazole molecule described with histidine force-field parameters, using the thermodynamic cycle in Figure 3. The transfer free energies along the thermodynamic cycle were then used to redefine the reference state, according to eq 16. Since  $N_\epsilon$  and  $N_\delta$  are chemically coupled, the transfer potential  $V^{\text{transf}}(\lambda_1, \lambda_2)$  was defined according to eq 33, which accounts for the dependency of the transfer free energies from the protonation state of the

respective other site. The transfer free energies were calculated via free-energy simulations (thermodynamic integration, eq 1). In a first step, the bond lengths and angles were changed from their force-field values in imidazole to those in histidine. In a second step, Lennard-Jones parameters, and, in a last step, charges (see Table 1s in the Supporting Information) were modified. Each free-energy simulation consisted of 18 independent simulations with  $\lambda$  values between 0 and 1. At each  $\lambda$  value, 100 ps of equilibration were followed by 300 ps of data collection. The integration was carried out numerically using the trapezoidal method. The error in  $\langle \partial H/\partial \lambda \rangle_\lambda$  was estimated using block averaging.<sup>63,64</sup>

**3.4.  $\lambda$  Probability Distribution and Free-Energy Profile.** In order to calculate the probability distribution  $p(\lambda)$  during the constant pH simulation, the  $\lambda$  interval was divided in 10 bins [ $\lambda_1, \dots, \lambda_i, \dots, \lambda_{10}$ ], and  $p(\lambda)$  at bin  $i$  was obtained as

$$p(\lambda_i) = \frac{n_i}{N} \quad (41)$$

where  $n_i$  is the time of the simulation during which  $\lambda$  visited bin  $i$  and  $N$  is the total simulation time.

The probability distribution of  $\lambda$ , which is given by the entropic term introduced by the use of the circular coordinate, was calculated as

$$p(\lambda_i) = \int_{\lambda_i}^{\lambda_{i+1}} p(\lambda) d\lambda \quad (42)$$

with

$$p(\lambda) = \frac{\exp[-\beta A(\lambda)]}{Z} \quad (43)$$

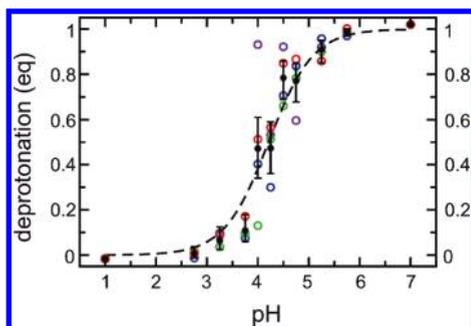
and  $A(\lambda)$  and  $Z$  being obtained from eqs 11 and 12, respectively.  $p(\lambda_i)$  was then used to obtain a free-energy profile as a function of the  $\lambda$  titration coordinate, with the free energy  $G(\lambda)$  at bin  $i$  being given by

$$G(\lambda_i) = -RT \ln p(\lambda_i) \quad (44)$$

## 4. RESULTS

To test the accuracy of the constant pH MD simulation method described above, we have calculated the titration curves of four compounds: glutamic acid, a Glu-Ala dipeptide, imidazole, and histidine. The effects of the choice of the barrier height of the biasing potential, the temperature coupling scheme, and the force field, on the simulation were also investigated.

**4.1. Glutamic Acid.** First, we asked if the constant pH MD simulation method is able to accurately reproduce the titration curve of glutamic acid. To this end, glutamic acid with neutral amino and carboxyl termini ( $-\text{NH}_2$  and  $-\text{COOH}$ , respectively) was solvated in water, and four constant pH simulations of 5 ns



**Figure 5.** Calculated titration curve of glutamic acid with neutral termini. The deprotonation of glutamic acid (in equivalents, eq) is plotted as a function of pH. At each pH, four simulations of 4.5 ns each were performed. Data from each of these simulations (colored dots), and from the average of the four simulations (black filled dots), are shown. Error bars denote estimates from the statistics of the observed transitions. The dashed line is a Henderson–Hasselbalch fit to the average data.

each (0.5-ns equilibration time and 4.5-ns data collection time) were carried out for 11 pH values between 1.0 and 7.0. Figure 5 shows the fractions of deprotonated acid (in equivalents) as a function of pH (i.e., the titration curve) for each of the four sets of simulations (colored dots), and their average (black dots), the latter together with error bars, which were determined from the statistics of the observed transitions, as described in the Methods section and the Supporting Information. As can be seen, the scatter of the four simulations agrees with the error estimate of the average. Note that, at the end points of the titration curve, values slightly below 0 or above 1 are observed, which is due to the use of a radius of  $r = 0.55$  for the circle covered by the angular coordinate  $\theta$  (see the Theory section). We chose  $r = 0.55$  to get 0 and 1, on average, at the protonated ( $\lambda < 0.1$ ) and deprotonated states ( $\lambda > 0.9$ ), respectively. Because of statistical fluctuations, however, slightly negative values and values above 1 occur. However, this is much better than averages of  $\langle \lambda \rangle = 0.05$  or  $\langle \lambda \rangle = 0.95$  for  $r = 0.5$ .

From a fit of the Henderson–Hasselbalch equation (eq 34) to the average deprotonation (the dashed line in Figure 5), the  $pK_a$  value was estimated to be  $4.21 \pm 0.14$ , which is consistent with the measured  $pK_a$  value of 4.25.<sup>60</sup>

For the chemists, we note that, in a titration experiment, the pH is usually measured as a function of the volume of a strong base (or acid) solution added to the analyte solution. In contrast, in the constant pH simulations, pH is a fixed parameter, whereas the equivalents of analyte (i.e., how much of the analyte supplies or reacts with one mole of hydrogen ions) is the quantity to be estimated. Therefore, the titration curves in Figure 5 are to be read as *inverted* titration curves, with respect to a typical experimental titration curve.

Figures 6A–C illustrate the effect of different barrier heights of the biasing potential (see eq 9). As expected, an increase of the barrier height by  $1 k_B T$  ( $\sim 2.5 \text{ kJ mol}^{-1}$ ) reduces the number of transitions by a factor of  $\sim 2.5$ – $3$ . Therefore, by adjusting the biasing potential, the transition rate can be optimized to ensure sufficient sampling of the physical end states. At the same time, the fraction of intermediate states remains small (between 30% with a barrier of  $3 \text{ kJ mol}^{-1}$ , and 10% with a barrier of  $7.5 \text{ kJ mol}^{-1}$ ). Overall, by adjusting the barrier, the statistical error of the constant pH simulation can be minimized.

Note that the effective barrier between the protonated and deprotonated states has a contribution from the entropic barrier

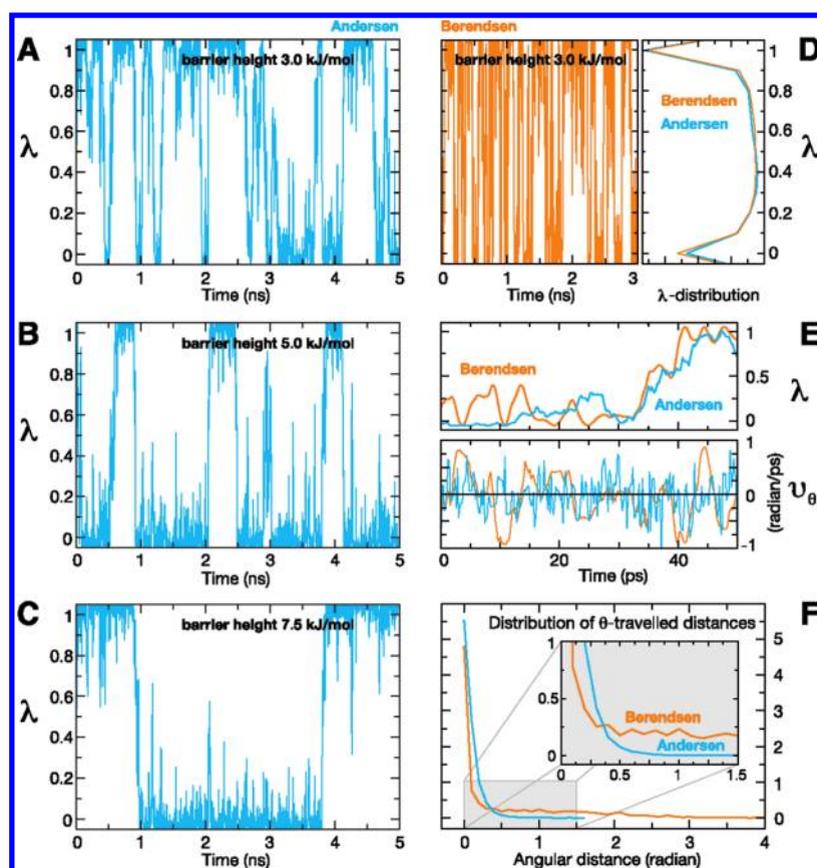
introduced by the use of an angular coordinate to perform the actual  $\lambda$ -dynamics (see the Theory section, Figure 1 and eq 11). This can be seen in Figure 7, which shows the free-energy profile as a function of the titration coordinate  $\lambda$  from an 18-ns constant pH simulation of glutamic acid in explicit solvent at pH 4.25 and with the barrier height of the biasing potential  $3 \text{ kJ mol}^{-1}$  (continuous line). The free energy at  $\lambda = 0.5$  is  $\sim 7 \text{ kJ mol}^{-1}$  more positive than that at  $\lambda = 0$  and  $\lambda = 1$ . When the biasing potential is subtracted from the simulation free-energy profile, we obtain the dotted line in Figure 1, which shows a residual barrier of  $\sim 4 \text{ kJ mol}^{-1}$ . This compares with the entropic barrier term introduced by the use of the angular coordinate  $\theta$  (broken dotted line in Figure 7).

To investigate the effect of the chosen temperature coupling scheme on sampling of the protonation states during the constant pH simulations, the following two variants were considered. In variant (i), the  $\lambda$  particle was coupled to a separate heat bath via the Andersen thermostat,<sup>42</sup> and the rest of the system was coupled to the Berendsen thermostat, whereas in variant (ii), all degrees of freedom were coupled to a common heat bath via the Berendsen thermostat.<sup>30</sup> Figures 6A and 6D compare typical  $\lambda$ -trajectories for the two variants. As can be seen, the number of transitions for the Berendsen variant is  $\sim 3$ – $4$  times larger than that for the Andersen method. Accordingly, the average residence time is  $\sim 3$ – $4$  times shorter for the Berendsen simulation ( $\sim 60 \text{ ps}$ ), compared to that for the Andersen simulation ( $\sim 200 \text{ ps}$ ). The probability distributions of  $\lambda$  with the two variants of the temperature coupling scheme are very similar (see right plot of Figure 6D). Figure 6E shows typical short-time (50 ps) traces of both simulations, with  $\lambda(t)$  shown in the top row, and respective velocities of the underlying angular coordinate ( $v\theta$ ) at the bottom. As can be seen, the  $\lambda$ -trajectories are similar, with the Berendsen variant showing somewhat larger oscillations at the end states. The velocities, in contrast, look very different, with a marked proportion of high-frequency fluctuations for the Andersen thermostat, which are absent for the Berendsen thermostat. Figure 6F quantifies this behavior, in terms of the distribution of angular distances covered by the circular coordinate  $\theta$  between successive velocity reversals. These distances are, on average, shorter for the Andersen thermostat (0.08 radians), as compared to the Berendsen thermostat (0.57 radians). In particular, the long tail for the Berendsen thermostat (up to 6 radians) shows that inertia-driven full circle motions do occur, which implies correlated transitions. This effect reduces the statistical accuracy and is not seen for the Andersen thermostat.

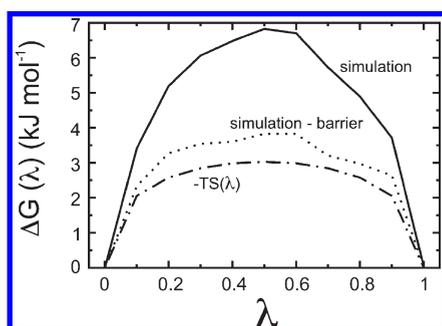
Overall, the Andersen temperature coupling scheme seems to provide a better tradeoff between residence times and the number of uncorrelated transitions. In particular, the  $\lambda$ -trajectories obtained with the Andersen variant showed a sufficiently long residence time at the physical end states, allowing the system to respond to the new protonation state. Because these features are crucial for constant pH simulations, the Andersen temperature coupling scheme has been used for all subsequent simulations.

**4.2. Glu-Ala Dipeptide.** The second system that we considered was the dipeptide Glu-Ala. This system has three interacting titratable sites—glutamic acid (Glu), amino terminus (N-terminus), and carboxyl terminus (C-terminus)—and, therefore, was chosen to test if our method is capable of describing  $pK_a$  shifts due to these interactions.

Constant pH simulations were carried out for 14 pH values between 1.0 and 11.0. Each trajectory covered a 20-ns simulation



**Figure 6.** Dynamics of the deprotonation variable  $\lambda$  of glutamic acid for different barrier heights of the biasing potential and different temperature coupling schemes of  $\lambda$ : (A, B, C)  $\lambda$  is plotted over time during constant pH simulations at pH 4.25 for three different barrier heights ((A)  $3.0 \text{ kJ mol}^{-1}$ , (B)  $5.0 \text{ kJ mol}^{-1}$ , and (C)  $7.5 \text{ kJ mol}^{-1}$ ) of the biasing potential using the Andersen temperature coupling scheme. (D) In the left-hand side of the panel,  $\lambda$  is shown during a constant pH simulation at pH 4.25 for a barrier height of  $3.0 \text{ kJ mol}^{-1}$ , using Berendsen temperature coupling; on the right-hand side of the panel, the  $\lambda$ -distributions of this simulation and of simulation (see panel A) are superimposed. (E) Variable  $\lambda$  and respective velocity  $v\theta$  (in radians/ps) during 50 ps of simulation with Andersen and Berendsen temperature coupling schemes. (F) Distributions of the angular distances (in radians) covered between velocity reversals by the  $\theta$ -variable, during the simulations depicted in panels A and D.

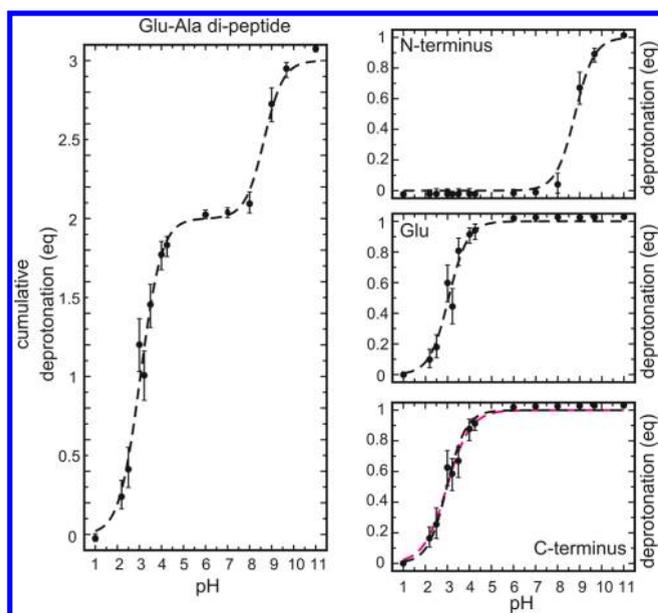


**Figure 7.** Relative free-energy profile as a function of the titration coordinate  $\lambda$  from the 18-ns constant pH MD simulation of glutamic acid in explicit solvent at pH 4.25 (continuous line). The biasing barrier potential (barrier height =  $3 \text{ kJ mol}^{-1}$ ) is subtracted from the simulation free-energy profile to yield the dotted line. The broken dotted line is the relative free-energy profile that is due to the circle entropy ( $-TS(\lambda)$ ; see text).

time with a 2-ns equilibration time and an 18-ns data collection time. Figure 8 shows the obtained titration curve (left graph), in which the cumulative deprotonation (in equivalents) of all three titratable sites is plotted as a function of pH, together with the individual contributions of the three sites (see right side of Figure 8).

Apparent  $\text{pK}_a$  values were estimated from a fit of a sum of three Henderson–Hasselbalch equations (eq 35) to the cumulative titration curve (Table 2). Similarly, the  $\text{pK}_a$  values of each of the titratable sites (site-specific  $\text{pK}_a$  values) were obtained by fitting the Henderson–Hasselbalch equation to the individual titration curves (see Table 2). Note that the apparent  $\text{pK}_a$  values are listed in Table 2, next to each titratable site, only for the sake of clarity, because they are defined in terms of deprotonation of the entire dipeptide.

The apparent and site-specific  $\text{pK}_a$  values are similar for the N-terminus, whereas, for Glu and for the C-terminus, there is a difference of 0.28 and  $-0.22 \text{ pK}_a$  units, respectively (see Table 2). The Henderson–Hasselbalch curve fitted the calculated deprotonation equivalents of the N-terminus and Glu (see top and center right of Figure 8) well, whereas the titration curve of the C-terminus (see bottom right of Figure 8) deviated slightly from the Henderson–Hasselbalch curve. In particular, the slope of the titration curve is shallower, as is indicative of interactions between titratable sites.<sup>48,49</sup> Since the N-terminus was constantly protonated below pH 7 (see top right in Figure 8), the interacting titratable sites were the C-terminus and Glu, which had similar  $\text{pK}_a$  values ( $\sim 3$ ). A fit of the Hill equation (eq 36) to the C-terminus titration curve (dashed magenta line in Figure 8, lower graph) recovers the  $\text{pK}_a$  value of 2.98 already obtained for the Henderson–Hasselbalch fit, and



**Figure 8.** Calculated titration curves of a Glu-Ala dipeptide: (left) titration curve of the dipeptide and (right) site-specific titration curves of the N-terminus, Glu, and C-terminus. The fitted Henderson–Hasselbalch curve (dashed line) and, for the C-terminus, the fitted Hill curve (dashed magenta line) are also shown. Error bars denote estimates from the statistics of the observed transitions.

**Table 2.** Calculated  $pK_a$  values of the Glu-Ala di-peptide

| titratable site | apparent $pK_a$ | site-specific $pK_a$ | reference $pK_a^a$ |
|-----------------|-----------------|----------------------|--------------------|
| N-terminus      | $8.66 \pm 0.13$ | $8.79 \pm 0.10$      | $8.0^{61}$         |
| Glu             | $3.33 \pm 0.08$ | $3.05 \pm 0.08$      | $4.25^{60}$        |
| C-terminus      | $2.76 \pm 0.07$ | $2.98 \pm 0.07$      | $3.5^{61}$         |

<sup>a</sup> Measured  $pK_a$  values of the isolated titratable sites (reference  $pK_a$  values) are listed for the sake of comparison.

it provides a better description of the titration behavior. The obtained Hill coefficient of  $n = 0.82$  indicates a certain degree of anticooperativity in the binding/unbinding of the proton. This is expected, as both the deprotonated C-terminus and Glu are negatively charged, and the release of a proton from one site will increase the affinity of the respective other site.

To quantify the interaction between the C-terminus and Glu in terms of a free energy, we determined the shift in the  $pK_a$  values of these two groups, which is due to the change in the protonation state of the respective other group. For that purpose, we selected, for each titratable site from the trajectories, those frames where the respective other site (C-terminus or Glu) was protonated or deprotonated. In the former case (trajectories where the opposite site was always protonated), we obtained the titration curve for the deprotonation of a site given the respective other was protonated (first *microscopic* titration curve), whereas in the latter (trajectories where the opposite site was always deprotonated), we obtained the titration curve given the other site was deprotonated (second *microscopic* titration curve). From a fit of the Henderson–Hasselbalch equation to the first and second *microscopic* titration curves, we obtained the *microscopic*  $pK_a$  values ( $pK_a^I$  and  $pK_a^{II}$ , respectively). For the C-terminus,  $pK_a^I = 2.89$  and  $pK_a^{II} = 3.05$ , and for Glu,  $pK_a^I = 2.95$  and  $pK_a^{II} = 3.11$ , which show a difference of  $0.16 pK_a$  units between

the first and second *microscopic*  $pK_a$  values for both the C-terminus and the Glu. Thus, the affinity of the two titratable sites for the proton increased upon deprotonation of the other site by  $\sim 1 \text{ kJ mol}^{-1}$ , which is of the same order as a simple estimate of the interaction energy (at the average distance of  $0.6 \text{ nm}$ , see below) from the Coulombic law ( $\sim 3 \text{ kJ mol}^{-1}$ ).

As the C-terminus and Glu became charged, the average distance between these two groups increased. In particular, this distance changed gradually from  $0.55 \text{ nm}$  to  $0.60 \text{ nm}$  between pH 1 (when both groups were protonated) and pH 6 (when both groups were negatively charged), and then more markedly from  $0.60$  to  $0.74 \text{ nm}$  between pH 8 and pH 11, when the N-terminus was mostly deprotonated, and the system had a net charge of  $-2$ .

As can be seen from Table 2 when comparing the site-specific and reference  $pK_a$  values, in all three cases, a shift in the  $pK_a$  value was observed, favoring the charged form of the titratable sites in the dipeptide. In particular, the  $pK_a$  of the N-terminus increased by almost  $1 pK_a$  unit, whereas the  $pK_a$  of the Glu and C-terminus decreased by  $1.2$  and  $0.5 pK_a$  units, respectively. The more-pronounced shifts in the  $pK_a$  value of the N-terminus and Glu suggest that these two groups interact favorably in their charged states. The average distance between the nitrogen of the N-terminus and the oxygens of the carboxyl group of Glu decreased from  $0.47 \text{ nm}$  to  $0.43 \text{ nm}$  between pH 2 and pH 6 and, beyond pH 8, increased again to  $0.47 \text{ nm}$ . The N-terminus and Glu were at the closest distance of  $0.43 \text{ nm}$  between pH 6 and pH 8, when both groups were mainly in their charged states. No significant salt-bridge formation was observed between these two groups ( $<15\%$  of simulation time). On average, the distance between the C-terminus and N-terminus was larger, and almost constant, between pH 1 and pH 8 (between  $0.58 \text{ nm}$  and  $0.59 \text{ nm}$ ).

**4.3. Force-Field Comparison: GROMOS96 and OPLSA.** To assess the sensitivity of the constant pH MD approach to the chosen force field, we calculated the titration curve and  $pK_a$  value of glutamic acid (with neutral termini) with a second force field. In particular, the OPLSA<sup>56</sup> and TIP4P<sup>57</sup> explicit water model was used. In addition, we calculated titration curves and  $pK_a$  values for a tripeptide of sequence Ala-Glu-Ala with GROMOS96<sup>50</sup> and OPLSA,<sup>56</sup> both with an SPC<sup>51</sup> explicit water model.

For glutamic acid with neutral termini, both force fields yielded very similar titration curves, with  $pK_a$  values very close to the reference  $pK_a$  value, as can be seen in Table 3 and in Figure s1 in the Supporting Information. This is expected, because the constant pH simulation is parametrized via the measured  $pK_a$  value of the reference state, which, in these simulations, was glutamic acid, such that any possible force-field bias should cancel.

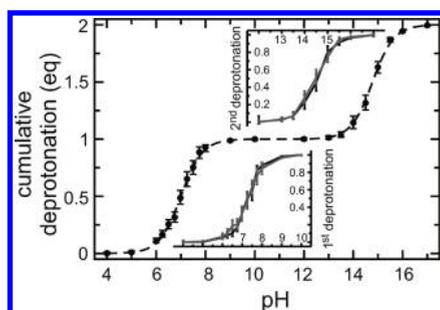
For the Ala-Glu-Ala tripeptide, at each of 15 pH values between 1 and 11, four constant pH simulations were performed for a total of 30 ns per pH value. Slight differences between the titration curves obtained with the two force fields are seen. In particular, the Glu and C-terminus titration curves differed most significantly (see Figure s2 in the Supporting Information). As can be seen in Table 3, the site-specific  $pK_a$  of the C-terminus is shifted by  $-0.4 pK_a$  units, with respect to the reference  $pK_a$  value in the GROMOS96 simulations, whereas it is shifted slightly by  $0.1 pK_a$  unit in the OPLSA simulations. For Glu, the site-specific  $pK_a$  value is shifted by  $0.2$ , with respect to the reference state, in the GROMOS96 constant pH simulations, whereas it is shifted by  $-0.2 pK_a$  units in the OPLSA simulations. Overall, the force-field sensitivity seems to be small.

**4.4. Imidazole.** The titratable sites considered above in the dipeptide simulations interacted only via electrostatics. The

**Table 3.** Calculated  $pK_a$  Values of a Single Glutamic Amino Acid with Neutral Termini ( $\text{NH}_2\text{-Glu-COOH}$ ), and Ala-Glu-Ala Tripeptide with GROMOS96<sup>50</sup> and OPLSA,<sup>56</sup> in Combination with TIP4P<sup>57</sup> and SPC<sup>51</sup> Water Molecules

| $\text{NH}_2\text{-Glu-COOH}$ |                 |                      |                 |                      |                    |
|-------------------------------|-----------------|----------------------|-----------------|----------------------|--------------------|
| titratable site               | GROMOS96 + SPC  |                      | OPLSA + TIP4P   |                      | reference $pK_a^a$ |
|                               | $pK_a$          |                      | $pK_a$          |                      |                    |
| Glu                           | 4.21 ± 0.14     |                      | 4.14 ± 0.07     |                      | 4.25 <sup>60</sup> |
| Ala-Glu-Ala                   |                 |                      |                 |                      |                    |
| titratable site               | GROMOS96 + SPC  |                      | OPLSA + SPC     |                      | reference $pK_a^a$ |
|                               | apparent $pK_a$ | site-specific $pK_a$ | apparent $pK_a$ | site-specific $pK_a$ |                    |
| N-terminus                    | 7.93 ± 0.08     | 8.05 ± 0.08          | 8.01 ± 0.10     | 8.15 ± 0.09          | 8.0 <sup>61</sup>  |
| Glu                           | 4.48 ± 0.13     | 4.46 ± 0.07          | 4.09 ± 0.18     | 4.04 ± 0.06          | 4.25 <sup>60</sup> |
| C-terminus                    | 3.14 ± 0.12     | 3.12 ± 0.11          | 3.55 ± 0.18     | 3.59 ± 0.07          | 3.5 <sup>61</sup>  |

<sup>a</sup> Measured  $pK_a$  values of the isolated titratable sites (reference  $pK_a$  values) are listed for the sake of comparison.



**Figure 9.** Titration curve of imidazole. The cumulative deprotonation (in equivalents, eq) of the two titratable sites ( $N_\epsilon$  and  $N_\delta$ ) is plotted as a function of pH. The dashed line is a fitted Henderson–Hasselbalch curve. The insets show the microscopic titration curves of sites  $N_\epsilon$  (black line) and  $N_\delta$  (gray line) for the first (bottom graph) and second (top graph) deprotonation reaction of imidazole. Error bars were determined from the statistics of the observed transitions.

chemical character of each site of the dipeptide (technically, its force-field parameters) was independent of the protonation state of the other sites, and the free energy of deprotonation of one site was affected only via Coulombic interactions with the other sites. Now, in contrast, we focus on two examples of “chemical coupling”, where two titratable sites interact also via chemical bonds. In this case, the chemical character, and, thus, the  $pK_a$  value of a site, is affected by any change in the protonation state of the respective *coupled* site. As a first example, we consider the two chemically coupled titratable sites  $N_\epsilon$  and  $N_\delta$  of imidazole (Figure 4). A second example, histidine, is discussed further below.

Constant pH simulations of imidazole were carried out for 22 pH values between pH 4 and pH 17. Each trajectory covered 20 ns, with 2 ns of equilibration and 18 ns of data collection. At pH values of 8–13, a barrier height of 0 kJ mol<sup>-1</sup> was used, as discussed further below. Figure 9 shows the titration curve of imidazole, in which the cumulative deprotonation of both  $N_\epsilon$  and  $N_\delta$  titratable sites is plotted as a function of pH. The first and second apparent  $pK_a$  values of imidazole were estimated by a Henderson–Hasselbalch fit as described above, and are listed as  $\text{Im}(N_\epsilon + N_\delta)$  in Table 4. The obtained apparent  $pK_a$  values of  $7.00 \pm 0.12$  and

**Table 4.** Calculated and Measured  $pK_a$  Values of Imidazole (Im) and Histidine (His)

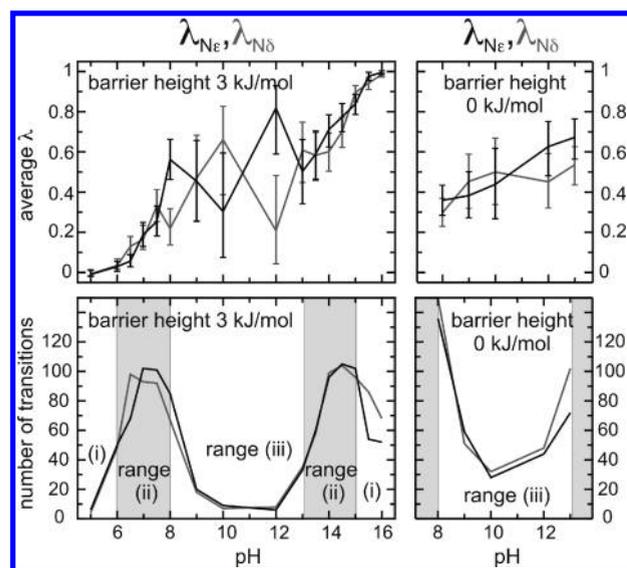
| titratable site                     | calculated $pK_a$               | measured $pK_a$     |
|-------------------------------------|---------------------------------|---------------------|
| Imidazole                           |                                 |                     |
| $\text{Im}(N_\epsilon + N_\delta)$  | $7.00 \pm 0.12, 14.78 \pm 0.08$ | $6.98,^{62} 14.7^1$ |
| $\text{Im}(N_\epsilon)$             | $7.29 \pm 0.08, 14.51 \pm 0.16$ | $7.28,^{62} 14.4^1$ |
| $\text{Im}(N_\delta)$               | $7.28 \pm 0.18, 14.46 \pm 0.18$ | $7.28,^{62} 14.4^1$ |
| Histidine <sup>a</sup>              |                                 |                     |
| $\text{His}(N_\epsilon + N_\delta)$ | $6.56 \pm 0.21$                 | $6.42^{62}$         |
| $\text{His}(N_\epsilon)$            | $7.18 \pm 0.23$                 | $6.92^{62}$         |
| $\text{His}(N_\delta)$              | $6.70 \pm 0.23$                 | $6.53^{62}$         |

<sup>a</sup> For histidine, only the first deprotonation reaction was investigated.

$14.78 \pm 0.08$  agree with the measured  $pK_a$  values of 6.98 (from ref 62) and 14.7 (from ref 1). Note that the measured value for the imidazole second apparent  $pK_a$  value, which was also used for the reference state, is replaced by the one of the chemically similar histidine, for which there are experimental data.<sup>1</sup>

The microscopic  $pK_a$  values of the  $N_\epsilon$  and  $N_\delta$  sites (see Table 4) were estimated from the microscopic titration curves. These were obtained, similar to the C-terminus and Glu of the dipeptide, by plotting the fraction of deprotonated acid at one site, given that the other site was protonated (bottom inset in Figure 9; black for  $N_\epsilon$ , gray for  $N_\delta$ ), or deprotonated (top inset in Figure 9; black for  $N_\epsilon$ , gray for  $N_\delta$ ). The first and second microscopic  $pK_a$  values were similar for  $N_\epsilon$  ( $7.29 \pm 0.08$  and  $14.51 \pm 0.16$ ) and  $N_\delta$  ( $7.28 \pm 0.18$  and  $14.46 \pm 0.18$ ). This is expected, because the two titratable sites of imidazole are equivalent by symmetry. Consistently, the difference between the apparent and microscopic  $pK_a$  values is approximately  $-(\log_{10} 2)$  and  $+(\log_{10} 2)$  for the first and second deprotonation reaction, respectively (see eqs 38 and 40, for the case where  $pK_{a,1} = pK_{a,2}$ ). This follows from the fact that the probability of deprotonating either two of the sites is twice the probability of deprotonating one of the sites.

Since the affinities of the  $N_\epsilon$  and  $N_\delta$  titratable sites are identical, one expects to observe, at every pH value, similar corresponding



**Figure 10.** Average values (top) and number of transitions (bottom) of imidazole titration coordinates  $\lambda_{N_\epsilon}$  (for site  $N_\epsilon$ , black) and  $\lambda_{N_\delta}$  (for site  $N_\delta$ , gray). Constant pH simulations (of 20 ns) were carried out at each pH at barrier heights as indicated. Error bars denote estimates from the statistics of the observed transitions. Ranges (i), (ii), and (iii) are as referenced in the main text.

average deprotonation levels, which provides an independent test of the statistical accuracy of the calculations. Accordingly, Figure 10 shows the average deprotonation  $\langle \lambda \rangle$  (top left), and the number of transitions (bottom left) for  $N_\epsilon$  and  $N_\delta$ , in black and gray, respectively, as a function of pH. Three ranges can be distinguished: (i) pH values close to 5 and 16, with similar  $\langle \lambda \rangle$  of the two sites, and few transitions; (ii) pH values close to 7 and 14, with  $\langle \lambda \rangle$  of  $N_\epsilon$  and  $N_\delta$  also similar, but many transitions; and (iii) pH between 8 and 13, with marked differences for the average deprotonation between the two sites, and again few transitions. To enhance sampling in this last regime, and, thus, statistical accuracy, we lowered the barrier height of the biasing potential from  $3 \text{ kJ mol}^{-1}$  to  $0 \text{ kJ mol}^{-1}$ , which is expected to increase the observed transitions, and repeated the calculations. As can be seen in the right side of Figure 10, the difference in average deprotonation is now significantly smaller, as, indeed, more transitions are observed. This example demonstrates how, by adjusting the barrier height, the transition frequency can be controlled and, thus, the accuracy can be enhanced.

Note that, in range (i), close to pH 5 and 16, also few transitions occur, but the accuracy is much higher than for range (iii) at pH 8–13. This is due to the fact that, in range (iii), statistical fluctuations can favor one of the singly deprotonated forms over the other, whereas in range (i), only one form of imidazole is (mainly) sampled, fully protonated at pH close to 5, and fully deprotonated at pH close to 16. Therefore, insufficient sampling can result in a large inaccuracy in range (iii), as compared to the more straightforward case of range (i), where only one form is sampled.

Similar to range (iii), in range (ii), more than one form of imidazole is significantly sampled (fully protonated and neutral forms, and fully deprotonated and neutral forms, at pH close to 7 and 14, respectively). In this range, the inaccuracy is also larger than that observed in range (i). However, in contrast to range (iii), many more transitions are observed and, thus, sampling is

**Table 5.** Free Energies of Transfer ( $\Delta G^{\text{transf}}$ ) of Each Imidazole Protonation State (see Figure 4) to the Corresponding Histidine State

| protonation state    | $\Delta G^{\text{transf}}$ ( $\text{kJ mol}^{-1}$ ) |
|----------------------|---|
| 00 (+)               | $-12.59 \pm 0.54$                                   |
| 10 ( $N_\delta+$ )   | $-0.04 \pm 0.61$                                    |
| 01 ( $N_\epsilon+$ ) | $-13.68 \pm 0.55$                                   |
| 11 (-)               | $-14.36 \pm 0.59$                                   |

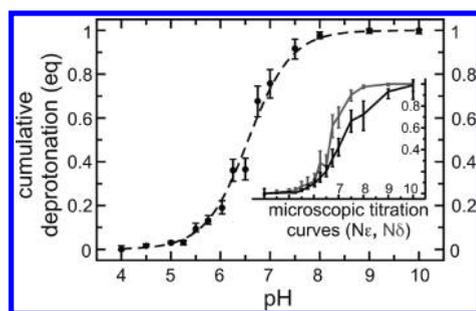
enhanced. This is due to the fact that the free-energy difference between the protonated and deprotonated states of  $\lambda$  is small at pH values close to the  $\text{p}K_a$  value, and the transition barrier is lower, implying more-frequent transitions.

**4.4.1. Histidine.** As a second example of chemical coupling, we considered histidine, which plays a crucial role in many biological processes, because its  $\text{p}K_a$  value is close to the physiological pH. Accordingly, its protonation state changes with its local electrostatic environment. Here, we considered only biologically relevant pH values ( $\text{pH} < 10$ ), because no accurate force-field parameters for the negatively charged, fully deprotonated form of histidine at  $\text{pH} > 10$  are available.<sup>47</sup>

In the previous section, we have studied imidazole, which is the chemical moiety of the histidine side chain. The difference in the measured  $\text{p}K_a$  values of histidine and imidazole is  $\sim 0.5 \text{ p}K_a$  units,<sup>62</sup> with histidine having lower affinity for the proton (see Table 4). Moreover, in histidine, the affinities of the two sites are not identical, as in imidazole, but differ with respect to each other, also by  $\sim 0.5 \text{ p}K_a$  units.<sup>62</sup> This situation enabled us to address the question of whether the constant pH simulation method is capable of quantitatively describing these differences (i.e., the effect of the presence of the backbone on the affinities for the proton of  $N_\epsilon$  and  $N_\delta$ ).

For this purpose, we parametrized the constant pH simulation, such that contributions to the proton affinities from the histidine backbone were not present in the reference state simulation, for which we used imidazole. Because of these contributions, the calculated  $\text{p}K_a$  value is expected to be equal to the measured  $\text{p}K_a$  value of histidine, and is expected to differ from the measured  $\text{p}K_a$  value of the reference imidazole compound. Prior to starting the constant pH simulations, however, the contribution to the affinities from the different force-field parameters of imidazole and histidine were calculated. The thermodynamic cycle in Figure 3 served this purpose (i.e., to compute the free energies of transferring imidazole parameters to histidine parameters for each of the protonation states). Table 5 shows the free energies that have been obtained. As can be seen, these are similar to each other (between  $-12.59 \text{ kJ mol}^{-1}$  and  $-14.36 \text{ kJ mol}^{-1}$ ), except for the neutral form (01). In this form,  $N_\epsilon$  is protonated, whereas  $N_\delta$ , which is two bonds away from the backbone  $C_\beta$ , is deprotonated. The free energies in Table 5 were then used to redefine the reference state (see eq 33) prior to starting the simulations.

Constant pH simulations of histidine were carried out for 15 pH values between pH 4 and pH 10. Each trajectory covered 20 ns of simulation time with 2 ns of equilibration time and 18 ns of data collection time. Similar to imidazole, for pH values between 8 and 10, a barrier height of  $0 \text{ kJ mol}^{-1}$  was used. Figure 11 shows the obtained titration curve. The calculated  $\text{p}K_a$  value of  $6.56 \pm 0.21$ , estimated via a Henderson–Hasselbalch fit (dashed line), agrees with the measured  $\text{p}K_a$  value of 6.42 (see Table 4). Thus,



**Figure 11.** Titration curve of histidine. The cumulative deprotonation (in equivalents, eq) of the two titratable sites ( $N_\epsilon$  and  $N_\delta$ ) is plotted as a function of pH. The dashed line is the fitted Henderson–Hasselbalch curve. The inset shows the microscopic titration curves of the  $N_\epsilon$  and  $N_\delta$  (in black and gray, respectively). Error bars were determined from the statistics of the observed transitions.

the  $\sim 0.5$  p $K_a$  units downward shift in the p $K_a$  value of histidine, with respect to the reference compound (imidazole), was calculated within the statistical error (see Table 4).

The inset of Figure 11 shows the microscopic titration curves, which were obtained as described above for imidazole and Glu-Ala. Henderson–Hasselbalch fits to these curves yielded microscopic p $K_a$  curves of  $7.18 \pm 0.23$  for  $N_\epsilon$  and  $6.70 \pm 0.23$  for  $N_\delta$ , which is consistent with a macroscopic p $K_a$  value of 6.56 (see eq 38). The difference of 0.48 p $K_a$  units between the two microscopic p $K_a$  values agrees with the measured value of 0.39 p $K_a$  units (see Table 4). Thus, in addition to decreasing the affinity of the two sites, the effect of the histidine backbone manifests itself with a shift in the p $K_a$  of the two titratable sites with respect to each other, with  $N_\epsilon$  having a higher affinity than  $N_\delta$ .

## 5. DISCUSSION AND CONCLUSIONS

We have developed a framework to describe changes in protonation states at constant pH, where the requirements of (i) sampling of the relevant  $\lambda$  configurational space, with conformations being the protonated ( $\lambda = 0$ ) and deprotonated ( $\lambda = 1$ ) states, (ii) control of the rate of transition between the two states, and (iii) fully atomistic description of the system are fulfilled. The method, which was implemented within the molecular dynamics (MD) package GROMACS,<sup>37–39</sup> is based on the  $\lambda$ -dynamics approach of Kong and Brooks,<sup>32</sup> and it follows, in the main lines, the constant pH simulation method by Brooks and co-workers.<sup>32,34,35</sup> A new general approach was developed to treat chemically coupled sites, and it was applied to describe the proton tautomerism of imidazole and histidine. In proteins, other examples of chemical coupling are coordinating residues around metal ions, such as that observed in copper binding sites<sup>65</sup> or zinc binding sites.<sup>66</sup>

In order to test whether, and under which conditions, the above-mentioned requirements are actually fulfilled by our method, constant pH simulations of four systems, glutamic acid, Glu-Ala dipeptide, imidazole, and histidine, were carried out. In the following, we will briefly discuss these results in light of the aforementioned requirements, and then we will address the questions of how accurately the calculated average protonation agreed with the measured p $K_a$  values, and whether the method is capable of describing interacting titratable sites. In particular, two types of interactions were considered: those between chemically uncoupled sites, which interact only via electrostatics, and those

between chemically coupled sites, for which a new coupling scheme was developed.

During the constant pH simulations, the average  $\lambda$  in the protonated and deprotonated states was found to be very close to values of 0 or 1, respectively, as required to describe the system in a physically realistic way. This was achieved by appropriately increasing the radius that defines the circular degree of freedom that is used. Similarly, sampling of the intermediate unphysical states was minimized by introducing suitable biasing potentials, in addition to the entropic barrier (of a few  $\text{kJ mol}^{-1}$ ) implied by the angular degree of freedom. It was shown that, for a  $3 \text{ kJ mol}^{-1}$  biasing potential barrier height, more than 70% of the simulation time was spent close to physical states ( $\lambda < 0.1$  and  $\lambda > 0.9$ ).

Adjusting the barrier of the biasing potential also allowed us to control the transition rate, as demonstrated for glutamic acid and imidazole. In particular, for imidazole, it was shown how the accuracy of the calculations at pH values between 8 and 13 was significantly enhanced by increasing the transition rate, thus achieving fast sampling of different protonation states.

In all systems investigated, a fully atomistic description was used, including an explicit solvent. Interestingly, we found that the average residence time at the protonated and deprotonated states is more than 2 orders of magnitude larger than that for comparable systems simulated with the  $\lambda$ -dynamics constant pH approach developed by Brooks and co-workers<sup>34,35</sup> with an implicit solvent. The choice of the thermostat is critical, as shown in Figure 6, where the average residence time is three times larger for the Andersen thermostat, compared to the Berendsen thermostat. Note that the transitions observed in the simulations with the Berendsen thermostat were partially correlated, which reduced the statistical accuracy. However, the thermostat alone does not seem to explain the differences between simulations in implicit and explicit solvents. Thus, the explicit description of water is likely to be crucial as well. We note that the fluctuations in the effective barrier for a transition are quite large due to the water dipoles. These effects, which are important for the kinetics of proton transfer, are not described in implicit water. It would certainly be interesting to study these in more detail.

For the first test system (glutamic acid), the calculated titration curve agreed very well with the measured one. Although this result may seem trivial, as the constant pH simulation was parametrized via the measured p $K_a$  value of glutamic acid (the reference state), it nevertheless shows that the effect of pH is taken into account correctly. As expected, an increase of the total simulation time from 4.5 ns to 18 ns significantly improved accuracy. By increasing the length of the simulation, on one hand, the number of transitions increases, and, on the other hand, a more extensive sampling of the configurational space of the side chain at a certain protonation state is achieved. Both factors enhanced the accuracy of the simulation. Note that adjusting the barrier height of the biasing potential allows one to study the relaxation effects that are due to the change in protonation state.

To study the interaction between titratable sites, we further considered the dipeptide Glu-Ala as a test case. Here, we expected the interactions in the dipeptide to shift the calculated p $K_a$  values, with respect to the values of the individual titratable sites (the reference states). Indeed, the calculated p $K_a$  values were all shifted to favor the charge states of the titratable sites. This was more evident for Glu and N-terminus, which moved closer to each other in the pH range at which they were mainly in their charged states. The C-terminus and Glu had rather similar p $K_a$  values ( $\sim 3$ ), and the individual contributions of these two

groups to the deprotonation of the dipeptide were distinguishable only in the site-specific titration curves. By analyzing the microscopic  $pK_a$  values, the interaction between these two titratable sites was estimated as  $\sim 1 \text{ kJ mol}^{-1}$ . The Hill coefficient of 0.82 for the titration curve of the C-terminus indicated anticooperativity in the system, in agreement with the microscopic  $pK_a$  values. We note that, by calculating the microscopic  $pK_a$  values, this anticooperativity was quantified here, in terms of free energy.

For Glu-Ala, the titratable sites interacted only via electrostatics. In contrast, in imidazole, which was the third system considered, the titratable sites  $N_\epsilon$  and  $N_\delta$  interacted *chemically*, because the affinity for the proton of one site is a function of the protonation state of the respective other site. To describe this type of interaction, a general approach was developed, in which each site is described by a titration coordinate  $\lambda$ , and coupling between the sites is explicitly taken into account. Note that this approach was applied here to describe the tautomerism of imidazole and histidine, but it can be used to describe chemical coupling between any two or more sites. Moreover, we showed that our general approach simplifies to the case of chemically uncoupled sites when interactions occur only via electrostatics. The approach of describing each site of a tautomer as a separate titratable site (or pseudo-site) is not new.<sup>67</sup> However, to avoid the occurrence of the double deprotonated state at pH 7, we do not introduce an arbitrary energy penalty.<sup>67</sup> Instead, the reference states of the pseudo-sites are coupled, such that they are a function of the protonation state of the titratable site. For example, in histidine, the reference  $pK_a$  value of one of the sites on the side chain increases from  $\sim 6$  to  $\sim 14$  as the respective other site deprotonates. Therefore, at pH 7, a second deprotonation is highly improbable. Alternatively, in the constant pH approach of Khandogin and Brooks,<sup>35</sup> tautomerism of titratable amino acids is described by considering three states only. In practice, only one titration coordinate is used, whereas an additional continuous coordinate controls the interconversion between the two tautomeric forms.<sup>35</sup> We note that, by appropriately choosing the protonation states, a three-state description also is obtained within our approach.

We did not use a tautomeric model for Glu in this work. However, it is straightforward to apply such a model to Glu as well, to allow for deprotonation/protonation of both oxygens on the carboxylic group. In particular, such a description is required in protein simulations, in which specific intramolecular interactions can significantly increase the barrier for rotation of the carboxylic group.

The obtained titration curve of imidazole agreed well with the measured one. To test the model, we simulated over a large pH range, to also observe the doubly deprotonated state. Although, at physiological pH, this form is quite unlikely to occur, it cannot be excluded that, in the presence of particular interactions, it plays a role as well.

Histidine was considered as a second example of chemically interacting sites. Here, we investigated the shift of the calculated proton affinity, with respect to the reference one, because of the presence of the backbone. For these simulations, the reference state was *transformed* to a similar one by means of the thermodynamic cycle shown in Figure 3. In general, the reference state is chosen such that the chemical character of the titratable site is similar in the reference and simulated states<sup>28</sup> (i.e., that all differences are described via electrostatics). This implies that one is restricted to those states for which experimental data are

available. Now we have proposed an approach that allows one to use a less similar state, therefore, broadening the range of accessible systems.

Finally, we would like to note that our constant pH approach will also be useful for determining protonation states from X-ray structures. A constant pH MD simulation is performed before the production run is started. During this equilibration phase, position restraints can be applied to the protein backbone, or heavy atoms, to keep the atomic coordinates close to the X-ray data. This procedure might be particularly useful for proteins, in which internal water molecules play a role in stabilizing protonation states.<sup>68</sup>

## ■ ASSOCIATED CONTENT

**S Supporting Information.** In the Supporting Information, Tables s1, s2, and s3 list the side-chain atomic charges of histidine and imidazole, the contribution of bonded terms to the deprotonation free energy of glutamic acid, and the coefficients of polynomial fits for the reference states, respectively. Figures s1 and s2 show the titration curves calculated with GROMOS96<sup>50</sup> and OPLSA<sup>56</sup> force fields, of Glu with neutral termini, and of the tripeptide Ala-Glu-Ala, respectively. The error estimate for the average protonation state via bayesian approach is also presented. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +49-551-2012300. Fax: +49-551-2012302. E-mail addresses: [hgrubmu@gwdg.de](mailto:hgrubmu@gwdg.de) (H.G.), [ggroenh@gwdg.de](mailto:ggroenh@gwdg.de) (G.G.).

### Present Addresses

<sup>†</sup>Institute of Computer Science, Georg August University, Göttingen, Germany.

## ■ ACKNOWLEDGMENT

The authors thank Berk Hess from the University of Stockholm, Sweden, for helpful discussions and Ulrike Gerischer for proofreading the manuscript.

## ■ REFERENCES

- (1) Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman and Company: New York, 1993.
- (2) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
- (3) Haas, J.; Vöhringer-Martinez, E.; Bögehold, A.; Matthes, D.; Hensen, U.; Pelah, A.; Abel, B.; Grubmüller, H. *Chem. Biol. Chem.* **2009**, *10*, 1816–1822.
- (4) Stouffer, A. L.; Acharya, R.; Salom, D.; Levine, A. S.; Costanzo, L. D.; Soto, C. S.; Tereshko, V.; Nanda, V.; Stayrook, S.; DeGrado, W. F. *Nature* **2008**, *451*, 596–599.
- (5) Schnell, J. R.; Chou, J. J. *Nature* **2008**, *451*, 591–595.
- (6) Tournaire-Roux, C.; Sutka, M.; Javot, H.; Gout, E.; Gerbeau, P.; Luu, D.-T.; Bligny, R.; Maure, C. *Nature* **2003**, *425*, 393–397.
- (7) Wiley, D. C.; Skehel, J. J. *Annu. Rev. Biochem.* **1987**, *56*, 365–394.
- (8) Popov, K.; Ronkkomaki, H.; Lajunen, L. H. J. *Pure Appl. Chem.* **2006**, *78*, 663–675.
- (9) Allen, R. I.; Box, K. J.; Comer, J. E. A.; Peake, C.; Tam, K. Y. *J. Pharm. Biomed. Anal.* **1998**, *17*, 669–741.
- (10) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225.
- (11) Ullmann, G. M.; Knapp, E.-W. *Eur. Biophys. J.* **1999**, *28*, 533–551.

- (12) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
- (13) Lolis, E.; Petsko, G. A. *Biochemistry* **1990**, *29*, 6619–6625.
- (14) Kozlov, A. G.; Lohman, T. M. *Proteins: Struct. Funct. Bioinf.* **2000**, *41*, 8–22.
- (15) Tamm, L. K.; Han, X. *Biosci. Rep.* **2000**, *20*, 501–518.
- (16) Cohen, F. S.; Melikyan, G. B. *Nat. Struct. Biol.* **2001**, *8*, 653–655.
- (17) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (18) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
- (19) Day, T. J. F.; Soudackov, A. V.; Cuma, M.; Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **2002**, *117*, 5839–5849.
- (20) Brancato, G.; Tuckerman, M. E. *J. Chem. Phys.* **2005**, *122*, 224507.
- (21) Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A. *J. Phys. Chem. A* **2006**, *110*, 631–639.
- (22) Lill, M. A.; Helms, V. *J. Chem. Phys.* **2001**, *114*, 1125–1132.
- (23) Cymes, G. D.; Ni, Y.; Grosman, C. *Nature* **2005**, *438*, 975–980.
- (24) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 1–7.
- (25) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. *J. Chem. Phys.* **2002**, *117*, 4184–4200.
- (26) Dlugosz, M.; Antosiewicz, M. *Chem. Phys.* **2004**, *302*, 161–170.
- (27) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. *Proteins: Struct. Funct. Bioinf.* **2002**, *47*, 469–480.
- (28) Börjesson, U.; Hünenberger, P. H. *J. Chem. Phys.* **2001**, *114*, 9706–9719.
- (29) Baptista, A. M. *J. Chem. Phys.* **2002**, *116*, 7766–7768.
- (30) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Nola, A. D. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (31) Mertz, J. E.; Pettitt, B. M. *Int. J. Supercomput. Appl. High Perform. Eng.* **1994**, *8*, 47–53.
- (32) Kong, X.; Brooks III, C. L. *J. Chem. Phys.* **1996**, *105*, 2414–2423.
- (33) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*.
- (34) Lee, M. S.; Salsbury, J. F. R.; Brooks III, C. L. *Proteins: Struct. Funct. Bioinf.* **2004**, *56*, 738–752.
- (35) Khandogin, J.; Brooks III, C. L. *Biophys. J.* **2005**, *89*, 141–157.
- (36) Wallace, J. A.; Shen, J. K. *Methods Enzymol.* **2009**, *466*, 455–475.
- (37) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (38) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (39) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. J. *Comput. Chem.* **2005**, *26*, 1701–1718.
- (40) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (41) van der Vegt, N. F. A.; Briels, W. J. *J. Chem. Phys.* **1998**, *109*, 7578–7582.
- (42) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- (43) Warshel, A.; Sussman, F.; King, G. *Biophysics* **1985**, *25*, 8368–8372.
- (44) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (45) Hwang, J.-K.; Warshel, A. *J. Am. Chem. Soc.* **1987**, *109*, 715–720.
- (46) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430–437.
- (47) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: GROMOS96 Manual and User Guide*; BIOMOS b.v.: Zürich, Switzerland, and Groningen, The Netherlands, 1996.
- (48) Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413–3419.
- (49) Klingen, A. R.; Bombarda, E.; Ullmann, G. M. *Photochem. Photobiol. Sci.* **2006**, *5*, 588–596.
- (50) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (51) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel: Dordrecht, The Netherlands, 1981; pp 331–342.
- (52) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (53) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8592.
- (54) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (55) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (56) Jorgensen, W. L.; Maxwell, S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (57) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (58) Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.
- (59) Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem.* **1996**, *100*, 1206–1215.
- (60) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285–297.
- (61) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. M. *Protein Sci.* **2006**, *15*, 1214–1218.
- (62) Tanokura, M. *Biochim. Biophys. Acta* **1983**, *742*, 576–585.
- (63) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461–466.
- (64) Hess, B. *J. Chem. Phys.* **2002**, *116*, 209–217.
- (65) Xie, Y.; Inoue, T.; Miyamoto, Y.; Matsumura, H.; Kataoka, K.; Yamaguchi, K.; Nojima, M.; Suzuki, S.; Kai, Y. *J. Biochem.* **2005**, *137*, 455–461.
- (66) Mehle, A.; Thomas, E. R.; Rajendran, K. S.; Gabuzda, D. *J. Biol. Chem.* **2006**, *281*, 17259–17265.
- (67) Baptista, A. M.; Soares, C. M. *J. Phys. Chem. B* **2001**, *105*, 293–309.
- (68) Damjanovic, A.; Brooks, B. R.; Garcia-Moreno, B. *J. Phys. Chem. A* **2011**, *115*, 4042–4053.