PROTEINS: Structure, Function, and Bioinformatics 56:738–752 (2004)

Constant-pH Molecular Dynamics Using Continuous Titration Coordinates

Michael S. Lee,[1] Freddie R. Salsbury, Jr.,[2] and Charles L. Brooks, III[3]*
[1]*Department of Cell Biology and Biochemistry, U.S. Army Medical Research Institute of Invectious Diseases, Frederick, Maryland*
[2]*Department of Physics, Wake Forest University, Winston-Salem, North Carolina*
[3]*Department of Molecular Biology (TPC6), The Scripps Research Institute, La Jolla, California*

ABSTRACT **In this work, we explore the question of whether pK_a calculations based on a microscopic description of the protein and a macroscopic description of the solvent can be implemented to examine conformationally dependent proton shifts in proteins. To this end, we introduce a new method for performing constant-pH molecular dynamics (PHMD) simulations utilizing the generalized Born implicit solvent model. This approach employs an extended Hamiltonian in which continuous titration coordinates propagate simultaneously with the atomic motions of the system. The values adopted by these coordinates are modulated by potentials of mean force of isolated titratable model groups and the pH to control the proton occupation at particular sites in the polypeptide. Our results for four different proteins yield an absolute average error of ~1.6 pK units, and point to the role that thermally driven relaxation of the protein environment in the vicinity of titrating groups plays in modulating the local pK_a, thereby influencing the observed $pK_{1/2}$ values. While the accuracy of our method is not yet equivalent to methods that obtain $pK_{1/2}$ values through the *ad hoc* scaling of electrostatics, the present approach and constant pH methods in general provide a useful framework for studying pH-dependent phenomena. Further work to improve our model to approach quantitative agreement with experiment is outlined. Proteins 2004;56:738–752.**

Key words: pK_a; proteins; ovomucoid; BPTI; RNase A; conformational change; free energy

INTRODUCTION

The pH of an aqueous protein solution and the consequent protonation states of the protein constituents are very important factors in the structure and function of proteins. For example, most biologically relevant proteins achieve their native fold in a narrow pH range near 7, and are unstable or non-native outside this range. In addition, the functionalities of active sites in many enzymes are sensitive to the protonation states of nearby residues. Likewise, accurate prediction of protonation states at physiological pH is integral in calculating accurate binding affinities of ligands to protein targets. Furthermore, pH modification is an important experimental tool in studying the folding pathways of small peptides and proteins.[1]

Despite the importance of pH in biological equilibria, simulation techniques for biomolecules generally consider pH in a relatively primitive fashion. Fixed protonation states are typically employed for each titratable group in the system. This can be particularly problematic for histidine residues, where it is not always clear which of the two imidazole nitrogens should be protonated at physiological pH. Also, simulations outside the pH = 7 range are very difficult to prepare because of uncertainty in the pK_a shifts from standard model reference pK_a values. Moreover, a fixed titration state is a severe approximation in systems where pH and conformational changes are coupled as in the β-amyloid peptide,[2] the prion protein,[3] or other small peptides.[4–6]

In the last decade, several techniques, which we term simple continuum models, have been developed to determine the pK_a shifts in proteins.[7–9] The general idea embodied in these calculations is to use an implicit water model, such as continuum electrostatic theory or a Langevin dipole water model, to estimate the pK_a shifts of a static structure due to electrostatic forces in the protein environment. These calculations were originally based on a single static X-ray structure.[7] However, it is becoming common practice to average the pK_a shifts over multiple conformations, using either several snapshots obtained by molecular dynamics (MD) runs or an ensemble of NMR structures.[10–12] With suitable (*ad hoc*) adjustments of the dielectric constant of the protein, these methodologies may achieve pK_a values on average less than 1 pK unit in error of experimental results.

Unfortunately, a number of difficulties and ambiguities plague the application of simple continuum models. First, an adjustable parameter known as the solute dielectric must be manually selected. This parameter essentially determines the magnitude of the pK_a shifts and the

Grant sponsor: National Institutes of Health; Grant numbers: GM57513 and GM37554 (to C. L. Brooks, III), and GM20334 (to F. R. Salsbury).

*Correspondence to: Charles L. Brooks, III, Department of Molecular Biology (TPC6), The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037. E-mail: brooks@scripps.edu

Received 21 May 2003; Accepted 4 January 2004

Published online 11 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20128

strength of the electrostatic interactions of a titratable group with its environment. To determine a macroscopic dielectric constant for a desired system without fitting to the experimental results requires a significant computational cost.[13] However, this quantity does not[9,14] necessarily have the same meaning as the optimal dielectric constant to use for calculating pK_a values. In addition, it is known that in a protein, the dielectric constant is a spatially varying function. The dielectric of different proteins and at various positions in the protein can range from ~10 to ~40.[9,13,15,16] Thus, choosing a single dielectric constant to use in calculations of pK_a shifts for all titratable groups in a protein is problematic.[14] Furthermore, the standard protocol is to assign the solute dielectric of each model compound to be the same as the protein. This procedure seems to be physically incorrect, since it is well known that the solvation energies of small compounds match experiment quite well using continuum theory with a dielectric value of one.[17] Yet another issue specific to MD- or Monte Carlo–based configurational averaging techniques is that in most methods, the protein configurations are biased toward the fixed protonation states that are selected in the simulation runs. Studies employing the linear response approximation (LRA) address this issue by configurational averaging in both the charged and uncharged states of a titratable group for calculation of the intrinsic pK_a.[9,18,19]

The primary purpose of this work is to revisit the question of whether a method that couples conformational dynamics to titration events and utilizes a macroscopic solvent with a dielectric constant equal to unity can provide accurate predictions of pK_a values in proteins. Coupling of conformational dynamics to titration events refers to weighting protein configurations based on the energetic favorability of certain protonation states. It also implies weighting protonation states by the probabilistic occurrence of certain conformational states. This physics should, in principle, be sufficient to account for microscopic relaxation of the protein atoms around each titratable group and hence obviate the need for a nonunity dielectric constant. Warshel and coworkers first showed that it is indeed possible to obtain reasonable estimates of pK_a values using LRA with either a Langevin dipole solvent model with $\epsilon_{solute} = 1$[9,18,19] or an all-atom solvent model.[9,19,20] Other methods, however, that couple protonation and conformation have been developed and reported in the last few years.[21–24] Aiming for a similar objective, almost a decade ago, Mertz and Pettitt[25] proposed a simple constant-pH method based on an open system Hamiltonian. More recently, Borjesson and Hunenberger[24] introduced a method by which titrations of groups are linked to a proton bath via a continuous titration parameter.

Two methods worthy of note are the ones of Bürgi et al.[21] and Baptista et al.[23] Bürgi et al.[21] developed a hybrid Monte Carlo/MD scheme in which titration events occur based on a simple Metropolis criterion. This stochastic switching criterion is derived from an estimate of the free energy difference between the protonated and unprotonated state. The free energy difference is approximated by a short run of thermodynamic integration, ~40 ps, for a single residue at a time. The main drawback of this approach is that one acceptance/rejection cycle for each of the titratable residues for a reasonably sized protein would, in total, take on the order of a nanosecond of simulation time. Nonetheless, the 40 ps thermodynamic integration period is essential to allow for rearrangement and relaxation of the protein and water molecules.

Contrasting the free energy approach of Bürgi et al.,[21] Baptista et al.[23] have developed a procedure in which an explicit water MD simulation is interrupted periodically by a continuum-based static Monte Carlo pK_a calculation. A random selection of one of the lowest energy protonation states in that calculation is then used in the next MD simulation cycle. This approach also avoids the issue of mixed titration states, since only endpoint states are considered. Nonetheless, sudden changes in the protonation state of a system will introduce discontinuous impulsive "shocks" in a MD simulation. This phenomenon can lead to spikes in the dynamical forces and possible numerical instabilities. Another problem is that the acceptance/rejection of a titration state is based upon an instantaneous protein conformation. Although this fact does not affect the ability to sample a proper distribution of conformations and titration states, it may cause the sampling of protonation states to be inefficient. Finally, the acceptance/rejection criterion is based on a continuum model calculation; thus, it appears one would again have to manually determine an appropriate value for the dielectric constant of the protein.

In this work, we develop a new constant pH molecular dynamics (PHMD) technique that incorporates many of the salient features of the aforementioned coupling approaches. The basis behind our method is a set of continuous titration coordinates that describes transitions between protonated and unprotonated groups. The continuous titration coordinate allows the protein to experience for short periods of time mixed states that can improve the likelihood of a full titration event for a group. For the purpose of obtaining equilibrium thermodynamic quantities, the titration coordinate can have an arbitrary meaning so long as the endpoints are the desired protonated and unprotonated states. For our approach, in particular, a simple linear interpolation of the charges and van der Waals (vdW) interactions is employed. This interpolation procedure is similar to the methods of Borjesson and Hunenberger[24] and Baptista et al.[22] Nonetheless, our definition of a continuous titration coordinate is fundamentally different, because it represents an instantaneous microstate rather than a fractional protonation population.

The most physically realistic simulation of titration events would involve modeling the exchange of a proton with surrounding water molecules. Indeed, Warshel and coworkers[26,27] have demonstrated that the simulation of explicit proton transfer between titratable groups and water molecules is tractable using a highly parameterized empirical valence bond (EVB) framework.[28] However, such an approach is infeasible for more than specialized

demonstration calculations. Since we restrict ourselves to a macroscopic description of the solvent in this work, it is unclear how to model proton transfer events explicitly.

As in all classical pH methods, the free energy of titrating reference model groups must be determined, since classical methods neglect two key energy terms of quantum chemical origin: the free energy of solvation of a proton in water and the free energy of formation of the proton–oxygen/proton–nitrogen bond. In our approach, model free energies are obtained by thermodynamic integration along the titration coordinate prior to the full system simulation. The model free energies only need to be calibrated once per force field.

There are many parallels between our new methodology and the λ-dynamics method of Kong and Brooks.[29] In λ-dynamics, a set of continuous variables modulates between two physical endpoints, while the intermediate states have no real physical basis. The λ variables propagate along with the atomic coordinates in a system described by an extended Hamiltonian. With this framework, it is possible to introduce biasing potentials in the λ coordinate to allow for dynamical transitions between the endpoints. It is also possible to introduce multiple competing ligands in a protein pocket. In this case, the λ variables directly measure the relative affinities of the ligands.[30] We borrow a number of the ideas from λ-dynamics in our new approach. In particular, as we describe below, we introduce biasing barriers between the physical endpoint states to enhance the residency time of these protonic configurations. We note that when we speak of proton residence times below, these do not have a physical meaning in the context of experimental properties but are a property we manipulate to expedite convergence of thermodynamically related properties.

In the next section, we introduce the theory behind our new constant PHMD method. We then apply this technique in the Results section to several peptides and proteins. Comparisons between computed $pK_{1/2}$ values and experimental $pK_{1/2}$ values are made. Difficulties that emerge in the current methodology are discussed in a final section, and possible means to address them are described.

THEORY

The primary objective of our approach is to run classical MD simulations on a protein under conditions of a particular (or changing) pH. To achieve an equilibrium description of the system at a given pH, a statistical mechanical model, where multiple groups in the protein titrate back and forth between protonated and unprotonated states in response to the specified pH and the protein environment, is developed. Consider the chemical process for proton association/dissociation of a single titratable group, A, in a protein with an unknown protonation free energy, ΔG_{exp} (protein):

$$AH_{(protein)} \leftrightarrow A^{-}_{(protein)} + H^{+}_{(aq)}. \quad (1)$$

This process cannot be directly described by a classical force field approach for two reasons. First, the quantum mechanical energy of breaking/forming the proton–oxygen or proton–nitrogen bond is not included in classical force fields. Furthermore, the solvation of a proton in water cannot be easily treated, because it is not clear from experiment and theory as to the exact value of the free energy of proton solvation.

These two problems can be addressed through the introduction a model compound, $A_{(aq)}$, which has the same titratable group as $A_{(protein)}$, yet has a known experimental pK_a^{exp}. The proton association/dissociation reaction of the model compound,

$$AH_{(aq)} \leftrightarrow A^{-}_{aq} + H^{+}_{aq}, \quad (2)$$

has a protonation free energy defined by

$$\Delta G_{\text{exp}}(\text{model}) = -\log(10) k_B T[pK_a^{\text{exp}} - \text{pH}], \quad (3)$$

where k_B is Boltzmann's constant and T is the temperature. Both of the reactions in Eqs. (1) and (2) can be simulated classically in an approximate yet identical fashion, such that the difference in calculated free energies is equal to the difference in experimental free energies:

$$\Delta G_{\text{exp}}(\text{protein}) - \Delta G_{\text{exp}}(\text{model}) = \Delta G_{class}(\text{protein}) - \Delta G_{class}(\text{model}), \quad (4)$$

which leads ultimately to an estimate of the experimental free energy of protonation for a particular site in the protein:

$$\Delta G_{\text{exp}}(\text{protein}) = \Delta G_{class}(\text{protein}) - \Delta G_{class}(\text{model}) + \Delta G_{\text{exp}}(\text{model}). \quad (5)$$

Using Eq. (5) as a guide, titratable groups in proteins are viewed simply as model compounds that are perturbed by the protein environment through mainly nonbonded interactions. This idea is key to simple dielectric pK_a methods, where pK_a shifts from the model are assumed to be composed of energetics differences between the model and protein.[7,31] Our approach utilizes the concept of a model compound in a similar fashion. Because an empirical force field Hamiltonian of a system is insufficient to describe the changes in protonation states of titratable groups with respect to pH, a biasing potential is introduced. This biasing potential has two components. The first component, analogous to the term $-\Delta G_{\text{class}}$ (model) in Eq. (5), is a potential of mean force (PMF), which is the negative of the free energy necessary to deprotonate a model compound. The second component of the biasing potential, parallel to ΔG_{exp} (model), biases the probability of titration events such that in the limit of simulating the model compound of a single titratable group in isolation, a standard titration curve can in principle be elicited.

In order to classically simulate the titratable groups, we first define a continuous titration coordinate, λ_i, which is bounded between 0 and 1, to govern the progress of protonation/deprotonation of a group labeled i. The $\lambda_i = 0$ condition corresponds to a completely protonated state, and the $\lambda_i = 1$ condition corresponds to a completely unprotonated state. Intermediate values of λ_i correspond to a mixing process (described below) that develops along

the pathway from protonation to deprotonation. In order to restrict each titration coordinate, λ_i, to the range [0,1], it is convenient to define λ_i as an implicit function of an underlying coordinate, θ_i, through the formula,

$$\lambda_i = \sin^2(\theta_i). \quad (6)$$

The coordinate, θ_i, can have unrestricted values and is ultimately the variable that is propagated in a dynamics simulation. With unrestricted coordinates, constructs such as Lagrange multipliers, wall potentials, and normalizations as used in λ-dynamics are unnecessary.[30]

In practice, when the titration coordinates are allowed to propagate dynamically, the precise endpoints $\lambda_i = 0$ and $\lambda_i = 1$ are rarely sampled. Therefore, three approximate macrostates are defined: protonated $(P) \equiv (0 \le \lambda_i < \lambda_P)$, unprotonated $(U) \equiv (1 \ge \lambda_i > 1 - \lambda_P)$, and mixed $(M) \equiv (1 - \lambda_P \ge \lambda_i > \lambda_P)$, where λ_P is an empirically adjustable parameter between 0 and 0.5. In our studies, we have arbitrarily selected a value of $\lambda_P = 0.1$ as a compromise between restricting the protonated and unprotonated states to a narrow range and allowing for sufficient statistical sampling. Given these definitions, the fractional population of unprotonated states, S_i^{unprot}, can be defined as

$$S_i^{unprot} = \frac{\rho_i^U}{\rho_i^P + \rho_i^U} = \frac{1}{1 + \rho_i^P/\rho_i^U} = \frac{1}{1 + \dfrac{\rho(\lambda_i < \lambda_P)}{\rho(\lambda_i > 1 - \lambda_P)}} \approx \frac{1}{1 + \dfrac{N(\lambda_i < \lambda_P)}{N(\lambda_i > 1 - \lambda_P)}}, \quad (7)$$

where ρ is the probability of finding a particular condition, and N is the number of times a condition is observed in an actual simulation. The mixed regime is not included in Eq. (7), because we do not assume that is has any inherent physical meaning. Nonetheless, the mixed regime is a critical pathway for simulated transitions to occur. These definitions should not be confused with the designation of titration coordinates in Borjesson and Hunenberger[24] and Baptista et al.[22] In those works, the titration coordinate directly represents titration probabilities; hence, all values between 0 and 1 have a physical basis.

Equating the free energy difference between the P and U regimes with their thermodynamic probabilities,

$$\exp\left(\frac{\Delta G_i^{P\to U}}{kT}\right) = \rho_i^P/\rho_i^U, \quad (8)$$

and then substituting Eq. (8) into Eq. (7) and noting that in the case of an isolated ideal model compound,

$$\Delta G_i^{P\to U} = \log(10) k_B T[\mathrm{pK}_a^i - \mathrm{pH}], \quad (9)$$

we obtain the Henderson–Hasselbach (HH) relation,[24,32]

$$S_i^{unprot}(\mathrm{pH}) = \frac{1}{1 + 10^{(\mathrm{pK}_a^i - \mathrm{pH})}}. \quad (10)$$

This relation provides an empirical measure of the ability of our constant pH scheme to produce a physically reasonable titration observable at least for the limiting case of an isolated model compound. Of course, for titratable groups in proteins, deviations from Eq. (10) are expected, since multiple groups are often physically coupled.

TABLE I. Charges for Protonated and Unprotonated States of All Titratable Groups Used

Name	Atom	Protonated	Unprotonated
Asp	CB	−0.21	−0.28
	CG	0.75	0.62
	OD1	−0.55	−0.76
	OD2	−0.61	−0.76
	HD2*	0.44	0.00
Glu	CG	−0.21	−0.28
	CD	0.75	0.62
	OE1	−0.55	−0.76
	OE2	−0.61	−0.76
	HE2*	0.44	0.00
Lys	CE	0.24	0.40
	HE1	0.08	−0.05
	HE2	0.08	−0.05
	NZ	−0.24	−0.98
	HZ1	0.28	0.34
	HZ2	0.28	0.34
	HZ3*	0.28	0.00
Tyr	CZ	0.11	−0.29
	OH	−0.54	−0.71
	HH*	0.43	0.00
His-δ	ND1	−0.51	−1.00
	HD1*	0.44	0.00
	CE1	0.32	0.30
	HD2	0.13	0.10
His-ε	NE2	−0.51	−1.00
	HE2*	0.44	0.00
	CD2	0.19	0.15
	HE1	0.18	0.13
C-ter	C	0.72	0.34
	OT1	−0.55	−0.67
	OT2	−0.61	−0.67
	HC2*	0.44	0.00
N-ter	N	−0.30	−0.97
	HT3*	0.33	0.00

All charges were derived from the PARAM22 force field[34] except for unprotonated lysine, tyrosine, and N-terminus and protonated C-terminus (see text for details).
*Indicates which proton disappears in a linear fashion from the vdW potential as the unprotonated state is approached. These charges were designed to be used with the PARAM22 force field, such that the net charge of each residue at the titration endpoints is an integer.

Part of the reason that we do not assume physical meaning for the mixed macrostate is that there is an arbitrary choice in how to transform between a protonated and unprotonated group. In our method, two mixing processes occur. One process is the linear interpolation of charge states[22,29] of the titratable group, i,

$$q_a = \lambda_i q_a^U + (1 - \lambda_i) q_a^P, a \in i, \quad (11)$$

where a indexes over all of the atoms in the titratable group, i, and q_a^U and q_a^P are the charges of these atoms at the unprotonated and protonated endpoints, respectively (see Table I). The other process is the linear attenuation of

the nonbonded vdW interaction[29] between the proton of the titratable group and the surrounding protein environment as the state proceeds from protonated to unprotonated. The modified vdW potential, U'_{vdW}, for the interactions of the proton with other atoms is

$$U'_{vdW} = \begin{cases} (1-\lambda_i)U_{vdW}(i,p) \\ (1-\lambda_i)(1-\lambda_j)U_{vdW}(i,j) \end{cases}, \quad (12)$$

where i and j are titrating protons, and p indexes all other atom types, and U_{vdW} is the standard classical 6,12 Lennard–Jones potential as used in CHARMM.[33] In our model, we do not take into account changes in the internal energy potentials between protonation states. Furthermore, we maintain the mass of the proton even in the unprotonated state. In addition, we do not allow the atomic radii of the titrating species to change between the protonated and unprotonated endpoints. This is perhaps a shortcoming, since the change in charge states between the protonated and unprotonated forms may lead to slightly different vdW radii for some of the atoms. For instance, the negative charge on the unprotonated carboxylate ion leads to slightly larger vdW radii for the oxygen atoms (1.77 Å) versus the protonated form (1.7 Å) in the PARAM22 force field.[34] Nonetheless, the protons in this force field are completely embedded inside heavy atoms. Thus, the dielectric boundary should not change significantly based on the presence or absence of protons. Overall, the nonproton vdW radii issues may affect the vdW potential, the dielectric boundary in the generalized Born (GB) implicit solvent model, and the surface area for the hydrophobic free energy term.

As stated above, an essential feature of our method is a biasing potential consisting of the negative of the model potential and a pH-dependent component. The model potential can be thought of as a way to correct for the arbitrary classical description of the titration pathway and endpoints. In the case of simulating an isolated model compound with its own PMF subtracted, the resultant PMF should be flat. Given a flat PMF, one can then build in a pH-dependence for the model compound that approximates the desired characteristics [e.g., Eq. (10)].

The model potential, U^i_{model}, is a PMF along the titration coordinate, λ_i, of an isolated model compound embedded in solvent. The potential is obtained in this work via thermodynamic integration,

$$U^i_{\text{model}}(\lambda) = \int_0^{\lambda} \left\langle \frac{\partial U_{system}(\lambda')}{\partial \lambda'} \right\rangle_{\lambda'} d\lambda'. \quad (13)$$

Because of our particular choice in describing the titration pathway and the fact that we utilize an implicit solvent, the PMF of a model compound is essentially quadratic in nature with respect to λ and, hence, can be fit to a two-parameter parabolic function of the form,

$$U^i_{\text{model}}(\lambda_i) = -A_i(\lambda_i - B_i)^2 = -A_i(\sin^2(\theta_i) - B_i)^2. \quad (14)$$

Given the PMF, the classical free energy, ΔG_{class} (model), of protonating the model compound can be defined as

$$\Delta G^i_{\text{class}}(\text{model}) = U^i_{\text{model}}(\lambda = 0) - U^i_{\text{model}}(\lambda = 1) = -A_i(2B_i - 1). \quad (15)$$

The other component of the biasing potential imparts pH dependence on the simulation. This component is derived by analogy to Eq. (3) and assumes that the chemical potential of adding a fractional proton to the aqueous environment is linear with respect to λ_i,

$$U^i_{\text{pH}}(\lambda_i) = \log(10)k_BT \times \lambda_i \times [\text{pK}^{ref}_{a,1} - \text{pH}]. \quad (16)$$

If one were to simulate an isolated model compound at pH = pK^{ref}_a, where Eq. (14) and the negative of Eq. (14) are added to a classical force field potential, converged thermodynamic sampling would lead to two results. First, the free energy difference between the two titration endpoints would be zero. Furthermore, for all points along the titration coordinate, the free energy would be constant. At pH $\neq$ pK^{ref}_a, Eq. (16) can be used to reproduce the HH titration curve for an isolated model compound. We demonstrate this in the Results section.

A practical consideration in our model is that the residency time in the protonated or unprotonated basin should be sufficiently long to allow for the protein conformation to relax. This criterion can be justified in part by the experimental observation that there exist minima and transition barriers associated with proton transfer between titratable groups and water molecules.[26] Furthermore, sampling of the protonated and unprotonated macrostates should be maximized to acquire converged thermodynamic properties in a reasonable amount of simulation time. Finally, it is desired that interacting titratable groups should see each other most of the time in their physically meaningful protonated/unprotonated macrostates. One way to impose these conditions is to introduce an energy barrier at the center of the titration path,

$$U^i_{barr}(\lambda_i) = -4\beta\left(\lambda_i - \frac{1}{2}\right)^2, \quad (17)$$

where the parameter β determines the maximal height of the barrier. A value for β of 1.25 kcal/mol was chosen for our work, because it provides a reasonable compromise between protonated/unprotonated residency time and sampling efficiency (see Results section). Because Eq. (17) is symmetric in λ_i, it lowers both endpoints equally and therefore has no effect on the relative energy of the endpoints. Although there is a physical precedent for employing a barrier term, the endpoint residency times that we incur via Eq. (17) and $\beta = 1.25$ are not expected to correlate with experimental kinetic data.

Given the constructs developed above, the complete potential, U_{total}, in our approach involving n titratable groups can now be elaborated,

$$U_{\text{total}} = U_{\text{internal}} + U_{\text{vdW}}(\lambda[\theta]) + U_{\text{elec}}(\lambda[\theta]) + U_{\text{GB}}(\lambda[\theta]) + U_{\text{nonpolar}} + \sum_{i=1}^{n} [- U^i_{\text{model}}(\lambda_i[\theta_i]) + U^i_{\text{pH}}(\lambda_i[\theta_i]) + U^i_{\text{barr}}(\lambda_i[\theta_i])], \quad (18)$$

where λ and θ are the composite vectors of the n titration coordinates. The U_{internal} term involves the bond, angle, and torsional energy terms of a classical force field, which in this work are not coupled to the titration coordinates. The potential U_{elec} is the classical Coulombic energy,

$$U_{\text{elec}} = k \sum_{ab} \frac{q_a q_b}{r_{a<b}}, \quad (19)$$

where a and b are atomic indices, $k \sim 332$ (Å kcal)/[mol (atomic charge)^2], q_a is the charge on atom a, and r_{ab} is the internuclear distance between atoms a and b. The function U_{GB} is a slight modification of the GB solvent potential of Still et al.,[35]

$$U_{GB} = -\frac{1}{2} k \left(\frac{1}{\varepsilon_{solute}} - \frac{1}{\varepsilon_{solv}}\right) \sum_{ab} \frac{q_a q_b}{\sqrt{r_{ab}^2 + \alpha_a \alpha_b \exp(-r_{ab}^2/K\alpha_a\alpha_b)}}, \quad (20)$$

where K is assigned a value of 8,[36] which provides improved correspondence to Poisson results compared to the original value of 4.[35] The Born radii, α_a, are obtained analytically via the GB molecular volume (GBMV2) method.[36] The precise expression for Born radii is somewhat involved for this discussion and can be found elsewhere.[36] In all cases, the charges, q_i, are implicit functions of the λ vector as specified in Eq. (11). In this study, the solvent dielectric, ϵ_{solv}, was set to 80, and the solute dielectric, ϵ_{solute}, was set to 1. Exceptions to this are the two *static* pK_a calculations, which permitted direct comparison to Poisson-based methods where ϵ_{solute} was set to 20 (see Results section). As mentioned in the Introduction, our hypothesis is that proper sampling over solute configurations, even with a continuum solvent representation, will alleviate the need to introduce an arbitrary scaling parameter, $\epsilon_{solute} > 1$.[18] The complete the description of our solvent model, a nonpolar surface area term, $U_{nonpolar} = \sigma S_A$, is also employed, where $\sigma = 20$ cal/(mol-Å^2), and S_A is the solvent-accessible surface area with a probe radius of 1.4 Å calculated using the algorithm specified in Lee et al.[36] However, the nonpolar term is not a function of the titration coordinate. Finally, we note that although we use a continuum, "implicit" solvent in the present calculations, the formalism described above is applicable to detailed atomic solvent models as well.

Essential to our constant PHMD approach is to use an extended Hamiltonian-like formalism to treat the underlying titration coordinates, θ_i, as fictitious particles with mass, M_i. We introduce a kinetic energy term for the titration coordinates,

$$T_{\text{PHMD}} = \frac{1}{2} \sum_i M_i \dot{\theta}_i^2, \quad (21)$$

that is analogous to the kinetic energy term used in the λ-dynamics approach of Kong and Brooks.[29] The value of M_i governs the responsiveness of the titration coordinate to the dynamical forces of the system. A small mass allows for fast propagation of titration processes, whereas a large mass will produce slower transitions. A mass larger than 2 amu appears to be necessary to guarantee adequate numerical precision in the integration of the dynamical simulation when a conventional timestep of 2 fs is used. We have chosen all M_i to be 10 amu as a compromise between the responsiveness of the titration coordinate and the precision of the numerical integration of the equations of motion.

METHODS

Model Compounds and Model Potential Determination

Model compounds in this work are the blocked form of the amino acids obtained by acetylating the N-terminus (ACE), and amidating the C-terminus (CT2). For the models of the titratable N-termini, an alanine was appended to the first residue, NH_3^+-*X*-A-CT2, because there were insufficient force field parameters in PARAM22 for the molecule, NH_3^+-*X*-CT2, where *X* is any amino acid. The PMF, U^i_{model}, for each model compound was determined via thermodynamic integration. Since the analytical form of our target function is given in Eq. (14), the coefficients, A_i and B_i, were obtained by fitting the derivative of the model function with respect to θ,

$$\frac{\partial U^i_{\text{model}}(\lambda_i)}{\partial \theta_i} = \frac{\partial[-A_i(\sin^2\theta_i - B_i)^2]}{\partial \theta_i} = 2 - A_i \sin(2\theta_i)(\sin^2\theta_i - B_i), \quad (22)$$

to ensemble-averaged derivatives, $\left\langle \frac{\partial E}{\partial \theta_i} \right\rangle$, which were obtained by MD simulations at fixed points along the θ coordinate. For our study, model compound simulations of 500 ps (100 ps equilibration/400 ps production) were performed at each of the following seven values of θ, which evenly cover the range between 0 and π/2: 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, and 1.4. It was determined that the C- and N-termini for each amino acid needed to have distinct model functions. The justification for this decision is that the 1-3 and 1-4 exclusions of the force field influence the electrostatic interactions between the titrating terminal groups and the side-chains of their residue.[37] One of the most notable examples of this issue can be seen in the comparison of the alanine and valine C-termini. The carbon atom of the titrating carboxylate group in alanine has no electrostatic interaction with any of the atoms in the methyl side-chain. In contrast, the same C atom in valine electrostatically interacts with all six of the hydrogens bonded to the C_γ's, which have a net total charge of +0.54. This disparity and others analogous to it lead to a large difference in the calculated classical free energy of protonation of the C-termini between alanine and valine, even though the residues are chemically very similar.

The method outlined in this work has the limitation that an atom can be involved in only one titratable group. This

restriction precludes certain representations of naturally occurring tautomers. In most cases, rotation about a single bond is sufficient to allow for tautomerism (e.g., aspartic acid and lysine).[38] However, one important exception is the histidine residue, which has two titratable sites: N_δ and N_ϵ. The precise description of histidine requires three competing tautomers: one doubly protonated form and two singly protonated forms. From a simple quantum chemical calculation, it is seen that each state has a unique set of atomic charges on the imidazole ring. Because the two titratable groups involve the same atoms, the proper charge state description of each tautomer is impossible in our model. However, a reasonable approximation that is tractable is to split the imidazole ring into two titratable halves: (N_δ-H, C_ϵ-H) and (N_ϵ-H, C_δ-H). A remaining issue, which is independent of the choice of classical charge description, is avoidance of double deprotonation, an occurrence known not to occur at biologically relevant pH values. A simple approach, which is employed by others,[38] is to introduce a strong energy penalty for the double deprotonation occurrence. In our method, an energy term, $K\lambda_1\lambda_2$, is added to the potential, where K has been set to 50 kcal/mol. As both λ's approach one (i.e., double deprotonation), the penalty takes full effect.

Table I illustrates the charge models used for all of the titratable groups in this work. Most of the atomic charges were derived from the PARAM22 force field. However, not all titration states exist in PARAM22. The missing states are unprotonated lysine, tyrosine, and N-terminus and protonated C-terminus. For unprotonated lysine and tyrosine, we extracted electrostatic potential-derived (ESP) charges[39] from HF/6-31G* calculations using the GAUSSIAN98 software.[40] For the unprotonated N-terminus, the remaining negative charge after removing the proton was placed on the nitrogen. For the protonated C-terminus, charges values and internal coordinate parameters were replicated from the analogous carboxylate side-chain of the Asp residue.

Simulation Protocols

Our approach has been implemented as a module (PHMD) in a development version of CHARMM.[33] In order to run the module, a formatted input file is required, which consists of sections for each model compound. Each section specifies a name, its pK_a^{ref}, the two parameters, A and B, charge states for the protonated/unprotonated forms, and labeling of the proton that will disappear in the vdW potential. Given this input file, the PHMD module identifies all of the matching groups in the system. It is also possible for the user to select only a subset of groups to titrate.

All of the simulations described in this work were performed with the all-atom PARAM22 force field[34] and GBMV2 implicit solvent model.[36] The GBMV2 model was parameterized to obtain finite-difference Poisson continuum electrostatic solvation energies with an average of ~1% error. To improve computational speed, Born radii were updated every other timestep. This adjustment does not appear to adversely affect simulation results (results not shown). A new peptide backbone potential energy map[41] was also used. This map was derived from high-level ab initio quantum calculations. The SHAKE algorithm was applied to hydrogen bonds and angles, so that a 2 fs timestep could be used. Nonbonded electrostatic and vdW cutoffs for protein simulations were applied with a switching function starting at 18 Å and ending at 20 Å. The seeds for random generation of initial velocities for each individual simulation were selected randomly. Therefore, a repeat simulation, as was done for two of the proteins, provided estimates of sampling errors.

All simulations in this work utilized a Nosé–Hoover[42,43] thermostat to ensure that the atomic velocities achieved a canonical distribution at a constant temperature of 298 K. We have found that it is important as well to thermostat the titration progression variables. Therefore, we couple the θ velocities to a three-mass (30, 50, and 70 amu) Nosé–Hoover chain.[44] This approach assures that the θ velocities also obtain a canonical distribution at the desired temperature of 298 K.

Modifications of the Force Field

One of the key features of the simulations in this work is the use of the GB implicit solvent method. While it has been shown[36] that the particular GB approach we use here is a faithful representation of the Poisson continuum dielectric electrostatic solvation energy,[45] it is not evident that the GB/Poisson implicit solvent model utilizing PARAM22 charges and radii accurately reproduces experimental results pertaining to solvation free energies of small molecular systems.[17] Actually, the PARAM22 force field[34] was originally optimized for use with the *explicit* water force field, TIP3P.[46] Therefore, minor modifications of the vdW radii and/or charges may be necessary to optimize the PARAM22 force field for use with the GB/Poisson model.

It is beyond the scope of this work to completely reparameterize the PARAM22 force field for the purposes of GB-based simulations. Nonetheless, we made one modification of a vdW radius, since it improved $pK_{1/2}$ results dramatically. It was seen that the carboxylate oxygens in the unprotonated forms of Asp, Glu, and the C-terminus were too readily forming hydrogen bonds with other groups in the protein. This occurrence perhaps meant that these oxygens rarely were exposed to the solvent and hence were not in a position to titrate (i.e., accept protons). To ameliorate this situation, we scaled the radii of these carboxylate oxygen atoms by 0.95 to enhance their self-polarization solvation energies. This modification was intended to achieve a better balance between solvent exposure and internal hydrogen bond formation.

The convergence of protonation state sampling may be slow if there are large conformational barriers in the protein; for instance, the carboxylate-proton dihedral (H-O-C-O) for aspartic acid, glutamic acid and the C-terminus has a sizeable barrier to rotation (~2 kcal/mol) between the E and Z conformations. This leads to a very slow switching rate and the need for extensive sampling to yield fully converged populations. Since our aim is to employ

simulation times on the order of 1 ns (at a given pH), we have resolved this particular issue by lowering this dihedral barrier to 0.25 kcal/mol. A drawback to this solution is that unphysical out-of-plane conformations may be stabilized in the protein environment. An alternate solution, not implemented in this work, would be to allow for competing E and Z tautomers.

Barriers can also arise solely from the nature of a particular protein environment. To overcome these barriers, it would be necessary to use advanced sampling techniques, such as replica exchange, which is known to significantly improve sampling efficiency.[47] However, improved sampling methodologies are beyond the scope of the present study.

Protein Data Set

Four proteins comprise the systems examined in this study: turkey ovomucoid (1OMT),[48] bovine trypsin inhibitor (1BPTI),[49] hen egg white lysozyme (193L),[50] and ribonuclease A (7RSA).[51] For the NMR entry, 1OMT, only the first structure was used and the hydrogen atoms were subsequently removed. For all structures, hydrogen atoms were added using the HBUILD facility in CHARMM.[33]

Statistical Analysis of Constant-pH Simulations

The primary objective in this work is to calculate $pK_{1/2}$ values for the titratable groups in a protein. We can estimate these $pK_{1/2}$ values by running simulations at different pH values and fitting the calculated unprotonated fraction, S_i^{unprot} [see Eq. (7)] to a more generalized version of the HH formula[32]:

$$S'_{HH}(\mathrm{pH}) = \frac{1}{1 + 10^{-W(\mathrm{pH}-\mathrm{pK}_{1/2})}}, \qquad (23)$$

where W is a width parameter. The values of W and $pK_{1/2}$ are derived from this fit. For a titratable group that has no interactions with other titratable groups, W has a theoretical value of 1. On the other hand, strongly interacting titratable groups will have values of W much different than 1 and will probably poorly fit to Eq. (23).

Moreover, it is important to determine whether the continuous titration variables remain at the approximate endpoints the majority of the time, such that unphysical mixed states do not interfere with the sampling of protein conformational space. The purity percentage, defined as $Q = 100\%^* (N_{prot} + N_{unprot})/N_{tot}$, where N_{tot} is the total number of snapshots, measures the percentage of time spent at the protonated and unprotonated macrostates. Values of Q around 70–80% at the $pK_{1/2}$ are probably optimal, since some mixed states must be sampled to facilitate switching between the two endpoints. Another important measure is the titration endpoint lifetime, which is approximated as the total simulation time divided by the number of times the titration coordinate, λ, passes above or below the halfway point, 0.5.

Alternative Free Energy Techniques for Calculating $pK_{1/2}$ Values

Besides the PHMD method, there are other means of obtaining $pK_{1/2}$ values from MD simulations. For instance, several researchers have shown that pK_a values can be approximated by free energy techniques such as LRA,[9,18,19] thermodynamic integration (TI),[52] and Gaussian fluctuation analysis.[53] In order to compare our PHMD results to an alternative framework using the same force field and GB potential, we performed TI on single titrating groups. Our TI procedure involved titrating a specific group, i, over 15 windows (θ_i = 0.1, 0.2, …, 1.5) for 60 ps each (10 ps equilibration/50 ps production) at pH = pK_i^{model}. All of the other titrating groups, j, were fixed at the protonation states ascertained by their pK_j^{model}. In the cases that pH = pK_j^{model}, the group j was protonated. The free energy of unprotonating the group and resultant pK_a shift from the model, were calculated by trapezoidal integration over the ensemble-averaged derivatives in each window. The pK_a's obtained by this technique are approximate since site–site interactions are not taken into account.

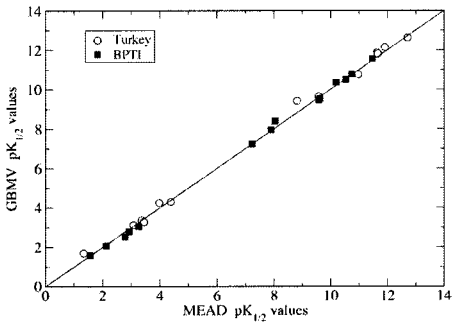

Fig. 1. Comparison of GBMV versus MEAD for calculated $pK_{1/2}$ values of various titratable groups in the static structures of two proteins: turkey ovomucoid (Turkey) and bovine pancreatic trypsin inhibitor (BPTI). Charges and vdW radii were obtained from the PARAM22 force field and Table I. A solid line indicating $y = x$ is displayed for comparison.

RESULTS AND DISCUSSION

Before studying the results of this method, it is important to verify that the GBMV2 model is capable of reproducing the static $pK_{1/2}$ results of the Poisson continuum model. The methods for static $pK_{1/2}$ determination for both continuum and GB approaches are described elsewhere (the MEAD program was used to computer continuum $pK_{1/2}$ values.[7,54] Figure 1 indicates excellent agreement between GBMV2 and MEAD $pK_{1/2}$ values for the two small proteins, 1OMT and 1BPI. The solute dielectric constant for this comparison was set to 20.[8,12]

The model compound parameters obtained by TI are presented in Table II. It is important to note that the calculated free energy of protonation for various termini have wide variation. This is due in large part to the above-mentioned interaction terms associated with the 1-3 and 1-4 exclusions found in most modern force fields.[37]

An initial test case for our PHMD method is the single titration of a blocked aspartic acid: ACE-Asp-CT2. The titration curve for blocked aspartic acid using a titration barrier of 1.25 kcal/mol is shown in Figure 2. The best-fit HH-type curve [Eq. (23)] has a near optimal W value of 1.03. However, the $pK_{1/2}$ value derived from the best fit is 4.3, which is somewhat higher than the model $pK_{1/2}$ value

TABLE II. Parameters for Model Potentials

Group	pK_a (exp)[a]	A[b]	B[b]	ΔG_{class}(model)[c]
Asp	4.0	67.74	0.2042	40.07
Glu	4.4	68.18	0.1741	44.44
Lys	10.4	78.43	0.6788	−28.05
Tyr	9.6	81.15	−0.0195	84.31
His-δ	6.6	84.65	0.3167	31.03
His-ε	7.0	89.17	0.3939	18.92
CT-Ala	3.8	81.27	0.1523	56.51
CT-Val	3.8	80.10	0.2042	47.39
CT-Leu	3.8	80.28	0.1796	51.44
CT-Asp	3.8	81.22	0.1579	55.57
CT-Cys	3.8	81.38	0.1963	49.43
NT-Ala	7.5	90.94	0.9950	−90.03
NT-Ser	7.5	91.51	1.1346	−116.14
NT-Arg	7.5	89.99	1.0363	−96.52
NT-Lys	7.5	89.71	1.0420	−97.25

[a]Values of pK_a (exp) are from the Nozaki and Tanford[65] except for His-δ and His-ε.[66]
[b]The method for calculation of the parameters A and B is described in the text.
[c]The classical free energy of protonating the model compound ΔG_{class}(model) is obtained via Eq. (15).

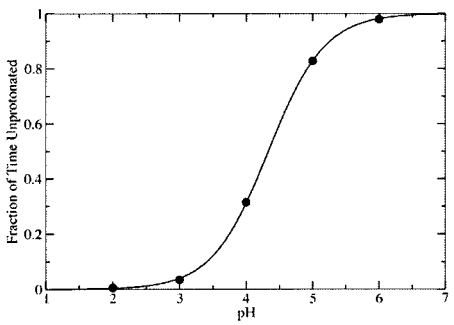

Fig. 2. Titration curve for blocked aspartic acid. Simulations at pH = 2, 3, 4, 5, and 6 were each run for 10 ns (β = 1.25 kcal/mol). Filled circles correspond to actual values obtained for each simulation. The solid curve indicates the optimal HH fit [Eq. (23)], where the calculated $pK_{1/2}$ value is 4.3 and W has a near optimal value of 1.03.

of 4.0. This discrepancy is expected to approach zero as the model and production simulation times approach infinity (i.e., as we achieve complete sampling). Figure 3 shows how the percentage of time spent in a pure state approaches 100% as the barrier height is increased. Since the mixed region ($0.1 < \lambda < 0.9$) is unphysical in our model, this graph clearly shows that it is important to include a barrier in the model potential. Figure 4 indicates that the mean lifetime of a titration endpoint increases exponentially as a function of the barrier height. On the one hand, longer lifetimes allow the protein time to relax around a titration endpoint. On the other hand, as the endpoint lifetimes increase, longer simulation times will be required to guarantee that all relevant configurations of protonation states have been adequately sampled. Thus, ultimately, a compromise between titration state purity and lifetime magnitude must be achieved. Our choice of a barrier height of 1.25 kcal/mol allows for a purity of ~70% and an endpoint lifetime of ~1.5 ps.

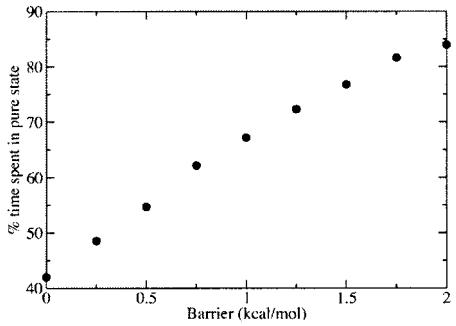

Fig. 3. Percentage of time that a titration state is pure ($\lambda < 0.1$ or $\lambda > 0.9$) as a function of barrier height for blocked aspartic acid. Simulations at pH = 4 were run for 14 ns at each value of the barrier height.

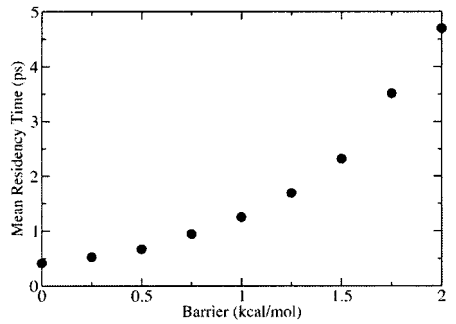

Fig. 4. Mean titration state lifetime as a function of barrier height for blocked aspartic acid. Results were extracted from the same simulations used in Figure 3. Lifetime is calculated as the total simulation time divided by the number of times λ crosses above and below 0.5. Lifetime values may be slightly overestimated, since snapshots of λ were only recorded every 10 timesteps (20 fs).

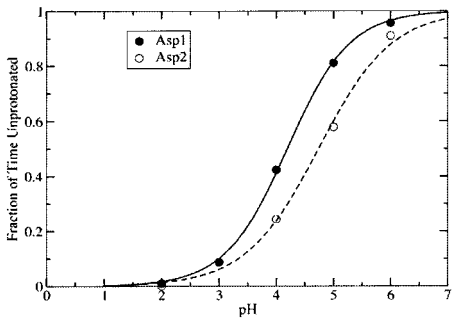

Fig. 5. Titration curves for both aspartic acid groups of the blocked Asp-Asp dipeptide. Simulations at pH = 2, 3, 4, 5, and 6 were each run for 10 ns. The curves indicate optimal HH fits where the W values were also allowed to vary [Eq. (34)]. The two calculated $pK_{1/2}$ values are 4.19 (W = 0.80) and 4.74 (W = 0.68).

Titration curves for the blocked Asp-Asp dipeptide are presented in Figure 5. Note that the two $pK_{1/2}$ values are shifted above and below 4.3, the simulated $pK_{1/2}$ value of the single residue Asp. There are two possible reasons for this behavior. First, the interaction of the two titratable groups is likely to cause $pK_{1/2}$ shifts in opposite directions. One can see a small deviation from ideal HH behavior, which may be evidence of interaction between the two sites. Because the calculation of the cross-correlation of the Asp1 and Asp2 λ trajectories at pH = 4 suggests only a

TABLE III. Calculated $pK_{1/2}$ Values of the 6 Acidic Groups in Turkey Ovomucoid

Residue	PHMD	MEAD	Null	Exp.
Asp7	4.5 (0.2)	3.1	4.0	2.7
Glu10	5.1 (0.2)	4.0	4.4	4.1
Glu19	2.7 (1.7)	1.3	4.4	3.2
Asp27	6.1 (1.0)	3.4	4.0	2.3
Glu43	6.0 (0.3)	4.4	4.4	4.8
CT-Cys56	4.6 (0.0)	3.4	3.8	<2.7
Avg. abs. error	1.2	0.8	1.0	

Simulations of 1 ns (where 200 ps was allotted to equilibration) were performed at pH = 0, 2, 4, 6 and 8. Two sets of PHMD simulations with different initial velocity seeds were run. The average value between the two sets is reported, with half the difference in parentheses. All titratable groups (including basic residues) were allowed to titrate. MEAD results were obtained with PARAM22 vdW radii and charges with Table I charges using $\varepsilon = 20$. Experimental values at 10m*M* KCl were obtained from Schaller and Robertson.[67]

TABLE IV. Calculated $pK_{1/2}$ Values for Bovine Pancreatic Trypsin Inhibitor

Residue	PHMD	TI	MEAD	Null	Exp.
NT-Arg1	7.6 (0.7)	8.1	7.2	7.5	8.1
Asp3	1.8 (0.6)	4.1	2.1	4.0	3.0
Glu7	4.1 (0.9)	5.1	2.9	4.4	3.7
Lys15	9.4 (0.2)	9.2	10.5	10.4	10.6
Lys26	9.9 (0.6)	10.2	10.7	10.4	10.6
Lys41	10.5 (0.6)	10.5	11.5	10.4	10.8
Lys46	9.4 (0.0)	10.8	10.2	10.4	10.6
Glu49	4.7 (0.2)	6.8	3.3	4.4	3.8
Asp50	2.2 (0.2)	4.4	1.6	4.0	3.4
CT-Ala58	4.1 (0.1)	3.1	2.8	3.8	2.9
Avg. abs. error	0.9	0.9	0.6	0.5	

Simulations at pH = 0, 2, 4, 6, 8, and 10 were each run 1 ns (200 ps of equilibration and 800 ps of production). Two sets of PHMD simulations with different initial velocity seeds were run. The average value between the two sets is reported, with half the difference in parentheses. TI results were obtained via the scheme elaborated in the Methods section. MEAD results were obtained with PARAM22 vdW radii and charges from Table I using $\varepsilon = 20$. Experimental values were obtained from Brown et al.[68,69] and Richarz and Wuthrich.[70]

TABLE V. Calculated $pK_{1/2}$ Values for Hen Egg White Lysozyme

Residue	PHMD	Null	PDLD	BKvG	Exp.
NT-Lys1	2.2	7.5			7.9 (0.1)
Lys1	10.2	10.4			10.8 (0.1)
Glu7	4.4	4.4	2.4	3.2	2.85 (0.25)
Lys13	11.8	10.4			10.5 (0.2)
His15	10.9	6.8			5.36 (0.07)
Asp18	1.1	4.0	1.6	3.3	2.66 (0.08)
Tyr20	8.3	9.6			10.3
Tyr23	10.7	9.6			9.8
Lys33	8.6	10.4			10.6 (0.1)
Glu35	5.8	4.4	4.3	12.3	6.2 (0.1)
Asp48	0.2	4.0	4.1	0.0	1.6 (0.4)
Asp52	5.8	4.0	3.6	8.7	3.68 (0.08)
Tyr53	11.8	9.6			12.1
Asp66	<−1.0	4.0	−0.3	1.5	0.9 (0.5)
Asp87	3.7	4.0	0.1	−0.2	2.07 (0.15)
Lys96	10.9	10.4			10.8 (0.1)
Lys97	9.9	10.4			10.3 (0.1)
Asp101	6.7	4.0	3.3	4.7	4.09 (0.07)
Lys116	9.6	10.4			10.4 (0.1)
Asp119	3.0	4.0	2.2	−0.5	3.20 (0.09)
CT-Leu129	1.6	3.8	n/a	3.3	2.75 (0.12)
Avg. abs. error	1.6	1.0			
Avg. abs. error (acidic groups only)	1.5	1.4	1.2	2.1	

Simulations at pH = 0, 2, 4, 6, 8, 10, and 12 were each run for 1 ns (200 ps of equilibration/800 ps of production). Experimental values were obtained from Baptista and Soares,[38] where averages and uncertainties were extracted from the range of values supplied in that work. BKvG corresponds to the results of Bürgi et al.,[21] which were averaged over their pH = 2, 3, and 4 simulations. PDLD corresponds to the results of Sham et al.[9]

minor anticorrelation of the two groups (results not shown), a more plausible explanation is that the eletrostatic environment of the two titratable groups is somewhat different from that of the model compound. Both sites see one of their blocked groups replaced with the peptide bond.

Constant pH simulations on four proteins have been performed to compare calculated and experimental $pK_{1/2}$ values (see Tables III–VI). The first two proteins, turkey ovomucoid (turkey) and BPTI, were run twice with different initial velocity seeds to estimate the convergence of $pK_{1/2}$ statistics for 1 ns simulations. We find that the difference between the two simulations for some residues can be as much as 2–3 pK units. The titration curves for the ten residues in BPTI using data from one of the 1 ns simulations are shown in Figure 6. In general, the deprotonation fractions fit well to the HH form, although some deviations exist, probably due to insufficient sampling.

It has been shown in previous works that free energy techniques such as LRA and TI can be used to determine intrinsic pK_α values and $pK_{1/2}$ values.[9,18–20,52] Therefore, an alternative technique, TI, was used to estimate $pK_{1/2}$ values to determine to what extent the errors in PHMD are dependent on the continuous titration scheme. TI results are presented for BPTI. We estimate the sampling error in the TI results to be <0.5 pK unit based on dividing the simulations in half (results not shown), although the true errors could be larger since the simulations were relatively short. Interestingly, the overall error with the TI approach is roughly the same. Nonetheless, the total simulation time for the TI calculations (9 ns) was 50% more than the PHMD calculations (6 ns). The TI and PHMD results are farther away from each other (average absolute deviation ~ 1.1 pK units) than either is to experiment (a.a.d. ~ 0.9 pK units). There are several possible explanations. First, neither method has achieved sampling convergence. Moreover, TI-based pK_α shifts are derived from the free energy difference between the two precise endpoints, whereas PHMD results are derived from protonation populations at approximate endpoints (see Theory section). Finally, the TI simulations do not explicitly take into account interactions between two or

TABLE VI. Calculated $pK_{1/2}$ Values for the "Acidic" (pK_a < 8) Residues of Ribonuclease A

Residue	PHMD	Null	Exp
NT-Lys 1	6.6	7.5	7.6
Glu 2	<−1.0	4.4	2.8
Glu 9	5.8	4.4	4.0
His 12	2.8	6.8	6.2
Asp 14	3.5	4.0	<2.0
Asp 38	2.4	4.0	3.1
His 48	7.7	6.8	6.0
Glu 49	6.4	4.4	4.7
Asp 53	4.5	4.0	3.9
Asp 83	7.4	4.0	3.5
Glu 86	5.9	4.4	4.1
His 105	10.8	6.8	6.7
Glu 111	5.8	4.4	3.5
His 119	7.5	6.8	6.1
Asp 121	3.3	4.0	3.1
CT-Val 124	<−1.0	3.8	2.4
Avg. abs. error	2.1*	0.7	

*Average absolute error assumes endpoints where there are only inequalities. Simulations at pH = 0, 2, 4, 6, and 8 were each run for 1 ns (200 ps of equilibration/800 ps of production). All titratable groups (including basic residues) were allowed to titrate. Experimental values were obtained from Antosiewicz et al.[8]

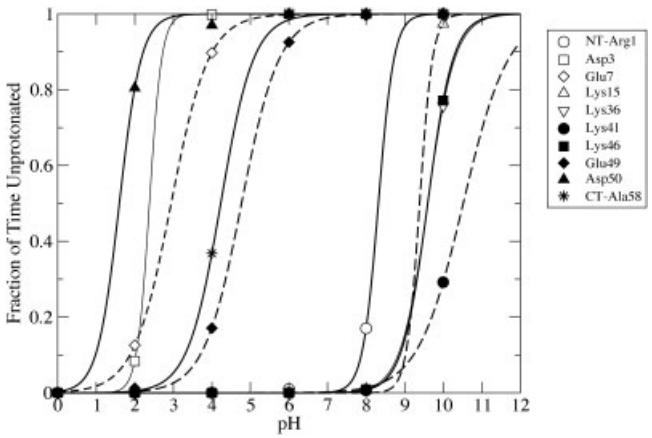

Fig. 6. Titration curves for the ten groups of BPTI from one of the two sets of simulations at pH = 0, 2, 4, 6, 8, and 10. Symbols correspond to the actual titration state observables, *S*. The curves are fitted via Eq. (23).

more titrating groups. This issue could be rectified by employing an (*ad hoc*) effective dielectric approximation[9] or by running multidimensional TI.

In the turkey ovomucoid and BPTI results (Tables III and IV), we included the MEAD static $pK_{1/2}$ results for comparison. This particular static continuum model has results that are on par with ours. As a note, our MEAD results do not reflect the state-of-the-art. Configurational averaging techniques can provide further improvements. For example, the work of van Vlijmen et al.,[10] which also uses a dielectric constant of 20, yielded $pK_{1/2}$ results with average absolute errors of about half those from the null hypothesis. Furthermore, configurational averaging over the charged and uncharged states using the LRA technique provides accurate results while requiring smaller dielectric constants.[9,14,19]

Results for lysozyme and RNase are presented in Tables V and VI. In Table V, we have included, for comparison, the averaged results of Bürgi et al.,[21] which only contained predictions for the acidic residues, Asp, Glu, and C-terminus. The constant pH method of Burgi et al.[21] is an approach where protonation states and conformational states of a protein are sampled in an explicit solvent environment. Their simulation of lysozyme entailed the explicit treatment of nearly 6000 water molecules. We also include the results of the protein dipoles–Langevin dipoles (PDLD) method of Sham et al.[9] In their method, intrinsic pK_a values were obtained via the LRA, which involved averaging the PDLD solvent energy over MD conformations generated in the charged and uncharged state of titratable group. Then, $pK_{1/2}$ values were obtained by interacting all titratable groups through an *ad hoc* effective dielectric function.[9] As one can see, our results have an accuracy on par with the method of Sham et al. and are somewhat more accurate than those from the method of Bürgi et al.

For all the protein results, the overall absolute $pK_{1/2}$ error is about 1.6 pK units. This is somewhat less accurate than the best empirical approaches based on scaled continuum electrostatic methods and closer to the accuracy achieved from the null hypothesis. In some cases, we predict the wrong titration state at the physiological pH of 7. This error usually occurs because the pK_a shift is either predicted to have the wrong sign or because it is too large. With its present parameterization, our approach is therefore not yet an optimal tool for assigning the appropriate protonation states at specified pH for modeling and simulation purposes. Nevertheless, our results indicate that significant microscopic relaxation of the protein conformation around the titratable groups occurs in response to pH changes and these are key in modeling pH-dependent phenomena. Had relaxation not occurred, the PHMD method would yield very large pK_a shifts (~10 pK units), analogous to values computed by setting the dielectric constant to 1 in a static pK_a method. In fact, we tested this assertion by running constant pH simulations on lysozyme while holding the heavy atoms fixed. We found that except for a few surface groups, the predicted pK_a shifts were very large and led to pK_a values beyond the pH values of the simulations.

The apparent paradox of our results and others who have attempted to simulate the solute microscopically[9,18,21,53] is that methods which incorporate more physics and are computationally more intensive do not necessarily lead to more accurate results. Warshel and coworkers suggest that microscopic approaches often have poorer accuracy because small pK_a shifts are derived from the summation of large opposing energetic quantities.[9] Furthermore, small errors (e.g., on the order of 1–2 kcal/mol), in the physical representation of the system will lead to prominent errors in the determination of small pK_a shifts. It is likely that schemes that dampen electrostatic terms through nonunity dielectric constants reduce the

effect of the errors in the physical representation. For example, some simple dielectric continuum models can achieve results better than the null hypothesis even though they avoid explicit representation of the proton and instead smear the positive charge onto the heavy atoms of the group.

There are many factors that may have contributed to the lack of quantitative agreement with experimental pK_a values. Limitations of the implicit solvent model, limitations of the force field, inadequate treatment of titratable group tautomers, sampling convergence, and theoretical limitations of the continuous titration scheme are just a few which warrant further investigation. We believe that one of the most important sources of error in our method is the GB implicit solvent model, which in its present form may not be accurate enough to reliably deduce small pK_a shifts when the solute dielectric is set to 1. The GBMV2 model that we used is one of the most accurate approximations to the molecular surface Poisson model in the literature to date.[36,55] However, the Poisson model, for which GB is based, is a continuum dielectric approximation that is imperfect. One the one hand, it has been shown that continuum solvent models can be quite reliable in reproducing the experimental solvation energy of small neutral organic compounds.[17] On the other hand, continuum solvent models tend to perform inadequately for most charged species, which are in fact the critical components of this work. For example, negatively charged aspartic and glutamic acid model compounds have solvation energies that, compared to explicit solvent results, are predicted too unfavorably by roughly 10 kcal/mol.[56,57] Studies[56,57] show that for charged species, certain atomic radii can be modified to achieve better correspondence between implicit electrostatic solvation energies and results derived from explicit water. In the case of the carboxylate group, a reduction of the atomic radii of the carboxylate oxygen by ~0.3 Å produces the desired solvation energy results.[56,57] However, the effectiveness of radii modification is unclear when the charged group is embedded in a protein.[57] Also, large radii modifications may adversely affect the GB charge-charge interactions [Eq. (20)]. In the Methods section, we mentioned that Asp, Glu, and C-termini residues did not titrate correctly unless the radii of the carboxylate oxygen were scaled by 0.95, corresponding to a radii reduction of ~0.1 Å. However, since we made radii modifications to the carboxylate oxygens in both the charged and neutral state, perhaps the smaller reduction was a compromise between no modifications (for the neutral state) and larger reductions for the charged state. Obviously, a more technically correct approach would be to allow the atomic radius of the carboxylate oxygen to linearly interpolate between the two titration endpoints. GB treatment of the positively charged amine groups of lysine and N-termini may also have similar problems. Since most of the amine groups in our protein test are surface exposed, the errors probably cancel between the model and protein environments. However, one notable exception is NT-Lys1 in lysozyme. The average Born radius of the N-terminal nitrogen (~3 Å) is larger the radius found in the isolated model compound (~2 Å). It is possible that the spuriously predicted pK_a value is due to an incorrect treatment of the desolvation penalty. Careful analysis and comparison to explicit solvent results is warranted.

Besides problems associated with the energetics of ionized states, implicit solvent models (or imperfect force fields in general) may cause distortions to the X-ray structure,[58] which can lead to compounding errors in the estimation of $pK_{1/2}$ values. A case in point is the His12 residue of ribonuclease, which has a calculated $pK_{1/2}$ value that is shifted tremendously lower than the experimental result. The X-ray structure of ribonuclease A reveals that the ϵ proton, which is bonded to N_ϵ, is exposed to the solvent in a narrow pocket. However, in MD simulations at $pH \geq 2$, N_ϵ becomes unprotonated and acts as a hydrogen bond acceptor to specific residues which approach only during dynamics (i.e., these residues were farther than 2 Å away in the X-ray structure). At pH = 2, for example, the N_δ-H group of His119 forms a hydrogen bond to the N_ϵ of His12. Furthermore, at pH = 6, the N_ϵ of His12 forms two hydrogen bonds: one with the peptide N-H group of Phe120, and one with the N_ϵ-H group of Gln11. Apparently, given the force field that we employed, the formation of these hydrogen bonds at both pH values is more energetically favorable than the solvation of the ϵ proton.

One approach to move toward more accurate treatment of the molecular system would be to add in a few layers of explicit water molecules in a hybrid solvent scheme.[57,59] The local interactions between titratable groups and explicit water molecules would be more physically realistic than the continuum description. Thus, this procedure might alleviate structural distortions that the generalized Born solvent model has induced. The caveat here, however, is that titration events would need to occur on a sufficiently slow timescale to allow for water rearrangements. In the mean-field treatment of GB/Poisson theory, the solvent instantaneously polarizes in response to changes in the titration state, and for the purpose of equilibrium calculations this is an asset. However, in an explicit solvent calculation, the nearby water molecules would have to rotate to interchange their roles as hydrogen bond donors and acceptors to the titrating group. The timescale for such an event has been estimated to be around 2–3 ps.[23] In our formalism, the titration timescale could be increased by either increasing the titration coordinate mass, M, or the barrier term, β. Of course, the fact that the closest water molecules have to make large rearrangements in response to a titration event is an artifact of a simplistic classical description. In a more realistic treatment, a protonated group could simply transfer its proton to the oxygen of a water molecule that had been a hydrogen bond acceptor to the proton.[26–28,60]

Another related source of potential error in our current model is the integration of empirical force field and implicit solvent. The combination of GB with the PARAM22 force field may not be consistent, since the force field was developed with the TIP3 explicit solvent model.[34] Another problem, which is general to most classical force fields, is

that the charges for each atom type are fixed regardless of the surrounding environment. Charges in fixed-charge force fields are often polarized from the *in vacuo ab initio* results to mimic solvent polarization.[61,62] While this prescription may be appropriate for the treatment of model compounds and surface residues, it is unclear whether it is valid for residues buried beneath the surface. Small changes in the charge distribution of a titratable group could amount to discrepancies of a few kilocalories per mole, which could easily skew predictions of pK_a values. Improvements such as fluctuating charge models[62,63] will hopefully resolve this issue.

Another area where our approach could be improved is in the treatment of tautomers of titratable groups. In a continuum model that utilizes a large solute dielectric constant and hence attenuates interactions between the titratable group and its local environment, this problem is subdued.[38] However, in a microscopic solute approach such as ours, a naïve treatment of isomerism may be problematic. A more elaborate approach would be to allow competing tautomers, in much the same way that λ-dynamics incorporates competing ligands for a binding pocket.[64] With this approach, it would be possible to treat the histidine residues correctly as four competing tautomeric states. This would also address the tautomeric issues concerning the carboxylate group of Asp, Glu, and C-terminus and the amine groups of Lys and N-terminus in a more aesthetic manner.

The case of Asp83 in ribonuclease A points out another issue for future consideration regarding our representation of the protonated carboxylate oxygen. Here, the calculated $pK_{1/2}$ value is too large by about 4 pK units. Upon closer inspection of the trajectory generated during the simulations at lower pH values, it appears that the Asp83 proton is stabilized in a hydrogen bond to the O_γ of Thr45 and, hence, the deprotonation event is pushed to a higher pH than expected. The caveat, however, is that the proton is *out-of-plane* from the carboxylate moiety of the Asp. Thus, it is evident that a reduction of the natural barrier for the E ↔ Z transition from 2.05 to 0.25 kcal/mol has led to the creation of a non-native hydrogen bond, which in turn, skews the calculated $pK_{1/2}$ value.

Convergence errors, which are problematic in any method where values are derived by sampling, also play a role in our present approach. Even with 1 ns of sampling, some of the pK_a values in the turkey and BPTI results deviated by more than 1 pK unit between separate runs. Sampling errors are expected to be even more evident for buried titratable groups, since protein relaxation may require overcoming large energy barriers. Also, it is difficult to know when a sufficient number of protonation and conformational states have been sampled to obtain converged results. In fact, simply looking at the number of possible protonation states, 2^n (where n is the number of titratable groups) for a given protein, one can see that complete sampling is likely intractable for a short simulation time. For example, lysozyme which has 22 titratable groups in our approach, has approximately 4 million protonation states. Of course, many of these states have extremely low probabilities at a given pH owing to their unfavorable energies. Sampling of conformational states of the protein is also a difficult problem given that averaging needs to take place over multiple plausible hydrogen bonding and side-chain packing patterns.

Given the limitations in quantitative accuracy of our current method and others which treat the solute microscopically,[9,18,21,53] what are the advantages of microscopic solute models compared to simpler models which employ electrostatic scaling through nonunity dielectric constants? First, microscopic solute methods provide a means of studying pH-dependent conformational changes. It is critical that they do not require the manual assignment of solute dielectrics, which may transform in nontrivial ways as a function of conformation. A prominent example would be a simulation of the denaturation pathway of a protein, where a buried residue moves from a low dielectric environment to a high dielectric environment upon surface exposure. Furthermore, manually assigning a dielectric constant is difficult in the cases of buried residues where no experimental pK_a data is available.[14] Also, microscopic solute methods can provide insight into how local protein structure rearranges and relaxes in response to titration events. Specifically, one can observe how hydrogen bonding and side-chain packing patterns change as a function of pH and protonation state. Finally, microscopic-solute pH methods provide a stringent test for force fields and implicit solvent models and thus can be used as a benchmark in the evaluation of newly emerging models.

CONCLUSIONS

In this work, we have presented a new technique to integrate pH with molecular dynamics simulations. Determination of $pK_{1/2}$ values for titratable groups was reasonably successful, given our aim of eliminating *ad hoc* dielectric constant scaling and instead relying on the microscopic conformational relaxation of the protein atoms.[9,18,19] Nonetheless, our results are not as reliable as those obtained by simple dielectric models in which electrostatics are reduced by an *ad hoc* scaling factor. We have pointed to possible ways to improve this model to move toward accurate pK_a predictions. These steps are the focus of ongoing studies and include (1) careful parameterization of the implicit solvent force field by modification of specific atomic radii, (2) explicit inclusion of all possible tautomers, and (3) inclusion of layers of explicit water molecules to improve the local physical description of the titratable groups. Another outstanding issue is the proper choice of titration coordinate barrier to obtain the best balance between sampling efficiency and physically realistic endpoints.

While further efforts must be invested to improve its accuracy, our new procedure provides a framework for studying pH-dependent conformational changes in proteins in a relatively efficient manner, particularly when coupled with efficient conformational sampling approaches such as replica-exchange MD.[47] Such pH-dependent phenomena for future study might include pH-based protein

unfolding/refolding, pH-dependent ligand binding, and pH-based biological activity.

ACKNOWLEDGMENTS

Computational support was provided, in part, by the Wake Forest University DEAC cluster.

REFERENCES

1. Yao J, Feher VA, Espejo BF, Reymond MT, Wright PE, Dyson HJ. Stabilization of a Type IV turn in a family of linear peptides in water solution. J Mol Biol 1994;243:736–753.
2. Kirshenbaum K, Daggett V. pH-dependent conformations of the amyloid beta(1-28) peptide fragment explored using molecular dynamics. Biochemistry 1995;34:7629–7639.
3. Alonso DO, DeArmond SJ, Cohen FE, Daggett V. Mapping the early steps in the pH-induced conformational conversion of the prion protein. Proc Natl Acad Sci USA 2001;98:2895–2899.
4. Ripoli DR, Vorobjev YN, Liwo A, Vila JA, Scheraga HA. Coupling between folding and ionization equilibria: effects of pH on the conformational preferences of polypeptides. J Mol Biol 1996;264: 770–783.
5. Ripoli DR, Vila JA, Villegas ME, Scheraga HA. On the pH-conformational dependence of the unblocked SYPYD peptide. J Mol Biol 1999;292:431–440.
6. Vila JA, Ripoli DR, Scheraga HA. Influence of lysine content and pH on the stability of alanine-based copolypeptides. Biopolymers 2001;58:235–246.
7. Bashford D Gerwert K. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. J Mol Biol 1992; 224:473–486.
8. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238:415–436.
9. Sham YY, Chu ZT, Warshel A. Consistent calculations of pKa's of ionizable residues in proteins: semi-microscopic and microscopic approaches. J Phys Chem B1997;101:4458–4472.
10. van Vlijmen HWT, Schaefer M, Karplus M. Improving the accuracy of protein pKa calculations: conformational averaging vs. the average structure. Proteins 1998;33:145–148.
11. Koumanov A, Karshikoff A, Friis EP, Borchert TV. Conformational averaging in pK calculations: improvement and limitations in prediction of ionization properties of proteins. J Phys Chem B 2001;105:9339–9344.
12. Gorfe AA, Ferrara P, Caflisch A, Marti DN, Bosshard HR, Jelesarov I. Calculation of protein ionization equilibria with conformational sampling: pKa of a model leucine zipper, GCN4 and barnase. Proteins 2002;46:41–60.
13. Simonson T, Brooks CL III. Charge screening and the dielectric constant of proteins: insights from molecular dynamics. J Am Chem Soc 1996;118:8452–8458.
14. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models. Proteins 2001; 44:400–417.
15. King G, Lee FS, Warshel A. Microscopic simulations of macroscopic dielectric constants of solvated proteins. J Chem Phys 1991;95:4366–4377.
16. Dominy BN, Minoux H, Brooks CL III. Electrostatic basis for the stability of thermophilic proteins. Proteins 2004. In press.
17. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free-energies using macroscopic solvent models. J Phys Chem 1994;98:1978–1988.
18. Russell ST, Warshel A. Calculations of electrostatic energies in proteins: the energetics of ionized groups in bovine pancreatic trypsin inhibitor. J Mol Biol 1985;185:389–404.
19. Lee FS, Chu ZT, Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. J Comp Chem 1993;14:161–185.
20. Warshel A, Sussman F, King G. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. Biochemistry 1986;25:8368–8372.
21. Bürgi R, Kollman PA, van Gunsteren WF. Simulating proteins at constant pH: an approach combining molecular dynamics and Monte Carlo simulation. Proteins 2002;47:469–480.
22. Baptista AM, Martel PJ, Petersen SB. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. Proteins 1997;27:523–544.
23. Baptista AM, Teixeira VH, Soares CM. Constant-pH molecular dynamics using stochastic titration. J Chem Phys 2002;117:4184–4200.
24. Borjesson U, Hunenberger PH. Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines. J Chem Phys 2001;114:9706–9719.
25. Mertz JE, Pettitt BM. Molecular dynamics at a constant pH. Supercomputer Applic High Perf Comput 1994;8:47–53.
26. Aqvist J, Warshel A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. Chem Rev 1993;93:2523–2544.
27. Sham YY, Muegge I, Warshel A. Simulating proton translocations in proteins: probing proton transfer pathways in the *Rhodobacter sphaeroides* reaction center. Proteins 1999;36:484–500.
28. Warshel A. Computer modeling of chemical reactions in enzymes and solutions. New York: Wiley; 1991.
29. Kong X, Brooks CL III. λ-Dynamics: a new approach to free energy calculations. J Chem Phys 1996;105:2414–2423.
30. Guo Z, Brooks CL III, Kong X. Efficient and flexible algorithm for free energy calculations using the λ-dynamics approach. J Phys Chem B 1998;102:2032–2036.
31. Warshel A. Calculations of enzymic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. Biochemistry 1981;20:3167.
32. Onufriev A, Case DA, Ullmann GM. A novel view of pH titration in biomolecules. Biochemistry 2001;40:3413–3419.
33. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4: 187–217.
34. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
35. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
36. Lee MS, Feig M, Salsbury FR Jr, Brooks CL III. New analytic approximation to the standard molecular volume model and its application to generalized Born calculations. J Comp Chem 2003; 24:1348–1356.
37. Grycuk T. Revision of the model system concept for the prediction of pKa's in proteins. J Phys Chem B 2002;106:1434–1445.
38. Baptista AM, Soares CM. Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins. J Phys Chem B 2001;105:293–309.
39. Besler BH, Merz KM, Killman PA. Atomic charges derived from semiempirical methods. J Comp Chem 1990;11:431–439.
40. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Salvador P, Dannenberg JJ, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA. Gaussian 98 (Revision A.11.1). Pittsburgh, PA: Gaussian, Inc.; 2001.
41. Feig M, Mackerell AD Jr, Brooks CL III. Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations. J Phys Chem B 2003;107:2831–2836.
42. Nosé, S. A Unified formation of the constant-temperature molecular-dynamics method. J Chem Phys 1984;81:511–519.
43. Hoover WG. Canonical dynamics—equilibrium phase–space distributions. Phys Rev A 1985;31:1695–1697.
44. Martyna GJ, Klein ML, Tuckerman M. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. J Chem Phys 1992; 97:2635–2643.
45. Gilson MK, Honig B. Calculation of the total electrostatic energy

of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins 1988;4:7–18.
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
47. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–151.
48. Hoogstraten CG, Choe S, Westler WM, Markley JL. Comparison of the accuracy of protein solution structures derived from conventional network-edited NOESY data. Protein Sci 1995;4:2289–2299.
49. Parkin S, Rupp B, Hope H. Structure of bovine pancreatic trypsin inhibitor at 125K: definition of carboxyl terminal residues Gly57 and Ala 58. Acta Crystallogr D Biol Crystallogr 1996;52:18–29.
50. Vaney MC, Maignan S, RiesKautt M, Ducruix A. High-resolution structure (1.33 Å) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus: data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission. Acta Crystallogr D Biol Crystallogr 1996;52:505–517.
51. Wlodawer A, Svensson LA, Sjolin L, Gilliland GL. Structure of phosphate-free ribonuclease A refined at 1.26 Å. Biochemistry 1988;27:2705–2717.
52. Mackerell AD Jr, Sommer MS, Karplus M. pH dependence of binding reactions from free energy simulations and macroscopic continuum electrostatic calculations: application to 2′GMP/3′GMP binding to ribonuclease T_1 and implications for catalysis. J Mol Biol 1995;247:774–807.
53. Del Buono GS, Figuerido FE, Levy RM. Intrinsic pKas of ionizable residues in proteins—an explicit solvent calculation for lysozyme. Proteins 1994;20:85–97.
54. Onufriev A, Bashford D, Case DA. A modification of the generalized Born model suitable for macromolecules. J Phys Chem B 2000;104:3712–3720.
55. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL III. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comp Chem 2004:25:265–284.
56. Nina M, Beglov D, Roux B. Atomic radii for continuum electrostatics based on molecular dynamics free energy simulations. J Phys Chem B 1997;101:5239–5248.
57. Lee MS, Olson MA. Evaluation of Poisson solution models using a hybrid explicit/implicit solvent method. Manuscript in preparation.
58. Calimet N, Schaefer M, Simonson T. Protein molecular dynamics with the generalized Born/ACE solvent model. Proteins 2001;45:144–158.
59. King G, Warshel A. A surface constrained all-atom solvent model for effective simulations of polar solutions. J Chem Phys 1989;91:3647–3661.
60. Sagnella DE, Tuckerman ME. An empirical valence bond model for proton transfer in water. J Chem Phys 1998;108:2073–2083.
61. Chipot C. Rational determination of charge distributions for free energy calculations. J Comp Chem 2003;24:409–415.
62. Patel S, Brooks CL III. CHARMM fluctuating charge force field for proteins: I. Parameterization and application to bulk organic liquid simulations. J Comput Chem 2004;25:1–15.
63. Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA. Fluctuating charge, polarizable dipole, and combined models: parameterization from ab initio quantum chemistry. J Phys Chem B 1999;103:4730–4737.
64. Banba S, Guo Z, Brooks CL III. New free energy based methods for ligand binding from detailed structure–function to multiple-ligand screening. In: Reddy MR, Erion MD, editors. Free energy calculations in rational drug design. Vol. 1. New York: Kluwer Academic/Plenum; 2001. p 195–223.
65. Nozaki Y, Tanford C. Examination of titration behavior. Methods Enzymol 1967;11:715–734.
66. Bashford D, Case DA, Dalvit C, Tennant L, Wright PE. Electrostatic calculations of side-chain pKa values in myoglobin and comparison with NMR data for histidines. Biochemistry 1993;32:8045–8056.
67. Schaller W, Robertson AD. pH, ionic strength, and temperature dependencies of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. Biochemistry 1995;34:4714–4723.
68. Brown LR, Marco AD, Wagner G, Wuthrich K. A study of the lysyl residues in the basic pancreatic trypsin inhibitor using 1H nuclear magnetic resonance at 360 MHz. Eur J Biochem 1976;62:103–107.
69. Brown LR, Marco AD, Wagner G, Wuthrich K. The influence of a single salt bridge on static and dynamic features of the globular solution conformation of the basic pancreatic trypsin inhibitor. Eur J Biochem 1978;88:87–95.
70. Richarz R, Wuthrich K. High-field ^{13}C nuclear magnetic resonance studies at 90.5 MHz of the basic pancreatic trypsin inhibitor. Biochemistry 1978;17:2263–2269.