BIOCHEMISTRY

 Volume 45, Number 31 August 8, 2006

Articles

Toward the Accurate First-Principles Prediction of Ionization Equilibria in Proteins†

Jana Khandogin and Charles L. Brooks, III*

Department of Molecular Biology, TPC6, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received April 11, 2006; Revised Manuscript Received June 7, 2006

ABSTRACT: The calculation of pK_a values for ionizable sites in proteins has been traditionally based on numerical solutions of the Poisson-Boltzmann equation carried out using a high-resolution protein structure. In this paper, we present a method based on continuous constant pH molecular dynamics (CPHMD) simulations, which allows the first-principles description of protein ionization equilibria. Our method utilizes an improved generalized Born implicit solvent model with an approximate Debye-Hückel screening function to account for salt effects and the replica-exchange (REX) protocol for enhanced conformational and protonation state sampling. The accuracy and robustness of the present method are demonstrated by 1 ns REX-CPHMD titration simulations of 10 proteins, which exhibit anomalously large pK_a shifts for the carboxylate and histidine side chains. The experimental pK_a values of these proteins are reliably reproduced with a root-mean-square error ranging from 0.6 unit for proteins containing few buried ionizable side chains to 1.0 unit or slightly higher for proteins containing ionizable side chains deeply buried in the core and experiencing strong charge−charge interactions. This unprecedented level of agreement with experimental benchmarks for the de novo calculation of pK_a values suggests that the CPHMD method is maturing into a practical tool for the quantitative prediction of protein ionization equilibria, and this, in turn, opens a door to atomistic simulations of a wide variety of pH-coupled conformational phenomena in biological macromolecules such as protein folding or misfolding, aggregation, ligand binding, membrane interaction, and catalysis.

Most proteins contain side chains that can gain or loose protons in response to environmental pH conditions (*1*). The protonation−deprotonation process leads to a change in the charge state and hence can play a critical role in the stability and function of the protein. For example, acid-induced denaturation of globular proteins can be explained by a net increase in the level of intramolecular repulsion among positively charged residues due to the protonation or charge neutralization of the acidic carboxylate groups (*2*). The ionization equilibrium also plays a vital role in the catalytic function of enzymes. A prominent example is the direct involvement of active site ionizable side chains, frequently of aspartates, glutamates, or histidines, in the generalized acid-base reactions of enzymes (*3*). The pH-dependent fibril formation of amyloidogenic proteins highlights another aspect of the functional role of ionization states in proteins (*4*). Finally, life depends on pH, which is largely regulated through the ionization states of protein histidine residues in our body (*5*).

The pH condition under which a particular ionizable side chain of the protein exists in equal populations of protonated and deprotonated states gives rise to the p$K_{1/2}$ or pK_a value

† Financial support from the National Institutes of Health (GM 57513) is greatly appreciated.

* To whom correspondence should be addressed. Phone: (858) 784-8035. Fax: (858) 784-8688. E-mail: brooks@scripps.edu.

10.1021/bi060706r CCC: $33.50
Published on Web 07/15/2006

of the side chain. In principle, a unique pK_a, termed the intrinsic pK_a, can be defined for an isolated ionizable side chain. In practice, however, due to solubility issues, an ionizable group is placed in a model compound or peptide, resulting in the so-called model compound pK_a and standard or normal pK_a (*3*). In a protein environment, some ionizable side chains can have substantially perturbed pK_a values due to desolvation effects, charge-charge interactions, and charge-dipole interactions. The experimental determination of strongly perturbed pK_a's is nontrivial because the variation in pH may induce a large conformational change or even unfolding of the protein. Over the past decade, Poisson−Boltzmann (PB) equation-based methods have become widely used in the theoretical prediction of protein pK_a's. These methods yield predictions with about 1 pK_a unit root-mean-square deviation (rmsd) from experimentally observed pK_a shifts [see a recent review by Bashford (*6*)]. Most recently, an empirical method that yields results in comparable agreement with experiment has been introduced (*7*). These methods, however, require the knowledge of a high-resolution structure of the target protein, since they do not allow for conformational flexibility (*6*). Although tuning the effective protein dielectric constant (*8*) and/or explicitly including multiple conformations (*9*, *10*) in the PB-based methods can modestly offset the lack of local reorganization of polar side chains, these methods break down in cases where conformation has a significant influence on pK_a (*6*).

In principle, the effect due to the flexibility of proteins can be addressed by more detailed or more realistic models such as the protein-dipole Langevin dipole method (*11*) and the recently described microscopic methods that incorporate protonation-deprotonation events into the molecular dynamics simulation via Monte Carlo (MC) sampling or coupling to a proton bath [see a recent review by Mongan et al. (*12*)]. The latter approach, termed constant pH molecular dynamics (PHMD), can be applied, in principle, not only in pK_a predictions for proteins with known structure but also in pH-dependent conformational dynamics and first-principles protein folding studies. One particularly interesting application area, inaccessible with the conventional fixed protonation state simulation, is to explore pH-induced partially folded conformational states, which has been gaining much attention in the pursuit of understanding the disease-related protein misfolding and aggregation phenomena.

Among all PHMD methods, the λ-dynamics (*13*) and generalized Born (GB) implicit solvent-based continuous constant pH molecular dynamics (CPHMD) method has shown particularly great promise due to its speed, stability of the dynamics trajectory, and faster convergence of protonation states relative to the MC-based discrete methods (*14*, *15*). The extension of the CPHMD method to explicitly include proton tautomerism has allowed the prediction of pK_a shifts with correct signs and a mean absolute error of $<$1.5 units for the carboxylate, histidine, and α-amino side chains in two benchmark proteins (*15*). Our previous work, however, also indicated two areas for further improvement (*15*). First, the overstabilization of salt bridge interactions due to the underlying GB model led to a systematic underestimation of pK_a's for carboxylate side chains. Second, the lack of extensive conformational sampling led to random errors as large as 0.5 pK_a unit in the model compound titrations. The latter problem has been found to be common for the class of methods that utilize a direct coupling between conformational and protonation equilibria in a microscopic fashion (*12*).

Salt ions are ubiquitous in biological environments. For example, Na^+ is the primary electrolyte that regulates the extracellular water level in the body. Under normal conditions, human plasma contains a Na^+ concentration of 136−145 mM. K^+, on the other hand, is the primary cation that helps pump the byproducts of cellular processes out of the cell. Salt ions are almost always present in the experimental studies of proteins. In particular, protein pK_a measurements are conducted in the presence of salt concentrations ranging typically from 50 to 300 mM. The influence of salt ions on the pK_a's of a protein occurs mainly through the Debye−Hückel screening of charge-charge interactions, binding of counterion to specific sites, and a change in water structure. The former can be quite large for a titrating side chain forming a solvent-exposed salt bridge or hydrogen bond contact with another charged or polar side chain, a scenario which occurs frequently for carboxylate side chains that exhibit depressed pK_a's (*15*). Thus, accounting for Debye−Hückel screening in CPHMD simulations will allow us to fine-tune theoretical predictions of protein pK_a's and improve quantitative comparison with experiment.

In this paper, we will address the aforementioned salt bridge and convergence problems in the CPHMD method by exploiting the newly improved GB parametrization, which better captures the balance between the solvation and intramolecular forces (*16*), and through the use of the replica-exchange (REX) enhanced sampling technique (*17*), which has been widely applied in first-principles protein folding studies (*18*). Also, we will account for salt effects in the CPHMD simulation by employing a linearized Debye−Hückel screening approximation in the GB solvation calculation, which has been shown to reproduce salt dependence in PB calculations (*19*). To examine the accuracy and robustness of the current methodology, CPHMD titration simulations are carried out for 10 proteins, which have experimentally characterized pK_a's for the carboxylate and histidine side chains and exhibit anomalously large shifts with respect to the model compound values. The major goal of this work is to demonstrate that the CPHMD method has matured into a practical tool for quantitative first-principles prediction of ionization equilibria in proteins.

METHODS

REX-CPHMD. Replica-exchange (REX), also called parallel tempering, is a generalized ensemble sampling technique that allows an increased rate of barrier crossing on the potential energy landscape through random walks in the temperature space (*17*). In a REX simulation, an ensemble of replicas is simulated over an exponentially spaced temperature range. The replicas adjacent in temperature are allowed to swap or exchange configurations according to the Metropolis criteria (*17*). Although a rigorous assessment of the computational acceleration in a REX simulation relative to the conventional canonical simulation has not yet been made, many applications in protein folding simulations have emerged (*18*). We make use of REX as a more efficient conformational sampling protocol to improve the sampling of protonation states and hence the convergence of CPHMD simulations.

The CPHMD method (*14*, *15*) is an extended Hamiltonian approach, where a set of titration coordinates is propagated simultaneously with the spatial coordinates. The linear attenuation of electrostatic and van der Waals interactions by the titration coordinates enables a coupling between the conformational and protonation states. Since the temperature dependence of an equilibrium constant is small and not of our interest, we assume that the change in pK_a at different temperatures is solely an effect of heat-induced conformational transitions (unfolding). Thus, the REX algorithm can be implemented for CPHMD simulations in a straightforward fashion such that both spatial and titration coordinates of adjacent replicas are subject to swapping with the appropriate temperature rescaling of the free energy of deprotonation for the model compound [ln $10k_BT$(pK_a−pH)]. We note that it is straightforward to couple both temperature and pH in a two-dimensional version of the REX protocol. Such an approach, although likely to further improve the convergence of pK_a calculations, is not explored here.

Calculation of pK_a Values. The pK_a's of proteins were computed by fitting the unprotonated fractions (*S*) obtained from REX−CPHMD simulations at various pH values (typically, pH 2, 4, 6, and 8) to the generalized Henderson−Hasselbach (HH) equation:

$$S^{\mathrm{unprot}} = \frac{1}{1 + 10^{n(\mathrm{p}K_a - \mathrm{pH})}} \quad (1)$$

where *n* is the Hill coefficient. The deviation of the value of *n* from unity reflects the degree of cooperativity (coupling) between groups that interact and ionize over the same pH range. The analytic functions of the potential of mean force (PMF) for the protonation of model compounds were derived following the procedure described in our previous work (*15*) using the following standard pK_a's (*20*, *21*): 4.0 for aspartate, 4.4 for glutamate, 3.8 for C-carboxyl, 7.5 for N-amino, 8.5 for cysteine, and 6.6 and 7.0 for Nδ and Nϵ titrations of histidine, respectively.

Modeling Salt Effects. The Debye-Hückel screening of charge-charge interactions is taken into account for low salt concentrations in the current GB model (*22*) by scaling the solvent dielectric constant ϵ by the factor $e^{-\kappa r}$, where $\kappa^2 = 1/\lambda^2 = 8\pi q^2 I/ekT$, λ is the Debye−Hückel length, and *I* is the ionic strength of the solution (*19*). Thus, the solvation energy becomes

$$\Delta G = -\frac{1}{2}\sum_{i,j}\left(1 - \frac{e^{-\kappa r_{ij}}}{\epsilon}\right)\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i\alpha_j \exp(-r_{ij}^2/F\alpha_i\alpha_j)}} \quad (2)$$

where r_{ij} is the distance between atoms *i* and *j*, q_i and q_j are the respective partial charges, and α_i and α_j are the respective effective Born radii, which can be roughly interpreted as the distances from the respective atoms to the dielectric boundary. The factor *F* is a parameter that is typically chosen to be 4. Thus, the net effect of Debye-Hückel screening is to increase the self-solvation energies of charged side chains and to weaken the electrostatic interactions between them.

pK_a's of Buried Residues. In the GBSW model (*22*), the underlying implicit solvent model in the current simulation, the dielectric boundary is defined by a set of van der Waals radii. As a result, the crevices between van der Waals spheres that are not accessible to solvent molecules are missing in the calculation of solute volume, leading to an underestimation of the effective Born radii for deeply buried atoms in the protein interior. Consequently, the solvation energies of these atoms are overestimated, and the electrostatic interaction energies between them are underestimated. These effects become more pronounced for buried charged residues when salt is present, because Debye-Hückel screening (eq 2) impacts the solvation and charge-charge interactions in the same direction. Consider a buried carboxylate residue that interacts with a buried cationic residue and a buried anionic residue. The overestimation of solvation and underestimation of repulsive interaction lead to a lower pK_a, whereas the underestimation of attractive interaction leads to a higher pK_a. For a buried histidine, a same set of interactions causes its pK_a to shift in the opposite directions.

Simulation Protocol. All REX-CPHMD simulations were conducted using the PHMD (*14*, *15*) and GBSW modules (*22*) within the CHARMM molecular dynamics program (version c33a1) (*23*). The replica-exchange protocol was enabled through the interface package MMTSB tool set [http://mmtsb.scripps.edu (*24*); REX−PHMD feature to be released]. We made use of the CHARMM22 all-atom force field for proteins (*25*), including the dihedral cross-term corrections (CMAP) (*16*, *26*), and the newly optimized GB input radii set (*16*).

Simulation details are as follows. The model compound simulations were conducted with four replicas at 298−550 K. For proteins, 16−24 replicas in the temperature range of 298−450 K were used such that the exchange ratio is ∼35−45%. Each replica was subjected to a constant pH, volume, and particle MD run. The SHAKE algorithm was applied to hydrogen bonds to allow a 2 fs time step, and a 22 Å distance truncation was applied to the nonbonded and GB energy/force evaluations. A replica-exchange attempt was allowed every 1000 dynamics steps. The total number of exchange attempts was 500, resulting in a total simulation time of 1 ns. In the GB calculations, a smoothing length of 0.6 Å at the dielectric boundary with 24 radial integration points up to 20 Å and 50 angular integration points were used. The nonpolar solvation energy was approximated from the solvent-accessible surface area (SA) calculation using a phenomenological surface tension coefficient of 0.005 kcal mol^{-1} Å^{-2}, derived from hydrocarbon transfer free energy data (*27*). Although more rigorous treatments are forthcoming, this surface area-dependent term has been shown to capture the major nonpolar solvation force in protein folding simulations (*16*). In the propagation of titration coordinates, the temperature was kept the same as the spatial coordinates. A titration barrier of 2.5 kcal/mol and a tautomer interconversion barrier of 2.5 kcal/mol were applied to suppress the fractional population of mixed titration and tautomeric states to below 20%.

Model Compound Titrations. Large random error and slow convergence represent the major bottlenecks in the discrete and continuous constant pH molecular dynamics simulations (*12*). Our previous work pointed out that the conventional single-temperature 4 ns CPHMD simulations of the blocked aspartate and histidine residues gave rms deviations in pK_a values ranging from 0.3 to 0.5 unit (*15*). Table 1 shows that 1 ns REX-CPHMD simulations using four replicas afford

Table 1: REX-CPHMD Simulations of Blocked Aspartic Acid and Histidine[a]

	run 1	run 2	run 3	run 4	run 5	mean (rmsd)
S(Asp)	0.46	0.46	0.45	0.66	0.40	0.49 (0.09)
pK_a	4.1	4.1	4.1	3.7	4.2	4.0 (0.16)
S(His)	0.72	0.65	0.55	0.69	0.72	0.67 (0.06)
pK_a	6.6	6.7	6.9	6.7	6.6	6.7 (0.12)

[a] Unprotonated fractions (*S*) in five independent 1 ns REX-CPHMD simulation runs of the blocked aspartic acid at pH 4 and histidine at pH 7. The pK_a's computed from the *S* values are also shown.

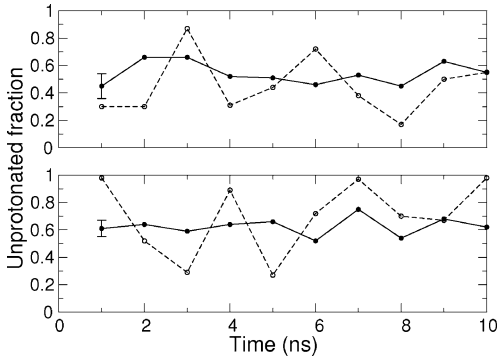

FIGURE 1: Unprotonated fractions computed from 1 ns windows in 10 ns MD trajectories for the blocked aspartate (top) and histidine (bottom) residues. The dashed line represents the single-temperature CPHMD simulation at 298 K, while the solid line represents the 298 K window in the REX-CPHMD simulation using four replicas spanning 298−550 K. The error bar on the first data point represents the rms deviation obtained from five independent simulation runs (Table 1).

rms deviations of 0.16 and 0.12 for the pK_a's of blocked aspartic acid and histidine, respectively. Figure 1 reveals that REX-CPHMD simulations also result in more rapid convergence in terms of protonation state samplings, which is readily seen from the smaller fluctuations in unprotonated fractions along the MD trajectories compared to those in the single-temperature simulations. Another significant computational advantage, due to the significant reduction in random errors and enhancement in convergence, is that the REX protocol allows us to reduce the total number of titration simulations in the determination of pK_a's for large proteins. For all proteins that were investigated in this work, a titration pH interval of 2 leads to good fits to the generalized Henderson-Hasselbach curve, which is in sharp contrast to the previous results using the single-temperature CPHMD simulations (*14*, *15*).

A set of titration simulations at the pH values equal to the expected model compound pK_a's was conducted to verify whether it gives a 50% protonated population (*15*). The only exception is the blocked histidine, which yielded a pK_a that is 0.2 pK_a unit higher than the expected value of 6.5 (Table 1). Thus, a corresponding correction was made to all the computed histidine pK_a's shown in the Results. Next, the pK_a's of model compounds at ionic strengths of 50 and 300 mM were verified using the same set of PMF functions derived for titrations in pure water. Both ionic strengths resulted in deviations of approximately −0.2 and 0.1 unit from the expected pK_a's of the blocked carboxylate and histidine, respectively.

Structure Preparation. All simulations were based on either crystal or average solution NMR structures. The PDB entries are 7RSA for RNase A, 1OMU for OMTKY3, 1CQU for NTL9, 2CI2 for CI2, 1LSA for HEWL, 1A2P for barnase, 2RN2 for RNase HI, 1BCX for xylanase, 1ERU and 1TRS for TRX-ox, and 1ERT and 1TRW for TRX-red. The simulations of TRX-ox and TRX-red in water were performed using the crystal structures 1ERU and 1ERT, respectively, while those in salt were performed using the average NMR structures 1TRS and 1TRW, respectively. The N-terminal and C-terminal residues were kept in their ionized states and subjected to titration except for the C-terminus of NTL9, which was blocked with an NH_2 group as in the experiment (*28*).

RESULTS

RNase A. The predictive power of the CPHMD method for protein pK_a's depends largely on the underlying solvent model. Our previous work (*15*) on bovine pancreatic ribonuclease A (RNase A) indicated that the overall overestimation of depressed pK_a's for carboxylate groups could be largely attributed to the undersolvation or overstabilization of solvent-exposed salt bridges, defined here as oppositely charged pairs of residues that are ≤4 Å apart in terms of the minimum distance between the side chain heavy atoms. In this work, we examine whether the new GB input radii for the charged and polar side chains can provide a better description of the strength of charge-charge interactions and improve the agreement of pK_a predictions with experiment.

Table 2 summarizes the computed pK_a values for RNase A from this and our previous work (*15*) in comparison to the NMR titration data. We first compare the simulations in pure water with experimental data obtained under minimal salt conditions. The rmsd error in this work is 0.7 unit as compared to 1.3 units in the previous work. Below we will analyze the factors that contribute to this improvement. The pK_a's of Glu2, Asp83, Asp121, and CT124 were computed to be more than 2.3 units lower than experimental values in the previous work. In particular, that of Asp83, which has a measured pK_a of 3.3, was computed to be −0.9 due to the overstablized hydrogen bond contacts with Arg85 via NE and NH2 (Table 3). A comparison with the crystal structure (PDB entry 7RSA) reveals that the former hydrogen bond is non-native and the latter represents one of the crystallographic conformations, in which Arg85 is pointed toward Asp83. The alternative conformation in the crystal structure, in which Arg85 is rotated away from Asp83, is not sampled. In this simulation, due to the enhanced solvation of arginine (the input radius of the amino nitrogen atom is adjusted from 2.13 to 1.7 Å in the new GB parametrization), the rotation of the Arg85 side chain is more flexible, which allows MD samplings around both conformations (Table 3). Consequently, the computed pK_a for Asp83 is increased by 1.4 units relative to that in the previous work.

One way to assess how close our simulation can reproduce the strength of oppositely charged side chain−side chain interactions observed in experiment is to compare the equilibrium distances from the MD trajectory to the available structure data. Our previous work revealed that the carboxylate residues, which gave overestimated pK_a shifts in RNase A, were closer in distance to the nearby cationic residues than the crystal structure, and those that formed salt bridges displayed small positional fluctuations. Table 3 compares

Table 2: Calculated and Experimental pK_a Values in RNase A[a]

	prev	REX			expt	
residue	no salt	no salt	60 mM salt	200 mM salt	≤60 mM salt	200 mM salt
Glu2	0.0	2.4	3.5	3.5	2.6	2.8
Glu9	2.6	2.8	3.9	3.9	−	4.0
His12[b]	5.1	6.6	7.8	7.8	6.0	6.2
Asp14[b]	3.1	0.8	1.7	2.8	1.8	≤2.0
Asp38	1.5	2.2	3.7	4.0	2.1	3.5
His48[b]	6.3	8.3	8.5	8.5	6.1	6.0
Glu49	5.0	3.7	3.8	3.7	4.3	4.7
Asp53	3.8	3.6	4.9	4.4	3.7	3.9
Asp83	−0.9	2.2	2.9	3.4	3.3	3.5
Glu86	3.9	3.2	4.6	4.4	4.0	4.1
His105	7.3	8.3	8.3	8.5	6.5	6.7
Glu111	3.7	3.3	3.5	3.6	−	3.5
His119	5.7	6.8	6.6	7.9	6.5	6.1
Asp121[b]	0.4	1.0	2.4	2.1	3.0	3.1
CT124	0.0	1.6	3.4	3.3	2.3	2.4
rmsd	1.7	0.8	1.1	1.2		
maximum	4.2	2.2	2.4	2.4		

[a] The prev column refers to the pK_a's obtained from the single-temperature CPHMD simulation (*15*) with the GBSW input radii parametrization by Nina et al. (*39*). The REX columns refer to the pK_a's obtained from the REX-CPHMD simulation with the GBSW parametrization by Chen et al. (*16*). The expt columns refer to the pK_a's determined via NMR titration at a minimal salt concentration (varying at different pHs but not exceeding 60 mM) by Baker et al. (*40*) or at 200 mM NaCl by Rico et al. (see ref 38 in ref *40*). The rmsd and maximum rows refer to the rms deviation and the maximum absolute error from the measurements conducted at equal or similar ionic strengths, respectively. The experimental pK_a's of Glu9 and Glu111 in 60 mM salt are assumed to be the same as those in 200 mM salt. [b] Completely or largely buried residues. A residue is defined as buried if the solvent-accessible surface computed from the crystal or NMR structure is less than 30 Å^2, corresponding to roughly 15−20% of the maximum solvent-exposed surface area for carboxylate or histidine side chains.

Table 3: Electrostatic Environment of Carboxylate Residues with Strongly Depressed pK_a Values in RNase A[a]

	REX	prev	cryst
Glu2 CD−Lys7 NZ	9.6 (3.4)	6.8 (1.3)	7.3
Glu2 OE2−Arg10 NE	10.9 (6.1)	3.9 (1.6)	2.8
Asp83 OD1−Arg85 NE	5.4 (1.0)	2.9 (0.5)	6.6 (8.8)[b]
Asp83 OD2−Arg85 NH2	5.1 (1.6)	2.9 (0.2)	3.1 (9.2)[b]
Asp121 OD1−Lys66 N	3.4 (0.7)	3.8 (1.0)	3.7
Asp121 OD1−His119 NE2	5.2 (1.2)	4.5 (1.7)	2.9
CT124 OT2−Lys104 NZ	5.1 (1.2)	4.3 (0.9)	4.7
CT124 OT1−His105 N	4.4 (2.1)	2.9 (0.3)	5.0
CT124 N−His105 O	3.6 (1.6)	2.8 (0.1)	2.9
CT124 OT1−His105 ND1	4.6 (1.1)	−	5.3

[a] REX and prev columns refer to the average distance and rms fluctuations (in parentheses) in angstroms of the residue pairs from this simulation (at pH 3) and previous simulations (*15*), respectively. The cryst column refers to the distances in the crystal structure (PDB entry 7RSA) obtained at pH 5.3. [b] In parentheses are the distances based on the second crystallographic conformation in which the Arg85 is rotated away from Asp83.

the distances between ion pairs (Glu2-Lys7, Glu2-Arg10, Asp83-Arg85, Asp121-Lys66, Asp121-His119, CT124-Lys104, and CT124-His105) from the previous and current simulations. In the current simulation, all four carboxylate residues are farther from the cationic residues and/or exhibit larger rms distance fluctuations. Also, on the basis of the crystal structure data, the electrostatic environment of these carboxylate residues seems to be more realistic as compared to the previous simulation. As a result, the average error for these residues is reduced to −1 unit from the previous value of −2.9 units (Table 2).

The salt effect is an important factor that modulates the strength of charge-charge interactions in solvent. Here and throughout the remainder of this paper, we will assess the extent to which the current approximate treatment for the Debye-Hückel screening (see Methods) captures the salt effects on pK_a shifts. The inclusion of an ionic strength of 60 mM in the simulation of RNase A reduces the pK_a shifts by more than 1 unit for the carboxylates that exhibit pK_a's of <3 in pure water simulations. In particular, the most depressed pK_a's of Glu2, Asp121, and CT124 are brought closer to the measured values under a similar salt condition as compared to the results with pure water (Table 2). Because the Debye-Hückel screening effects are exponentially dependent on the ionic strength, a further increase in ionic strength from 60 to 200 mM is expected to have a much smaller effect than the increase in ionic strength from 0 to 60 mM. This is indeed the case for both the computed and measured pK_a's for these carboxylate groups.

Let us now consider the computed pK_a's of histidines. Arg33, Asp14, and His48 form a salt-linked triad, where the former is partially buried and latter two are completely buried. In the crystal structure, Asp14 is linked to Arg33 via a salt bridge and His48 via both a side chain-side chain salt bridge and a hydrogen bond between the backbone carbonyl oxygen of Asp14 and ND1 of His48. In the simulation with water, an additional hydrogen bond is formed between the carboxylate oxygen of Asp14 and the ND1 atom of His48, which likely leads to the 1 unit underestimation of the pK_a for Asp14. The inclusion of salt brings the computed value closer to experiment, although at an ionic strength of 200 mM the correction is too large by 0.8 unit. In contrast, the computed pK_a of His48 in water is 1.8 units higher than experiment and is not affected by the salt screening effect.

His12 and His119 are two electrostatically coupled residues. While His12 is completely buried and interacts with a helix dipole that favors the charged state, His119 is mostly exposed and interacts with buried residue Asp121, partially buried residue Glu111, and solvent-exposed residue Lys7. The computed pK_a of His12 in water shows a positive deviation of 0.6 unit from experiment. The inclusion of ionic strengths of 60 and 200 mM further increases the discrepancy by more than 1 unit, most likely due to the weakened interaction with His119. The computed pK_a of His119 shows a similar behavior. While in water the overestimation is <1 unit, the inclusion of an ionic strength of 200 mM results in a positive deviation of 1.8 units. Apparently, the Debye−Hückel screening changes the aforementioned balance among several opposing electrostatic interactions. His105 is a solvent-exposed residue. Its ionization state is modulated by the balance between the attractive interactions with the C-terminal carboxylate and the backbone carbonyl of Lys104 and the repulsive interaction with the side chain of Lys104. The computed pK_a is too high by 1.8 units and is not affected by the inclusion of salt.

OMTKY3. Turkey ovomucoid third domain (OMTKY3) is a 56-residue serine protease inhibitor that has five carboxylate residues and one histidine, the protonation states of which have been determined by NMR in the absence of the protein target under minimal salt conditions (at ∼15 mM

Table 4: Calculated and Experimental pK_a Values in OMTKY3[a]

	prev	REX		expt
residue	no salt	no salt	50 mM salt	~10 mM salt
Asp7	3.2	2.5	3.2	2.7
Glu10	2.6	3.5	3.7	4.1
Glu19	2.6	2.3	3.0	3.2
Asp27[b]	4.0	3.3	3.7	2.3
Glu43	4.3	4.0	4.3	4.8
His52	–	7.3	7.0	7.5
CT56	1.0	1.1	1.9	2.5
rmsd	1.1	0.8	0.6	
maximum	1.7	1.4	1.4	

[a] Experimental pK_a's were determined via NMR titration for the carboxyl groups at 10 mM KCl by Schaller et al. (*29*) and for the N-amino and histidine groups at 15 mM KCl by Forsyth et al. (*30*). The rmsd and maximum rows refer to the rms deviation and maximum absolute error from experiment, respectively. [b] Buried residue as defined in Table 2.

KCl) (*29*, *30*). Table 4 summarizes the computed pK_a's for OMTKY3 in comparison with our previous work (*15*) and experimental data. The rms error is reduced to 0.8 unit in our new simulation with pure water from 1.1 in the previous work. Appreciable improvement can be seen for coupled residues Asp7 and Glu10 as well as for largely buried residue Asp27, mainly due to the enhanced conformational sampling. This work predicts the pK_a of Asp7 to be 2.5, 1 unit lower than that of Glu10 (3.5), in better agreement with experiment (2.7 for Asp and 4.1 for Glu10) than the previous work, where the predicted pK_a of Asp7 was 0.6 unit higher than that of Glu10. The pK_a of Asp27 is nontrivial to predict because it is largely buried and interacts with both Lys29 and Tyr31 via a charge-charge interaction and a partially buried hydrogen bond interaction. Our previous work showed an incomplete convergence in the protonation state sampling for Asp27 (*15*) and gave an overestimation of its pK_a by 1.7 units, which is reduced to 1 unit in the simulation presented here. The underestimation of buried hydrogen bond interaction may be responsible for the overestimation of the pK_a of Asp27.

Although the NMR titration experiment was conducted with a stock solution that contained a salt concentration of 10 mM, the actual ionic strength was slightly higher due to the adjustment of the pH value via addition of HCl or NaOH (*29*). Thus, we decided to repeat the simulation with an ionic strength of 50 mM. As expected, the pK_a's of Glu19 and CT56, which are computed to be below 2.5 in simulations with pure water, are now shifted slightly higher, closer to the measured values (Table 4). As a result, the rms error is improved by 0.2 unit. It is, however, worth noting that the pK_a on largely buried Asp27 is shifted farther from the experimental value, likely because the ion screening amplifies the oversolvation of buried electrostatic interactions.

NTL9. The 56-residue N-terminal domain of ribosome protein L9 (NTL9) has six carboxyl residues, all of which have pK_a's determined by NMR at a salt concentration of 100 mM (*28*). REX-CPHMD simulations in pure water yield negative pK_a shifts ranging from 1.4 to 3.3 units for the carboxylate residues, indicating overall very strong attractive charge-charge interactions (Table 5). The rms deviation from experiment is 1.7 units. Inclusion of an ionic strength of 100 mM in the simulations results in an increase of all the pK_a's and a reduction in the rms error to 0.7 unit. It is, however, worth noting that Asp23 still displays a negative deviation of 1.4 units from the measured value. A closer examination of the simulation trajectory and NMR data reveals a possible cause being the lack of sampling in the energetically less favorable conformational regions. In our simulation, Asp23 is always pointing toward the α-amino group (Figure 2, left), whereas two of the total of 18 entries of the NMR ensemble (PDB entry 1CQU) show an orientation in which Asp23 is rotated away from the α-amino group through the χ_1 rotation (Figure 2, right). Apparently, sampling of this rare conformational state requires a significantly longer simulation time, since an additional 1 ns simulation does not make any appreciable difference (Table 5). In fact, the 2 ns simulations give very similar pK_a values with the largest deviation being 0.3 unit for Glu54.

Table 5: Computed and Experimental pK_a Values in NTL9[a]

	REX		expt
residue	no salt	100 mM salt	100 mM salt
Asp8	1.4	2.4 (2.3)	3.0
Glu17	1.8	3.7 (3.7)	3.6
Asp23	0.7	1.9 (1.8)	3.1
Glu38	3.1	3.5 (3.4)	4.0
Glu48	3.0	3.9 (4.0)	4.2
Glu54	2.4	3.5 (3.2)	4.2
rmsd	1.7	0.7 (0.8)	
maximum	2.4	1.2 (1.3)	

[a] In parentheses are the pK_a's obtained from the 2 ns simulations. Experimental pK_a's were obtained via NMR titration by Kuhlman et al. (*28*).

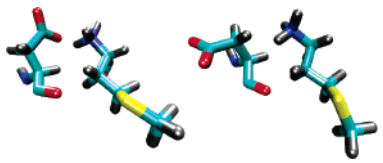

FIGURE 2: Orientation of Asp23 with respect to the α-amino group in NTL9. (Left) Predominant conformation in the NMR ensemble (PDB entry 1CQU), in which the carboxylate of Asp23 is pointing at the α-amino group. (Right) A less populated conformation in which the carboxylate of Asp23 is rotated away from the α-amino group.

CI2. Barley chymotrypsin inhibitor 2 (CI2), a serine protease inhibitor from barley seeds, is an 83-residue protein with 10 carboxylate groups. Our simulation is based on the crystal structure of the free inhibitor with the first 18 residues cleaved (*31*). The pK_a's of CI2 have been experimentally determined at both minimal, 50 mM, and high ionic strength, 200 mM. We will first consider the simulation in water. A comparison of the computed pK_a's in water with the experiment at minimal salt concentration shows a rms deviation of 0.83 unit (Table 6). The MD trajectory at pH <3 reveals that Glu26 forms a salt bridge contact with Lys24, resulting in a negative pK_a shift. This interaction is likely stronger in simulations with pure water than under experimental conditions. The pK_a's for Asp52 and Asp55 are computed to be the same as the model compound value but are 1.5 and 1 unit above and below experimental values, respectively. The MD trajectory at pH 3 reveals that Asp52 interacts with both Asp55 and Lys53, suggesting the requirement of more extensive conformational sampling. Indeed, when the simulation is extended to 2 ns, the two pK_a's start to split in the directions toward the experimental values,

Table 6: Computed and Experimental pK_a Values in CI2[a]

	REX			expt	
residue	no salt	50 mM salt	200 mM salt	50 mM salt	200 mM salt
Glu4	3.5 (3.3)	3.9	4.5 (4.2)	2.9	3.2
Glu7	2.6 (2.3)	2.9	2.8 (2.8)	2.9	3.3
Glu14	3.3 (2.8)	3.3	3.5 (3.5)	3.5	4.2
Glu15	3.6 (3.4)	4.0	3.6 (3.6)	2.8	3.8
Asp23	3.5 (3.6)	3.4	4.1 (4.0)	2.4	−
Glu26	2.5 (2.7)	2.9	3.5 (3.4)	3.7	3.9
Glu41	3.3 (2.9)	3.5	3.5 (3.2)	3.1	3.5
Asp45[b]	3.3 (2.8)	3.0	3.4 (3.7)	3.6	3.8
Asp52	4.0 (3.5)	3.7	4.1 (4.0)	2.5	2.8
Asp55	4.0 (3.7)	3.9	4.8 (4.6)	5.0	5.0
rmsd	0.8 (0.8)	0.8	0.7 (0.7)		
maximum	1.5 (1.2)	1.2	1.3 (1.2)		

[a] Results from the additional 1 ns simulations are listed in parentheses. Experimental pK_a's were obtained via NMR by Tan et al. (*32*). [b] Buried residue as defined in Table 2.

although the magnitude of the splitting still remains too low. Another factor that may contribute to the discrepancy with experiment is that these residues display large coordinate errors in the crytal structure (*31*).

The titration of Glu14 and Glu15 resembles that of Asp52 and Asp55. In this case, Glu14 electrostatically interacts with both Glu15 and Lys17. In the first 1 ns simulation, the pK_a of Glu15 is higher than that of Glu14 by 0.3 unit. An additional 1 ns of conformational sampling further decreases the pK_a of Glu14 and enlarges the difference between the two to 0.6 unit. It is curious that the order of pK_a's for these two residues is the opposite of that determined by experiment. In the crystal structure, the carboxylate of Glu14 faces the amino side chain of Lys17 with the distance between the CD and NZ atoms being 6.2 Å while Glu15 is rotated away to solvent with the distance to Lys17 of 8.6 Å. Thus, the difference between the electrostatic environment of Glu14 and Glu15 as discerned from the crystal structure and MD trajectory as well as its resemblance to that of Asp52 and Asp55 leads us to believe that the computed values of 2.8 and 3.4, respectively, are well justified and the experimentally determined pK_a's of 3.5 and 2.8, respectively, might be switched due to difficulties in resolving pK_a's from the biphasic titration data (*32*).

Inclusion of an ionic strength of 200 mM in the simulation does not change the pK_a's very much except for those of Glu4, Glu26, and Asp55, where the latter two are increased by 0.8 unit to be closer to experimental values. The rms error at an ionic strength of 200 mM is 0.7 unit, 0.1 unit lower than that in water.

HEWL. Hen egg white lysozyme (HEWL) is a 129-amino acid secretory enzyme that catalyzes the hydrolysis of polysaccharides found in many bacterial cell walls. It has one histidine and nine carboxylate residues, among which Glu35 and Asp52 are directly involved in catalysis. Our computed pK_a's at an ionic strength of 100 mM have a rms error of 0.7 unit.

Glu35, Asp52, and Arg114 form an interacting triad. The former two residues are largely buried, whereas the latter is exposed to solvent. The order of pK_a's for Glu35 and Asp52 reflects the depth of their buried side chains from the surface of the protein and is well reproduced in simulations in both pure water and salt solution. Due to the ion screening effect on their attractive interactions with Arg114, both pK_a's are increased relative to that in the simulation in pure water. The same argument can be raised for the salt-induced change in the pK_a's for the partially buried residues, His15 and Asp66, where the charged state of His15 is destabilized by the nearby solvent-exposed residues, Arg14 and Lys96, whereas that of Asp66 is stabilized by Arg54 and Arg68.

RNase HI. Escherichia coli ribonuclease H (RNase HI) is a 155-residue endonuclease that cleaves the RNA strand of a RNA-DNA hybrid. It has 20 carboxyl and five histidine groups. The catalytic site consists of three carboxylate residues: Asp10, Glu48, and Asp70. The computed pK_a's with an ionic strength of 100 mM give a rms deviation of 0.9 unit from experiment. Asp10, Asp70, and Asp134 form a cluster of interacting residues. Our simulation reveals that Asp10 and Asp70 are completely and partially buried, respectively, while Asp134 sticks out to solvent and makes a salt bridge contact with Arg138. The computed pK_a's for Asp10 and Asp70 are in close agreement with the NMR titration data, which showed a two-step titration around 6.1 and 2.6. However, the computed pK_a of Asp134 is 1.9 units below experiment, the largest deviation for RNase HI.

His114 is sequestered from solvent and is interacting with solvent-exposed residue His62. Inclusion of salt increases the pK_a for both, reflecting the screening of their mutual repulsive interaction. The pK_a for His114 is computed to be too high, most likely due to the desolvation effect being underestimated. His124 is located on a loop and fluctuates between a conformation that allows the interaction with the active site carboxylates of Asp10, Asp70, and Asp134 and a conformation in which His124 is pointing to solvent. In the crystal structure (PDB entry 2RN2), which is used as the starting structure of the current simulation, His124 has a χ_1 angle of 59°, corresponding to the solvent-exposed conformation. However, in a second crystal structure (PDB entry 1RNH), His124 has a χ_1 angle of −89°, corresponding to the solvent-sequestered conformation, which is the predominant state observed in the MD trajectory at pH 6. The negative deviation from the experimental pK_a of His124 may be attributed to the underestimation of the buried salt bridge interactions with the active site carboxylates, since the inclusion of salt weakens this interaction and moves the pK_a even lower relative to experiment.

Xylanase. Xylanase from *Bacillus circulans* is a 185-residue enzyme that catalyzes the hydrolysis of xylan, a heteropolysaccharide found in plant cell walls. It has nine carboxylate and two histidine groups, among which His149 displays one of the lowest known pK_a's for histidines. Our calculated pK_a's in water have a rms deviation of 0.8 unit from the NMR data obtained in 25 mM Na_3PO_4. His149 is deeply buried with its ND1 atom, accepting a hydrogen bond from the OG atom of Ser130 and its NE2 atom making electrostatic contacts with both Asp83 and Asp101, which causes the interaction with Asp101 to be stronger. Thus, it is plausible that our simulation shows 85% of titration occurring on ND1, in full agreement with the hydrogen−deuterium exchange data, which indicated the protection of NE2 (*33*), although the authors did not consider the interaction with the two aspartates. Xylanase undergoes acid-induced denaturation at pH <2.3 (*33*). Thus, the underestimation of our computed pK_a for His149 is likely larger than 0.6, suggesting again the underestimation of the desolvation

effect. In contrast to His149, His156 is exposed to solvent and exhibits a very small pK_a shift, consistent with experiment. Our simulation reveals that Glu172 is largely buried and makes electrostatic contact with Arg112. Its computed pK_a shows a positive shift of 0.9 unit, which is 1.4 units smaller than experiment, due to the underestimation of the desolvation effect. An accurate prediction of the pK_a for Glu83 is nontrivial, since it is buried and interacts with both Arg136 and Asp101. The calculated value is 1.2 units too high relative to experiment. It is interesting that Asp101 displays a pK_a shift as large as −3.4 units according to our simulation, although it does not interact strongly with any positively charged residue. A closer examination of the MD trajectory reveals that both carboxylate oxygens are hydrogen bond acceptors for the OG1 atom and the backbone amide nitrogen of Thr145.

Barnase. Barnase is a ribonuclease from *Bacillus amyloliquefaciens* that consists of 108 residues, 13 of which are carboxylates and two of which are histidines. The inhibition of barnase by the protein barstar is largely based on the complementarity to the positive electrostatic potential resulting from active site residues Lys27, Arg59, Arg83, Arg87, and His102, which are in turn stabilized through electrostatic interactions with Asp54, Glu73, Asp75, and Asp86. The computed pK_a's in water deviate with a rmsd of 1.0 unit from experiment. The pK_a's for Asp8, Asp12, Asp22, Glu73, and Asp86 exhibit negative deviations of greater than 1 unit, whereas those for Asp75, Asp93, and Asp101 exhibit positive deviations of greater than 1 unit. Asp8, Asp12, and Arg110 form a solvent-exposed salt triad. Therefore, inclusion of an ionic strength of 50 mM in the simulation brings the pK_a's of Asp8 and Asp12 closer to experiment.

Asp22 is a surface residues that makes a salt bridge contact with the positively charged N-amino group and long-range electrostatic interactions with Lys17 and Lys47. The former interaction is a result of an artificial N-terminal group, since the crystal structure we used is a truncated form (by two residues) of barnase. Consequently, the computed pK_a for Asp22 is 1.2 units below experiment. Glu60 is another solvent-exposed residue that interacts with positively charged Arg57 and Lys60. The computed pK_a in water is almost identical to the measured one, but inclusion of salt increases the pK_a to approximately the model compound value, resulting in a positive deviation from experiment.

Asp75 is a completely buried residue that makes a buried salt bridge contact with Arg83 as well as with Arg87. The calculated pK_a in pure water and salt solution is 1 and 2 units above the experimental value, respectively, likely due to the underestimation of the strength of the buried electrostatic interaction. The same argument is valid for explaining the positive pK_a deviations for partially buried Asp93 and Asp101, since the former forms partially buried double hydrogen bonds with the NE and NH2 atoms of Arg69 and the latter with the OG1 and backbone amide atom of Thr100.

Barnase has two histidines, one of which (His18) is situated at the end of the N-terminal helix and exhibits a positive pK_a shift larger than 1 unit in experiment, in good agreement with our calculation. Moreover, since the MD trajectory reveals that His18 is distant from negatively charged side chains, its pK_a shift must be a result of stabilization due to the dipole of the α-helix as suggested by Šali et al. (*34*).

TRX. Human thioredoxin (TRX) is a 105-residue oxireductase that functions as a general reductant for disulfides in proteins. It has one histidine and 18 carboxylate groups, among which Asp26 exhibits one of the largest positive pK_a shifts known for carboxylates. The computed pK_a's show rms deviations of 1.0 and 1.3 from experimental data for the reduced and oxidized forms of TRX, respectively. The pK_a of Asp26 is underestimated by 2.6 and 4.5 units for the reduced and oxidized forms, respectively.

Asp26 is a deeply buried residue. Its positive pK_a shift can be attributed to the destabilization of the charged state due to three factors, namely, the desolvation effect, buried repulsive interaction with Glu56, and a buried hydrogen bond with the OG atom of Ser28. The latter interaction is represented in the solution NMR structures of TRX (1TRS and 1TRW), where the side chains of Asp26 and Ser28 are pointing toward each other in alignment with strand β3. In crystal structures 1ERT and 1ERU, however, Ser28 is rotated away and not hydrogen bonded with Asp26. In our simulation, independent of the starting conformation (crystal or solution structure), Ser28 rotates away from Asp26 and makes hydrogen bond contact with Glu56 instead. Thus, it seems that both the underestimation of desolvation and the lack of hydrogen bond contact with Ser28 are the main contributors to the insufficient positive pK_a shift for Asp26, although it is noteworthy that our single-temperature simulation was unable to capture the pK_a shift at all (data not shown). The difference between the experimental pK_a's for the oxidized and reduced forms is not captured, most likely because the rearrangement due to the reduction of Cys32 and Cys35 is not sufficiently captured in this simulation. Nevertheless, our simulation is able to reproduce the large negative pK_a shift on Cys32. While interacting with Asp26, Glu56 also makes a solvent-exposed salt bridge contact with Lys39. The latter is the main cause for its depressed pK_a, which is increased when salt is included.

DISCUSSION

By making use of the replica-exchange enhanced sampling protocol and the improved parametrization of the GBSW implicit solvent model, this work has addressed two major issues raised in our previous work (*15*), namely, large statistical error and overestimation of pK_a shifts for carboxylates that form solvent-exposed salt bridge contacts with lysine or arginine side chains. Titration simulations of model compounds demonstrated a more rapid convergence of protonation states and a reduction of random error from as large as 0.5 to <0.16 unit with the same amount of total CPU time (4 ns). In the simulations of OMTKY3 and RNase A, for which the pK_a's of the carboxylate and histidine groups have been experimentally measured under minimal salt conditions (≤50 mM), the rms deviation between the computed values in water and experiment is 0.8 unit, well below those from our previous work. This improvement is mainly due to the more realistic electrostatic environment of carboxylate residues as a result of enhanced self-solvation of lysine and arginine side chains, which leads to the weakening of the solvent-exposed interactions with carboxylates.

To further improve the physical realism of CPHMD simulations and to allow quantitative comparison with

experimental data obtained in salt solutions, we have taken into account the salt effects by employing an approximate Debye-Hückel screening function in the solvation and protonation energy evaluation. Simulations of OMTKY3, CI2, and xylanase showed that the effect due to the presence of an ionic strength of 50 mM (minimal salt condition) is negligible, whereas the simulations of NTL9, HEWL, RNase HI, and TRX demonstrated that the ion screening effect due to an ionic strength of 100 mM is significant and can result in a difference in the rms deviation of as large as 1 unit. Since the Debye-Hückel screening effect is exponentially dependent on the charge-charge distance ($e^{-\kappa r}$), the salt effect on pK_a is expected to be the largest for carboxylates that form tightly bound solvent-exposed salt bridges. Thus, the most depressed carboxylate pK_a's in NTL9, HEWL, RNase HI, TRX, and RNase A are elevated and brought closer to the measured values. Inclusion of salt, however, does not improve the agreement of the pK_a with experiment for buried residues, as discussed later. Thus, the rms errors for barnase and RNase A obtained from the simulations with salt are equal or slightly larger than those in water.

The high accuracy and robustness of REX−CPHMD simulations for the first-principles determination of protein pK_a's have been demonstrated through a comparison of theoretical and experimental titration studies of 10 proteins, ranging from 56 to 185 amino acid residues in size. These proteins all exhibit anomalously large pK_a shifts. Our titrating groups are comprised of 119 carboxylate, 15 histidines, and one cysteine, among which His149 of xylanase and Asp26 of TRX display the largest pK_a shifts known to date. The titration of lysine, arginine, and tyrosine is left out because they titrate at a much higher pH (>10) and generally give very small pK_a shifts (*10*). Thus, the titrating residues we chose provide one of the most stringent test sets for theoretical pK_a prediction methods. In Table 8, we summarize the rms deviations between the calculated and experimental pK_a's according to the residue and solvent exposure types. For a total of 135 titrating groups, the rms error relative to measurement is 1 unit, or 0.8 unit for the solvent-exposed residues and 1.5 units for the largely or completely buried residues. These results are comparable to those of the PB-based methods that use an empirically adjusted protein dielectric constant (*8*) and/or are combined with multiple protein conformations (*9*, *10*), and the method based on empirically screened Coulomb potential (*35*), as well as the fully empirical structure-based prediction approach (*7*). It is, however, important to stress the fundamental difference between the CPHMD method and the structure-based macroscopic methods. Namely, the CPHMD method, in principle, does not require a high-resolution structure for the determination of the side chain ionization equilibria.

Larger rms errors for the pK_a's of buried residues and detailed analysis of their electrostatic environment have indicated a problem due to underestimation of the desolvation and buried charge−charge interactions, which is a known deficiency in the underlying implicit solvent model that employs van der Waals surface as the dielectric boundary. The error due to underestimation of desolvation energy increases with the depth of the burial. Thus, the largest computational error is for Asp26 from human thioredoxin, where the positive pK_a shift is underestimated by 2.6 and 4.5 units for the oxidized and reduced forms, respectively.

Table 7: Calculated and Experimental pK_a Values in HEWL, RNase HI, Barnase, Xylanase, and Thioredoxin[a]

residue	salt	water	expt	residue	salt	water	expt
HEWL				barnase			
Glu7	3.2	2.2	2.9	Asp8	2.9	1.5	2.9
His15[b]	6.2	4.7	5.7	Asp12	3.0	2.0	3.8
Asp18	3.3	2.3	2.7	His18	7.9	7.9	7.9
Glu35[b]	5.5	4.9	6.2	Asp22	2.1	1.6	3.3
Asp48	3.5	2.6	2.5	Glu29	4.3	4.2	3.8
Asp52[b]	4.7	4.0	3.7	Asp44	3.8	3.4	3.4
Asp66	1.9	1.1	2.0	Asp54	2.4	2.0	2.2
Asp87	2.7	0.8	2.1	Glu60	4.2	3.3	3.4
Asp101	4.0	1.2	4.1	Glu73	2.1	1.0	≤2.1
Asp119	2.5	3.3	3.2	Asp75[b]	5.4	4.1	3.1
CT129	2.7	2.0	2.7	Asp86	3.5	2.8	4.2
rmsd	0.6	1.2		Asp93[b]	4.3	3.3	<2.0
maximum	1.0	2.9		Asp101[b]	4.1	3.4	≤2.2
				His102	6.8	6.2	6.3
RNase HI				rmsd	1.1	1.0	
Glu6	3.8	2.4	4.5	maximum	2.3	1.8	
Asp10[b]	5.7	4.8	6.1	TRX-ox			
Glu32	3.5	3.0	3.6	Glu6	4.0	4.0	4.9
Glu48[b]	3.5	2.1	4.4	Glu13	3.8	3.7	4.4
Glu57	3.3	3.1	3.2	Asp16	3.6	2.7	4.2
Glu61	3.3	0.8	3.9	Asp20	3.4	1.2	3.8
His62	6.1	5.6	7.0	Asp26[b]	5.6	5.5	8.1
Glu64	3.4	2.2	4.4	Cys32[b]	–	–	–
Asp70	2.3	2.6	2.6	Glu47	3.7	2.8	4.3
His83	6.2	5.4	5.5	Glu56	2.7	0.2	3.3
Asp94	2.9	2.0	3.2	Asp58[b]	5.3	5.7	5.2
Asp102	3.1	1.9	<2.0	Asp60	4.3	5.4	2.7
Asp108	2.4	−1.0	3.2	Asp61	4.5	3.3	3.9
His114[b]	6.0	4.9	<5.0	Asp64	3.0	1.9	3.2
Glu119	3.4	1.8	4.1	Glu68	3.8	3.5	5.1
His124	5.0	5.9	7.1	Glu70	4.3	3.4	4.8
His127	6.8	7.1	7.9	Glu88	3.8	3.3	3.6
Glu129[b]	4.6	3.1	3.6	Glu95	3.5	3.1	4.1
Glu131	4.0	3.4	4.3	Glu98	3.9	2.9	3.9
Asp134	2.2	0.5	4.1	Glu103	4.0	4.0	4.5
Glu135	3.6	3.2	4.3	rmsd	1.0	1.6	
Glu147	4.1	3.8	4.2	maximum	2.5	2.6	
Asp148	3.1	2.9	<2.0	TRX-red			
Glu154	4.0	3.0	4.4	Glu6	4.5	3.9	4.8
CT155	3.1	1.5	3.4	Glu13	4.2	4.1	4.4
rmsd	0.9	2.3		Asp16	3.2	3.0	4.0
maximum	1.9	4.2		Asp20	3.5	0.9	3.8
xylanase				Asp26[b]	5.7	5.4	9.9
Asp4		3.2	3.0	Cys32[b]	6.9	-	6.3
Asp11		2.4	2.5	Glu47	3.8	2.7	4.1
Glu78[b]		4.5	4.6	Glu56	2.3	1.0	3.3
Asp83[b]		3.2	<2.0	Asp58[b]	5.5	5.0	5.3
Asp101		0.6	<2.0	Asp60	4.9	5.3	2.8
Asp106[b]		3.7	2.7	Asp61	4.1	3.7	4.2
Asp119		3.6	3.2	Asp64	2.7	1.9	3.2
Asp121		3.3	3.6	Glu68	3.6	3.4	4.9
His149[b]		2.7	<2.3	Glu70	3.9	3.9	4.6
His156		7.2	6.5	Glu88	3.5	2.9	3.7
Glu172[b]		5.3	6.7	Glu95	3.8	3.5	4.1
CT185		2.5	2.7	Glu98	3.6	3.1	3.9
rmsd		0.8		Glu103	4.1	3.7	4.4
maximum		1.4		rmsd	1.3	1.7	
				maximum	4.2	4.5	

[a] Computed pK_a's under the experimental salt conditions are listed adjacent to those obtained in pure water. Experimental pK_a values were obtained at room temperature under the following salt conditions: HEWL at 100 mM NaCl (*41*, *42*), barnase at 50 mM KCl (*37*) for the carboxylate groups and in water for the histidines (*34*), RNase HI at 100 mM NaCl for the carboxylate groups (*43*) or in water for histidines (*44*), xylanase at 25 mM Na_3PO_4 (*33*, *45*), and thioredoxin at 100 mM Na_3PO_4 (*46*). [b] Buried residues as defined in Table 2.

Theoretical prediction of the pK_a for a buried side chain that forms a strong electrostatic interaction with another buried charged residue is nontrivial because it requires the desol-

Table 8: Summary of rmsd Values between Computed and Experimental pK_a Values According to Residue and Solvent Exposure Types[a]

residue	*N*	any	exposed	buried
carboxyl	119	0.9	0.8	1.5
His	15	1.2	1.1	1.5
Cys	1	0.6	−	0.6
all	135	1.0	0.8	1.5

[a] Computed pK_a's refer to the simulations at the experimental (or close to experimental) ionic strength. For RNase A and CI2, the rmsd is calculated between the simulation and experimental data at 60 and 200 mM salt, respectively.

vation effect, attractive and/or repulsive charge-charge interactions, to be accurately captured at the same time. Inclusion of ionic strength via the approximate Debye-Hückel screening function tends to move the computed pK_a farther from the measured value in the presence of buried charge-charge or hydrogen bond interactions. This is likely the reason for the increased overestimation of pK_a's for Asp27 in OMTKY3 and residues Asp75, Asp93, and Asp101 in barnase.

Methods that explicitly sample the protonation states of protein ionizable side chains have been plagued with convergence problems (*12*). The work presented here has demonstrated that sufficient convergence of the protonation states can be achieved within a computationally tractable time by enhancing sampling of conformational degrees of freedom. As shown in the model compound simulations as well as for NTL9 and CI2, increasing CPU time beyond 1 ns per replica does not noticeably shift the protonation state populations. This is consistent with the recent success in the application of short-time REX-GB simulations for NMR structure refinement (*36*). Extensive protonation and conformational state sampling are particularly critical in the prediction of pK_a's for the buried or coupled titrating residues because in the former case conformational rearrangement is slow while in the latter case fluctuation in the protonation state population is large. Through the use of the enhanced conformational sampling protocol, improved agreement with experimental pK_a is seen for the buried residues, Asp27 of OMTKY3 and Asp26 of TRX. The enhanced sampling protocol has also allowed the experimental pK_a's of coupled titrating residues to be reproduced to a reasonable degree, such as Asp7 and Glu10 of OMTKY3, Asp52 and Asp55 of CI2, and Asp54, Asp75, and Asp86 of barnase, although the splitting between the two pK_a's seems to be underestimated.

As the accuracy of theoretical pK_a predictions approaches that of experimental measurements, small deviations in pK_a values can be dominated by the difference in solvent and protein conditions, measurement errors, and the differences in model compounds representing the standard pK_a's. For example, the reported ionic strength used in the titration experiments is often based on that of the stock solution. The actual ionic strength varies slightly due to the addition of HCl or NaOH for the adjustment of pH value. This is the case in the determination of pK_a's for OMTKY3 (*29*). Accurate measurement of pK_a's from NMR titrations relies upon the observation of chemical shifts for both protonated and deprotonated states. Consequently, the pK_a of a residue that is not fully titrated at a low pH where the protein unfolds cannot be accurately determined due to the lack of baseline representing the fully protonated state, as is the case for Glu73, Asp93, and Asp101 in barnase (*37*). Experimental determination of pK_a's for coupled ionizable residues is difficult. Fitting of ideal titration curves to the NMR chemical shift data of these residues leads to poorly resolved pK_a's, as seen for Asp54, Asp75, and Asp86 in barnase (*37*) or Glu14 and Glu15 in CI2 (*32*). Finally and perhaps most importantly, theoretical pK_a's are computed relative to the standard pK_a's, which are taken from a limited number of titration measurements on the derivatives of single amino acids or small peptides (*3*). The reported values of these standard pK_a's vary depending on the environment surrounding the specific side chain in the model compound or peptide (*3*). On the other hand, calculation of the standard pK_a in the simulation is based on a blocked single amino acid. Although we estimate the error of this type to be within 0.2 unit of the carboxyl and histidine side chains, it is much larger for rare groups such as the thiol group of cysteine, which has a measured standard pK_a ranging from 8.4 to 9.1 (*3*, *21*, *38*). Thus, some variation in measured and calculated pK_a values will result from differing reference compounds used in the experimental analysis. Better agreement between experiment and theory may be achieved by using the same model compound or peptide in the simulation. This is, however, often not practical due to the lack of force field parameters representing the compound or peptide used in the experiment.

This work has shown that the experimental pK_a's of proteins can be reliably reproduced by first-principles REX-CPHMD simulations. The rms error is ~0.6 unit for proteins containing none or very few buried ionizable side chains, such as OMTKY3, NTL9, CI2, and HEWL, ~0.8 unit for proteins containing a few buried groups, such as RNase A, RNase HI, and xylanase, and 1.0 unit or slightly higher for proteins containing ionizable side chains deeply buried in the core and experiencing strong charge-charge interactions. When the aforementioned discrepancy between the experimental and simulation conditions is taken into account, this level of agreement suggests that the REX-CPHMD method is maturing into a practical tool, which can be applied in exploring a wide variety of pH-dependent conformational phenomena in biological macromolecules, such as protein folding and misfolding, aggregation, ligand binding, membrane insertion, and catalysis.

REFERENCES

1. Matthew, J. B., Gurd, F. R., Garcia-Moreno, E. B., Flanagan, M. A., March, K. L., and Shire, S. J. (1985) pH-dependent processes in proteins, *CRC Crit. Rev. Biochem. 18*, 91−197.
2. Creighton, T. E. (1993) Chemical properties of polypeptides, in *Proteins: Structure and Molecular Properties*, 2nd ed., pp 1−48, W. H. Freeman and Company, New York.
3. Harris, T. K., and Turner, G. J. (2002) Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites, *IUBMB Life 53*, 85−98.
4. Kelly, J. W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior, *Curr. Opin. Struct. Biol. 6*, 11−17.
5. Burton, R. F. (2002) Temperature and acid base balance in ectothermic vertebrates: The imidazole alphastat hypotheses and beyond, *J. Exp. Biol. 205*, 3587−3600.
6. Bashford, D. (2004) Macroscopic electrostatic models for protonation states in proteins, *Front. Biosci. 9*, 1082−1099.
7. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pK_a values, *Proteins 61*, 704−721.

8. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) Prediction of pH-dependent properties of proteins, *J. Mol. Biol. 238*, 415−436.
9. You, T. J., and Bashford, D. (1995) Conformation and hydrogen ion titration of proteins: A continuum electrostatic model with conformational flexibility, *Biophys. J. 69*, 1721−1733.
10. Georgescu, R. E., Alexov, E. G., and Gunner, M. R. (2002) Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins, *Biophys. J. 83*, 1731−1748.
11. Sham, Y. Y., Chu, Z. T., and Warshel, A. (1997) Consistent calculations of pK_as of ionizable residues in proteins: Semi-microscopic and microscopic approaches, *J. Phys. Chem. B 101*, 4458−4472.
12. Mongan, J., and Case, D. A. (2005) Biomolecular simulations at constant pH, *Curr. Opin. Struct. Biol. 15*, 157−163.
13. Kong, X., and Brooks, C. L., III (1996) λ-dynamics: A new approach to free energy calculations, *J. Chem. Phys. 105*, 2414−2423.
14. Lee, M. S., Salsbury, F. R., Jr., and Brooks, C. L., III (2004) Constant-pH molecular dynamics using continuous titration coordinates, *Proteins 56*, 738−752.
15. Khandogin, J., and Brooks, C. L., III (2005) Constant pH molecular dynamics with proton tautomerism, *Biophys. J. 89*, 141−157.
16. Chen, J., Im, W., and Brooks, C. L., III (2006) Balancing solvation and intramolecular interactions: Toward a consistent generalized Born force field, *J. Am. Chem. Soc. 128*, 3728−3736.
17. Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding, *Chem. Phys. Lett. 314*, 141−151.
18. Nymeyer, H., Gnanakaran, S., and García, A. E. (2004) Atomic simulations of protein folding using the replica exchange algorithm, *Methods Enzymol. 383*, 119−149.
19. Srinivasan, J., Trevathan, M. W., Beroza, P., and Case, D. A. (1999) Application of a pairwise generalized Born model to proteins and nucleic acids: Inclusion of salt effects, *Theor. Chem. Acc. 101*, 426−434.
20. Nozaki, Y., and Tanford, C. (1967) Examination of titration behavior, *Methods Enzymol. 11*, 715−734.
21. Jeng, M.-F., Holmgren, A., and Dyson, H. J. (1995) Proton sharing between cysteine thiols in *Escherichia coli* thioredoxin: Implications for the mechanism of protein disulfide reduction? *Biochemistry 34*, 10101−10105.
22. Im, W., Lee, M. S., and Brooks, C. L., III (2003) Generalized Born model with a simple smoothing function, *J. Comput. Chem. 24*, 1691−1702.
23. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) Charmm: A program for macromolecular energy minimization and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
24. Feig, M., Karanicolas, J., and Brooks, C. L., III (2004) MMTSB tool set: Enhanced sampling and multiscale modeling methods for applications in structure biology, *J. Mol. Graphics Modell. 22*, 377−395.
25. MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B 102* (18), 3586−3616.
26. Feig, M., MacKerell, A. D., Jr., and Brooks, C. L., III (2003) Force field influence on the observation of π-helical protein structures in molecular dynamics simulations, *J. Phys. Chem. B 107*, 2831−2836.
27. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate calculation of hydration free energies using macroscopic solvent models, *J. Phys. Chem. 98*, 1978−1988.
28. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896−4903.
29. Schaller, W., and Robertson, A. D. (1995) pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain, *Biochemistry 34*, 4714−4723.
30. Forsyth, W. R., Gilson, M. K., Antosiewicz, J., Jaren, O. R., and Robertson, A. D. (1998) Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain, *Biochemistry 37*, 8643−8652.
31. McPhalen, C. A., and James, M. N. G. (1987) Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds, *Biochemistry 26*, 261−269.
32. Tan, Y.-J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) Perturbed pK_a-values in the denatured states of proteins, *J. Mol. Biol. 254*, 980−992.
33. Plesniak, L. A., Connelly, G. P., Wakarchuk, W. W., and Mcintosh, L. P. (1996) Characterization of a buried neutral histidine residue in *Bacillus circulans* xylanase: NMR assignments, pH titration, and hydrogen exchange, *Protein Sci. 5*, 2319−2328.
34. Šali, D., Bycroft, M., and Fersht, A. R. (1988) Stabilization of protein structure by interaction of α-helix dipole with a charged side chain, *Nature 335*, 740−743.
35. Mehler, E. L., and Guarnieri, F. (1999) A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins, *Biophys. J. 75*, 3−22.
36. Chen, J., Im, W., and Brooks, C. L., III (2004) Refinement of NMR structures using implicit solvent and advanced sampling techniques, *J. Am. Chem. Soc. 126*, 16038−16047.
37. Oliveberg, M., Arcus, V. L., and Fersht, A. R. (1995) pK_a values of carboxyl groups in the native and denatured states of barnase: The pK_a values of the denatured state are on average 0.4 units lower than those of model compounds, *Biochemistry 34*, 9424−9433.
38. Kortemme, T., and Creighton, T. E. (1995) Ionisation of cysteine residues at the termini of model α-helical peptides. Relevance to unusual thiol pK_a values in proteins of the thioredoxin family, *J. Mol. Biol. 253*, 799−812.
39. Nina, M., Beglov, D., and Roux, B. (1997) Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations, *J. Phys. Chem. B 101*, 5239−5248.
40. Baker, W. R., and Kintanar, A. (1996) Characterization of the pH titration shifts of ribonuclease A by one- and two-dimensional nuclear magnetic resonance spectroscopy, *Arch. Biochem. Biophys. 327*, 189−199.
41. Takahashi, T., Nakamura, H., and Wada, A. (1992) Electrostatic forces in two lysozymes: Calculations and measurements of histidine pK_a values, *Biopolymers 32*, 897−909.
42. Bartik, K., Redfield, C., and Dobson, C. M. (1994) Measurement of the individual pK_a values of acidic residues of hen and turkey lysozymes by two-dimensional ^{1}H NMR, *Biophys. J. 66*, 1180−1184.
43. Oda, Y., Yamazaki, T., Nagayama, K., Kanaya, S., Kuroda, Y., and Nakamura, H. (1994) Individual ionization constants of all the carboxyl groups in ribonuclease HI from *Escherichia coli* determined by NMR, *Biochemistry 33*, 5275−5284.
44. Oda, Y., Yoshida, M., and Kanaya, S. (1993) Role of histidine 124 in the catalytic function of ribonuclease HI from *Escherichia coli*, *J. Biol. Chem. 268*, 88−92.
45. Joshi, M. D., Hedberg, A., and Mcintosh, L. P. (1997) Complete measurement of the pK_a values of the carboxyl and imidazole groups in *Bacillus circulans* xylanase, *Protein Sci. 6*, 2667−2670.
46. Qin, J., Clore, G. M., and Gronenborn, A. M. (1996) Ionization equilibria for side-chain carboxyl groups in oxidized and reduced human thioredoxin and in the complex with its target peptide from the transcription factor nfκB, *Biochemistry 35*, 7−13.

BI060706R