Constant pH Molecular Dynamics in Generalized Born Implicit Solvent

JOHN MONGAN,[1,2] DAVID A. CASE,[1,3] J. ANDREW McCAMMON[1,4,5]

[1]The Center for Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Dr., La Jolla, California 92093-0365

[2]Bioinformatics Program, University of California at San Diego, 9500 Gilman Dr., La Jolla, California 92093-0365

[3]Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, California, 92037

[4]Department of Chemistry and Biochemistry, and Department of Pharmacology, University of California at San Diego, 9500 Gilman Dr., La Jolla, California 92093-0365

[5]Howard Hughes Medical Institute, 9500 Gilman Dr., La Jolla, California 92093–0365

Received 22 March 2004; Accepted 12 August 2004

DOI 10.1002/jcc.20139

Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: A new method is proposed for constant pH molecular dynamics (MD), employing generalized Born (GB) electrostatics. Protonation states are modeled with different charge sets, and titrating residues sample a Boltzmann distribution of protonation states as the simulation progresses, using Monte Carlo sampling based on GB-derived energies. The method is applied to four different crystal structures of hen egg-white lysozyme (HEWL). pK_a predictions derived from the simulations have root-mean-square (RMS) error of 0.82 relative to experimental values. Similarity of results between the four crystal structures shows the method to be independent of starting crystal structure; this is in contrast to most electrostatics-only models. A strong correlation between conformation and protonation state is noted and quantitatively analyzed, emphasizing the importance of sampling protonation states in conjunction with dynamics.

 J Comput Chem 25: 2038–2048, 2004

Key words: constant pH molecular dynamics; generalized Born; pKa prediction; Monte Carlo; lysozyme

Introduction

Protein structure and function are strongly dependent on solvent pH. This dependence is due to changes in the predominant protonation state of titratable groups (chiefly side chains of certain amino acids and termini of peptide chains) as solvent pH changes. The protonation state of a titratable group is determined by the solvent pH, and the relative acidity of the group, measured by its pK_a. The instantaneous pK_a of a given group is influenced by its electrostatic environment, which is determined by the protein conformation and protonation state of other titratable groups. Protonation state, in turn, has a strong effect on protein conformation, principally due to the charge differences between different protonation states.

Due to the tight coupling between protein conformation and protonation state, the importance of solvent pH in molecular dynamics (MD) simulations of proteins has long been recognized. Traditionally, treatment of pH in MD has been limited to setting a constant protonation state for each titratable group. This approach has many drawbacks. First, assigning protonation states requires knowledge of pK_a values for the protein's titratable groups. Second, if any of these pK_a values are near the solvent pH there may be no single protonation state that adequately represents the ensemble of protonation states appropriate at that pH. Finally, because the assumed protonation states are constant, this approach decouples the dynamic dependence of pK_a and protonation state on conformation.

The solution pH is an important extrinsic thermodynamic variable, analogous to temperature or pressure, that is readily controlled experimentally and has considerable spatial and temporal

Correspondence to: J. Mongan; e-mail: jmongan@mccammon.ucsd.edu

Contract/grant sponsor: NSF

Contract/grant sponsor: NIH; contract/grant number: GM57513, GM31749

Contract/grant sponsor: NSF Supercomputer Centers and Accelrys, Inc.

This article includes Supplementary Material available from the authors upon request or via the Internet at http://www.interscience.wiley.com/jpages/0192-8651/suppmat.

variation in living organisms. It is natural to seek simulation methods that allow the user to directly specify the pH as an input variable. In the past decade, a number of models have been proposed for performing MD at constant pH with dynamic protonation states. The earlier of these employed protonation parameters that were allowed to vary continuously between protonation states.[1,2] This is problematic because a continuous model of protonation states is not equivalent to sampling from discrete states. Especially in the case of strongly coupled titratable groups, where titration curves are highly nonlinear, the continuous model may lead to inappropriate estimates of physical observables.[3] Lee et al. have recently addressed the problem of nonphysical intermediate protonation states in the continuous model by adding an energy barrier along the continuous protonation state coordinate that favors values corresponding to fully protonated or deprotonated states.[4]

Other than the method described by Lee et al., most of the recent work in constant pH MD has involved discrete protonation states; three such models have been developed. Each of these uses Monte Carlo sampling to select protonation states based on calculated energy differences between the possible protonation states. The methods devised by Baptista et al.[5] and Walczak and Antosiewicz[6] use Poisson–Boltzmann (PB) methods to calculate protonation energies, with the former employing explicit solvent Newtonian dynamics and the latter constant dielectric Langevin dynamics between the Monte Carlo steps. One drawback to these approaches is that the potentials used for dynamics are not consistent with those used to choose protonation states in the Monte Carlo steps. Bürgi et al. avoid the consistency problem by calculating transition energies using thermodynamic integration (TI) under the same explicit solvent conditions they use for dynamics.[7] However, the TI calculations perturb the dynamics trajectory even when the Monte Carlo step is rejected. Furthermore, both PB and TI calculations are computationally expensive, limiting the number of Monte Carlo steps that can be evaluated.

We propose a model using generalized Born (GB) implicit solvation[8] that combines the best aspects of these discrete protonation state models. The same GB electrostatics are used for calculating protonation state transition energies and dynamics, so the potentials are consistent. Furthermore, calculation of transition energies using GB is fast, and there is no need for solvent equilibration, so sampling is fast. This model is tested using simulation of hen egg-white lysozyme, examining convergence, stability, agreement with experimental pK_a values, and correlation between conformation and protonation. Close agreement between predicted and experimental pK_a values suggest that this method accurately samples protonation states, providing a more physically realistic basis for studying dynamics of systems with titratable groups.

Theory and Methods

Algorithm

The proposed method employs GB solvated MD, with periodic Monte Carlo sampling of protonation states. Between Monte Carlo steps, the system evolves according to standard generalized Born solvated MD.[9,10] This sampling scheme and the justification for it are essentially the same as those described by Baptista et al.,[5] with the exception that there is no solvent equilibration step because the MD is conducted in implicit solvent.

Table 1. Reference pK_a Values for Titratable Side Chains.

Residue	$pK_{a,ref}$
Aspartate	4.0
Glutamate	4.4
Histidine-δ	6.5
Histidine-ε	7.1
Lysine	10.4
Tyrosine	9.6

Reference pK_a values for aspartate, glutamate, lysine, and tyrosine were taken from Bashford et al.[37] The reference pK_a values for histidine were taken from Kyte.[38]

At each Monte Carlo step, a titratable site and a new protonation state for that site are randomly chosen. A transition free energy for the protonation or deprotonation is calculated according to

$$\Delta G = k_B T(\mathrm{pH} - \mathrm{pK_{a,ref}})\ln 10 + \Delta G_{\mathrm{elec}} - \Delta G_{\mathrm{elec,ref}} \qquad (1)$$

where k_B is the Boltzmann constant, T is temperature, pH is the specified solvent pH, $pK_{a,ref}$ is the pK_a of the appropriate reference compound (see later and Table 1), ΔG_{elec} is the electrostatic component of the free energy calculated for the titratable group in the protein, and $\Delta G_{\mathrm{elec,ref}}$ is the electrostatic component of the transition free energy for the reference compound, a free dipeptide amino acid described later. This equation is based on a division of the total transition free energy into electrostatic and nonelectrostatic portions. The nonelectrostatic transition free energy comprises all free energy contributions not accounted for in the GB electrostatics, including the quantum mechanical bond free energy and proton solvation free energy. It is difficult to calculate the nonelectrostatic transition free energy, but it can be assumed to have approximately the same value independent of electrostatic environment. Under this assumption, a reference compound with known pK_a can be introduced to cancel the nonelectrostatic portion of the transition free energy, resulting in eq. (1). The electrostatic portion of the transition free energy (ΔG_{elec}) is calculated by taking the difference between the potential calculated with the charges for the current protonation state and the potential calculated with the charges for the proposed state; because there is no need for solvent equilibration, this is done in a single step. Equation (1) can then be used to calculate the total transition free energy, as all other terms are known. This method of calculating transition free energies is similar to that employed by Bürgi et al.,[7] except that in our proposed model only charges change between different protonation states, while they change van der Waals radii as well. Changing van der Waals radii may be added in a further refinement of this model, but good results are seen with changing only charges.

The total transition free energy, ΔG, is used as the basis for applying the Metropolis criterion to determine whether the transi-

tion will be accepted. If the transition is accepted, MD is continued with the titratable group in the new protonation state; if not, MD continues with no change to the protonation state.

Computationally, the time to evaluate a Monte Carlo step is less than that required for an MD step, so constant pH MD using this approach is only slightly slower than traditional constant protonation state GB MD.

Under this model, the total charge on the molecule is generally nonzero and changes when a titratable group changes protonation state. Because GB solvation does not employ periodic boundary conditions and the free energy associated with introducing and solvating a charge are included in the nonelectrostatic portion of the transition free energy accounted for above, the changing total charge does not present a problem.

Molecular Dynamics

MD was performed using a prerelease version of AMBER 8,[9] modified to implement the algorithm described above (these modifications have been included in the released version of AMBER 8). The ff99 force field[11] was employed. The first GB model developed by Onufriev, Bashford, and Case[10,12,13] (igb = 2) was used for solvation. Salt concentration (Debye–Hückel based) was set at 0.1 M. The cutoff for nonbonded interactions and computation of effective Born radii was 30 Å. Solute temperature was weakly coupled to a Berendsen temperature bath[14] at 300 K with a time constant of 2 ps. Lengths of bonds including hydrogen were constrained using SHAKE. The time step was 2 fs.

Protonation State Models

Titratable group models were developed for the side chains of aspartate, glutamate, histidine, lysine, and tyrosine. Protonation states for a given group differ only in partial charges. When a group's protonation state changes, charges on all of its side-chain atoms are changed to reflect the new state. Titratable hydrogens have zero charge in the deprotonated state. The titratable hydrogens in aspartate, glutamate, and tyrosine have zero van der Waals radius in the AMBER force field, so when their charge is zero they have no nonbonded interactions with the system, although they retain defined positions. Van der Waals radii for titratable hydrogens on lysine and histidine were left unchanged at their ff99 value of 0.6 Å. This does not seem to substantially affect results, because pK_a predictions for amine (LYS) and carboxylic acid (ASP and GLU) residues were of comparable quality (see Results and Discussion).

Partial charges were taken from the protonated and deprotonated residue definitions in the AMBER 99 force field. This force field does not define a deprotonated tyrosine; charges for the deprotonated tyrosine were calculated using Antechamber[15] with the RESP charge method based on HF/6-31G* calculations conducted with Gaussian 98.[16] Although the largest charge changes in these charge sets are concentrated near the titratable proton, every atom has some charge difference between protonated and deprotonated forms. If peptide backbone charges are changed when the protonation state changes, it is not possible to use a single reference free energy. Due to the 1–4 electrostatic interactions defined in the AMBER force field, backbone atoms have specific electrostatic interactions with side-chain atoms of neighboring residues. To avoid this problem, backbone charges were fixed at the values defined for the protonated state across all protonation states. A charge correction was added to the beta carbon of the deprotonated state such that the total charge difference between protonated and deprotonated states was 1.

In this charge-change only model of protonation states, a deprotonated group can gain a proton only at the location of a zero charge "ghost" proton. A titratable group may unrealistically favor the deprotonated state if its ghost proton rotates into an unfavorable position for protonation. This problem is especially severe for carboxylic acids, where the *syn* location for the proton is much more favorable than the *anti*. When a ghost proton moves into the *anti* position it is unlikely to protonate, and unlikely to move until it protonates, because no forces act on a ghost proton. This problem is addressed by building a carboxylic acid model with two protons on the oxygen, kept 180 degrees apart by an improper torsion. Because rotation of the carboxylic group to exchange the oxygen atoms is also slow, we define two protons on each oxygen of the carboxylic groups. The protonation state charge sets are defined such that no more than one of the four protons has a nonzero charge at any time.

Reference Compounds

Reference free energy differences for the titratable groups were calculated for single amino acids as dipeptide (blocked) molecules, having the sequence acetyl–amino acid–methyl amine. A titration of the dipeptide reference compound was performed with solution pH set to $pK_{a,ref}$. (Reference pK_a values are listed in Table 1.) $\Delta G_{elec,ref}$ was adjusted based on the results of this titration to give equal populations in the protonated and deprotonated states for titrations of the reference compound having pH equal to pK_a. For the simple case of a titratable group with only two protonation states, $\Delta G_{elec,ref}$ should be equal to the free energy difference calculated between the states by thermodynamic integration (TI). Calculations of $\Delta G_{elec,ref}$ were checked by performing TI between the protonated and deprotonated states using the parameters described earlier to calculate ΔG_{TI} according to

$$\Delta G_{TI} = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad (2)$$

where V is the potential and λ is the coupling parameter between the charges for the protonated and deprotonated states. Eleven equally spaced values were used for λ. At each value of λ, the reference compound was equilibrated for 40 ps and sampled for 1.6 ns. The free energy difference was calculated in Mathematica[17] by numerical integration of a fourth degree polynomial fit to the $(\partial V/\partial \lambda)$ values. In all simple cases ΔG_{TI} matched $\Delta G_{elec,ref}$ to within 0.05 kcal/mol. This consistency is, of course, expected, and is really just a check on the correctness of our implementation of the Monte Carlo algorithm. It has recently been demonstrated that ΔG_{TI} values calculated using the GB model adopted here are similar to those computed using explicit solvent models.[18]

Calculations for the carboxylic residues were complicated by having four protonated states defined (*syn* and *anti* on each of the

oxygens). For these residues, ΔG_{TI} (which was calculated between the deprotonated state and one of the *syn* protonated states) differed from $\Delta G_{elec,ref}$ by approximately $k_B T$ ln 2 due to the statistical effects of multiple protonated states (only the two *syn* sites see appreciable populations). In addition to balancing the relative populations of the protonated and deprotonated states, it is important that the relative proton affinities of the *syn* and *anti* states are correct. Based on relative populations of the *syn* and *anti* states in test titrations of the model compound, the free energy of the *syn* state was calculated to be 1.6–1.9 kcal/mol lower than the *anti* state. This is in close agreement with quantum mechanical calculations and experimental estimates,[19] so it was assumed that the force field accurately accounts for the free energy difference between the *syn* and *anti* states, and no adjustment was made to the relative energies of these states.

Test System Molecular Models

Hen egg white lysozyme (HEWL) was selected as the test system because it is well studied[7,20,21] and has a number of residues with pK_a values that differ markedly from their reference values. Structures 1AKI, 1LSA, 3LZT, and 4LYT from the PDB were selected as starting crystal structures. The structures were chosen to facilitate comparison with earlier work[7] and provide a diversity of crystal properties. The structures are from orthorhombic, tetragonal, triclinic, and monoclinic space groups, respectively. 3LZT is at high resolution (0.92 Å), 4LYT is at low resolution (2.5 Å), and 1LSA has crystal contacts that have been problematic in earlier studies.

Each structure was prepared using WHAT IF[22] to optimize the hydrogen bond network[23] (by flipping side chains of HIS, ASN, and GLN) and strip crystal waters. Hydrogens were added to the structures using the LEaP module of AMBER. They were then minimized with 100 steps of steepest descent followed by 100 steps of conjugate gradient using the SANDER module of AMBER and the MD parameters described earlier.

Simulations starting from the 1AKI structure were performed at 0.5 pH increments from pH 2.0 to 4.0 with aspartates and glutamates titrating, from pH 4.5 to 6.5 with aspartates, glutamates, and histidine titrating, and from pH 9.0 to 12.0 with tyrosines and lysines titrating. Simulations starting from 1LSA, 3LZT, and 4LYT were performed at 1.0 pH increments from pH 2.0 to 7.0 with all aspartates, glutamates, and histidines titrating and from 9.0 to 12.0 with all tyrosines and lysines titrating. There were 10 fs between Monte Carlo steps. Nontitrating residues were fixed at their most probable protonation states (protonated for basic residues and deprotonated for acidic residues). Protonation state models for terminal residues have not yet been created, so terminal residues are fixed at their most likely neutral pH protonation state in all simulations: protonated for the N-terminus and LYS-1 side chain and deprotonated for the C-terminus. This approximation is expected to have little effect on the titrating sidechains. The C-terminal residue is approximately 10 Å from the nearest acid-pH titrating group and the N-terminal residue is nearly 15 Å from the nearest basic-pH titrating group, so direct interactions are small. There may also be an indirect interaction in the high pH simulations due to perturbation of the conformations sampled because the N-terminus (experimental pK_a of 7.8–8.0[21]) is held in the protonated state. The C-terminus is sufficiently acidic (experimental pK_a of 2.63–2.87[21]) that the indirect interaction should be negligible.

pK_a Prediction Calculations

Constant pH simulations can be analyzed in a fashion entirely analogous to that used for experiments that give protonation information for individual side chains as a function of pH. As long as the protonation fraction is a monotonic function of pH, the pK_a of a side chain can be defined as the pH value for which the protonated and deprotonated populations are equal. The special case of an ideal titratable group having no interactions with other titratable groups has a sigmoidal titration curve, and behavior characterized by the Henderson–Hasselbalch (HH) equation

$$pK_a = pH - \log_{10}\left(\frac{[A^-]}{[HA]}\right) \quad (3)$$

Following the reasoning of Baptista et al.,[5] the system is assumed to be ergodic, so the ratio of time that a titratable group spends in the protonated and deprotonated states can be used as a ratio of concentrations. This can be combined with the pH according to eq. (3) to yield a prediction of the pK_a. When a titratable group has sufficiently weak interactions with other titratable groups, its behavior is well described by the HH equation, and pK_a values calculated from simulations at different pH values will differ only by random error. As interactions increase, the HH equation will not adequately describe the titration curve.

Titration data are often represented in a Hill plot, where $\log_{10}(|A^-|/HA)$ is plotted vs. pH. A titration curve for a titratable group governed by the HH equation has the form of a straight line with a Hill coefficient (slope) of 1. Titrating groups with non-HH behavior will have Hill coefficients that differ from 1. The Hill coefficient can be determined by linear regression of the titration data points on a Hill plot. Because the coefficient calculated by regression may differ from 1 due to random error or non-HH titration behavior, a *t*-test should be used to decide whether the Hill coefficient suggests statistically significant non-HH behavior. In such cases, further simulations can be conducted to plot a full titration curve.

Results and Discussion

A method for constant pH MD simulations should be computationally efficient and capable of reproducing accurate titration curves. Furthermore, when applied to macromolecules, an ideal method would yield pK_a predictions in close agreement with experimental values, converge rapidly to these predictions and maintain the stability of the trajectory. As previously mentioned, the proposed method is only slightly more computationally expensive than traditional GB MD. Here, we evaluate how well the proposed method meets the remainder of these criteria, and investigate non-HH behavior and conformation-protonation state correlations suggested by the results.

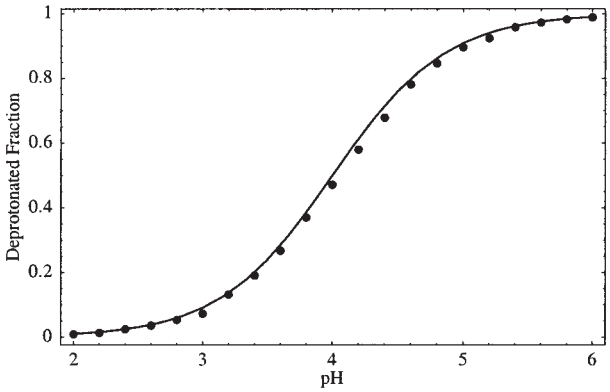

Figure 1. Deprotonated fraction for 5-ns simulations of aspartate model compound at pH values between 2 and 6. Solid line represents Henderson–Hasselbalch titration curve for residue with pK_a of 4.0. The small systematic error apparent near the midpoint is due to error in the value calculated for $\Delta G_{elec,ref}$.

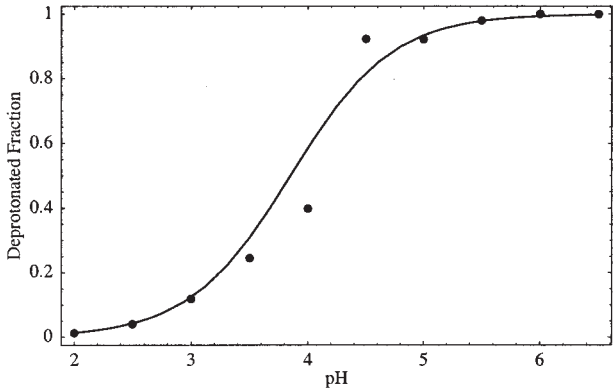

Figure 3. Deprotonated fraction for GLU-7 from 1-ns simulations starting at 1AKI at pH values between 2.0 and 6.5 (data also presented in Table 3). Solid line represents best fit titration curve, which has pK_a of 3.85.

Convergence

Small systems, such as the reference compounds, converge to the relative protonation state populations predicted by the HH equation within a few nanoseconds of simulated time. For example, a titration curve for 5-ns simulations of the aspartate model compound, shown in Figure 1, closely matches the predicted titration curve.

Convergence in larger systems, such as HEWL, is much more difficult to achieve. As seen in Figure 2, the predicted pK_a value for most residues stabilizes within a few hundred picoseconds. This stabilization does not represent convergence; the same residue may stabilize at a significantly different pK_a value if a different random seed is used in an otherwise identical simulation.

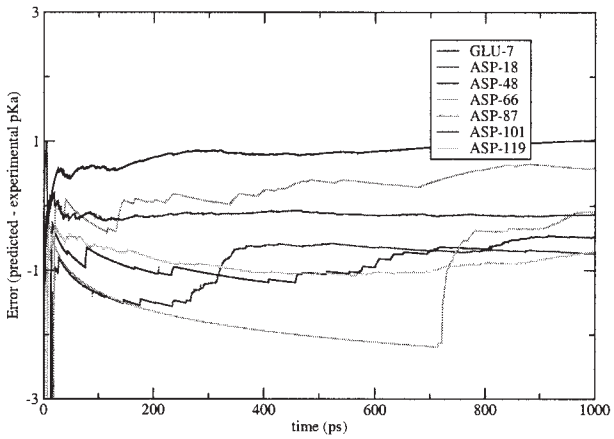

Figure 2. Time evolution of predicted pK_a for acidic residues in simulation of HEWL at pH 3.0, starting from structure 1AKI. Each point represents the predicted pK_a calculated from all protonation data collected up to that time in the simulation. Residues ASP-52 and GLU-35 do not converge due to H-bonding issues and large offset (see text), respectively, and are not shown on this plot.

Because the random error due to incomplete convergence may produce larger effects than those caused by a small change in pH, titration curves for titrating groups in proteins can be noisy. This is demonstrated by comparing the protein titration curve seen in Figure 3 to the model compound data of Figure 1, noting that the protein data are for GLU-7, one of the better converged residues in HEWL. Despite the precision problems posed by these random errors, pK_a predictions are generally fairly accurate, and the impact of noise can be reduced by combining results from multiple simulations, as seen later.

We believe that the major limiting factor on convergence is conformational sampling. As shown later, the instantaneous pK_a is strongly dependent on conformation, so if two simulations sample conformation space differently, it should be expected that they would have differing protonation state populations. Sufficiently complete conformational sampling is achievable for small systems, but is currently computationally infeasible for systems the size of HEWL.

Accepting that complete conformational sampling is out of reach for HEWL, titrations of 1 ns were performed to allow sufficient time for predicted pK_a stabilization, if not convergence. For the simulation shown in Figure 2, each of the titrating residues was evaluated for a protonation state transition an average of 11,000 times, of which between 160 and 840 transitions were accepted.

Simulation Stability

Because this method involves instantaneous changes in protonation state, which result in nonphysical discontinuities in energy and force, we examined system stability across protonation state changes. When the protonation state changes, there is a discontinuous change in total energy reported by AMBER equal to ΔG_{elec}. Most of this energy change represents transfer of energy between the energy modes governed by the force field and those outside the scope of the force field (e.g., quantum mechanical energy of the bond and solvation free energy of the proton). The remainder of the change represents Boltzmann sampling of the energy levels of

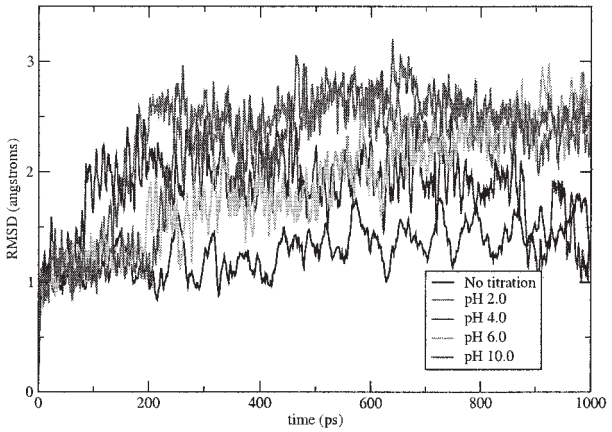

Figure 4. Alpha carbon RMSD from crystal coordinates for residues 4–125 of 1AKI structure of HEWL at five different solvent pH values. Plots are representative of behavior of other simulations. RMSD for constant protonation state (no titration) MD under similar conditions is shown for comparison.

different protonation states. Even when temperature regulation is removed, there is no trend to the changes in total energy. Average kinetic energy does follow the fluctuations in total energy, so temperature fluctuations are increased somewhat by this method. However, even in the worst case of a very small system (the reference compound) with no temperature regulation, root-mean-square temperature fluctuations were only about twice as high for a simulation where protonation state changed rapidly as they were for a simulation where protonation state was fixed. In the more common case of a larger system with temperature regulation and slower protonation state changes, temperature fluctuations are fairly small—on the order of 5 K for the HEWL titrations described here.

Conformational stability, as well as energetic stability, is of interest in biomolecule simulations. Experimentally, HEWL is a stable protein across a wide range of solvent pH values, and this should be reflected in the trajectories of the titrations. Figure 4 compares alpha carbon root-mean-square deviation (RMSD) vs. crystal structure for titrations starting from 1AKI at a range of pH values to a nontitrating trajectory. Most importantly, this plot shows that after an initial relaxation period, RMSD stabilizes for each trajectory. One simulation was continued to 3 ns to confirm stability: there were no significant excursions beyond an RMSD of 2.5 Å. In general, RMSD for titrating trajectories increased more rapidly and stabilized at higher values than for the nontitrating trajectory. Traditional MD trajectories are known to be biased toward conformations that are compatible with their fixed protonation state;[24] it seems reasonable that allowing protonation states to change would reduce this bias and allow greater conformational sampling, producing a higher RMSD. One might also expect that simulations conducted at pH values close to the pH of the crystal structure would tend to sample conformations closer to the crystal structure. Indeed, Figure 4 shows that the pH 4.0 trajectory stabilizes at the lowest RMSD relative to the 1AKI crystal structure, which was solved at pH 4.5.

pK_a Predictions

Although the primary aim of this work is to improve the physical realism of dynamics, it is difficult to validate the quality of simulated dynamics as a function of pH, because HEWL appears to have no major structural changes in the pH range considered here. HEWL was selected because it has been well studied, so it provides a good test system for determining whether protonation states are accurately sampled, and accurate sampling is a prerequisite for simulating pH-dependent conformational changes. Therefore, pK_a values, which can be calculated from the simulations using eq. (3), are compared to experimentally measured values as the primary quality measure used to validate the proposed method.

In comparing predicted pK_a values to experimental measurements, it is useful to have a method for combining predictions from simulations conducted at different solvent pH values into a single composite pK_a prediction, taking account of their relative accuracies. This is commonly achieved by plotting the data at each pH on a Hill plot and performing linear regression. In regression, each data point is weighted according to the inverse of its variance (more properly, the variance of the distribution from which the data point is drawn). It is expected that the variance of a data point will be dependent on the absolute difference between the pH at which it was taken and the predicted pK_a. This follows from eq. (3), which becomes increasingly sensitive to small changes in the number of time steps spent in each protonation state as the quotient of these numbers becomes very large or very small. It is computationally infeasible to run sufficient simulations to determine a separate variance model for each titratable group, so data for all titratable groups were pooled to determine a global variance model for the method. Figure 5 shows a scatter plot of absolute difference between the pH and predicted pK_a (offset) vs. pK_a prediction error. The running (windowed) variance line on this plot shows that variance is roughly uniform when the offset is less than 2.0 pH units, and increases rapidly outside of this range. Because there are insufficient data to empirically determine a variance at each offset,

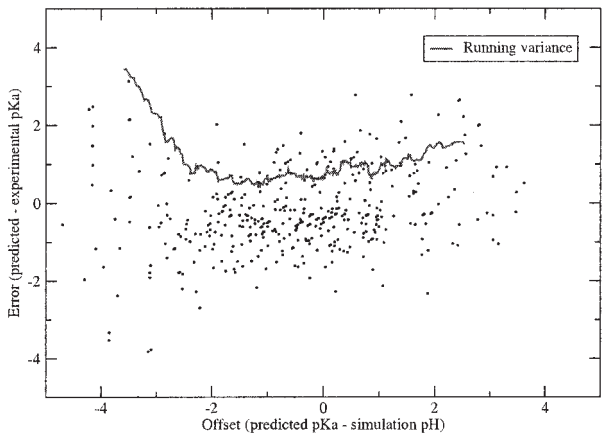

Figure 5. Scatter plot of difference between predicted pK_a and solvent pH (offset) vs. difference between predicted and experimental pK_a (prediction error). Points represent all predictions with offsets between −5 and 5. Variance is calculated with a window size of 40 data points.

Table 2. pK_a Predictions for Acidic Residues of HEWL, Calculated from 1-ns Simulations Starting from 1AKI at Specified pH.

	pH 2.0	pH 2.5	pH 3.0	pH 3.5	pH 4.0	pH 4.5	pH 5.0	pH 5.5	pH 6.0	pH 6.5	Comp.	Exp.
ASP-18	0.97	2.02	1.92	2.05	1.98	2.12	2.20	1.93	$-\infty$	$-\infty$	1.83	2.66
ASP-48	2.84	0.75	2.03	1.29	1.94	0.92	2.17	1.96	$-\infty$	$-\infty$	1.87	2.50
ASP-52	2.15	2.04	−0.14	2.05	$-\infty$	2.38	$-\infty$	2.54	$-\infty$	$-\infty$	2.08	3.68
ASP-66	2.57	1.51	1.92	0.40	$-\infty$	$-\infty$	$-\infty$	2.82	$-\infty$	$-\infty$	2.00	2.00
ASP-87	3.14	2.17	2.66	2.36	3.12	0.84	$-\infty$	$-\infty$	$-\infty$	$-\infty$	2.69	2.07
ASP-101	3.56	3.76	3.96	3.90	3.80	3.67	4.07	3.85	3.69	4.09	3.80	4.09
ASP-119	3.70	2.29	2.50	2.40	2.36	2.46	2.49	2.69	3.15	$-\infty$	2.60	3.20
GLU-7	3.89	3.88	3.87	3.99	4.18	3.42	3.93	3.82	2.68	3.39	3.85	2.85
GLU-35	3.87	5.97	6.48	5.27	5.65	5.60	6.78	5.31	5.19	5.36	5.32	6.20
HIS-15	—	—	—	—	6.26	5.87	6.69	6.69	6.71	6.26	6.45	5.71

Composite pK_a is the average of predictions with absolute offset less than 2.0 (see text for discussion). Where experimental pK_a values[21] were given as a range, the midpoint of the range is used; where only an upper bound was given, the upper bound is used.

a simplified variance model is drawn from these data: uniform variance for offsets less than 2.0 pH units and very high (effectively infinite) variance for larger offsets. This leads to uniform weights for the small offset data points and zero weight for those with large offset. Because the non-HH behavior in this system is small, and in general does not affect the predicted pK_a, the number of free parameters in the fit is reduced by restricting the slope of the fitted line to 1. The composite pK_a calculated by the fitting process described here can be determined by the mathematically equivalent operation of averaging all pK_a predictions with an offset less than 2.0 pH units.

Tables 2 and 3 show pK_a values predicted from 1-ns simulations starting from the 1AKI structure of HEWL. In general, the composite predictions show close correspondence to experimental data, and variation between simulations at different solvent pH values is small. A few problematic cases are worthy of mention. ASP-52 has a hydrogen bond to ASN-46 in all four crystal structures. While this bond is maintained, ASP-52 is prevented from protonating. In most simulations, this hydrogen bond is stable throughout all or nearly all of the simulation, leading to a very low predicted pK_a. The simulations yielding the best results for ASP-52 (pH 2.0, 2.5, and 3.5) were those in which ASP-52 and ASN-46 were dissociated for much of the simulation. A similar effect of hydrogen bonding leading to erroneously low pK_a predictions was seen with GLU-35, but to a lesser extent. Results for the three tyrosine residues were markedly poorer than results for the other residue types. This may be due to slower conformational sampling due to steric hindrance of the large aromatic ring, and ignoring the effect of the polarizability of the aromatic ring.

The overall quality of the pK_a value predictions can be measured by the RMS error of predicted pK_a values relative to experimental values, which is 0.86 for simulations starting from the 1AKI structure, as shown in Table 4. RMS error for null model predictions, where each residue's pK_a is predicted to be equal to the reference value given in Table 1, are also shown in Table 4. The current method gives predictions that are an overall improvement on the null model, and superior for each type of titrating residue, except lysine. Prediction results for lysine are actually more accurate than for any other residue (RMS error 0.64), but due

Table 3. pK_a Predictions for Basic Residues of HEWL, Calculated From 1-ns Simulations Starting from 1AKI at Specified pH.

	pH 9.0	pH 9.5	pH 10.0	pH 10.5	pH 11.0	pH 11.5	pH 12.0	Comp.	Exp.
TYR-20	11.21	10.73	9.91	11.86	11.02	10.82	10.84	10.86	10.3
TYR-23	11.54	12.02	12.43	11.45	11.38	11.08	11.27	11.30	9.8
TYR-53	10.88	10.38	11.08	11.41	10.48	11.18	10.89	10.90	12.1
LYS-13	9.77	9.12	9.54	10.08	9.38	9.38	9.34	9.58	10.5
LYS-33	9.47	9.93	9.44	10.04	9.36	9.55	9.62	9.63	10.6
LYS-96	9.86	9.70	9.96	10.22	10.13	8.98	9.96	9.97	10.8
LYS-97	9.55	9.87	10.08	10.15	10.13	10.17	10.21	10.02	10.3
LYS-116	9.97	10.14	10.27	10.14	10.33	10.11	10.20	10.17	10.4

Composite pK_a is the average of predictions with absolute offset less than 2.0 (see text for discussion). Where experimental pK_a values were given as a range, the midpoint of the range is used; where only an upper bound was given, the upper bound is used.

Table 4. RMS Errors of Predicted pK_a Values From Experimental Values.

	All structures	1AKI	1LSA	3LZT	4LYT	Null model
All residues	0.82	0.86	0.77	0.88	0.95	1.19
Aspartates	0.69	0.80	0.72	0.86	0.78	1.34
Glutamates	0.88	0.94	0.97	1.01	0.54	1.68
Histidine	0.21	0.74	0.11	0.10	0.01	0.69
Tyrosines	1.29	0.88	1.10	1.23	1.69	1.50
Lysines	0.64	0.83	0.57	0.51	0.78	0.21

All structures refers to composite pK_a predictions using data from all simulations on 1AKI, 1LSA, 3LZT, and 4LYT. Null model RMS errors are provided for comparison; in the null model, all residues are predicted to have the reference pK_a values given in Table 1.

to the very small shifts of these residue's pK_a values from reference values, the null model RMS error is very low.

The results reported here are a significant improvement on the explicit solvent TI-based constant pH MD pK_a predictions reported for 1AKI HEWL by Bürgi et al.,[7] which have RMS error of 2.8–3.8 and seem to be far from convergence in 3-ns titrations. They are also more accurate than Lee et al.'s continuous protonation state results for HEWL, which had RMS error of 1.31 for nonterminal residues.[4] The constant pH method employing Poisson–Boltzmann protonation state sampling described by Walczak and Atosiewicz had RMS error of 0.81–1.12 (depending on parameters) in application to ovomucoid third domain.[6] In our opinion, ovomucoid third domain represents an easier prediction problem than HEWL, as it has fewer residues with pK_a values that are substantially shifted relative to reference pK_a values; these strongly shifted residues have the greatest errors in the Walczak and Atosiewicz method. It is difficult to compare the quality of the proposed GB constant pH MD method to that of Baptista et al.,[5] as they have only reported on application of their method to succinic acid, and not to proteins.

The Poisson–Boltzmann-based pK_a prediction methods employed by Baptista et al. and Walczak and Atosiewicz for protonation state sampling have a long history[25] and continue to be an active area of research.[21,24,26] Non-PB-based electrostatics methods have also found success.[27,28] When the best of these electrostatics-only methods are applied to crystal structures they provide somewhat more accurate predictions of pK_a values (RMS error of 0.5–0.7) in less computer time than the proposed GB constant pH MD method. Although these methods are fairly accurate, they can be very sensitive to details of the crystal structure because all atomic positions are fixed, and they often produce widely varying pK_a value predictions for different crystal structures of the same protein.[20] Models (PB and non-PB-based) that allow for some conformational rearrangement have much less dependence on crystal structure,[29–31] and a dynamics-based method should be immune to these effects. This was tested by running simulations starting from three additional crystal structures (PDB) identifiers 1LSA, 3LZT, and 4LYT). These structures were chosen for maximum diversity of crystal characteristics, as described earlier. As seen in Table 5 and summarized in Table 4, pK_a value predictions were highly consistent across the four structures, with a total variation in RMS error of only 0.18 pH units. This stands in contrast to a recent electrostatics study of these structures, which yielded RMS errors of 1.01, 1.44, 1.15, and 2.03 for 1AKI, 1LSA, 3LZT, and 4LYT, respectively.[20]

Non-Henderson–Hasselbalch Behavior

Titrating residues were tested for non-HH behavior [titration curves that do not match the sigmoidal shape defined by eq. (3)] using the procedure described earlier; these results are shown in Table 6. First, it should be noted that, as in most proteins, the magnitude of non-HH behavior is small—in all significant cases, it is less than 0.35 deviation in pK_a prediction for every 1 unit change in solution pH. Furthermore, any error will tend to be canceled by the opposing effects of predictions made from simulations with pH above and below the residue's pK_a, so ignoring non-HH behavior in the lysozyme pK_a calculations above is a reasonable approximation. The results in Table 6 also justify the use of only single-site MC moves for this system—none of the interactions are strong

Table 5. Composite pK_a Predictions for Simulations Starting From the 1AKI, 1LSA, 3LZT, and 4LYT Crystal Structures.

	1AKI	1LSA	3LZT	4LYT	Exp.
ASP-18	1.83	1.69	2.38	2.55	2.66
ASP-48	1.87	2.48	2.04	2.38	2.5
ASP-52	2.08	2.68	1.75	2.65	3.68
ASP-66	2.00	1.18	—	2.19	2.0
ASP-87	2.69	2.66	2.32	3.34	2.07
ASP-101	3.80	3.74	3.76	3.96	4.09
ASP-119	2.60	2.45	2.17	1.98	3.2
GLU-7	3.85	3.72	3.89	3.58	2.85
GLU-35	5.32	5.14	5.23	5.97	6.2
HIS-15	6.45	5.82	5.61	5.70	5.71
TYR-20	10.86	10.82	10.98	11.62	10.3
TYR-23	11.30	11.42	11.39	12.21	9.8
TYR-53	10.90	11.25	10.84	11.10	12.1
LYS-13	9.58	9.87	9.97	9.56	10.5
LYS-33	9.63	9.66	9.94	9.47	10.6
LYS-96	9.97	10.33	10.15	10.04	10.8
LYS-97	10.02	10.04	9.94	9.86	10.3
LYS-116	10.17	10.12	10.19	10.12	10.4

No value is shown for ASP-66 in the 3LZT simulations because none of the predictions had offsets with magnitude less than 2.

Table 6. Hill Coefficients for Titration Data Determined by Linear Regression.

Residue	Hill coefficient	*p*-Value
GLU-7	1.006	0.897
LYS-13	0.966	0.755
HIS-15	0.816	0.100
ASP-18	0.763	0.149
TYR-20	0.740	0.131
TYR-23	0.955	0.864
LYS-33	0.939	0.446
GLU-35	**0.670**	**0.017**
ASP-48	0.995	0.984
ASP-52	0.838	0.528
TYR-53	0.908	0.393
ASP-66	1.003	0.993
ASP-87	**0.751**	**0.039**
LYS-96	**0.796**	**<0.001**
LYS-97	**0.865**	**0.004**
ASP-101	**0.885**	**0.030**
LYS-116	**0.908**	**0.002**
ASP-119	1.076	0.575

p-Value is the significance level at which the Hill coefficient differs from one. Residues with *p*-values less than 0.05 are indicated in boldface.

enough for any titratable group to block protonation state changes in a nearby group. Nevertheless, some interaction between titrating residues leading to non-HH behavior is expected for lysozyme, and it is reassuring that the proposed method reproduces these effects.

The bold lines in Table 6 indicate which residues have statistically significant non-HH behavior. LYS-96 and LYS-97 interact with each other due to their obvious proximity in both primary and tertiary structure. LYS-116 projects toward TYR-23 (titrating N to O distance 7.7 Å in 1AKI); the non-HH effect on TYR-23 is presumably lost in noise due to the poor tyrosine results. ASP-87 most likely interacts with HIS-15, which reaches 90% confidence for non-HH behavior. ASP-101 does not appear to have specific interactions with any single titrating group. However, weak interactions have statistical significance for ASP-101 because it is one of the best converged residues, and as such has little noise. The analysis for GLU-35 is dominated by three data points with very negative errors, representing simulations in which GLU-35 was significantly H-bonded. When these outliers are eliminated, the *p*-value rises to 0.7.

Conformation-Protonation Correlation

A major motivation for the implementation of this method is the idea that protonation state and conformation are strongly coupled, such that they cannot be adequately studied in isolation (e.g., electrostatics-based pK_a predictions and traditional molecular dynamics). Results of these simulations support this idea: we examine the protonation and conformation of ASP-18 in the pH 2.5 simulation starting from 1AKI as an example.

Essential dynamics (ED)—principal component analysis (PCA) of trajectory data—is a useful technique for separating functionally important, slow, large-scale motions from local fluctuations.[32] Projections of a particular snapshot from the trajectory onto the most significant principal components (the largest eigenvectors) can be used as a dimensionally reduced representation of a molecular conformation. Best results in correlating conformation to protonation for a particular residue are obtained when the atoms included in the PCA are limited to the residue and its immediate neighbors. The results presented here are based on PCA[33] of ASP-18 and all atoms within 7.5 Å of ASP-18 in the 1AKI crystal structure. In Figure 6, position of the data points represents conformation (projection onto the two largest eigenvectors) while shade represents degree of protonation (darker is more protonated). A strong qualitative association between conformation and protonation is apparent: the conformational cluster in the upper left is almost entirely deprotonated, while the cluster in the lower right is predominantly protonated.

Plots such as Figure 6 illustrate correlation between conformation and protonation, but deriving quantitative data directly from such a plot necessitates partitioning into conformational clusters which, if done by hand, is subjective and effectively limited to two or three dimensions. Clustering algorithms provide an objective means for identifying clusters and can operate in high-dimensional spaces that are not readily visualized. A *k*-means clustering algorithm[34] (Euclidean distance) was employed to conformationally cluster one thousand 1-ps snapshots, represented by their projections onto the 10 largest eigenvectors, and pK_a was calculated separately for each cluster. There is no obvious choice for the value of *k* (the number of clusters): values that are too low may force distinct conformations with different protonation characteristics into a single cluster, while values that are too high may divide what should be a single cluster into two clusters. It seems

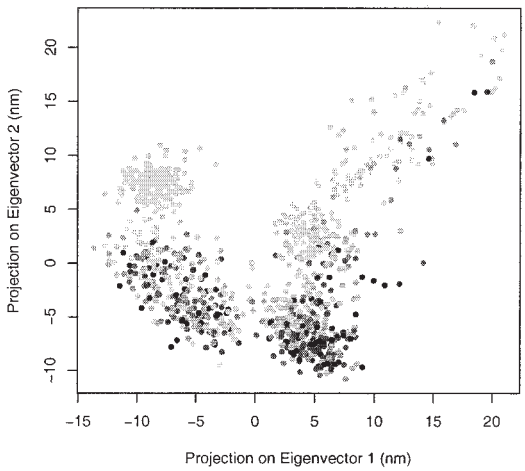

Figure 6. Coupling of conformation and protonation are illustrated in this plot, where spot location represents conformation and shade represents protonation. Specifically, principal component analysis was performed on a 1-ns trajectory at pH 2.5 beginning from 1AKI. Only atoms within 7.5 Å of ASP-18 were included in the analysis. This plot shows the projections of 1-ps snapshots from the trajectory onto the first two (largest eigenvalues) principal components. Shading represents fraction of time spent protonated in the 1-ps window surrounding the snapshot: black represents fully protonated, lightest gray represents fully deprotonated.

Table 7. pK_a Values Calculated for Conformational Clusters.

	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
1	1.95	1.48	0.85	0.91	0.75	0.55
2	2.07	1.95	1.49	1.53	1.39	1.39
3		2.15	2.15	1.64	1.61	1.61
4			2.24	2.26	1.66	1.66
5				2.35	2.24	2.16
6					2.34	2.27
7						2.34

Clusters were generated using the *k*-means algorithm with Euclidean distances. Data points to be clustered were projections of trajectory snapshots onto the first 10 principal components (see text). pK_a values were calculated for each cluster from the combined protonation state data for each snapshot assigned to the cluster.

prudent to try a range of values, increasing k until it is clear that further increases will not identify clusters with unique protonation properties. For instance, in Table 7 it is clear that k of 2 or 3 is too small to separate distinct protonation properties. $k = 4$ identifies a very acidic cluster with pK_a of less than 1.0, a cluster with pK_a near 1.5 and two more clusters with pK_a greater than 2.0. Increasing k beyond 4 serves only to nonproductively subdivide these clusters.

Mapping the clusters from principal component space back to atomic coordinates provides a means to identify the physical basis for the protonation behaviors exhibited by the different clusters. The centroid of each cluster is taken as the representative of the conformations in the cluster. The process of projecting a snapshot onto the principal components is reversed to generate atomic coordinates in Cartesian space from the centroid coordinates in principal component space.

Representations of the $k = 4$ cluster centroids in atomic coordinates are illustrated in Figure 7. These images show that in this trajectory, LYS-13 adopts three distinct conformations, with varying distances from ASP-18 leading to a difference in pK_a of 1.4 between cluster 1 and cluster 4. The dramatic differences in the protonation states sampled in these conformational clusters demonstrates the coupling of the protonation state and conformation, and emphasizes the need to use techniques that maintain this coupling by analyzing protonation state in conjunction with dynamics.

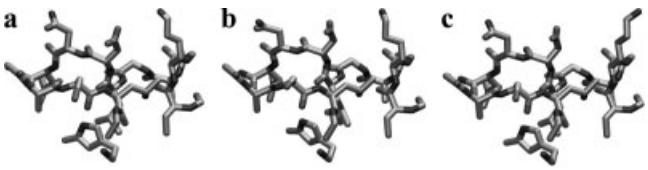

Figure 7. Rendered images[36] representing atomic coordinates corresponding to centroids of clusters 1, 2, and 4 (left to right, having pK_a values 0.85, 1.49, 2.24) from $k = 4$ clustering in Table 7. ASP-18 is top center, ASN-19 is in the upper left, and LYS-13 is in the upper right. Note that bond lengths and angles are somewhat distorted due to the averaging effects of taking the centroid.

Summary

The method described here, combining GB MD with Monte Carlo sampling of discrete protonation states, provides a computationally efficient means for performing constant pH MD. As evidenced by close agreement between predicted and experimental pK_a values, this method accurately samples protonation states while producing a conformationally and energetically stable trajectory. Convergence is rapid for small molecules, but much slower for larger biomolecules. Slow convergence is due to slow conformational sampling and, in systems that have more strongly interacting titratable groups than HEWL, barriers to moving in protonation-state space. We plan to improve the method by combining it with techniques to accelerate conformational sampling, and by incorporating multiple titration site Monte Carlo moves into the protonation state sampling. Results also stand to benefit from continuing improvement in generalized Born models.[35]

The analysis of correlation between conformation and protonation state in these results illustrates the strong coupling between these aspects of molecular configuration; the ability to sample protonation states concurrently with conformation is an important step in improving the physical realism of MD simulations. We hope that the accurate constant pH MD achieved by this method will facilitate the study of pH and protonation state-dependent dynamics that have been inaccessible with traditional MD.

Acknowledgments

We thank Jens Erik Nielsen and Alexey Onufriev for their valuable advice and suggestions. John Mongan is supported by the Medical Scientist Traning Program, the Taft Family of La Jolla through the Taft Fellowship, and Burroughs Wellcome through the La Jolla Interfaces in Science Predoctoral Fellowship.

References

1. Börjesson, U.; Hünenberger, P. H. J Chem Phys 2001, 114, 9706.
2. Baptista, A. M.; Martel, P. J.; Petersen, S. B. Proteins 1997, 27, 523.
3. Onufriev, A.; Case, D. A.; Ullmann, G. M. Biochemistry 2001, 40, 3413.
4. Lee, M. S.; Salsbury, F. R.; C. L. Brooks, I. Proteins 2004, 56, 738.
5. Baptista, A. M.; Teixeira, V. H.; Soares, C. M. J Chem Phys 2002, 117, 4184.
6. Walczak, A. M.; Antosiewicz, J. M. Phys Rev E Stat Nonlin Soft Matter Phys 2002, 66, 051911.
7. Burgi, R.; Kollman, P. A.; Gunsteren, W. F. V. Proteins 2002, 47, 469.
8. Bashford, D.; Case, D. A. Annu Rev Phys Chem 2000, 51, 129.
9. Case, D. A.; Darden, T.; Cheatham, T. E., III; Simmerling, C.; Wang, J.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Duke, R. E.; Crowley, M.; Brozell, S.; Luo, R.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A.; Amber 8, University of California, San Francisco, 2004.
10. Onufriev, A.; Case, D. A.; Bashford, D. J Comput Chem 2002, 23, 1297.
11. Wang, J.; Cieplak, P.; Kollman, P. A. J Comput Chem 2000, 21, 1049.
12. Onufriev, A.; Bashford, D.; Case, D. A. J Phys Chem B 2000, 104, 3712.

13. Onufriev, A.; Bashford, D.; Case, D. A. Proteins 2004, 55, 383.
14. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
15. Wang, J.; Wang, W.; Kollman, P. A. Am Chem Soc Abstr 2001, 220, 135.
16. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head–Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98 (revision a.11.4), Gaussian Inc.: Pittsburgh, PA, 1998.
17. Wolfram Research Inc.; Mathematica, version 5, Champaign, IL, 2003.
18. Simonson, T.; Carlsson, J.; Case, D. A. J Am Chem Soc 2004, 126, 4167.
19. Chen, J. L.; Noodleman, L.; Case, D. A.; Bashford, D. J Phys Chem 1994, 98, 11059.
20. Nielsen, J. E.; McCammon, J. A. Protein Sci 2003, 12, 313.
21. Demchuk, E.; Wade, R. C. J Phys Chem 1996, 100, 17373.
22. Vriend, G. J Mol Graph 1990, 8, 52.
23. Hooft, R. W.; Sander, C.; Vriend, G. Proteins 1996, 26, 363.
24. Wlodek, S. T.; Antosiewicz, J.; McCammon, J. A. Protein Sci 1997, 6, 373.
25. Gilson, M. K.; Honig, B. H. Proteins Struct, Funct Genet 1988, 3, 32.
26. Nielsen, J. E.; Vriend, G. Proteins 2001, 43, 403.
27. Warshel, A. Biochemistry 1981, 20, 3167.
28. Mehler, E. L.; Guarnieri, F. Biophys J 1999, 77, 3.
29. Sham, Y. Y.; Chu, Z. T.; Warshel, A. J Phys Chem B 1997, 101, 4458.
30. Alexov, E. G.; Gunner, M. R. Biophys J 1997, 72, 2075.
31. Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. Biophys J 2002, 83, 1731.
32. Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Proteins Struct, Funct Genet 1993, 17, 412.
33. Lindahl, E.; Hess, B.; van der Spoel, D. J Mol Model 2001, 7, 306.
34. de Hoon, M. J. L.; Imoto, S.; Nolan, J.; Miyano, S. Bioinformatics 2004, 20, 1453.
35. Feig, M.; Onufriev, A.; Lee, M.; Im, W.; Case, D. A.; Brooks, C. L. J Comput Chem 2004, 25, 265.
36. Humphrey, W.; Dalke, A.; Schulten, K. J Mol Graph 1996, 14, 33.
37. Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. Biochemistry 1993, 32, 8045.
38. Kyte, J. Structure in Protein Chemistry; Garland Publishing, Inc.: New York, 1995, p. 64.