JCTC Journal of Chemical Theory and Computation

Coupling Constant pH Molecular Dynamics with Accelerated Molecular Dynamics

Sarah L. Williams,*,[†] César Augusto F. de Oliveira,[†,§] and J. Andrew McCammon[†,‡,§,ll]

Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, California 92093-0365, Center for Theoretical Biological Physics, University of California San Diego, La Jolla, California 92093, Howard Hughes Medical Institute, University of California San Diego, La Jolla, California 92093-0365, Department of Pharmacology, University of California San Diego, La Jolla, California 92093-0365

Abstract: An extension of the constant pH method originally implemented by Mongan et al. (*J. Comput. Chem.* **2004,** *25,* 2038−2048) is proposed in this study. This adapted version of the method couples the constant pH methodology with the enhanced sampling technique of accelerated molecular dynamics, in an attempt to overcome the sampling issues encountered with current standard constant pH molecular dynamics methods. Although good results were reported by Mongan et al. on application of the standard method to the hen egg-white lysozyme (HEWL) system, residues which possess strong interactions with neighboring groups tend to converge slowly, resulting in the reported inconsistencies for predicted pK_a values, as highlighted by the authors. The application of the coupled method described in this study to the HEWL system displays improvements over the standard version of the method, with the improved sampling leading to faster convergence and producing pK_a values in closer agreement to those obtained experimentally for the more slowly converging residues.

Introduction

It is well-known that the structure and function of a protein are highly dependent on the pH of its surrounding aqueous environment due to pH-mediated changes in the protonation state of titratable residues. The protonation state of a titratable residue in a protein is determined by its pK_a and the solution pH, the former being a measure of the relative acidity of the residue, which is influenced by interactions with neighboring residues, including titratable residues. These changes in protonation equilibrium, which are essentially of electrostatic nature, are closely linked with the conformation and are fundamental in the definition of the often-narrow pH range for the functioning protein, beyond which unfavorable conformational change and denaturation of the protein structure may occur.

The important pairing of protonation state and protein conformation is not accounted for in standard molecular dynamics (MD) simulations. Currently, the majority of standard simulations of biological systems use fixed, predetermined protonation states for titratable residues, which are generally based on the pK_a values of the isolated residue in solution. In addition, protonation states are usually assigned during the preparation of the system and are not changed throughout the standard MD simulation. This method of protonation state assignment is a severe approximation, as the pK_a values of titratable residues are frequently shifted from that of the model residue in solution, making the assignment a nontrivial task. Furthermore, protonation states are not single constant values; they are subject to the changing electrostatic environment surrounding the titratable group. Therefore, incorporating pH as an input variable in MD simulations is highly desirable, as it would allow a more

* Corresponding author phone: 858-822-0168; fax: 858-534-4974; e-mail: swilliam@mccammon.ucsd.edu.
[†] Department of Chemistry & Biochemistry.
[§] Howard Hughes Medical Institute.
[‡] Center for Theoretical Biological Physics.
[ll] Department of Pharmacology.

Coupling Constant pH

J. Chem. Theory Comput., Vol. 6, No. 2, 2010 **561**

accurate study of pH-coupled conformational phenomena, such as reaction mechanisms, ligand binding, and the determination of the structure and function of proteins as a function of pH.

Over approximately the past 15 years, several methods have been proposed which enable MD to be carried out at a constant pH with changing protonation states. These constant-pH MD (CpHMD) methods can be largely classified into two categories, discrete[1−4] and continuous.[5−7] Several reviews have been published which compare and contrast the different methods.[8−10] In the following paragraphs, a brief description of some of these methods is given.

Continuous protonation state models, such as that of Börjesson and Hünenberger[6,7] and Baptista et al.,[5] consider protonation state as a continuous titration parameter, which advances simultaneously with the atomic coordinates of the system. However, these methods use a mean-field approximation, whereby they do not take into account any interaction with other nearby titratable residues that may occur, and the titratable groups are represented by fractional, nonphysical protonation states, intermediate between the protonated and unprotonated forms.[11,12] These factors cause the models to perform poorly for tightly coupled residues[7] and result in inadequate estimation of physical observables. The more recent work of Lee et al.[13] overcomes the issues with unphysical fractional protonation states with the use of λ dynamics with the addition of an artificial titration barrier along the continuous titration coordinate between the fully protonated and deprotonated end points. This has the effect of forcibly lengthening simulation time in the fully protonated or deprotonated values. The authors report good correlation between the predicted and experimental pK_a values for the hen egg-white lysozyme (HEWL), turkey ovomucoid, and bovine pancreatic trypsin inhibitor, although convergence issues were encountered for these systems, and even for the simpler aspartate model. Khandogin and Brooks[14] developed an extension to this method, a two-dimensional λ-dynamics method using GBSW[15] solvation. The two dimensions are the two reaction coordinates: the deprotonation process and the interconversion between proton tautomers, to account for proton tautomerism in simulations involving histidine and carboxyl residues. The authors observe significant quantitative improvement over the previous work of Lee et al.[13] and note that the method could be further improved with enhanced sampling and an improved solvent model. In other work by the same group,[16] the continuous titration method is coupled with replica exchange (coupled method known as REX-CPHMD), used with an improved GB solvent model[17] and a salt-screening function, to achieve more accurate predictions of pK_a shifts when applied to 10 test protein systems, all possessing residues with significant pK_a shifts.

The majority of the more recent studies have involved the use of discrete protonation state models, which avoid the nonphysical intermediate charge states. These methods use MD simulations for conformational sampling, with periodic sampling of discrete protonation states through trial Monte Carlo (MC) moves. The main differences between these methods lie in their choice of solvent model

and the protocol for updating the protonation states.[1−4] The methods employing explicit solvent are computationally expensive, and MC trial moves are attempted relatively infrequently, causing long convergence times for systems with multiple titration sites. Both Bürgi et al.[18] and Baptista[1] et al. developed methods using explicit solvent. Baptista et al. used Poisson−Boltzmann (PB) electrostatics for the calculation of protonation state energies to be used for the MC test. However, the PB calculations are time-consuming and introduce a solvent potential different from that used for the explicit-solvent dynamics. Bürgi et al. avoid the discrepancy in the potentials with the use of thermodynamic integration (TI) under the same explicit solvent conditions as used for the dynamics, to determine the transition energies for the MC test. However, this has the effect of perturbing the trajectory, even when the MC trial is rejected, since the final trajectory is formed from the concatenation of the TI segments. In addition, the length of time over which the TI calculations are performed (∼20 ps) makes their significance uncertain, and the expense of explicit solvent is a probable contributor to the apparent poor convergence of the simulations.[8] Stern introduces a method whereby the issues associated with instantaneous protonation state change when using explicit solvent are circumvented, with the use of a hybrid Monte Carlo procedure.[19] In this method, the trial moves comprise relatively short MD trajectories, which employ a time-dependent potential energy that interpolates between the old and new protonation states. The method has been successfully applied to acetic acid in water but has not yet been applied to protein systems.

Methods employing implicit solvent for both the dynamics and MC steps include the work of Dlugosz and Antosiewicz,[2,3] who use PB calculations in the calculation of transition energies, and the analytical continuum electrostatics (ACE/GB) model of CHARMM for dynamics. Again, this method has the problems associated with the expense of PB calculations and the mismatch in potentials used, although the method reports fair agreement with experiment when applied to a heptapeptide derived from the ovomucoid third domain and succinic acid. Mongan et al. use GB solvation for both the MC steps and the dynamics,[4] therefore avoiding the discrepancy in the potentials used, with pK_a predictions agreeing well with experimental results on application to the HEWL system, although convergence issues are noted for some residues of the system.

In this work, we propose an extension to the constant pH model of Mongan et al., whereby the methodology is coupled with the enhanced sampling technique of accelerated molecular dynamics (aMD) to increase the sampling and alleviate the reported convergence issues. This version of the method has been implemented in AMBER8 and has been applied to the popular test case, the HEWL system. The results show improvement in the sampling compared with the standard Mongan et al. method, with pK_a results close

to those obtained experimentally for the more problematic, more slowly converging residues.

Background

The standard constant pH (Mongan et al.) and aMD methods (de Oliveria et al.) coupled in this study are described in detail elsewhere, and so only outlines of the techniques are given here. The CpHMD method described here differs from the standard method in that the sections of conventional MD are replaced with the enhanced sampling technique, aMD. The combined method is denoted as CpHaMD in this work.

The method employs GB-solvated aMD with periodic MC sampling of protonation states, also using the same GB electrostatics. At each MC step, a titratable residue and a new protonation state are chosen at random, with the total transition energy, ΔG, being used as the Metropolis criterion for the decision of protonation state. The calculation of ΔG is shown in eq 1, where k_B is the Boltzmann constant, T is the temperature, pH is the specified solvent pH, $pK_{a,ref}$ is the pK_a of the reference compound, ΔG_{elec} is the electrostatic energy component of the titratable residue, and $\Delta G_{elec,ref}$ is the electrostatic component of the transition energy for the reference compound. If the MC move is accepted, the protonation state of the residue will change to the new state and MD is continued; if not, the simulation will continue with the residue remaining in the unchanged protonation state.

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

In the previous implementation of the method, conventional MD was employed between the MC steps. In the version of the method reported in this work, standard MD is replaced with aMD. As mentioned previously, a limitation of constant pH methods is often convergence,[8,16] therefore implying the performance of the method may be improved by the use of enhanced sampling techniques, as shown by Khandogin and Brooks with their REX-CPHMD method.[16] Here, a recently modified version of the dual-boost aMD method (referred to as aMDtT[b] in the literature) by de Oliveira et al. is used (a modification of the Hamelberg et al. aMD methodology[20]), which has been found to be useful in improving the accuracy and convergence of TI simulations. This approach increases conformational sampling through the modification of the energy landscape by lowering energy barriers while leaving the potential surface in the vicinity of the minima unchanged. The energy barriers are reduced through the application of a boost potential, $\Delta V(r)$, to the true potential surface, $V(r)$, in cases where the true potential exceeds a predefined energy level, E. The boost potential is implemented in the method according to eq 2, where α modulates the shape of the modified potential (lower values of α result in a flatter modified potential, and higher values approach the unmodified potential).

$$\Delta V(r) = \begin{cases} \dfrac{(V(r) - E)^2}{\alpha + (V(r) - E)}, & V(r) \geq E \\ 0, & V(r) < E \end{cases} \quad (2)$$
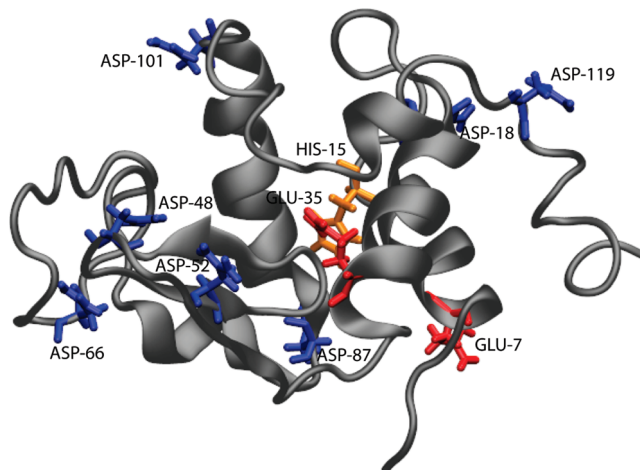

Figure 1. The HEWL enzyme (PDB ID: 1AKI) with titratable groups highlighted in liquorice style (aspartates, blue; glutamates, red; and histidine, orange).

In cases where the true potential exceeds the boost energy level E, the boost potential is subtracted from the true potential, and the simulation is performed on this modified potential surface $V^*(r) = V(r) - \Delta V(r)$. At times where the true potential lies below the boost energy level, E, the simulation is performed on the true potential, $V^*(r) = V(r)$.

In this work, the dual-boost approach[21] has been used in order to increase the sampling of both the torsional degrees of freedom and the atomic packing arrangements. The first boost is applied to only the torsional terms, $V_t(r)$, and the second boost is added to the total potential energy, $V_T(r) = V_0(r) + V_t(r)$ (eq 3).

$$V^*(r) = \{V_0(r) + [V_t(r) - \Delta V_t(r)]\} - \Delta V_T(r) \quad (3)$$

The correct canonical averages of an observable, in this case pK_a, are calculated from configurations sampled on the modified potential energy surface and are fully recoverable by reweighting each point in configuration space by eq 4.

$$\exp\{-\beta[\Delta V_t(r) + \Delta V_T(r)]\} \quad (4)$$

Test Case: Hen Egg White Lysozyme

HEWL is a 129-residue monomeric single-domain enzyme which catalyzes the hydrolysis of polysaccharides found in many bacterial cell walls (see Figure 1). The enzyme is known to possess several residues with pK_a values significantly shifted from the model isolated residue values.[22,23] Additionally, it is a well-known example of an enzyme which employs a proton donor and a catalytic nucleophile (Asp 52 and Glu 35)[24] within the active site, located in a cleft between an all-α and a β-rich region. Owing to extensive experimental study of this system, and the challenging nature of the pK_a shifts of some of the residues, HEWL has been a popular test system for many of the pK_a calculation methods. In this study, focus is placed on the acidic residues of this enzyme, which have been experimentally determined to possess the most significant pK_a shifts of the system.

Molecular Dynamics Simulations

The standard CpHMD and coupled CpHaMD methods have been implemented in the AMBER 8 molecular dynamics

Coupling Constant pH

J. Chem. Theory Comput., Vol. 6, No. 2, 2010 **563**
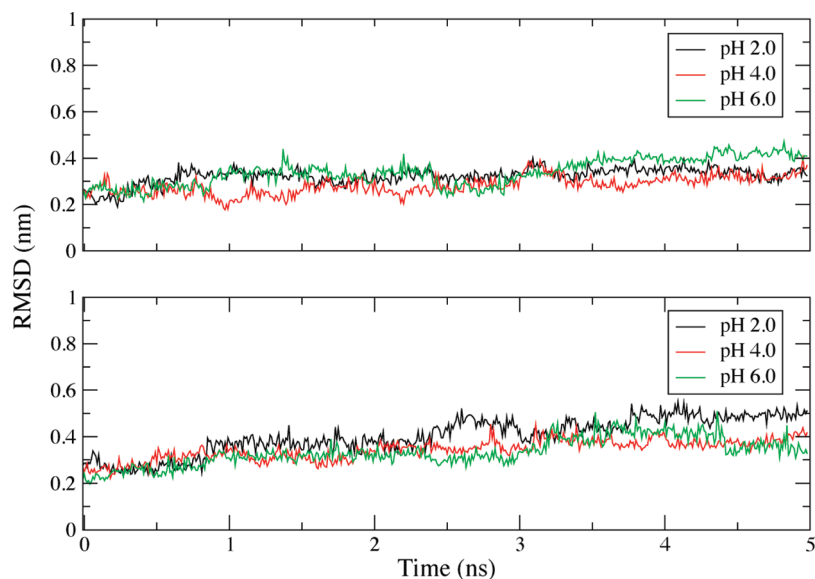

Figure 2. The RMSD of CA atoms with respect to the crystal structure, over the duration of 5 ns simulations carried out at pH 2.0, 4.0, and 6.0 using CPH-aMD (lower plot) and CpHMD (upper plot).

program. This study follows from the work of Mongan et al., and the parameters used match those used in their work. All simulations described employed the AMBER99 force-field[25] and the GB solvent model[26−28] (igb=2). Salt concentrations were set at 0.1 M, and a 30 Å cutoff value for nonbonded interactions and effective Bond radii calculations was used. The SHAKE algorithm was used to constrain all bonds involving hydrogen, allowing a time step of 2 fs to be used. The temperature was maintained at 300 K using the Berendsen temperature coupling method with a time constant of 2 ps. A period of 10 fs of MD or aMD separated the MC trials.

For the HEWL system, values for the boost energy, E, applied to the torsional degrees of freedom and the total potential energy were estimated on the basis of the average torsional and total potential energies and the root mean square (RMS) deviation in these energies over CpHMD simulations carried out on the unmodified potential at the pH of interest. The parameter, E, was calculated from subtracting the sum of twice the RMS deviation from the average potential and torsional energies. The value of the α parameter for the total potential energy was estimated to be ∼5 kcal/atom, and for the torsional potential, a value of ∼30% of the average dihedral potential energy, obtained from the simulation carried out on the unmodified potential, was found to be efficient.

All simulations were started from the minimized 1AKI (PDB ID) crystal structure, as prepared by Mongan et al. (details given in ref 4). Simulations of 5 ns in length were carried out in the pH range 2−6.5 at 0.5 pH unit intervals, using both CpHMD and CpHaMD methods. GLU and ASP residues were set to titrate from pH 2.0 to pH 6.5, with the addition of HIS residues from pH 4.5 to pH 6.5. HIS residues were not set to titrate for the most acidic simulations, as they are likely to remain in the protonated state in this pH range, as indicated by its model pK_a value of 6−7.[29] Models for the terminal residues have not been developed yet for this system, so these residues were set to their neutral

protonation states. This approximation has been deemed to be sufficient for these simulations, as explained in the prior work on this system, by Mongan et al. All nontitrating residues were set to their expected protonation states.

Extended Simulations. To further investigate the effects of CpHaMD, simulations at pH 3 and pH 6.5 were extended to 40 ns in triplicate, the further two simulations initialized from different random seeds, and generated from re-equilibration of the minimized structure. The pH values were chosen as they represented two different regions of the acidic pH range, pH 3 being close to the experimental pK_a values for the majority of the residues and pH 6.5 being more challenging in obtaining convergence and accurate pK_a evaluations due to the many residues of interest being in the deprotonated state.

Principal Component Analysis (PCA). Details of PCA can be found elsewhere in the literature.[30,31] The GROMACS analysis program,[32] g_covar, was used for the calculation and diagonalization of the covariance matrix, with the analysis of the resultant eigenvectors performed using g_anaeig. The covariance matrix of positional fluctuation was calculated for atoms of the residue of interest and atoms in the vicinity, within a distance 7.5 Å, from the 12 concat-enated trajectories of 40 ns (CpHMD: pH 3.0, three simulations; pH 6.5, three simulations; CpHaMD: pH 3.0, three simulations; pH 6.5, three simulations). The two-dimensional plots were generated from the projection of the trajectories onto the first two eigenvectors.

Results

Simulation Stability. Initially, simulations of 5 ns in length were performed (as described in the Molecular Dynamics Simulations section). Figure 2 monitors the root-mean squared deviation (RMSD) of the Cα atoms, with respect to the crystal structure, over the duration of the 5 ns simulations at pH 2.0, 4.0, and 6.0. Simulations employing the standard CpHMD and the adapted CpHaMD method are

Table 1. pKa Predictions of Titratable Residues of the HEWL Enzyme over the Acidic pH Range[a]

residue	simulation pH										av.	exptl. value
	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		
ASP-18	2.62	2.92	2.47	2.24	2.41	2.34	2.58	2.38	−	−	2.50	2.66
	(2.50)	(2.99)	(2.53)	(1.69)	(2.07)	(2.27)	(2.15)	(2.78)	(2.26)	(−)	(2.36)	
ASP-48	2.70	**0.35**	2.83	2.38	3.34	1.98	1.96	3.06	**1.32**	2.01	2.19	2.50
	(2.39)	(2.23)	(2.70)	(2.82)	(2.86)	(3.24)	(**0.30**)	(3.42)	(2.66)	(1.80)	(2.44)	
ASP-52	**1.99**	1.17	2.36	2.17	2.14	−0.1	1.34	3.27	3.37	4.71	2.24	3.68
	(**2.18**)	(**2.45**)	(**2.52**)	(**1.33**)	(**1.72**)	(**2.45**)	(**2.25**)	(2.87)	(−)	(**2.57**)	(**2.26**)	
ASP-66	2.71	**3.28**	2.47	2.15	**3.12**	2.14	2.62	2.31	2.52	2.93	2.63	2.00
	(2.89)	(2.79)	(2.87)	(2.59)	(**3.44**)	(**3.43**)	(**1.79**)	(−)	(2.14)	(2.50)	(2.72)	
ASP-87	2.51	2.25	2.38	1.76	2.25	2.69	2.43	2.85	**3.29**	3.21	2.56	2.07
	(2.42)	(2.41)	(2.91)	(2.80)	(2.70)	(1.84)	(3.00)	(2.21)	(3.23)	(**3.18**)	(2.67)	
ASP-101	3.46	3.42	**2.82**	3.17	3.93	**2.88**	3.45	3.73	−	3.63	3.50	4.09
	(3.19)	(3.29)	(**2.39**)	(3.12)	(**2.59**)	(3.24)	(3.48)	(3.96)	(3.33)	(3.91)	(3.25)	
ASP-119	**2.07**	3.55	2.25	2.52	2.50	2.35	2.64	**1.92**	2.36	**1.21**	2.34	3.20
	(2.73)	(**2.08**)	(2.21)	(2.90)	(2.36)	(2.45)	(2.17)	(3.06)	(**2.00**)	(2.40)	(2.44)	
GLU-7	3.62	3.77	3.61	3.66	3.70	3.67	3.77	3.81	3.85	3.56	3.70	2.85
	(3.64)	(3.53)	(3.73)	(3.63)	(3.70)	(3.60)	(3.80)	(3.68)	(4.11)	(4.12)	(3.75)	
GLU-35	5.51	5.76	6.06	5.61	**5.02**	6.22	**4.92**	5.13	**4.72**	4.54	5.35	6.20
	(**4.65**)	(**4.75**)	(**4.76**)	(5.79)	(**4.17**)	(6.33)	(5.51)	(**3.05**)	(5.91)	(5.46)	(**5.04**)	
HIS-15	NM	NM	NM	NM	NM	3.94	5.20	5.48	6.52	**7.25**	5.68	5.71
						(**4.09**)	(5.47)	(4.85)	(**7.28**)	(**7.45**)	(5.83)	

[a] Results generated using the standard constant-pH methodology (lower values) and using the aMD-modified approach (upper values). Average values (av.) were calculated for each residue from 5 ns simulations performed at the indicated pH values. The pKa of HIS-15 was not measured (NM) at lower pH values. Where a value is missing (−), the pKa of that residue was unable to be measured owing to zero transitions in protonation state occurring over the duration of the simulation. Values highlighted in bold are >1 pKa unit from the experimentally reported range.[34]

shown to be reasonably stable, with no major domain motion over the 5 ns. This is in agreement with experimental evidence; the HEWL enzyme has been experimentally reported to be stable over a wide range of pH values, including a pH stability screen carried out in the range of pH 3−8 which revealed HEWL to be very stable at pH 4, 5, and 8.[33]

pK_a Predictions Calculated from 5 ns Simulations. A summary of pK_a predictions for residues titratable over the acidic pH range, calculated from the set of 5 ns simulations performed in this study, is shown in Table 1. At each pH value, the predicted pK_a was calculated according to the Henderson−Hasselbach equation, with the ratio of time that a titratable group spends in the protonated and deprotonated states used as a ratio of concentrations. For CpHaMD simulations, each state is reweighted by eq 4, before the ratio of concentrations is calculated. For comparison with experimental pK_a values, a composite pK_a value for each residue was obtained from the combination and averaging of the individual pK_a values generated at the various pH conditions by the CpHaMD and CpHMD simulations.

Over the pH range studied, the simulations predict pK_a values of titratable residues which correlate well with experimental values (good correlation is deemed as <1 pK_a unit deviation from the experimental range given in ref 34). However, both methodologies predict pK_a values which deviate more than 1 pK_a unit from the experimental results for several residues (highlighted in bold in Table 1), and significant variation is observed for some residues between predictions made for the same residue at different pH values (for example, the calculated pK_a for ASP-52 at pH 4.5 is −0.1, and that at pH 5.5 is 3.27), indicating a lack of convergence for these residues.

Mongan et al.[4] employed the constant-pH method for a set of 1 ns simulations of the HEWL system and reported

pK_a predictions for a range of pH values. As observed in this study, they also obtain good overall calculated pK_a values with inconsistency in predictions obtained from different simulations for some of the more strongly interacting titratable groups. They suggest one limitation of the method to be its inability to sufficiently sample conformational space, as, due to the dependence of instantaneous pK_a on conformation, limited conformational sampling would restrict the accuracy of pK_a prediction, especially for the more slowly converging residues of the system (e.g., the more buried Asp-52 and Glu-35 residues). In this study, measurement of the calculated pK_a over the duration of the simulations reported here (Figure 3) shows that 5 ns is still an insufficient simulation time to observe convergence for all residues, even for simulations using the enhanced sampling methodology.

The most problematic case across the pH range is shown to be ASP-52, one of the catalytically crucial residues, residing within the active site (see Figure 1). Residues located within the protein experience a very different electrostatic environment from the isolated model residue, resulting in significant shifts in pK_a. It is common for such residues to form strong interactions with residues in the vicinity, which often causes sampling problems with the use of conventional MD, owing to the slower convergence of these residues. In addition to convergence issues, deficiencies in the GB solvent model used have been highlighted in the literature.[16] For buried residues, the GB model underestimates the desolvation energy and buried charge−charge interactions owing to neglect of the solvent excluded volume. Although an improved GB model would certainly improve results, this study is only focused on the sampling issues associated with the constant pH method and improvement of results with the use of enhanced sampling. In the case of ASP-52, an interaction with ASN-46 is shown to persist for the duration of several simulations, causing the calculated pK_a to be
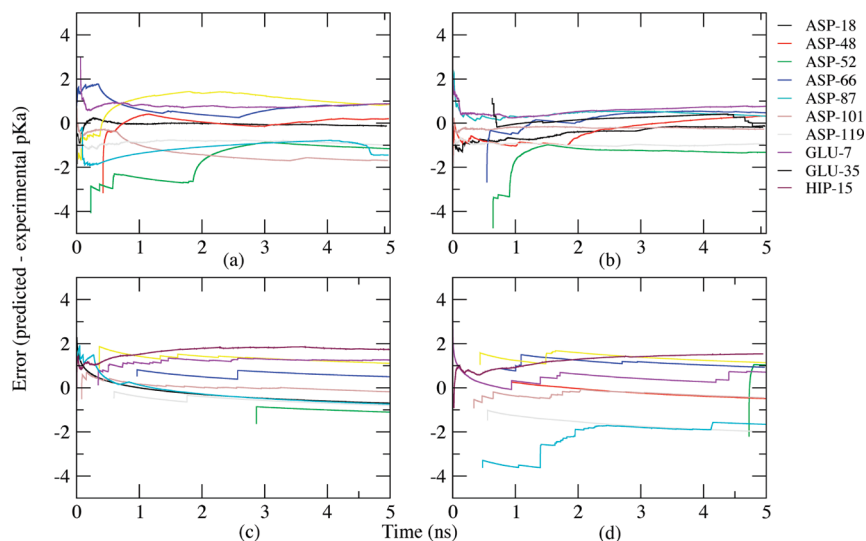
Coupling Constant pH

J. Chem. Theory Comput., Vol. 6, No. 2, 2010 **565**

Figure 3. pK_a values of titratable residues over the duration of 5 ns simulations employing CpHMD (top two plots) and CpHaMD (lower two plots) at pH 3.0 (plots a and c) and pH 6.5 (plots b and d).

Table 2. RMS Error of pKa Values Calculated from 5 ns Simulations over the pH Range of 2–6.5, with Respect to the Mid-Point Experimental pKa Value

	RMS error		
	CpHaMD	CpHMD	null model
all residues	0.73	0.80	1.39
aspartates	0.75	1.46	1.34
glutamates	0.85	1.04	1.68
histidine	0.03	0.12	0.69

Table 3. RMS Error Values Calculated for Acidic Residues of HEWL Listed in Table 1[a]

methodology	RMS error
null model	1.39
Bürgi et al. (2002)[18]	2.97
Lee et al. (2004)[13]	2.28
Mongan et al. (2004)[4]	0.82
Khandogin and Brooks (2006)[16]	0.65,[b] 1.19[c]
Machuqueiro and Baptista (2008)[35]	PME: 0.79–0.93;[d] GRF: 0.84
this work	CpHMD: 0.8; CpHaMD: 0.73[e]

[a] Values calculated with respect to the mid-point of the experimental range given in ref.[34] [b] Calculated from simulations carried out in the implicit solvent model in the presence of salt effects. [c] Carried out in the implicit solvent model in the absence of salt effects. [d] Range of values for the dielectric constant used. [e] Calculated from 5 ns CpHaMD simulations.

notably lower than the experimentally reported pK_a range of 3.6–3.76. As also observed in the constant-pH study of Mongan et al., persistent interactions with neighboring residues also exist with GLU-35, but to a lesser degree when compared with ASP-52.

RMS Error Analysis. The RMS error was calculated for groups of residue types, with respect to the midpoint of the experimental value range[34] (Table 2). The RMS error was also calculated for the null model, against which both methods are shown to perform considerably better.

As stated earlier, although the simulations do not appear to have reached convergence for all residues, overall, the CpHaMD method is indicated to predict pK_a values which are closer to experimental results, confirmed by the lower RMS error values reported in Table 2. The RMS error for the single histidine residue included in the calculations has the lowest error and is reported experimentally to only have a small shift in pK_a from the model compound. This histidine residue resides on the surface, away from the active site, and possesses interactions with neighboring residues, including Thr-89, which are overestimated in a few of the simulations, resulting in the higher predicted pK_a values. The groups of aspartates and glutamates both contain residues exhibiting larger pK_a shifts and reside in more buried locations of HEWL, resulting in the relatively higher RMS errors.

In addition, the results of both the CpHMD and CpHaMD methods are shown to perform well on comparison with those achieved using other CpHMD methodologies (Table 3),

where the leading results have RMS errors in the range of 0.65–0.8 for the acidic residues. The constant-pH methods achieving good results when applied to the HEWL system include the more recent method of Khandogin and Brooks,[16] who combine the constant-pH method with replica exchange and an improved GB implicit solvent model, attaining pK_a predictions with a RMS error of 0.65 and 1.19 (RMS error calculated from simulations including and excluding salt effects, respectively). Machuqueiro and Baptista[35] achieve a RMS error value of 0.84 with the inclusion of proton tautomerism in their CpHMD method, and incorporating the generalized reaction field (GRF) for the treatment of long-range electrostatics. The RMS error increases to 0.79–0.93 when changing to the particle-mesh Ewald (PME) algorithm. Bürgi et al.[18] use constant-pH with TI and MC for the protonation state determination and achieve only qualitative pK_a results, denoted by a RMS error of 2.97, which has been attributed to inadequate simulation time for convergence.

Good results have also been achieved for HEWL using Poisson–Boltzmann-based pK_a calculations.[36–42] However, no good method has yet been developed which accounts for significant conformational change, and generally, the current methods are likely to be insufficient in cases where conformational change has a large influence on residue pK_a.[11] The
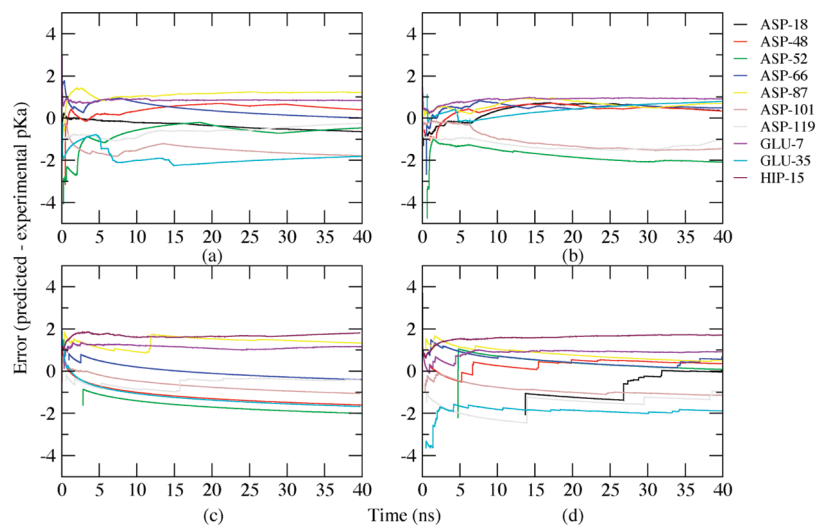
Figure 4. pK_a values of titratable residues over the duration of 40 ns CPMD (top two plots) and CpHaMD (lower two plots) simulations at pH 3.0 (plots a, c) and pH 6.5 (plots b, d).

Table 4. Average pKa Predictions Calculated from Three 40 ns Simulations Using CpHMD and CpHaMD Methods[a]

residue	CpHMD		CpHaMD		exptl. value
	pH 3.0	pH 6.5	pH 3.0	pH 6.5	
ASP-52	**2.47** (1.19)	**1.67** (−)	3.73 (0.67)	3.62 (0.78)	3.68

[a] The standard deviation is noted in brackets. (−) indicates pKa prediction only possible in one of three simulations.

CpHMD methods, such as those described here, are attractive since they incorporate flexibility and offer the ability to study the dynamics of pH-dependent phenonoma.

Extended Simulations. Simulations were extended to 40 ns and performed in triplicate for two pH values in different regions of the acidic pH range (pH 3.0 and pH 6.5), increasing the opportunity for conformational change and, thus, to test whether a further increase in conformational sampling would improve the accuracy of the pK_a prediction for the more challenging residues.

As Figure 4 demonstrates, the extended simulations carried out using the CpHaMD method appear to converge faster and progress closer toward the experimental values for a larger number of residues when compared with simulations carried out using the standard CpHMD method. This is especially pronounced for simulations carried out at pH 3.0, which is expected, since the majority of the experimental pK_a values lie closer to this pH. At pH 6.5, the majority of the measured residues are in the deprotonated state, and the calculation of the pK_a becomes very sensitive as a result of the relative magnitude of unprotonated to protonated states becoming very small, due to the considerably smaller number of transitions between protonation states.

The pK_a predictions for the previously mentioned problematic residue, ASP-52, are significantly improved using the CpHaMD method, with all six simulations (three simulations at pH 3 and three simulations at pH 6.5) generating values within 1 pK_a unit of the experimental range (see Table 4). A greater variation, indicated by the larger standard deviations, is observed between the

simulations employing the standard CpHMD method, with no value at all calculated for two of the three simulations carried out at pH 6.5. For these simulations, the residue has a strong tendency to become stuck in the deprotonated state, owing to a continued hydrogen bond with residue ASN-46. This interaction is present in the crystal structure and was noted to persist in the study using the standard version of this constant pH method.[4] ASN-46 resides in a loop region of HEWL, which does not display significant mobility in the CpHMD simulations. However, in simulations employing CpHaMD, this region undergoes increased conformational motion, as highlighted by the larger sampling area of the CpHaMD simulations, shown in the two-dimensional plots, generated from PCA analysis (Figure 5). This increased loop motion facilitates the dissociation of the interaction between the two residues, as when the loop moves away, the interaction is lost and ASP-52 is able to interchange to the protonated state (Figure 6). Within CpHaMD simulations, the aforementioned interaction is observed to repeatedly dissociate and reform, depending on local conformational change, illustrated by the number of transitions occurring between the protonated and deprotonated forms of the residue throughout the simulation (example shown in Figure 7). Over the three CpHaMD simulations at pH 6.5, an average of 181 transitions were recorded, whereas for the one CpHMD simulation at pH 6.5, for which a pK_a could be calculated, only two transitions took place throughout the 40 ns of CpHMD simulation. For less problematic residues, the number of transitions is far higher during CpHMD simulations at 6.5, with >10 000 transitions recorded for some residues.

Overall, the initial application of this newly coupled aMD enhanced sampling technique to the standard constant pH methodology of Mongan et al. signifies the CpHaMD technique to be promising in improving the convergence of constant-pH simulations, providing more accurate pK_a pre-

Coupling Constant pH

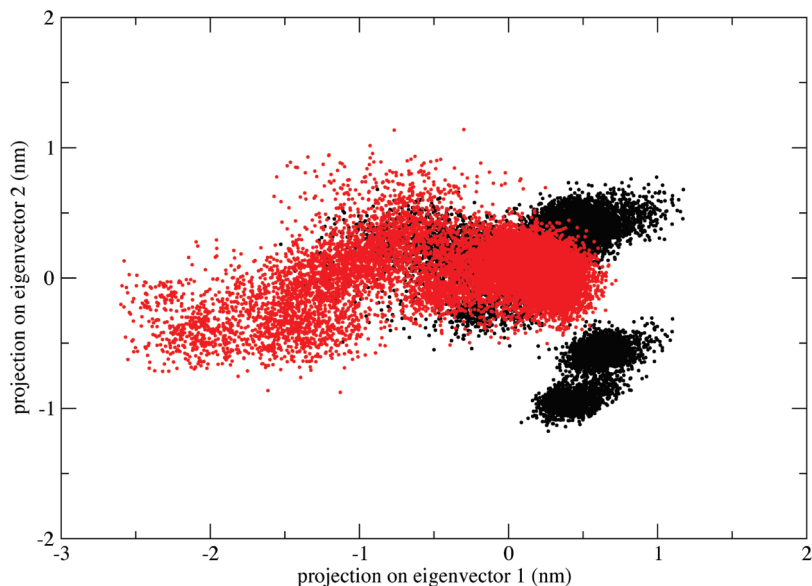*J. Chem. Theory Comput., Vol. 6, No. 2, 2010* **567**

Figure 5. Conformational sampling of residues (502 atoms) within 7.5 Å of ASP-52 demonstrated by PCA analysis. Eigenvectors generated from the concatenation of trajectories of simulations carried out at pH 6.5. Red: sampling from simulation carried out using CpHaMD. Black: sampling from simulation carried out using CpHMD.
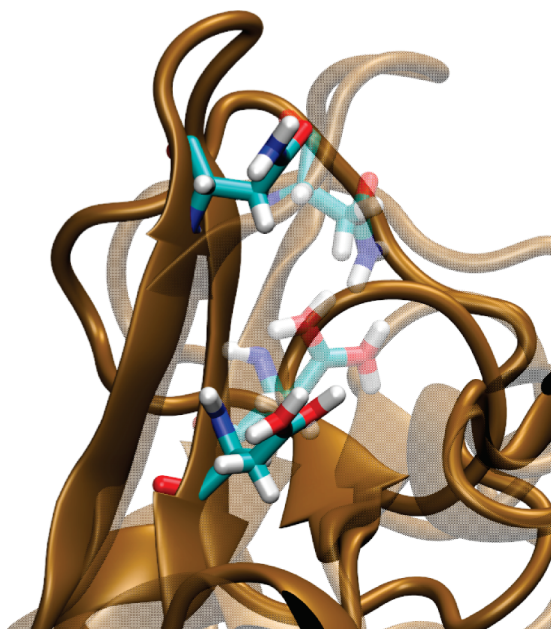

Figure 6. Motion of loop allowing the dissociation of the ASP-52−ASN-46 interaction in simulations employing CpHaMD methodology.

dictions and dynamics of titratable residues at a range of pH conditions.

Conclusions

This study has introduced a new technique whereby the CpHMD method of Mongan et al.[4] has been coupled with the aMD enhanced sampling method of de Oliveira et al. and Hamelberg et al.[20,43] This coupled technique substitutes the conventional MD employed in the standard CpHMD method with aMD, a method previously demonstrated to enhance sampling by lowering the energy barriers of the energy
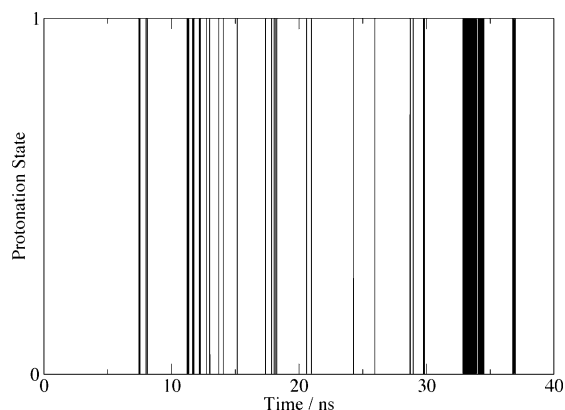

Figure 7. Transitions between protonated (1) and deprotonated (0) states of ASP-52 over a 40 ns CpHaMD simulation at pH 6.5.

landscape, while leaving the minima unchanged, with the capability of fully recovering the correct canonical averages of observables, in this case, pK_a. CpHaMD utilizes the same GB implicit solvation, with Monte Carlo sampling based on GB-derived energies as used in the standard method.

The initial results generated in this study show the CpHaMD method to more efficiently sample conformational space compared with the standard CpHMD method, resulting in faster convergence of constant pH simulations and improved agreement of calculated pK_a values with those obtained experimentally. In addition, the calculated RMS error between the predicted and experimental pK_a values of the acidic residues of HEWL demonstrate the CpHaMD methodology to generate results close to the leading results reported in the literature for other CpHMD methods. Owing to the improved conformational sampling, this method has proved to be advantageous over the CpHMD method in obtaining more accurate and consistent pK_a predictions for the more buried residues of the system, which are typically more problematic to obtain owing to their slow convergence. This has been highlighted by the considerably

improved results of the most problematic residue of HEWL, the catalytically important ASP-52, where the enhanced conformational motion observed in the vicinity of this residue in simulations utilizing CpHaMD clearly demonstrates the link between protonation state and conformation.

From this initial study, the RMS error measured between the calculated and experimental results are close to the leading results reported in the literature for HEWL. It is hoped that this method would assist in the study of biomolecular systems, in gaining more accurate thermodynamics and capturing important pH-coupled conformational events in a more time-efficient manner.

References

(1) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. *J. Chem. Phys.* **2002**, *117*, 4184–4200.

(2) Dlugosz, M.; Antosiewicz, J. M.; Robertson, A. D. *Phys. Rev. E* **2004**, *69*, 021915.1021915.10.

(3) Dlugosz, M.; Antosiewicz, J. M. *Chem. Phys.* **2004**, *302*, 161–170.

(4) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.

(5) Baptista, A. M.; Martel, P. J.; Petersen, S. B. *Proteins: Struct., Funct., Genet.* **1997**, *27*, 523–544.

(6) Borjesson, U.; Hunenberger, P. H. *J. Chem. Phys.* **2001**, *114*, 9706–9719.

(7) Borjesson, U.; Hunenberger, P. H. *J. Phys. Chem. B* **2004**, *108*, 13551–13559.

(8) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163.

(9) Chen, J. H.; Brooks, C. L.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140–148.

(10) Baker, N. A.; Bashford, D.; Case, D. A. In *New Algorithms for Macromolecular Simulation*; Leimkuhler, B., Chipot, C., Elber, R., Laaksonen, A., Mark, A., Schlick, T., Schütte, C., Skeel, R., Eds.; Springer: New York, 2006; Vol. 49, pp 263–295.

(11) Bashford, D. *Front. Biosci.* **2004**, *9*, 1082–1099.

(12) Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192–2198.

(13) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738–752.

(14) Khandogin, J.; Brooks, C. L. *Biophys. J.* **2005**, *89*, 141–157.

(15) Im, W. P.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(16) Khandogin, J.; Brooks, C. L. *Biochemistry* **2006**, *45*, 9363–9373.

(17) Chen, J. H.; Im, W. P.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.

(18) Burgi, R.; Kollman, P. A.; van Gunsteren, W. F. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 469–480.

(19) Stern, H. A. *J. Chem. Phys.* **2007**, *126*, 164112.

(20) de Oliveira, C. A. F.; Hamelberg, D.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 175105–175112.

(21) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*.

(22) Bartik, K.; Redfield, C.; Dobson, C. M. *Biophys. J.* **1994**, *66*, 1180–1184.

(23) Takahashi, T.; Nakamura, H.; Wada, A. *Biopolymers* **1992**, *32*, 897–909.

(24) Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. *Nature* **2001**, *412*, 835–838.

(25) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(26) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297–1304.

(27) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(28) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383–394.

(29) Kyte, J. In *Structure in Protein Chemistry*; Garland Publishing, Inc: New York, 1995; p 64.

(30) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. *J. Phys. Chem.* **1996**, *100*, 2567–2572.

(31) Mu, Y. G.; Nguyen, P. H.; Stock, G. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 45–52.

(32) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(33) Yeh, A. P.; McMillan, A.; Stowell, M. H. B. *Acta Crystallogr., Sect. D* **2006**, *62*, 451–457.

(34) Demchuk, E.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 17373–17387.

(35) Machuqueiro, M.; Baptista, A. M. *Proteins: Stuct., Funct., Bioinf.* **2008**, *72*, 289–298.

(36) Beroza, P.; Case, D. A. *J. Phys. Chem.* **1996**, *100*, 20156–20163.

(37) Yang, A. S.; Honig, B. *J. Mol. Biol.* **1993**, *231*, 459–474.

(38) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.

(39) Gibas, C. J.; Subramaniam, S. *Biophys. J.* **1996**, *71*, 138–147.

(40) van Vlijmen, H. W. T.; Schaefer, M.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 145–158.

(41) Baptista, A. M.; Soares, C. M. *J. Phys. Chem. B* **2001**, *105*, 293–309.

(42) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophys. J.* **2002**, *83*, 1731–1748.

(43) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.