pK_a Analysis for the Zinc-Bound Water in Human Carbonic Anhydrase II: Benchmark for "Multiscale" QM/MM Simulations and Mechanistic Implications†

Demian Riccardi and Qiang Cui*

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Avenue, Madison, Wisconsin 53706

Received: January 26, 2007; In Final Form: March 16, 2007

To quantitatively explore the applicability of the generalized solvent boundary potential (GSBP) based QM/MM approach as a "multiscale" framework for studying chemical reactions in biomolecules, the structural and energetic properties of the Human Carbonic Anhydrase II (CAII) are analyzed and compared to those from periodic boundary condition (PBC) simulations and available experimental data. Although the atomic fluctuations from GSBP based simulations are consistently lower compared to those from PBC simulations or crystallographic data, the fluctuations and internal coordinate distributions for residues in the proximity of the active site as well as diffusion constants of active-site water molecules are fairly well described by GSBP simulations. The pK_a of the zinc-bound water, calculated with a SCC-DFTB/MM-GSBP based thermodynamic integration approach, agrees well with experiments for the wild type CAII. For the E106Q mutant, however, a 9 pK_a unit downward shift relative to the wild type is found in contrast with previous experiments that found little change. This dramatic discrepancy signals a possible change in the mechanism for the interconversion between CO_2/HCO_3^- in the E106Q mutant, which may be similar to the bicarbonate mediated mechanism proposed for the Co^{2+} substituted CAII (*J. Am. Chem. Soc.* **2001**, *123*, 5861).[1] The study highlights pK_a analyses as a valuable approach for quantitatively validating the computational model for complex biomolecules as well as for revealing energetic properties intimately related to the chemical process of interest.

I. Introduction

With rapid developments in computational hardwares and novel computational algorithms, hybrid quantum mechanical/molecular mechanical (QM/MM) simulations[2,3] have become increasingly popular in the last two decades.[4−12] In recent years, there has been great interest in pushing forward the QM/MM techniques in a "multiscale" framework in order to quantitatively analyze reactive processes in very large biomolecular systems, such as ion pumping in membrane proteins and peptide synthesis in the ribosome.[13] The most straightforward implementation in this context is to treat the reactive fragments with QM, the immediate environment (e.g., within 20−25 Å) with MM, and the rest with continuum electrostatics. Although this scheme was envisioned many years ago by a number of researchers, a flexible implementation applicable to biomolecules has only been reported in recent years. For example, we have implemented the generalized solvent boundary (GSBP) condition approach of Roux and co-workers[14] in a QM/MM framework.[15] A related but different formulation based on the boundary element approach has been reported by York and co-workers.[16] With simpler QM methods, such a framework has been explored by Warshel and co-workers in their pioneering studies.[17]

As discussed in the original work,[14] the GSBP approach treats a small region (e.g., a 20 Å spherical region) in complete microscopic details while including the effects (largely electrostatic) due to atoms further away and the bulk environment (solution and/or membrane) with continuum electrostatics at the Poisson−Boltzmann level. In such a way, the GSBP approach is as computationally efficient as the popular "stochastic boundary condition"[18] but with better defined approximations and is potentially well suited for studying localized chemical processes (e.g., ligand binding, enzyme catalysis) in very large biomolecules. Our recent studies found that the QM/MM-GSBP protocol produced encouraging results at both the qualitative and quantitative levels for a number of biomolecular systems; these include active-site dynamics in human carbonic anhydrase II (CAII) in comparison to experimental observations and previous classical simulations,[15] water profile in the channel of aquaporin in comparison to explicit membrane-solvent simulations,[19] and pK_a values for titritable groups in the T4 lysozyme in comparison to experimental measurements.[20] In a study related to the current work,[21] QM/MM-GSBP calculations were compared to QM/MM-Ewald simulations for the active-site properties of CAII (see Supporting Information in ref 21) that included the distribution and diffusion constant of active-site water molecules as well as the flexibility of the proton acceptor (His 64). Overall, the agreement was again very encouraging except that the diffusion of water is slowed down near the inner/outer boundary in the GSBP simulations, as was expected.

To fully explore the applicability and potential limitation of the GSBP based QM/MM approach, it is important to study how structural and energetic properties of biomolecules depend on the size of the mobile region. Bearing this goal in mind, we continue to explore the properties of CAII, which is a small zinc enzyme that catalyzes the interconversion between CO_2 and bicarbonate (HCO_3^-). As discussed in previous studies (also see section II),[12,15,21] CAII is an ideal system for benchmarking QM/MM methods because of its small size and rich experimental background.[22−24] In particular, due to the importance of long-range proton transfer in its functional cycle, the properties of the CAII active site are sensitive to simulation

† Part of the "DFTB Special Section".

* Corresponding author. E-mail: cui@chem.wisc.edu.

10.1021/jp070699w CCC: $37.00
Published on Web 05/17/2007

details such as the treatment of long-range electrostatics, which makes a stringent test of QM/MM protocols possible.

Another important motivation for the current work is to illustrate the value of pK_a simulations for establishing a quantitative understanding of the electrostatic interaction network in the CAII active site. Although electrostatics have been recognized to be crucial in enzyme systems,[25] especially those that involve long-range charge/proton transfers, a quantitative analysis has been largely limited to the level of continuum electrostatics. By combining the thermodynamic integration technique with the QM/MM-GSBP protocol,[20] we illustrate how a microscopic analysis of pK_a values of critical groups can serve as a powerful benchmark for the simulation protocol and, at the same time, offer new mechanistic insights.

In the following, we first summarize the computational details for the QM/MM-GSBP simulations of CAII with two different mobile-regions within 20 and 25 Å. In section III, we analyze the active-site behavior and the pK_a of the zinc-bound water. In section IV, we draw a few conclusions regarding the applicability of the QM/MM-GSBP protocol in the context of enzyme simulations and highlight the value of pK_a analysis for studying chemical reactions in enzymes.

II. Computational Methods

Under proper buffer conditions,[26] the rate-limiting step in the catalytic cycle of CAII has been shown to be the proton transfer between the zinc-bound water in the active site and His 64.[22] In this investigation, we focus on the protonation state of the zinc-bound water while the His 64 is kept in the neutral state. Following Toba et al.,[27] the zinc-bound water and zinc-bound hydroxide states are referred to as CHOH and COH, respectively. All the simulations in this work employ a hybrid quantum mechanics and molecular mechanics (QM/MM[3,4,17]) approach. The standard second-order self-consistent charge density functional tight-binding approach (SCC-DFTB[28,29]) is used for the QM region, which consists of the zinc,[30] its three histidine (His 92, His 94, and His 119), and H_2O/OH^- ligands. The SCC-DFTB approach is chosen on the basis of its overall balance of computational efficiency and accuracy; the reader is referred to recent reviews[12,31,32] for more complete discussions. Specifically for the CAII system, gas phase benchmark calculations have been carried out for the model zinc compound that is identical to the QM region used here;[21] the SCC-DFTB approach gives an error of 7.0 kcal/mol for the proton affinity of the zinc-bound water compared to B3LYP/6-311+G**//B3LYP/6-31G* calculations. Although this significant error would make absolute pK_a prediction difficult,[20] we note that SCC-DFTB gives a very similar error (~8 kcal/mol) for the proton affinity of 4-methylimidazole. Therefore, the SCC-DFTB approach used here is expected to provide a balanced treatment for the relative pK_a of the proton donor and acceptor groups in CAII, which is confirmed by the fact that SCC-DFTB/MM PMF calculations give a nearly thermoneutral reaction energy for the proton transfer,[21] in agreement with experimental findings.[23] For the MM atoms, the CHARMM 22 forcefield[33] is used. Link atoms are placed between the C_α and C_β atoms of the MM and QM regions, respectively, to complete the valence of the quantum boundary atoms; the subtleties associated with the treatment of the QM/MM frontier[34−37] for pK_a calculations are discussed below. MM bonding terms are maintained between the QM and MM atoms across the boundary. For the van der Waals parameters for the QM atoms, the standard CHARMM parameters are used. As shown in our previous benchmark study,[38] these standard parameters work well for SCC-DFTB compared to an optimized set based on a small set of hydrogen-bonded complexes. Because we focus on relative pK_a between the zinc-bound water and 4-methyl imidazole, the effect of the QM van der Waals parameters is expected to be even smaller.

A. Generalized Solvent Boundary Potential (GSBP). The GSBP setup is similar to that described in our earlier publications.[12,15] The crystal structure (PDB code 2CBA[39]) with the "in'' rotomer of H64 is centered with the zinc atom at the origin, and additional water molecules are added for proper solvation; hydrogen atoms are added with HBUILD in CHARMM.[40] Two sizes of the inner region, 20 and 25 Å, with a smooth dielectric interface with the outer region are prepared. In the outer region, the dielectric interface between protein and bulk solution (with a dielectric constant of 1 and 80, respectively) is defined by the atomic radii of Roux and co-workers;[41] the inner region has a dielectric constant of 1. Because the outer region atoms are held fixed, their dielectric "constant" should, in principle, be larger than 1. However, because the number of protein atoms in the outer region is very small (e.g., 529 atoms for the 25 Å inner-region setup) in a medium-size enzyme such as CAII, changing the value of the dielectric constant for the outer region only has a small effect. For example, very modest changes were found in the previous SCC-DFTB/MM-GSBP based pK_a simulation for T4-lysozyme when the outer region dielectric constant is changed from 1.0 to 4.0.[20] The reaction field matrix is evaluated using 400 spherical harmonics. The static field due to the outer region is evaluated with the linear Poisson−Boltzmann approach using a focusing scheme that places a 56 Å cube of fine grid (0.4 Å) into a larger 132 Å cube of coarse grid (1.2 Å). Although the effect of salt ions in the bulk can be straightforwardly taken into account in the GSBP framework using a Debye−Hückel model, our previous studies[20] indicate that the effect of salt is minimal on the computed pK_a for a group in the center of the protein. Therefore, the ionic strength is set to zero in the current simulations; this is supported by the analysis here (see below) that the pK_a for the zinc-bound water is largely dominated by a few charged groups very close to the active site.

During the MD simulations, all atoms in the outer region (along with some atoms at the edge of the inner region[14,15]) are held fixed and provide anchors for the system. The inner region is further partitioned into Newtonian and Langevin regions:[18] all atoms between 16 (18) and 20 (25) Å are treated with Langevin MD and the rest are treated with Newtonian MD; all non-hydrogen atoms in the Langevin region are harmonically restrained with force constants corresponding to the B-factors in the PDB file. The Langevin atom list is updated heuristically, a 2 fs time step is used with SHAKE[42] applied to all bonds involving hydrogen, and the temperature is maintained at 300 K. Water molecules are kept within an 18 (23) Å radius by a weak spherical boundary potential. Classical electrostatic interactions and van der Waals interactions are calculated with extended electrostatics.[40]

B. pK_a Calculations. The pK_a of the zinc-bound water is calculated with the thermodynamic integration approach using the dual topology single coordinate (DTSC[43,44]) scheme, which has been reported in detail in several recent publications.[12,20,44] The dominant contribution in this approach is the free energy associated with converting the acidic proton to a dummy atom (D), i.e., from the CHOH to the CDOH state. The corresponding free energy derivative is given by

$$\frac{\partial \Delta G_{C(D/H)OH}}{\partial \lambda} = \langle U_{elec}^{CDOH}(X_{C(D/H)OH}) - U_{elec}^{CHOH}(X_{C(D/H)OH}) \rangle_{\lambda} \quad (1)$$

which is the average QM/MM energy difference (along with small contributions from bonded terms associated with the dummy atom[44]) averaged for a given simulation window with a specific coupling parameter λ; $X_{C(D/H)OH}$ emphasizes that *one* set of coordinates is used for both protonation states. The free energy contribution is determined via integrating the converged values from eq 1 with respect to λ from 0 to 1. As discussed previously, the method is formally exact because the free energy is a state function, although negligible errors arise in practical simulations due to constrained hydrogen bond lengths;[43] the contribution due to the van der Waals interaction between the dummy atom and the environment is often negligible[44] and therefore not included here. Simulations are carried out at $\lambda =$ 0.0, 0.25, 0.5, 0.75, and 1.0.

To gain insights into the contribution of various residues and water to the pK_a value, perturbative analysis is carried out in which the electrostatic contribution to ($\partial \Delta G_{C(D/H)OH}/\partial \lambda$) is analyzed systematically. Specifically, the contribution from a specific group *i* to ($\partial \Delta G_{C(D/H)OH}/\partial \lambda$) is evaluated as

$$\frac{\partial \Delta G_{C(D/H)OH}^{(i)}}{\partial \lambda} = \langle \Delta^{(i)} U_{elec}^{CDOH}(X_{C(D/H)OH}) - \Delta^{(i)} U_{elec}^{CHOH}(X_{C(D/H)OH}) \rangle_{\lambda} \quad (2)$$

where

$$\Delta^{(i)} U_{elec}^{C(D/H)OH} = U_{elec}^{C(D/H)OH} - U_{elec,(i)}^{C(D/H)OH} \quad (3)$$

Here $U_{elec,(i)}^{C(D/H)OH}$ is the electrostatic energy for the C(D/H)OH state *without* the contribution from group i. Integration of ($\partial \Delta G_{C(D/H)OH}^{(i)}/\partial \lambda$) over λ in the range (0, 1) gives the perturbative contribution of group i to the pK_a; it is a perturbative estimate because the original trajectories with the full electrostatic interactions are used in the equilibrium average.

In addition to the wild type CAII, the E106Q mutant is also analyzed. The E106Q mutation results in a 1000-fold reduction of k_{cat}, but only a ~10 fold reduction for k_{cat}/K_M;[45,46] the pK_a of the zinc-bound water was estimated to be ~6.9 from the pH profile of k_{cat}/K_M (Figure 3 of ref 45). This is very surprising considering that mutating the negatively charged Glu 106, which is in the immediate neighborhood of the zinc-bound water (Figure 1), should decrease the pK_a of the zinc bound water significantly. Estimates of this pK_a based on the pH profile of k_{cat} have not been carried out (Silverman, private communication), and therefore, the computation of the zinc-bound water pK_a for the E106Q mutant is of great interest.

The mutant simulations are based on the same wild type X-ray structure with Glu 106 replaced by a glutamine. This is justified by the observation that the X-ray structure of E106Q (PDB code 1CAZ[46]) is very similar to the wild type; e.g., the Cα RMSD is 0.14 Å. Because the mutation occurs in the inner region and the same inner/outer region partition is used in the mutant simulations (at either 20 or 25 Å), the reaction field matrix and outer region electrostatic potentials calculated for the wild type system can be used *without* any change; this is a particular attractive feature of the GSBP setup.[14]

Two and four independent sets of simulations are run for 20 and 25 Å inner regions, respectively, for both the WT and the E106Q mutant. For the 20 Å runs, ~0.4−1 ns of sampling is carried out for each λ window; for the 25 Å simulations, the sampling time for each λ window averages about 0.7 ns. The free energy derivative in each independent λ window is determined with a block averaging scheme that uses statistical tools to identify the boundary between equilibrating and equilibrated regions and to determine the mean and variance (P. König, unpublished).

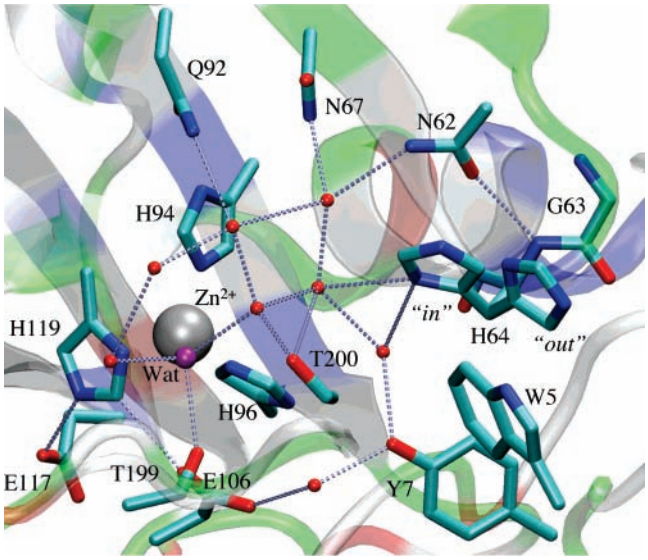

Figure 1. Active site of CAII rendered from the crystal structure (PDB ID: 2CBA[39]). All dotted lines correspond to hydrogen-bonding interactions with distances ≤3.5 Å. E117 and E106 are in close proximity to H119, and E106 also interacts with T199 through the presumed hydroxyl proton of T199 (not shown for clarity). H64 is resolved to partially occupy both the "in'' and "out'' rotameric states.

To make comparisons with the experimental value, the free energy shift (ΔpK_a) is calculated relative to 4-methylimidazole (4-MI) in solution (pK_a = 7);[47] the possibility of using a very different molecule as the reference is a unique feature for QM/MM based simulations[20] and not possible with conventional MM based methods.[48] We note that although in principle our QM/MM based approach can produce absolute pK_a values,[44] many important factors need to be taken into consideration for a reliable prediction, as systematically analyzed in our previous benchmark study of amino acid sidechains in solution.[20] For the purpose of this work, which focuses on the reliability of the GSBP approach and how the results depend on the size of the inner region and conformational sampling, a relative pK_a computation is sufficient. As mentioned above, the SCC-DFTB parametrization used here treats the zinc-bound water and 4-methylimidazole in a balanced manner; thus using the 4-methylimidazole as the reference system is a sensible choice. Another related point concerns whether the pK_a of water is a good reference in the current context, as suggested by one of the referees. Because the property of water is changed significantly when it is bound to a zinc ion, we believe that water is a less relevant reference system than 4-methylimidazole, which is a good model for the proton acceptor in CAII.

III. Results and Discussion

A. Active-Site Flexibility. As shown in Figure 2a, the root-mean-square-fluctuations (RMSF) calculated from SCC-DFTB/MM-GSBP simulations are substantially lower than both values converted based on the B-factors from the PDB data and those from periodic boundary condition (PBC) simulations (described in the Supporting Information of ref 21). This is true not only for atoms restrained during the simulations but also for inner region atoms that are not explicitly subjected to any restraints. This damping effect is most striking for the 20 Å inner-region

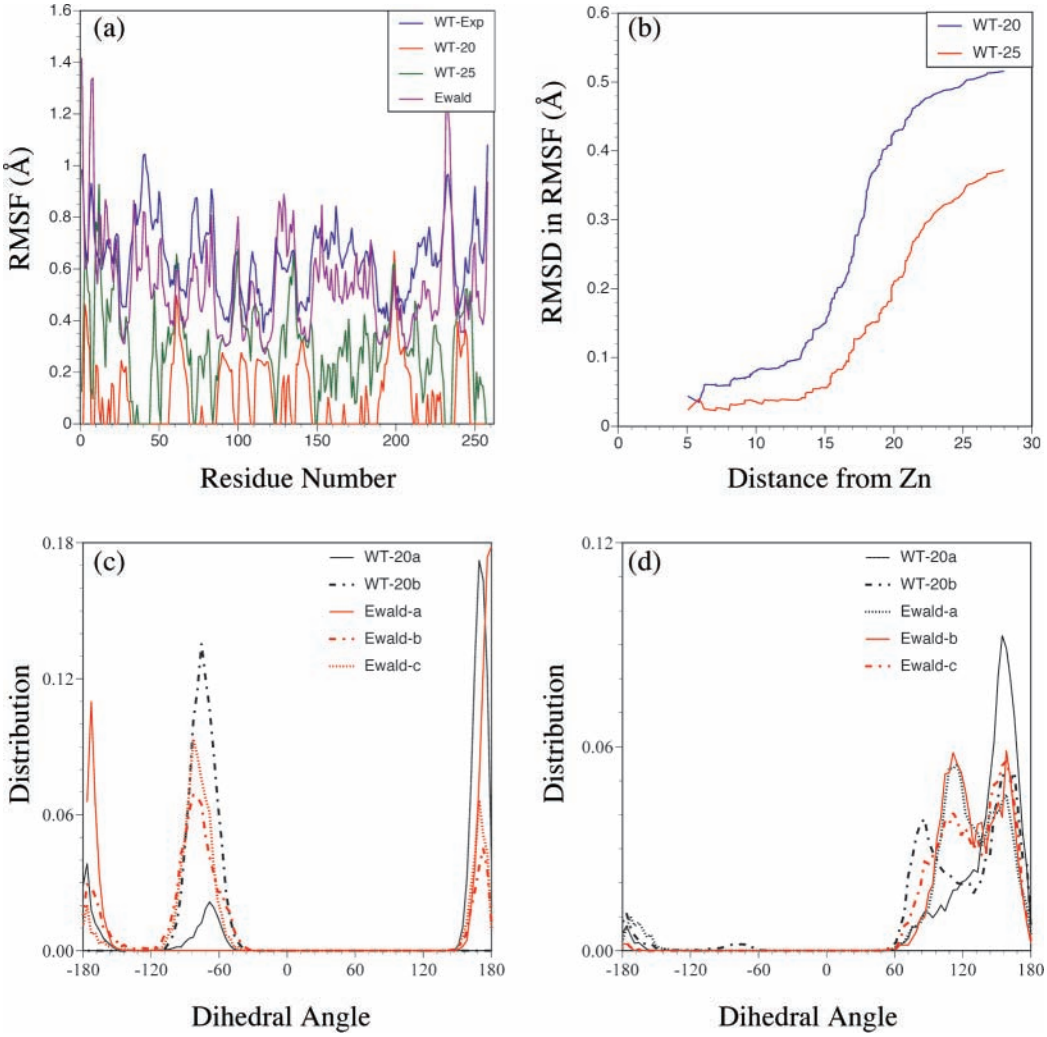

Figure 2. Flexibility of residues from different simulations and the X-ray data. (a) Root-mean-square fluctuations (RMSF) of C_α for the CHOH state plotted as functions of residue number (the trend is the same for the COHH simulations, data not shown). The results shown are averaged over the multiple independent simulations; those for the X-ray data (2CBA[39]) are converted from the Debye−Waller *B* factors using the expression: $B = (8\pi^2/3)\langle\Delta r^2\rangle$. (b) The root-mean-square differences between the RMSFs calculated from GSBP simulations and those from Ewald simulation, for atoms within a certain distance from the zinc, plotted as functions of distance from the zinc ion. Note that the center of the sphere in GSBP simulations is the position of the zinc ion in the starting (crystal) structure. (c, d) Side-chain dihedral angle (χ_1) distributions for (c) His 64 and (d) Gln 92 from independent sets of GSBP (WT-20a/b) and Ewald (Ewald-a/b/c) simulations.

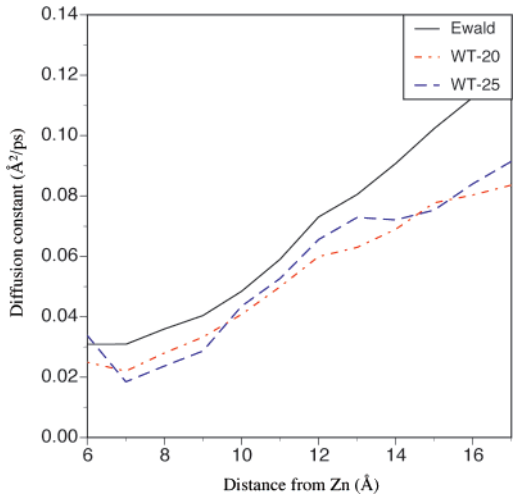

Figure 3. Diffusion constant for TIP3P water molecules as a function of the distance from the zinc ion in different simulations.

simulations, where even the largest RMSF is smaller than 0.5 Å. With a smaller number of atoms restrained, the 25 Å inner-region simulations better reproduce the X-ray data for some regions although still give substantially quenched fluctuations for many regions, such as between Val 160 and Asp 180, which is part of a helix-turn-β sheet motif on the surface.

To better compare the RMSF values from different simulations, it is instructive to examine the degree of quench in RMSF as a function of distance to the zinc ion. In Figure 2b, the root-mean-square differences (RMSDs) between the RMSFs calculated from GSBP simulations and those from the Ewald simulations, for atoms within a certain distance (based on Cα) from the zinc, are plotted against the distance from zinc. Evidently, the atomic fluctuations close to the zinc ion are, in fact, rather well reproduced in both GSBP calculations. The RMSD only starts to increase steeply when the buffer region (which are harmonically restrained as in stochastic boundary simulations[18]) is approached. For residues within 13.5 Å from the zinc, for example, the RMSD between the RMSFs from the 20 Å inner-region GSBP simulations and those from the Ewald simulations is 0.11 Å. The RMSDs of atomic RMSF in the 25 Å inner-region GSBP simulations are generally smaller than those from the 20 Å setup, even for atoms very close to

TABLE 1: Representative Results from Statistical Analyses Used To Determine the Values and Statistical Errors of $\partial\Delta G_{CH(D)OH}/\partial\lambda$ for the pK_a Calculation of the Zinc-Bound Water in CAII[a]

	set 1			set 2		
λ	length[b]	block size[c]	$\partial\Delta G_{CH(D)OH}/\partial\lambda$[d]	length[b]	block size[c]	$\partial\Delta G_{CH(D)OH}/\partial\lambda$[d]
0.00	1.4 (0.3)	13 (86)	212.6 (1.0)	1.3 (0.6)	15 (51)	211.1 (1.1)
0.25	1.1 (0.6)	5 (89)	184.0 (0.7)	1.1 (0.8)	6 (44)	176.1 (1.1)
0.50	1.4 (0.6)	8 (103)	156.8 (0.8)	1.2 (0.2)	9 (112)	144.9 (0.6)
0.75	1.1 (0.7)	6 (65)	123.0 (1.4)	1.1 (0.3)	8 (94)	113.9 (1.1)
1.00	1.4 (0.6)	14 (58)	69.0 (1.6)	1.2 (0.6)	18 (33)	67.5 (1.8)

[a] As examples, results from two independent sets of 20 Å inner-region simulations for the E106Q mutant of CAII are shown. [b] Total simulation length (in ns) for each λ window, the number in parentheses is the length of equilibration identified by trend analysis. [c] Number without any parentheses is the size of the block (in ps), number with parentheses is the number of blocks; these are determined after the equilibration sections of the trajectories are removed. [d] In kcal/mol; the number in parentheses is the statistical error evaluated based on block average.

TABLE 2: Free Energy of Deprotonation ($\Delta G_{CH(D)OH}$) Calculated from Independent Thermodynamic Integration Simulations[a]

simulation[b]	1	2	3	4	avg[c]
4-MI	152.4 (0.99)	152.5 (0.99)			152.5 (0.1)
WT-20	158.0 (0.99)	160.6 (0.96)			159.3 (1.8)
WT-25	156.1 (0.97)	156.8 (0.98)	158.9 (0.99)	159.3 (0.94)	157.8 (1.6)
E106Q-20	149.1 (0.98)	142.7 (0.99)			145.9 (4.5)
E106Q-25	146.3 (0.98)	146.9 (0.98)	146.4 (0.98)	143.5 (0.99)	145.8 (1.5)

[a] For each simulation, the value of $\Delta G_{CH(D)OH}$ (in kcal/mol) is determined based on integrating the linear fit of the free energy derivatives (the R^2 values for the linear fits are shown in parentheses) with respective to λ from 0 to 1. [b] 4-MI: 4-methyl-imidazole in solution; "20" and "25" indicate the size (in Å) of the inner region in the GSBP setup. [c] Number in parentheses is the standard deviation.

the center of the sphere (Figure 2b); to reach the same RMSD of 0.11 Å, for example, the region extends to 17 Å from the zinc ion.

These comparisons suggest that additional relevant observables that characterize the active-site flexibility, *assuming* that collective structural fluctuations do not play any major functional role (otherwise an active-site based simulation is not appropriate), are the distributions of internal coordinates of active-site residues. As shown in Figure 2c,d, which are representative for active-site residues, the χ_1 distribution can be different among independent trajectories for a specific boundary condition, but the overall trend is very consistent between even the 20 Å GSBP simulations and the Ewald simulations.

Finally, because the water molecules in the active site play an important role in modulating the proton-transfer pathways and energetics, it is instructive to compare the distribution and diffusion of these water molecules from different simulations. As shown in the previous work,[21] even a 20 Å inner-region simulation reproduces the distribution of water molecules within 17 Å from the zinc ion in close agreement with PBC simulations; the simulations with a larger (25 Å) inner region with different protonation states of the proton donor (zinc-bound water) and acceptor (His 64) also give similar results (data not shown). For the diffusion constants (Figure 3), the two sets of GSBP simulations in fact give rather similar results and both underestimate the values compared to the PBC simulations in ref 21, especially for water molecules close to the inner/outer boundary. The basic trend as a function of distance from the zinc ion, however, is well reproduced in both sets of GSBP simulations.

B. Statistical and Sampling Errors. Because multiple independent pK_a computations are carried out for both the WT and E106Q mutants, these data provide the opportunity to illustrate the statistical and sampling errors associated with pK_a calculations. As discussed in the Computational Methods, the free energy derivative for each λ window is determined in the *forward* direction using statistical tools that establish the block size, average, and statistical error for a trendless (equilibrated) region of the data in an automated fashion. Although a *reverse* cumulative analysis proposed recently[49] can also be used, the automated approach is advantageous when there are large sets of data.

The block sizes are found to vary from 2 to 18 ps, and the discarded (equilibrating) regions contained anywhere from 50 to 800(!) ps. The statistical errors associated with each free energy derivative are from 0.3 to 2 kcal/mol with most values around 1 kcal/mol. As an example of the analysis, the results from two 20 Å inner-region simulations for E106Q (denoted as "E106Q-20") are shown in Table 1. A linear response to deprotonation is evident from the linear fits of the free energy derivatives with respect to λ, which yield R^2 values typically ≥ 0.97 (Table 2). Integrating the linear equations (from 0 to 1) for each independent FEP computation yields the net free energy change, $\Delta G_{C(D/H)OH}$; though independent runs typically give similar $\Delta G_{C(D/H)OH}$ values, E106Q-20 simulations yield the largest deviation of 6.4 kcal/mol. It is clear that although the free energy derivatives converge with *statistical* errors $\sim$1 kcal/mol, there are differences between the two independent runs on the order of 10(!) kcal/mol for some λ windows (compare columns 4 and 7 of Table 1). Therefore, the free energy derivatives for these runs appear to have equilibrated to sample different regions of the configuration space. Inspection of the trajectories suggests that the difference is likely due to the different orientations of the Thr 199 side chain sampled in separate simulations, which leads to substantial variation in the interaction between Thr 199 and the zinc-bound water and therefore change in the free energy derivatives.

Overall, as seen in Table 2, the results for $\Delta G_{C(D/H)OH}$ agree well between the 20 and 25 Å inner-region setups with both sets of averages falling within the standard deviations. Moreover, as shown in Figure 4, the free energy derivatives (average of independent runs) and the linear fits also agree well between 20 and 25 Å runs within both the WT and E106Q systems. These observations support the expectation that the flexibility of the enzyme distant from the titration site does not significantly affect its pK_a; it is possible that this only holds for rather rigid enzymes such as CAII, whereas for enzymes with more floppy motifs, specific structural changes may propagate over a

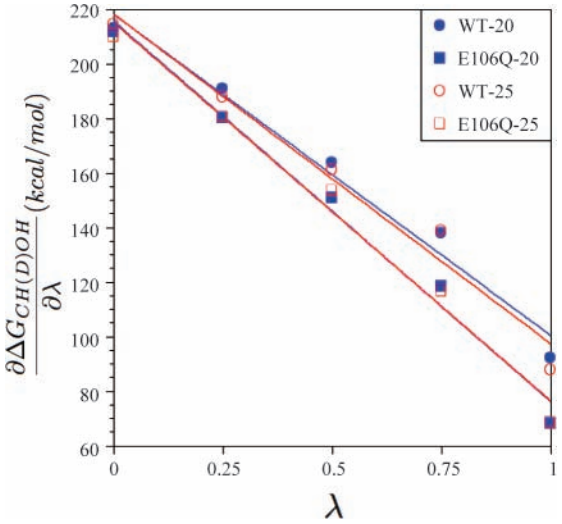

Figure 4. Linear fit of the of the free energy derivatives with respect to λ for the WT and E106Q mutants of CAII computed with 20 and 25 Å inner-region GSBP simulations.

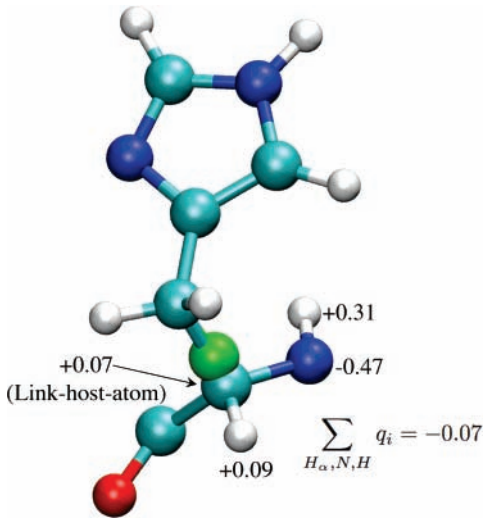

Figure 5. Example of QM/MM partitioning across the Cα and Cβ atoms for a histidine residue, where the side chain plus a link atom (in green) attached to Cβ are treated with QM (SCC-DFTB). The partial charges for the host atom (Cα) and its group in the CHARMM force field are shown. With the standard "link-host-atom" exclusion scheme, which is used to generate all trajectories here, the QM region interacts with all atoms in the group except Cα, and therefore interacts with a net charge of −0.07. The "link-host-group" exclusion scheme (EXGR) avoids this artifact[37] by excluding all QM/MM interactions involving the link-host-group (Cα, Hα, main chain NH).

significant distance to modulate the pK_a of active-site groups (C. L. Brooks and co-workers, private communication).

C. Comparison with Experiment. When comparing the calculated pK_a (or pK_a shift relative to the 4-methylimidazole in solution) to the experimental value, it is important to carefully consider the effect of the QM/MM frontier. As discussed in detail in previous studies,[36,37] the simple link-atom scheme with excluding only the partial charge on the "link-host-atom" (e.g., Cα for the zinc-bound His in the current study; see Figure 5), which is the default option in CHARMM, may produce large (on the order of 10 kcal/mol!) errors for the proton affinity even if the deprotonation site is several covalent bonds away from the QM/MM frontier. This is because the "link-host-atom" exclusion approach leaves a net charge (−0.07) due to the remaining atoms within the same group of backbone atoms (containing C_α, H_α, N, and its bound H; see Figure 5), which makes a significant electrostatic contribution to the calculated proton affinity due to the associated change in charge. Zeroing out all charges within the group ("link-host-group" exclusion, EXGR) avoids the spurious QM/MM electrostatic interaction and generally gives better proton affinity values.[37] For the present CAII simulations, re-evaluating the free energy derivatives with the "link-host-group" exclusion for the three zinc-bound His residues at configurations sampled using the "link-host-atom" exclusion scheme changes the free energy derivatives by 8−9 kcal/mol despite that the QM/MM frontiers are far from the zinc-bound water; this effect varies little (<0.5 kcal/mol) with the value of λ. With this effect taken into account, the calculated pK_a value for the zinc-bound water in the WT CAII is in decent agreement with experiment: the value is 7.1 (5.4) for the 20 (25) Å inner-region simulations, as compared to the experimental value of around 7.[22] We note that an important reason for the SCC-DFTB/MM simulation to produce such good agreement with experiments is that the SCC-DFTB approach treats the zinc-bound water and 4-methylimidazole in a balanced manner (see above) and we used 4-methylimidazole in solution as the reference in the pK_a calculations.

Relative to the WT CAII, the E106Q mutant is found to reduce the pK_a for the zinc-bound water by around ∼9 pK units. This result is reasonable considering that the mutation neutralizes a negative charge near the zinc-bound water. Strikingly, the pK_a determined experimentally from the pH profile of k_{cat}/K_M yielded no shift.[45] Considering that there is little structural change between the WT and E106Q mutant,[46] the calculation result suggests that there may be a change in the mechanism for the step manifested by k_{cat}/K_M. As described in previous work,[22] k_{cat}/K_M is associated with the reaction of the zinc-bound hydroxide with the CO_2 in the hydration direction and the dehydration of the zinc-bound bicarbonate in the reverse direction. Therefore, one possible mechanistic change is that the bicarbonate plays a more active role and the titration result reflects the pK_a for the total complex of the zinc, the bound water, and the bicarbonate. A similar scenario was proposed to explain the behaviors of the cobalt substituted CAII[1] where, unlike the Zn(II)-containing CAII, k_{cat}/K_M depends on the concentration of bicarbonate. In addition, the presence of the acetic acid bound to the zinc in the mutant structure[46] lends additional support for a zinc ion that strongly favors a negatively charged species. Nevertheless, additional investigations such as the titration of k_{cat} as done in ref[1] should be carried out for the E106Q mutant.

D. Dissecting the Contributions from Water and Protein to pK_a. In solution, binding of a divalent ion may shift the pK_a of water to be significantly lower than 7. In CAII, the zinc-bound water has a pK_a of nearly 7, which makes the proton transfer to the acceptor, His 64, nearly thermoneutral. This pK_a match is likely of functional importance because the hydration of CO_2 in CAII needs to be reversible for it to play its physiological role. Therefore, it is of great interest to understand what factors dictate the pK_a of the zinc-bound water. We explore this by performing the perturbative analysis described in the Computational Methods; overall, the results are fairly consistent between the 20 and 25 Å inner-region simulations (see Figures 6 and 7) and therefore only the results from the 20 Å simulations will be discussed explicitly.

1. Water Contribution. As a reference, consider the deprotonation of 4-methylimidazole in solution. The contribution to $\Delta G_{CH(D)OH}$ is from the surrounding water by definition, and the inner-region (18 Å) contribution is 40.6 kcal/mol (see Table 3). The outer region (dielectric continuum) contributes 9.2 kcal/

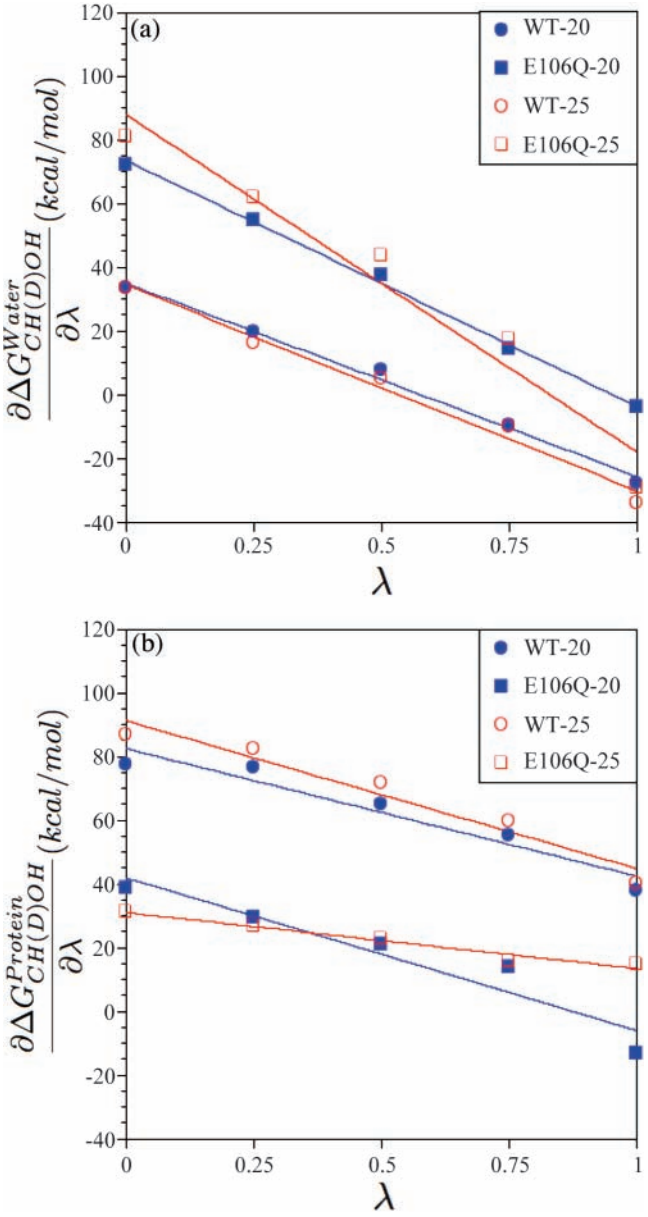

Figure 6. Linear dependence of the (a) water (only those in the inner region) and (b) protein (in both the inner and outer regions) electrostatic contributions to the free energy derivatives as functions of λ. See Table 3 for the values integrated over λ.

mol to the process, which is consistent with the Born correction[50] for a charge $+1 \rightarrow 0$ process in an 18 Å sphere.

The net contribution from water to $\Delta G_{CH(D)OH}$ in CAII is small for the WT enzyme, only 2.1 kcal/mol in the 25 Å simulations. The reorganization of the water, which is related to the variation of the water contribution as a function of λ, however, is substantial; it is $\sim+35$ kcal/mol for $\lambda = 0$ and ~-30 kcal/mol for $\lambda = 1$ (see Figure 6a). In other words, water molecules respond significantly to the change in the protonation state of the zinc-bound water. In a previous study,[12] we found that water molecules within 7.5 Å from the zinc ion have significant contribution to the energetics of proton transfers; with different water orientations sampled using different charge-distributions for the reactive moieties, the proton transfer can be either highly exothermic, highly endothermic or thermoneutral.

Interestingly, the water contribution in the E106Q mutant is systematically larger. The net contribution is about 20 kcal/mol (Table 3). The reorganization energy, however, is rather similar; the difference in the water contribution between $\lambda = 0$ and $\lambda = 1$ is also about 70 kcal/mol. Further looking at the behavior of the water contribution as a function of the distance from the zinc ion (see Figure 7a), it is clear that the water contribution grows much more rapidly for the E106Q mutant from the active site and continues to increase out toward the boundary between the inner and outer region. These aspects of the water contribution are expected considering that the overall change in charge upon deprotonation is the same, but the overall charge of the systems are different for the WT and E106Q.

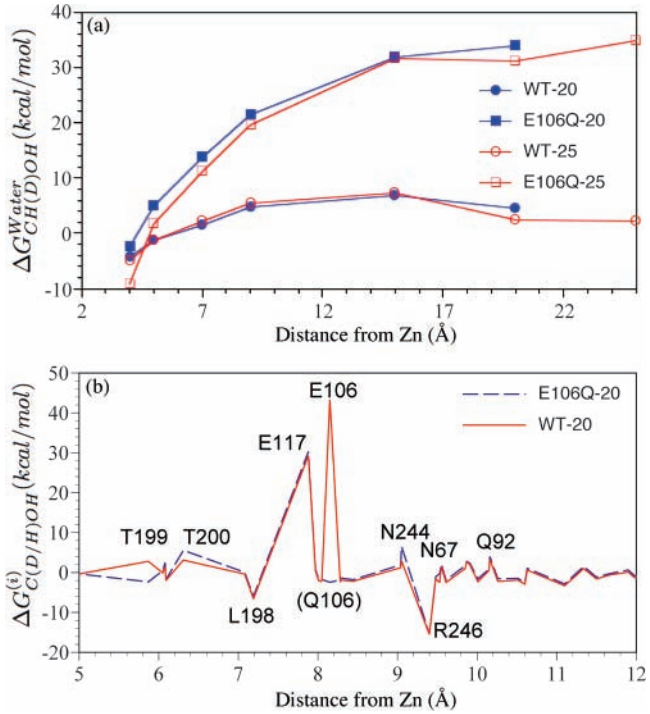

Figure 7. Contribution to the electrostatic component of free energy of deprotonation ($\Delta G_{CH(D)OH}$) from water and protein atoms (only MM atoms are considered) in the WT and E106Q mutant CAII based on perturbative analysis (integration of eq 2 over λ). (a) Cumulative contribution from water as a function of distance from the zinc ion. (b) Contribution from individual residues plotted against the distance (Cα) from the zinc ion. Note the striking similarity between the WT and E106Q results, except, apparently, for residue 106.

2. Protein Contribution. For the WT CAII, the protein atoms make a major contribution to $\Delta G_{CH(D)OH}$ and is on the order of 60 kcal/mol (see Table IIIb). The protein reorganization during deprotonation, which is reflected by the variation of the protein contribution to $\Delta G_{CH(D)OH}$ as a function of λ, is substantially smaller than water. As shown in Figure 6b, the difference in the protein contribution between $\lambda = 0$ and 1 is only $\sim$30 kcal/mol, nearly half of the value for water. This suggests that the groups making large contributions to $\Delta G_{CH(D)OH}$ have limited flexibility. Indeed, looking at the protein contribution by residue (Figure 7b) reveals that three charged groups (Glu 106, Glu 117, and Arg 246) make the dominant contributions; they are either fully buried or semi-buried (Arg 246) inside the protein.

The E106Q mutant has a protein contribution that is $\sim$30 kcal/mol smaller than the WT although the reorganization (with some deviation between the 20 and 25 Å simulations) is rather similar. Strikingly, the perturbative contributions from different residues (Figure 7b) are almost identical in the E106Q and WT enzymes except, obviously, for residue 106. In other words, the difference between the protein contributions in the WT and E106Q is due almost entirely to Glu 106. The loss of the large contribution of $\sim$40 kcal/mol from Glu 106 in E106Q according to the perturbative analysis is partially compensated by the larger

TABLE 3: Calculated pK_a's of the Zinc-Bound Water in the Wild Type and E106Q Mutant of CAII Using Thermodynamic Integration and Various Contributions Based on a Perturbative Analysis (Eq 2)[a]

calculation	pK_a[b]	$\Delta G_{CH(D)OH}$[c]	$\Delta\Delta G^{EXGR}$[d]	$\Delta G^{Protein}_{CH(D)OH}$[e]	$\Delta G^{Water}_{CH(D)OH}$[e]
4-MI	7.0	152.5 (0.1)			40.6 (0.1)
WT-20	7.1	151.3 (1.8)	−8.0	62.5 (0.9)	4.6 (1.8)
WT-25	5.4	149.0 (1.6)	−8.8	68.0 (0.9)	2.1 (1.9)
E106Q-20	−3.0	137.5 (4.5)	−8.5	34.9 (3.0)	18.0 (0.2)
E106Q-25	−3.4	136.9 (1.5)	−8.9	34.9 (1.3)	22.3 (2.2)

[a] The free energies and contributions are in kcal/mol; the numbers in parentheses are statistical uncertainties. [b] The pK_a is calculated using 4-methylimidazole in solution (4-MI) as the reference compound; i.e., $pK_a^{CAII} = 7.0(\text{ exp.}^{47}) + [\Delta G^{CAII}_{CH(D)OH} - \Delta G^{4-MI}_{CH(D)OH}]/1.370$. [c] Compared to the average values in Table 2, the values here include the contribution from the EXGR correction ($\Delta\Delta G^{EXGR}$). [d] The effect due to switching the QM/MM frontier treatment from "link-host-atom" exclusion to "link-host-group" exclusion (EXGR). [e] The net contribution from protein (both inner and outer atoms) and water molecules (only those in the inner region) based on the perturbative analysis (integration of eq 2 over λ from 0 to 1).

water contribution; thus the net pK_a shift caused by the E106Q mutation is only ∼9 pK_a unit. This result clearly highlights the importance of water in modulating the energetics of processes in biomolecules.

IV. Concluding Remarks

To meet the challenge of studying chemical processes in large biomolecular systems, "multiscale" QM/MM methods[15,16,25] have been developed in recent years. In these first generation of methods, atoms far (e.g., ∼20−25 Å) from the active site are fixed and treated with continuum electrostatics. Whether this type of protocol can faithfully describe the energetics and dynamics of the active site needs to be quantitatively explored, considering the recent surge of interests in the roles of "dynamics" and long-range effects in enzyme catalysis.[51−54] In this work, using a small globular enzyme, carbonic anhydrase, as an example, we do so for the QM/MM-GSBP protocol that has been recently implemented in our group[15]

The atomic fluctuations from GSBP simulations with different sizes of mobile region are consistently lower in value than those from periodic boundary simulations, which is not surprising. As shown in many previous studies that invoke a mode decomposition of protein motions,[55] the atomic fluctuations are dominated by low-frequency modes, which tend to be highly collective in nature. With some atoms constrained or even fixed in space, these collective modes are not present in the GSBP (or any stochastic boundary) based simulations. Therefore, even the fluctuations for atoms not explicit subjected to any restraint are quenched. However, for atoms in close proximity of the active site (e.g., 13 (17) Å from the zinc ion in a GSBP simulation with 16 (18) Å unrestrained inner region), the GSBP based QM/MM approach is found to produce atomic fluctuations (RMSD < 0.1 Å) and internal coordinate distributions in fairly good agreement with unconstrained periodic boundary simulations. The diffusion coefficients for water molecules in the active site are also well reproduced. More importantly, both GSBP simulations with different sizes of the inner region give consistent pK_a values for the zinc-bound water (*relative* to 4-methylimidazole in solution) and compare favorably with the experimental value. With all the observations in our recent studies[15,19−21] and results from this work taken together, we conclude that the QM/MM-GSBP approach is indeed well suited to analyze chemical reactions in the active site of globular enzyme systems with a compact structure where large-scale structural transitions are not involved in the chemical step.

In addition to providing a quantitative validation of the computational model, pK_a calculations and analyses can reveal energetics properties intimately related to the reaction of interest, which, in turn, may lead to new mechanistic insights. For the WT CAII, the analysis of the electrostatic contributions reveals that the enzyme predominately uses two glutamates (Glu 106, Glu 117) near the zinc to modulate the pK_a of the zinc-bound water. Notably, the calculated pK_a for this group in the E106Q mutant does not seem to agree with experimental measurements.[45] Considering the successes for the WT CAII and previous applications to T4-Lysozyme and small molecules in solution,[20] the large discrepancy (9 pK_a units!) leads to the hypothesis that a change in mechanism occurs in the E106Q mutant for the interconversion between CO_2 and HCO_3^{-1}, which is characterized by the measurement of k_{cat}/K_M. Considering the reverse direction (dehydration of bicarbonate), as seen in Co(II)-containing CAII,[1] the bicarbonate may play a more active role in the protonation of the zinc-bound hydroxide. If this is the case, the question of whether the bicarbonate binds first, yielding a pentacoordinated zinc, or promotes the protonation of the zinc-bound hydroxide via its anionic character is interesting and should be explored further both theoretically and experimentally.

From the technical perspective, the current study also brings up two important points. First, although statistical tests indicate that all thermodynamic integration windows have been properly equilibrated and the simulations have reached quasi-convergence, free energy derivatives in independent simulations of the same λ value can differ substantially (by as much as 10 kcal/mol for some E106Q windows) due to locally trapped sampling in regular molecular dynamics simulations. Therefore, it is of great value to integrate enhanced sampling techniques with thermodynamic integration based pK_a calculations. This is particularly crucial for pK_a calculations involving groups with open-shell characters or transition metals, for which high-level QM methods are likely needed. Second, although the perturbative analyses of the pK_a results are very informative regarding important protein/water contributions, the estimated contributions should be taken with great care due to the "vertical" nature of the analysis (i.e., trajectories for the system with complete interactions are used). For example, the contribution from Glu 106 in the WT CAII is estimated to be ∼40 kcal/mol based on perturbative analyses, but the calculated shift in pK_a upon the E106Q mutation is only ∼13 kcal/mol based on the actual simulations for the E106Q mutant. Apparently, water molecules in the mutant are able to compensate a large portion of the effect caused by the mutation and such contribution cannot be captured in a perturbative analysis.

In short, the GSBP based QM/MM approach provides a promising "multiscale" framework for analyzing chemistry in very large biomolecules. However, in addition to potential contributions from collective motions and other long-range effects, other issues such as the protonation state of buried

titritable residues and ambiguity in the choice of a proper dielectric model for the fixed outer region atoms may significantly impact the reliability of such calculations. In this context, we highlight that pK_a calculations are extremely valuable for both quantitatively validating the computational model but also reveal essential energetic properties relevant to the reaction of interest. In the investigation of proton pumping in complex biomolecules, for example, where electrostatics are crucial[21,56] and major ambiguities exist concerning the titration states of various groups, we argue that pK_a analyses of key residues are an indispensable step before the proton-transfer pathways can be explored.

Acknowledgment. The studies were partially supported from the National Science Foundation (MCB-0314327,CHEM-CAREER-0348649). We greatly acknowledge P. Koenig for useful discussions and providing his implementation of the statistical tools. Q.C. also acknowledges an Alfred P. Sloan Research Fellowship and discussions with Prof. H. Guo on many CAII related topics. Computational resources from the National Center for Supercomputing Applications at the University of Illinois are greatly appreciated.

References and Notes

(1) Tu, C.; Tripp, B. C.; Ferry, J. G.; Silverman, D. N. *J. Am. Chem. Soc.* **2001**, *123*, 5861.
(2) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(3) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(4) Lipkowitz, K. B., Boyd, D. B., Eds. Gao, J. In *Reviews in Computational Chemistry VII*; VCH: New York, 1995.
(5) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
(6) Shurki, A.; Warshel, A. *Adv. Prot. Chem.* **2003**, *66*, 249.
(7) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
(8) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279.
(9) Merz, J. K. M. *Curr. Opin. Struct. Biol.* **1993**, *3*, 234.
(10) Zhang, Y. K.; Liu, H. Y.; Yang, W. T. *J. Chem. Phys.* **2000**, *112*, 3483.
(11) Cui, Q.; Karplus, M. *Adv. Prot. Chem.* **2003**, *66*, 315.
(12) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G.; Xu, D.; Guo, H.; et al. *J. Phys. Chem. B* **2006**, *110*, 6458.
(13) Cui, Q. *Theor. Chem. Acc.* **2006**, *116*, 51.
(14) Im, W.; Bernéche, S.; Roux, B. J. Chem. Phys. **2001**, *114*, 2924.
(15) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905.
(16) Gregersen, B. A.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 536.
(17) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solution*; Wiley: New York, 1991.
(18) Brooks, C. L., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159.
(19) Koenig, P.; Ghosh, N.; Hoffman, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548 (Donald G. Truhlar Festschrift).
(20) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715.
(21) Riccardi, D.; König, P.; Prat-Resina, X.; Yu, H.; Elstner, M.; Frauenheim, T.; Cui, Q. *J. Am. Chem. Soc.* **2006**, *128*, 16302.
(22) Silverman, D. N.; Lindskog, S. *Acc. Chem. Res.* **1988**, *21*, 30.
(23) Silverman, D. N. *Methods Enzymol.* **1995**, *249*, 479.
(24) Silverman, D. N. *Biochim. Biophys. Acta.* **2000**, *1458*, 88.
(25) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425.
(26) Silverman, D. N.; Tu, C. K. *J. Am. Chem. Soc.* **1975**, *97*, 2263.
(27) Toba, S.; Colombo, G.; Merz, K. M. J. *J. Am. Chem. Soc.* **1999**, *121*, 2290.
(28) Elstner, M.; Porezag, D.; Jungnickel, G.; Elstner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
(29) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
(30) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565.
(31) Elstner, M.; Frauenheim, T.; Suhai, S. *THEOCHEM* **2003**, *632*, 29.
(32) Elstner, M. *Theo. Chem. Acc.* **2006**, *116*, 316.
(33) MacKerell, A. D., Jr.; Bashford, D.; Bellot, M.; Dunbrack, R. L.; Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(34) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
(35) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
(36) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chattfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.
(37) Konig, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.
(38) Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.
(39) Håkansson, K.; Carlsson, M.; Svensson, L. A.; Liljas, A. *J. Mol. Biol.* **1992**, *227*, 1192.
(40) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(41) Nina, M.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.
(42) Rychaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327.
(43) Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643.
(44) Li, G.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 14521.
(45) Liang, Z.; Xue, Y.; Behravan, G.; Jonsson, B.; Lindskog, S. *Eur. J. Biochem.* **1993**, *211*, 821.
(46) Xue, Y.; Liljas, A.; Jonsson, B.; Lindskog, S. *Proteins: Struct. Funct. Genet.* **1993**, *17*, 93.
(47) Lide, D. R., Ed. *CRC Handbook Chemistry and Physics*, 85th ed; CRC Press: Boca Raton, FL, 2005.
(48) Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167.
(49) Yang, W.; Bitetti-Putzer, R.; Karplus, M. *J. Chem. Phys.* **2004**, *120*, 2618.
(50) Roux, B.; Yu, H.; Karplus, M. *J. Phys. Chem.* **1990**, *94*, 4683.
(51) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196.
(52) Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2006**, *106*, 3055.
(53) Kovrigin, E. L.; Loria, J. P. *Biochemistry* **2006**, *45*, 2636.
(54) Olsson, M. H. M.; Parson, W. W.; Warshel, A. *Chem. Rev.* **2006**, *106*, 1737.
(55) Haywards, S.; Go, N. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223.
(56) Kato, M.; Pisliakov, A. V.; Warshel, A. *Proteins: Struct. Funct. Bioinform.* **2006**, *64*, 829.