

# **pK<sub>a</sub> Calculations in Solution and Proteins with QM/MM Free Energy Perturbation Simulations: A Quantitative Test of QM/MM Protocols**

**Demian Riccardi, Patricia Schaefer, and Qiang Cui\***

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

*Received: April 4, 2005; In Final Form: June 16, 2005*

The accuracy of biological simulations depends, in large part, on the treatment of electrostatics. Due to the availability of accurate experimental values, calculation of pK<sub>a</sub> provides stringent evaluation of computational methods. The generalized solvent boundary potential (GSBP) and Ewald summation electrostatic treatments were recently implemented for combined quantum mechanical and molecular mechanics (QM/MM) simulations by our group. These approaches were tested by calculating pK<sub>a</sub> shifts due to differences in electronic structure and electrostatic environment; the shifts were determined for a series of small molecules in solution, using various electrostatic treatments, and two residues (His 31, Lys 102) in the M102K T4-lysozyme mutant with large pK<sub>a</sub> shifts, using the GSBP approach. The calculations utilized a free energy perturbation scheme with the QM/MM potential function involving the self-consistent charge density functional tight binding (SCC-DFTB) and CHARMM as the QM and MM methods, respectively. The study of small molecules demonstrated that inconsistent electrostatic models produced results that were difficult to correct in a robust manner; by contrast, extended electrostatics, GSBP, and Ewald simulations produced consistent results once a bulk solvation contribution was carefully chosen. In addition to the electrostatic treatment, the pK<sub>a</sub> shifts were also sensitive to the level of the QM method and the scheme of treating QM/MM Coulombic interactions; however, simple perturbative corrections based on SCC-DFTB/CHARMM trajectories and higher level single point energy calculations were found to give satisfactory results. Combining all factors gave a root-mean-square difference of 0.7 pK<sub>a</sub> units for the relative pK<sub>a</sub> values of the small molecules compared to experiment. For the residues in the lysozyme, an accurate pK<sub>a</sub> shift was obtained for His 31 with multiple nanosecond simulations. For Lys 102, however, the pK<sub>a</sub> shift was estimated to be too large, even after more than 10 nanosecond simulations for each  $\lambda$  window; the difficulty was due to the significant, but slow, reorganization of the protein and water structure when Lys 102 was protonated. The simulations support that Lys 102 is deprotonated in the X-ray structure and the protein is highly destabilized when this residue is protonated.

## **1. Introduction**

Electrostatic interactions are central to biological processes and provide an effective way to correlate structure with function.<sup>1–3</sup> Enzyme catalysis, protein–protein interactions, redox processes, and proton transport are a few examples where these interactions play a major role. The long-range nature of electrostatics complicates computational investigations of highly charged macromolecular systems, and approximations are necessary to make the calculations efficient. To validate such approximations, sensitive benchmark calculations are required. Although careful benchmark calculations have been carried out extensively in classical simulations,<sup>3–4</sup> such validations have been rarely done in the context of hybrid quantum mechanical and molecular mechanics (QM/MM) simulations.<sup>5–10</sup> This is an important void to fill for pushing the limit of QM/MM methods toward quantitative investigations of biological systems, in particular, enzymes.<sup>11,12</sup>

In the typical implementation,<sup>13</sup> the interaction between QM and MM regions consists of van der Waals, electrostatic, and bonded terms,

$$\hat{H}^{\text{QM/MM}} = \hat{H}_{\text{vdW}}^{\text{QM/MM}} + \hat{H}_{\text{elec}}^{\text{QM/MM}} + \hat{H}_{\text{bonded}}^{\text{QM/MM}} \quad (1)$$

Recently, we showed that free energy calculations are rather

insensitive to the van der Waals parameters for QM atoms<sup>14</sup> as long as the parameters are reasonable, due mainly to error cancellation effects. Another issue often raised in the QM/MM community concerns the treatment of the QM/MM boundary, for which numerous investigations and developments have been made using link atoms,<sup>13,15–17</sup> frozen orbitals,<sup>18</sup> generalized hybrid orbitals,<sup>19</sup> or pseudobonds.<sup>20</sup> Several studies<sup>16,21,18</sup> clearly showed that certain link-atom schemes (e.g., the single-link-atom approach) should be avoided in the calculation of quantities such as proton affinity, although other schemes often give similar results. For the energetics of reactions that conserve charge, the results are even less sensitive to the frontier treatments.<sup>17,21</sup> Since electrostatic interactions often dominate free energy changes in biomolecules, carefully treating  $\hat{H}_{\text{elec}}^{\text{QM/MM}}$  is of the utmost importance. Recently our group has implemented the generalized solvent boundary potential (GSBP) method<sup>22,23</sup> and Ewald sums<sup>24</sup> for combined QM/MM simulations. To quantitatively test these electrostatic models and to evaluate the accuracy of QM/MM protocols, reduction potential and pK<sub>a</sub> calculations provide useful probes due to their sensitivity to electrostatics and the availability of reliable experimental data. Here, we report investigations into the successes and limitations of various QM/MM protocols as applied to pK<sub>a</sub> calculations for titratable groups in solution and protein systems.

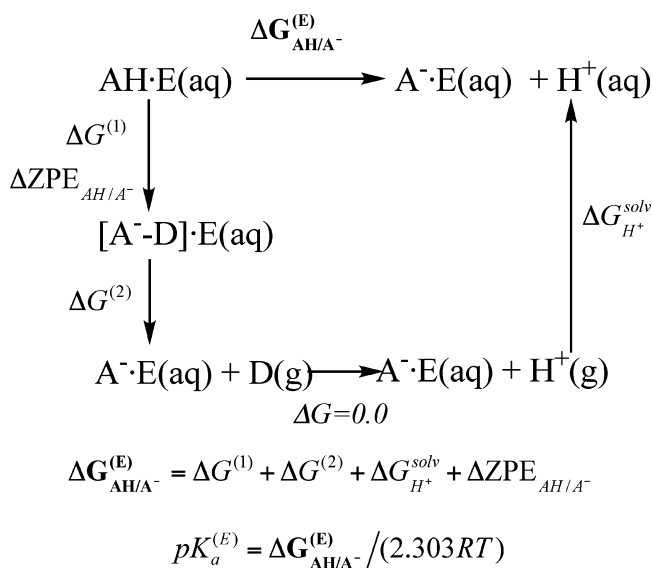
Application of these methods to reduction potential problems will be reported separately.<sup>25</sup>

Titrateable groups are often important to both structure and function of biomolecules, as reflected by the fact that protein stability and function depend sensitively on pH.<sup>26</sup> For example, adopting a specific protein structure may require an ionic residue to either form a salt-bridge or to attain the neutralized protonation state to stabilize a titrateable group in a hydrophobic core.<sup>27,28</sup> Placing titrateable groups in the active site to catalyze reactions via localized acid–base reactions is a common catalytic scheme;<sup>26</sup> algorithms have been developed to predict catalytic residues by identifying residues with significant  $pK_a$  shifts<sup>29</sup> and have met with moderate success. Finally,  $pK_a$  shifts of titrateable groups and water molecules coupled with redox chemistry form the basis of proton pumping in biomolecules involved in energy transduction.<sup>30–33</sup>

Not surprisingly, there has been a great deal of activity in calculating  $pK_a$  values in both the quantum chemistry and simulation fields. Quantum chemistry calculations coupled with a dielectric continuum,<sup>34</sup> hybrid QM/QM'/dielectric calculations, and QM/MM/dielectric calculations (as in recent applications of the effective fragment potential<sup>35</sup>) have been used with notable success.<sup>36,37</sup> The majority of  $pK_a$  studies for protein systems have utilized classical electrostatics calculations<sup>38,39,34</sup> by solving the Poisson–Boltzmann equation,<sup>40</sup> although there have been semimicroscopic and microscopic simulations as well.<sup>41–46</sup> Defining the protein dielectric constant ( $\epsilon_p$ ) and boundary is central to the continuum electrostatics calculations but is not straightforward. Antosiewicz et al.<sup>39</sup> calculated 60  $pK_a$  values in seven proteins and found  $\epsilon_p = 20$  gave the best results when compared to experiment; they justified the high dielectric constant as compensating for aspects lacking from their model, such as specific ion-binding sites and conformational relaxation. Studies by Warshel and co-workers<sup>43</sup> have shown that the appropriate “dielectric constant” depends on the electrostatic treatment and should be determined from discriminative benchmarks that, in addition, weed out unreliable treatments of electrostatics. In fact, they<sup>43</sup> suggested using two dielectric constants to describe the interaction of the titrateable residue with the protein charged groups ( $\epsilon_{\text{eff}}$ ) and permanent dipoles ( $\epsilon_d$ ), where  $\epsilon_d$  will be much lower than  $\epsilon_{\text{eff}}$  for more microscopic methods. These complications make continuum electrostatic based methods, with either quantum or classical treatment for the titrateable group, not well suited for the prediction of accurate  $pK_a$  for groups that may induce significant conformational rearrangements in the environment upon changing the protonation state. For those cases, microscopic models based on molecular dynamics and free energy perturbation<sup>41,44–46</sup> are more appropriate, although reproducing significant experimental  $pK_a$  shifts is very challenging due to limitations in the force field and the often slow convergence of free energy simulations. In a recent investigation, Simonson et al.<sup>44</sup> computed  $pK_a$  shifts relative to a model compound in solution for three aspartate residues in two proteins using both explicit solvent and the generalized Born model;<sup>47,48</sup> they found that the AMBER force field gave qualitatively correct shifts for all three, while the CHARMM force field predicted only two to have the correct shift direction.

In this investigation, we apply the free energy perturbation approach to calculate  $pK_a$  values with combined QM/MM potentials for small molecules in solution and two residues in the M102K T4-lysozyme mutant with large  $pK_a$  shifts. We will compare periodic boundary conditions and spherical boundary conditions, with and without including long-range electrostatics

## SCHEME 1



explicitly in the solution simulations. Since a change in protonation state is a chemical event, free energy simulation with QM/MM is, in principle, more reliable than simulations with nonpolarizable empirical force fields. Due to the uncertainty in several technical issues related to QM/MM simulations as mentioned above, however, one would not expect to necessarily acquire better agreement with experiment than semi-macroscopic (and even macroscopic) models where the various parameters (e.g., dielectric constant and atomic radii) can be empirically adjusted to implicitly include conformational relaxation, specific counterion binding sites, and induced dipoles lacking from these models. The overall goal of this investigation is to rigorously test typical QM/MM protocols in order to provide a clear picture of the accuracy and relative importance of various factors in determining the accuracy. The results will provide useful guidance to the quantitative application of QM/MM simulations in general.

In the following, we first describe the computational methods and simulation details in section 2; the results and discussion will be presented in section 3. We summarize this work with conclusions in section 4.

## 2. Computational Methods

Our general approach to calculating the free energy of deprotonation<sup>49</sup> is summarized in Scheme 1. First, the solute (or the protein titrateable group) is switched from a protonated QM molecule to a deprotonated QM molecule bound to an MM dummy atom. The dummy atom has the same van der Waals parameters as the ionizable QM proton, zero charge and typical MM bonded parameters (bond, angle, and dihedral terms). The zero point energy (ZPE) difference between protonation states is calculated using gas-phase models. The second step removes the van der Waals interaction and the MM bonded terms associated with the dummy atom. Finally, the total free energy change ( $\Delta G_{\text{AH/A}^-}^{(\text{E})}$ ) and absolute  $pK_a$  can be calculated by including the solvation free energy of the proton. The proton solvation energy is difficult to measure or compute accurately<sup>50–52</sup> but unnecessary when considering  $pK_a$  shifts relative to a reference molecule. With the current protocol,<sup>49</sup> the  $pK_a$  shift ( $\Delta pK_a$ ) between two systems is given by

$$\Delta\Delta G_{\text{AH/A}^- \rightarrow \text{BH/B}^-}^{(E)} = \Delta\Delta G_{\text{AH/A}^- \rightarrow \text{BH/B}^-}^{(1)} + \Delta\Delta G_{\text{AH/A}^- \rightarrow \text{BH/B}^-}^{(2)} + \Delta\Delta\text{ZPE}_{\text{AH/A}^- \rightarrow \text{BH/B}^-} \quad (2)$$

and

$$\Delta\text{p}K_a = \frac{\Delta\Delta G_{\text{AH/A}^- \rightarrow \text{BH/B}^-}^{(E)}}{2.303RT} \quad (3)$$

where AH and BH represent two different solutes. The  $\text{p}K_a$  shift due to different environments (solution or protein) can be calculated similarly,

$$\Delta\Delta G_{\text{S-AH/A}^- \rightarrow \text{E-AH/A}^-}^{(E)} = \Delta\Delta G_{\text{S-AH/A}^- \rightarrow \text{E-AH/A}^-}^{(1)} + \Delta\Delta G_{\text{S-AH/A}^- \rightarrow \text{E-AH/A}^-}^{(2)} \quad (4)$$

where S- and E- represent solution and protein environments, respectively;  $\Delta\Delta\text{ZPE}$  is calculated from gas-phase models and therefore does not contribute in eq 4. In the following, different components that contribute to  $\text{p}K_a$  in the dual topology single coordinate scheme will be briefly reviewed to define terms used in the discussion; the reader is referred to refs 49 and 8 for detailed development. The perturbative corrections used to improve computed  $\text{p}K_a$  values, electrostatic treatments in the SCC-DFTB/MM framework, and specific simulation setups will then be described.

## 2.1. Dual Topology Single Coordinate (DTSC) Approach to $\text{p}K_a$ Calculations.

**2.1.1. Calculation of  $\Delta G^{(1)}$ .** The free energy for converting the proton to a dummy atom ( $\Delta G_{\text{AH/A}^-}^{(1)}$  as described above) may be calculated with the dual topology single coordinate scheme<sup>8</sup> in a QM/MM framework. This method is based on thermodynamic integration<sup>53</sup> that uses a series of independent explicit solvent simulations that reversibly convert the potential energy of the system from AH ( $\lambda = 0$ ) to  $\text{A}^- - \text{D}$  ( $\lambda = 1$ ). Since QM potentials are not separable, the DTSC scheme computes a scaled potential due to both QM states (AH and  $\text{A}^- - \text{D}$ ) at the same configuration,

$$U(\mathbf{X}_{\text{AH}}, \mathbf{X}_{\text{C}}; \lambda) = (1 - \lambda)\langle \Phi_{\text{AH}} | \hat{H}_{\text{AH}}(\mathbf{X}_{\text{AH}}) + \hat{H}_{\text{AH/C}}(\mathbf{X}_{\text{AH}}, \mathbf{X}_{\text{C}}) | \Phi_{\text{AH}} \rangle + \lambda \langle \Phi_{\text{A}^- - \text{D}} | \hat{H}_{\text{A}^- - \text{D}}(\mathbf{X}_{\text{AH}}) + \hat{H}_{\text{A}^- - \text{D/C}}(\mathbf{X}_{\text{AH}}, \mathbf{X}_{\text{C}}) | \Phi_{\text{A}^- - \text{D}} \rangle + U_{\text{CC}}(\mathbf{X}_{\text{C}}) \quad (5)$$

where C indicates the environment (e.g., water) and the notations emphasize that the same set of coordinates is used for both AH and  $\text{A}^- - \text{D}$ . The QM/MM interaction energies do not depend on  $\lambda$ , and the free energy derivative is readily determined from eq 5. If the same set of van der Waals parameters is used for the interaction between the MM and the two QM states, only the electrostatic term contributes explicitly to the free energy derivative:

$$\frac{\partial G(\mathbf{X}_{\text{AH}}, \mathbf{X}_{\text{C}}; \lambda)}{\partial \lambda} = \left\langle \frac{\partial U(\mathbf{X}_{\text{AH}}, \mathbf{X}_{\text{C}}; \lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (6)$$

$$= \langle U_{\text{elec}}^{\text{A}^- - \text{D/MM}}(\mathbf{X}_{\text{AH}}, \mathbf{X}_{\text{C}}) - U_{\text{elec}}^{\text{AH/MM}}(\mathbf{X}_{\text{AH}}, \mathbf{X}_{\text{C}}) \rangle_{\lambda} \quad (7)$$

where  $U_{\text{elec}}^{\text{AH/MM}}$  ( $U_{\text{elec}}^{\text{A}^- - \text{D/MM}}$ ) is the total electronic energy of the protonated (deprotonated) QM solute, including the electrostatic interaction with the MM environment. The value of  $\lambda$  defines the solute–environment interaction over which the difference between the potential energies for both end states is averaged. The free energy ( $\Delta G^{(1)}$ ) is then calculated from the free energy

derivatives by integrating from  $\lambda = 0.0$  to 1.0. The method is formally exact because the free energy is a state function, although negligible errors will arise in practical simulations due to constrained bond lengths involving hydrogen atoms.

Two technical issues generally arise in free energy simulations: the number of  $\lambda$  windows and the length of simulation for each  $\lambda$ . Since the change in protonation state is dominated by electrostatics, for which linear response approximation is often found applicable,<sup>54</sup> a small number of  $\lambda$  windows can often be used. Although this is true for many solution cases, the situation in protein systems can be more complex. For example, the change in protonation state of a specific residue (especially those buried in the interior) can be associated with significant conformational changes in the protein, which makes the linear approximation inappropriate; a recent example that emphasizes this scenario has been reported in a free energy perturbation study of  $\text{p}K_a$  shifts of aspartatic acid residues in thioredoxin.<sup>44</sup>

As to the length of simulation, sufficient configurational sampling is critical to the accuracy of computed free energy difference. Because the reorganization of the environment due to charge changes may take a long time, convergence of the free energy (derivative) is a major practical issue. For complex biological systems, proper equilibration is often difficult to quantify, and including nonequilibrium configurations into the averaging will increase uncertainty. A useful protocol was recently proposed by Yang and co-workers,<sup>55</sup> where the cumulative averaging was initiated from the most recent point back to the early stage of the simulation. Termed “reverse cumulative averaging”, this method utilizes block averaging to identify the equilibration region which may then be discarded from the averaging; this was applied in this investigation.

**2.1.2.  $\Delta\text{ZPE}$ , van der Waals, and Internal Contributions to  $\text{p}K_a$ .** In addition to the electrostatic component, there are contributions to the  $\text{p}K_a$  from zero point energy difference, as well as bonded and van der Waals terms associated with the dummy atom in going from the protonated to the deprotonated states (see Scheme 1). Zero point energy difference ( $\Delta\text{ZPE}$ ) can be significant (−5 to −10 kcal/mol); in this study  $\Delta\text{ZPE}$  values are calculated for each QM state of interest with standard harmonic frequency calculation using gas-phase models. The bonded and van der Waals contributions due to the annihilation of the dummy atom are associated with  $\Delta G^{(2)}$  (see Scheme 1),

$$\Delta G^{(2)} = \Delta G_{\text{D}}^{\text{vdW}} + \Delta G_{\text{D}}^{\text{bond}} = - \int_0^1 d\lambda \langle U_{\text{D}}^{\text{vdW}} \rangle_{\lambda} + \Delta G_{\text{D}}^{\text{bond}} \quad (8)$$

where the van der Waals term is calculated from a second thermodynamic integration calculation in which the van der Waals interactions involving the dummy atom are gradually switched to zero. The bonded term associated with the MM angle and dihedral of the dummy atom is estimated from local configurational integrals,<sup>56–58</sup>

$$\Delta G_{\text{D}}^{\text{bond}} = -k_{\text{B}}T \ln \frac{V_0 \Lambda_{\text{d}}^{-1}}{r_{\text{A}^- - \text{D}}^2 \sin \theta_{\text{A}^- - \text{D}} \sqrt{(2\pi k_{\text{B}}T)^2 / K_{\theta} K_{\tau}}} + \frac{5}{2} k_{\text{B}}T \quad (9)$$

where  $r_{\text{A}^- - \text{D}}$  ( $\theta_{\text{A}^- - \text{D}}$ ) is the distance (angle) that characterizes the connection between the dummy atom and  $\text{A}^-$ . The quantities  $V_0$  and  $\Lambda_{\text{d}}$  are the molar volume and thermal wavelength associated with the dummy atom under the standard state;  $\Lambda_{\text{d}}$  contributes here because SHAKE was used to constrain the  $\text{A}^- - \text{D}$  bond. The magnitude of the bonded term is usually between −3 and −6 kcal/mol.



**2.2. Gas-Phase and QM/MM Coupling Corrections.** An approximate density functional theory method (SCC-DFTB)<sup>59</sup> was chosen for the free energy calculations based on its balance of accuracy and speed. Although SCC-DFTB was shown to be more accurate than commonly used semiempirical methods such as AM1 and PM3 for many type of chemical processes,<sup>60</sup> further corrections are needed for quantitative  $pK_a$  values. In this investigation, two types of corrections were considered. First, the SCC-DFTB/MM energy was corrected with B3LYP/MM to attain more accurate description for both the QM component and QM/MM interactions (vide infra). Second, the gas-phase proton affinity of the titratable group was corrected using B3LYP<sup>61–63</sup> and CCSD<sup>64</sup> calculations. Both adiabatic energy difference and an average vertical energy difference (for 50 solute configurations taken from an MD trajectory) between protonated and deprotonated QM states were evaluated. The B3LYP and CCSD calculations were carried out with the 6-31+G(d,p) and 6-311+G(d,p) basis sets,<sup>65,66</sup> respectively.

While the gas-phase correction is straightforward, the QM/MM correction warrants further discussion. The SCC-DFTB/MM method calculates the electrostatic interaction between QM and MM atoms using Mulliken charges on the QM atoms,<sup>67</sup>

$$\hat{H}_{\text{elec}}^{\text{SCC/MM}} = \sum_{I \in \text{MM}} \sum_{J \in \text{QM}} \frac{q_I \Delta q_J}{r_{IJ}} \quad (10)$$

where  $\Delta q_J$  is the Mulliken charge on QM atom J. This formulation is consistent with the way that charge relaxation is achieved for QM atoms in the SCC-DFTB method.<sup>59</sup> Since the QM electron distribution is represented by atom-based point charges, this approximation does not allow MM charges to interact with different orbitals on a QM atom differently. The more rigorous QM/MM coupling treatment includes the contribution from MM point charges in one electron integrals, which was done in the B3LYP/MM calculations.<sup>68</sup> Therefore, correcting SCC-DFTB/MM results based on B3LYP/MM calculations improves not only the QM level itself but also the way that MM region interacts and polarizes the QM region. The correction was done based on a straightforward one-step free energy perturbation,

$$\Delta G_{\text{B3LYP-SCC}} = -kT \ln \langle e^{-\beta(U_{\text{B3LYP/MM}} - U_{\text{SCC/MM}})} \rangle_{\text{SCC/MM}} \quad (11)$$

which was done at both end states ( $\lambda = 0, 1$ ). The difference between the perturbative correction at the two end states gives the B3LYP/MM correction to the free energy of deprotonation. Since only a small number (100) of snapshots from SCC-DFTB/MM trajectories was used, a second-order cumulant expansion<sup>69</sup> was used to improve the numerical stability of the perturbation calculation,

$$\Delta G_{\text{B3LYP-SCC}} = \langle U_{\text{B3LYP/MM}} - U_{\text{SCC/MM}} \rangle_{\text{SCC/MM}} - \frac{\beta}{2} [ \langle (U_{\text{B3LYP/MM}} - U_{\text{SCC/MM}})^2 \rangle_{\text{SCC/MM}} - \langle U_{\text{B3LYP/MM}} - U_{\text{SCC/MM}} \rangle_{\text{SCC/MM}}^2 ] \quad (12)$$

As discussed extensively in the literature,<sup>70,71</sup> such one-step perturbation is effective only if the configuration distribution at the two levels overlap significantly; this was assumed to be the case considering the previous observation<sup>60</sup> that SCC-DFTB often gives reliable geometries and energetics compared to B3LYP. Whether such an assumption is valid in practical cases can only be established by test calculations, as was done here. Although the accuracy of the perturbative correction can

certainly be improved by performing perturbation between B3LYP/MM and SCC-DFTB/MM at snapshots from B3LYP/MM trajectories, using either a linear response model<sup>54</sup> or more sophisticated weighted-average,<sup>72</sup> we did not pursue this option here considering the high cost of B3LYP/MM MD simulations.

**2.3. Electrostatics Methods.** As discussed above, the appropriate treatment of electrostatic interactions is central to accurate free energy calculations. Two general protocols based on the periodic boundary condition and finite-size boundary condition were used here. With Ewald sums and related approaches,<sup>24</sup> the periodic boundary condition allows treatment of electrostatics without truncation. The imposed periodicity, however, may cause artifacts in the simulations of highly charged systems such as biomolecules.<sup>73–78</sup> To minimize these artifacts, a large number of explicit solvent molecules are required for highly polar systems; therefore, the size and shape of solvated biological molecules that may be simulated is restricted. When the region of interest is localized in space, the finite-size boundary condition provides an efficient alternative that is conceptually more appropriate for the infinite dilute condition, which is often most relevant. The accuracy of such calculations, however, depends critically on the treatment of bulk dielectric response and boundary polarization.<sup>11,9,79</sup> For QM/MM simulations, since most implementations do not impose any truncation on the QM/MM interactions, MM/MM interactions also have to be treated in a balanced manner to avoid overpolarization of the system.<sup>10</sup> In this study, we tested both types of boundary conditions in SCC-DFTB/MM free energy perturbation calculations. For the finite boundary simulations of  $pK_a$  in solution, the conventional stochastic boundary condition with two different electrostatic models was used; the first model used the force-shift (Fshift) cutoff,<sup>80</sup> and the second used an extended electrostatics (Ext) treatment,<sup>80</sup> which avoids truncation by approximating electrostatic interactions beyond a radius with group based multipolar expansions (dipole and quadrupole moments). In addition, we recently implemented the generalized solvent boundary potential (GSBP) approach<sup>23</sup> in the SCC-DFTB/MM framework,<sup>10</sup> which provides a reliable and efficient protocol for treating electrostatics for chemical processes that occur in solvated macromolecules. Although preliminary calculations<sup>10</sup> on human carbonic anhydrase II suggest that GSBP based SCC-DFTB/MM simulations are more consistent with available experimental data compared to stochastic boundary simulations with either force-shift or extended electrostatics,  $pK_a$  calculations are a more stringent and quantitative test. For reliable periodic boundary simulations, we implemented the Ewald sum approach in the SCC-DFTB/MM framework in the program CHARMM.<sup>81</sup> Since Ewald summation in the QM/MM framework has become popular only very recently,<sup>5,6</sup> we briefly describe our implementation in the following. For the details of the GSBP implementation in the QM/MM framework, see ref 10.

**2.3.1. Ewald Sum with SCC-DFTB/MM.** The slowly converging Coloumbic energy of a charge-neutral periodic system,

$$E_{\text{Coul}} = \frac{1}{2} \sum_n \sum_{i,j=1}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \quad (13)$$

sums over all charge–charge ( $q_i q_j$ ) interactions at lattice-site dependent separations ( $\mathbf{r}_{ij} + \mathbf{n}$ ). The Ewald sum is a numerical approach that efficiently evaluates the Coloumbic energy by adding and subtracting a screening potential that rapidly converges in the reciprocal (Fourier) space. The total electrostatic energy is split into real space and reciprocal space terms,

$$E_{\text{Ewald}} = E_{\text{real}} + E_{\text{recip}} - E_{\text{self}}$$

$$= -\sum_{i,j=1}^N \sum_{|\mathbf{n}=0}^{\infty} q_i q_j \frac{\text{erfc}(\kappa|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} + \frac{1}{2V} \sum_{i,j=1}^N \sum_{k \neq 0} q_i q_j \frac{4\pi}{k^2} \times$$

$$\exp\left(-\frac{k^2}{4\kappa^2}\right) \cos(k\hat{\mathbf{k}} \cdot \mathbf{r}_{ij}) - \frac{\kappa}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \quad (14)$$

As described above, SCC-DFTB Mulliken charges are used to calculate the electrostatic interaction with MM point charges variationally. The Mulliken charges can be introduced into the Ewald sum equations to calculate the QM/QM and QM/MM lattice sum contributions to the total Ewald potential energy in a straightforward way,

$$E_{\text{Ewald}}^{\text{total}} = E_{\text{Ewald}}^{\text{QM}} + E_{\text{Ewald}}^{\text{QM/MM}} + E_{\text{Ewald}}^{\text{MM}} \quad (15)$$

The  $E_{\text{Ewald}}^{\text{QM}}$  term corresponds to the Ewald (real, reciprocal, and self) interaction between intramolecular and intermolecular-image QM Mulliken charges; due to the cutoff used in the SCC-DFTB matrix elements,<sup>59</sup> all long-range electrostatic interactions involve Mulliken charges and are, therefore, straightforward to handle considering the typically small number of QM atoms. The  $E_{\text{Ewald}}^{\text{QM/MM}}$  term is the Ewald interaction between QM Mulliken charges ( $\Delta q_i$ ) and MM point charges,

$$E_{\text{Ewald}}^{\text{QM/MM}} = E_{\text{real}}^{\text{QM/MM}} + E_{\text{recip}}^{\text{QM/MM}}$$

$$= \sum_{i \in \text{QM}} \sum_{j \in \text{MM}} \sum_{|\mathbf{n}=0}^{\infty} \Delta q_i q_j \frac{\text{erfc}(\kappa|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} +$$

$$\frac{1}{V} \sum_{i \in \text{QM}} \sum_{j \in \text{MM}} \sum_{k \neq 0} \Delta q_i q_j \frac{4\pi}{k^2} \exp\left(-\frac{k^2}{4\kappa^2}\right) \cos(k\hat{\mathbf{k}} \cdot \mathbf{r}_{ij})$$

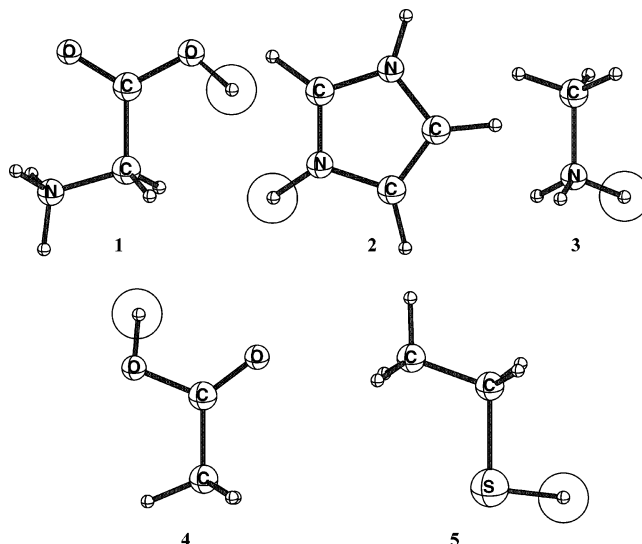
$$= \sum_{i \in \text{QM}} \Delta q_i \phi_{\text{Ewald}}^{\text{MM}}(\mathbf{r}_i) \quad (16)$$

where the self-interaction term is no longer present (because there are no  $i = j$  terms in the reciprocal sum). The  $E_{\text{Ewald}}^{\text{QM}}$  and  $E_{\text{Ewald}}^{\text{QM/MM}}$  terms give rise to new contributions to the Hamiltonian matrix elements for the SCF procedure. As shown in eq 16, it is useful to define the Ewald potential due to all MM charges at the position of QM atom  $i$ ,  $\phi_{\text{Ewald}}^{\text{MM}}(\mathbf{r}_i)$ ,

$$\phi_{\text{Ewald}}^{\text{MM}}(\mathbf{r}_i) = \sum_{j \in \text{MM}} \sum_{|\mathbf{n}=0}^{\infty} q_j \frac{\text{erfc}(\kappa|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} + \frac{1}{V} \sum_{j \in \text{MM}} \sum_{k \neq 0} q_j \frac{4\pi}{k^2} \times$$

$$\exp\left(-\frac{k^2}{4\kappa^2}\right) \cos(k\hat{\mathbf{k}} \cdot \mathbf{r}_{ij}) \quad (17)$$

because this quantity is a constant (for each QM atom) during the SCF iteration and therefore only has to be computed once before SCF; a similar scheme was also adopted in the QM/MM implementation of Ewald by York and co-workers.<sup>6</sup> In the presence of a large number of MM atoms, the loop over MM charges in computing  $\phi_{\text{Ewald}}^{\text{MM}}(\mathbf{r}_i)$  is conveniently distributed over many nodes with Message Passing Interface. To compute force, the derivative of  $E_{\text{Ewald}}^{\text{QM/MM}}$  with respect to the QM and MM coordinates are needed. For calculating the QM derivatives, the derivatives of  $\phi_{\text{Ewald}}^{\text{MM}}(\mathbf{r}_i)$  with respect to  $\mathbf{r}_i$  are evaluated in a straightforward manner. For calculating the MM derivatives, it is efficient to define  $\phi_{\text{Ewald}}^{\text{QM}}(\mathbf{r}_j)$ , the Ewald potential due to



**Figure 1.** The five small molecules studied here for  $pK_a$  values in solution. The acidic proton, which is mutated into a dummy atom in the QM/MM free energy perturbation calculations, is circled for each molecule. (1) Imidazole,  $pK_a = 7.0$ ; (2) methylamine,  $pK_a = 10.7$ ; (3) glycine,  $pK_a = 2.4$ ; (4) acetic acid,  $pK_a = 4.8$ ; (5) ethanethiol,  $pK_a = 10.6$ . For molecules 1–3, the charge changes from +1 to 0 upon deprotonation, while the charge changes from 0 to −1 for molecules 4 and 5.

all QM Mulliken charges on the  $j$ th MM atom, similar to  $\phi_{\text{Ewald}}^{\text{MM}}(\mathbf{r}_i)$ ,

$$\phi_{\text{Ewald}}^{\text{QM}}(\mathbf{r}_j) = \sum_{i \in \text{QM}} \sum_{|\mathbf{n}=0}^{\infty} \Delta q_i \frac{\text{erfc}(\kappa|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} +$$

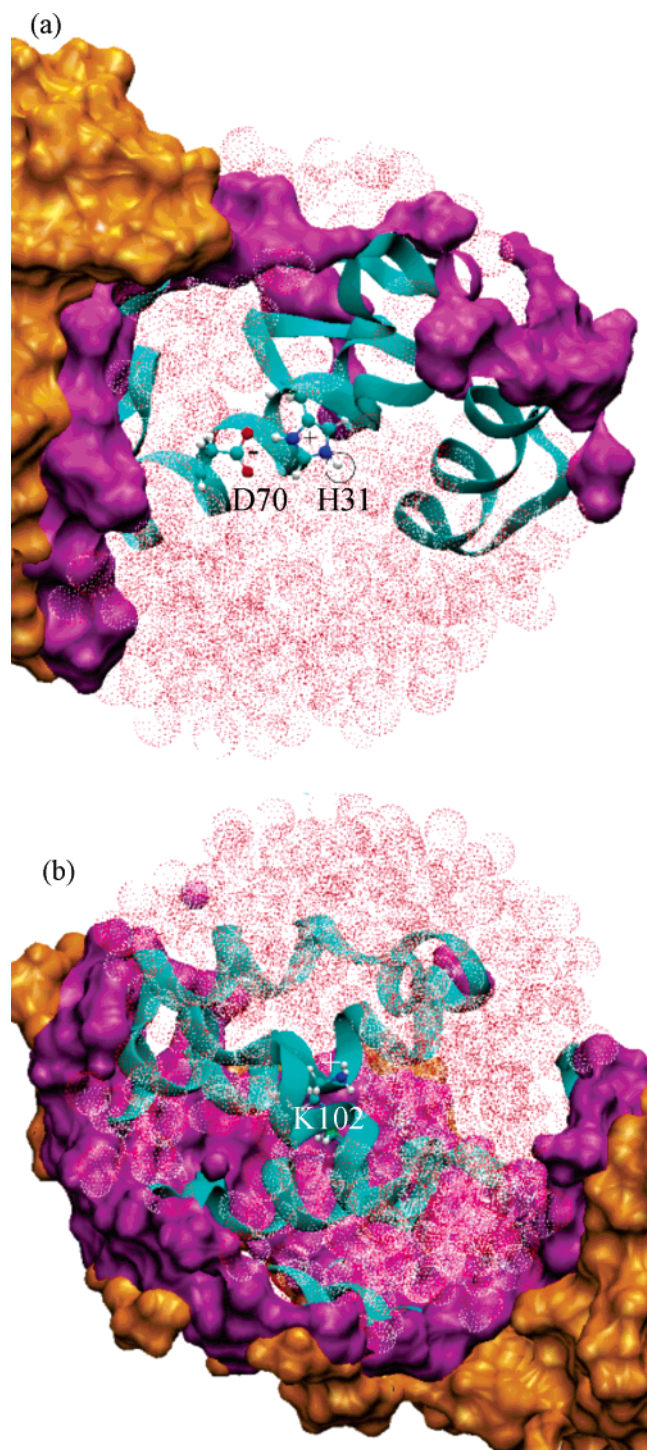
$$\frac{1}{V} \sum_{i \in \text{QM}} \sum_{k \neq 0} \Delta q_i \frac{4\pi}{k^2} \exp\left(-\frac{k^2}{4\kappa^2}\right) \cos(k\hat{\mathbf{k}} \cdot \mathbf{r}_{ij}) \quad (18)$$

However, since the number of QM atoms is typically smaller than the number of wave vectors used in the reciprocal sum, it is more efficient to distribute different sections of wave vector sums to different nodes in parallel calculations. For this study, the standard Ewald treatment discussed above is employed for QM/QM and QM/MM electrostatic interactions, while the MM/MM electrostatic interactions are determined with particle-mesh-Ewald<sup>82</sup> as implemented in CHARMM by Brooks and co-workers.<sup>75</sup> For a system that is not charge-neutral, the total volume correction term<sup>75</sup> contains QM and MM contributions,

$$E_{\text{volume}} = -\frac{\pi}{2\kappa^2 V} \left[ \sum_{i \in \text{QM}} \Delta q_i + \sum_{j \in \text{MM}} q_j \right]^2 \quad (19)$$

**2.4. Simulation Setup.** **2.4.1. Small Molecules.** The  $pK_a$  shifts between a series of small molecules (see Figure 1) were calculated with finite-size and periodic boundary conditions using the thermodynamic integration approach described above (see Scheme 1). This series of five small molecules was chosen to represent titratable groups of amino acids. The free energy derivatives were calculated until convergence was achieved for three  $\lambda$  values ( $\lambda = 0, 0.5$ , and  $1.0$ ), after the initial equilibration period of 80 ps. In addition, the convergence was tested and the free energy derivatives with the corresponding statistical error were determined with the reverse cumulative average.<sup>55</sup> Three windows were expected to be sufficient due to the linear response of the homogeneous solvent. The linear fit of the free





**Figure 2.** Generalized solvent boundary potential (GSBP) setup for the  $pK_a$  calculation of (a) His 31 and (b) Lys 102 in the M102K mutant of T4 lysozyme. The system was centered on (a) the acidic nitrogen in His 31 and (b) the acidic nitrogen in Lys 102; the solvated system was partitioned into fixed (orange), harmonically constrained (purple), and reaction regions (cyan) embedded in a dielectric continuum (shown schematically with diagonal lines).

energy derivatives was then integrated to yield the free energy of converting the acidic proton to a dummy atom ( $\Delta G^{(1)}$ ). The ZPE, internal, and van der Waals contributions were then calculated to complete the thermodynamic cycle (Scheme 1).

In preparing the finite-size boundary condition simulations, each small molecule (treated with SCC-DFTB) was solvated in a 16 Å sphere with approximately 600 TIP3P water molecules<sup>85</sup> subject to a deformable boundary potential;<sup>84</sup> the

**TABLE 1: Gas Phase Proton Affinities (in kcal/mol) Calculated at the SCC-DFTB, B3LYP, and CCSD Levels<sup>a</sup>**

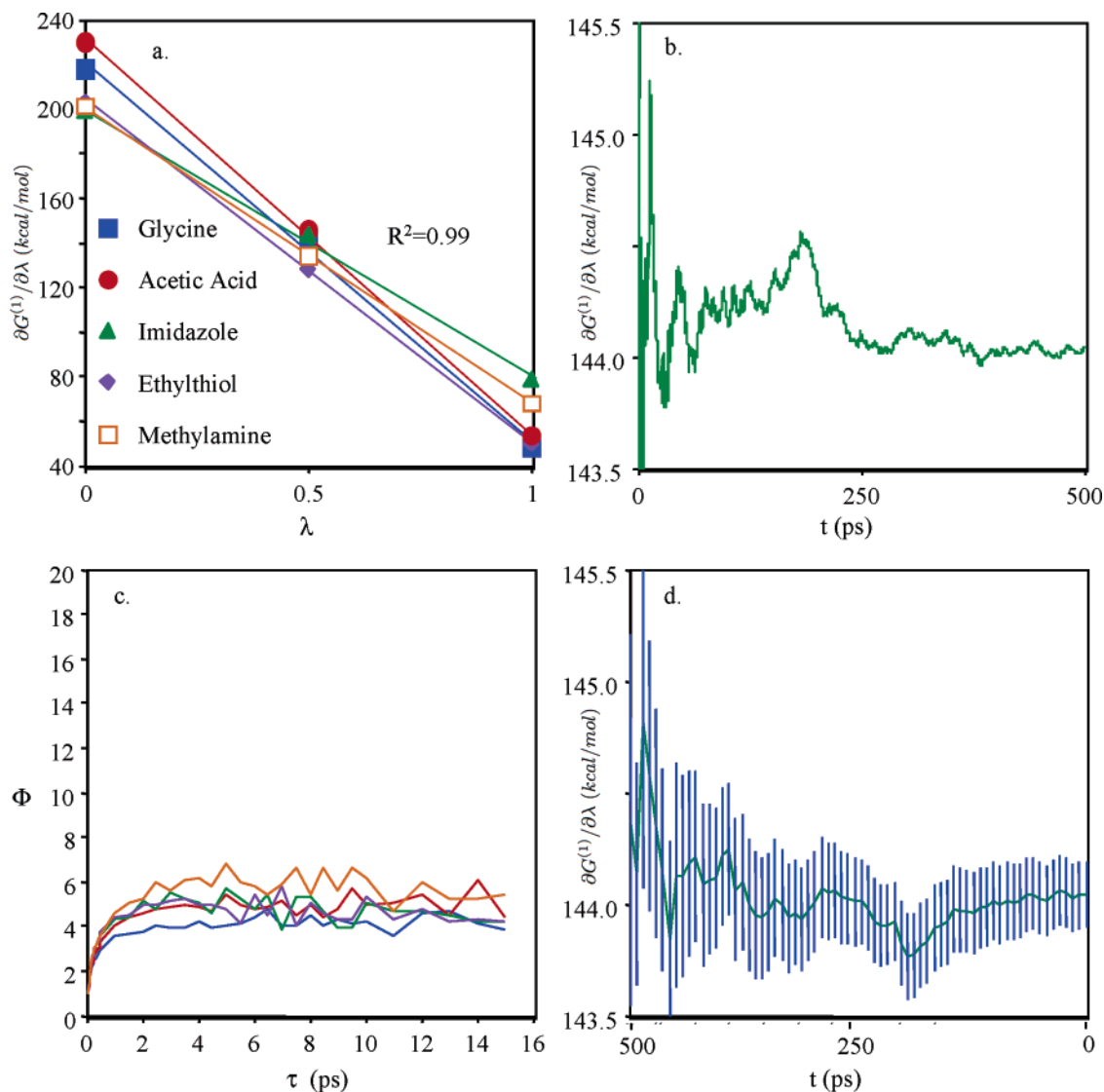
	SCC-DFTB	B3LYP	CCSD <sup>a</sup>
glycine	250.5	241.7	248.2
acetic acid	366.4	353.0	359.5
imidazole	231.3	234.0	235.7
ethanethiol	355.7	358.7	367.0
methylamine	206.7 (218.2) <sup>b</sup>	223.4	226.0

<sup>a</sup> Basis set used in the B3LYP and CCSD calculations is 6-31+G(d,p) and 6-311+G(d,p), respectively. The calculations were carried out at the SCC-DFTB geometries. <sup>b</sup> The value in parentheses was obtained with a set of modified N–H repulsive potential in the SCC-DFTB, which was found to systematically improve proton affinities for nitrogen-containing molecules (M. Elstner, private communication). However, the standard set of repulsive potential was used in all the  $pK_a$  calculations to illustrate the effectiveness of the QM/MM and QM corrections (see text).

system was partitioned into an inner, Newtonian MD region (within 13 Å), and an outer, Langevin MD region (13–16 Å)<sup>85</sup> with a temperature bath at 300 K. Bonds involving hydrogen were constrained with the SHAKE algorithm,<sup>86</sup> allowing 2 fs time steps for MD propagation. Nonbond interactions within the water sphere were either reduced to zero at 12 Å using a force-shift (Fshift) scheme or treated with an extended electrostatics (Ext) model where groups beyond 12 Å interact as multipoles.<sup>80</sup> The center of mass of the solute was restrained to the center of the water sphere with a weak harmonic potential (force constant of 10 kcal/Å<sup>2</sup>) that contributes, on average, about 1 kcal/mol to the total energy of the system. Long-range electrostatics due to the bulk were estimated with either finite difference Poisson–Boltzmann calculations (using snapshots from the MD simulations) or the Born correction. In addition, free energy simulations were also carried out with the GSBP approach with the extended electrostatics treatment of the Coloumbic interactions in the inner region. In the GSBP setup, 400 spherical harmonics (through the 20th order) were used to expand the charge distribution of the entire system (i.e., the outer region contains no atoms). The reaction field matrix, **M**, was constructed for the 16 Å water sphere placed in an 18 Å  $\epsilon = 1$ , sphere.

In preparing the periodic boundary simulations, each small molecule was solvated with approximately 920 TIP3P<sup>83</sup> water molecules in a 30.2 Å cube. The box dimensions and temperature were maintained within the isothermal isobaric ensemble.<sup>87</sup> Nonbond interactions were treated with Ewald sums for the QM (SCC-DFTB) and QM/MM interactions while particle-mesh–Ewald<sup>82</sup> was used for MM (TIP3P) atoms. The convergence of the energy and forces were used to determine the Ewald sum parameters ( $\kappa = 0.45$  and  $k_{\max} = 6$ ) and PME parameters ( $\kappa = 0.33$ , B-spline interpolation order of 10, and 32 grid points for the charge mesh). No charge corrections were added to the system.<sup>75</sup>

The  $\Delta ZPE$  and internal contributions for each small molecule were calculated, as described in section 2.1.2. The van der Waals contribution ( $\Delta G^{(2)}$ , Scheme 1) was expected to be small<sup>49,14</sup> and was calculated for only imidazole and acetic acid with the stochastic boundary set up; two sets of van der Waals parameters were used for the acidic proton/dummy atom ( $\epsilon = -0.04$  kcal/mol,  $\sigma = 0.28$  Å, and  $\epsilon = -0.02$  kcal/mol,  $\sigma = 1.32$  Å) to determine a range of possible contributions.<sup>14</sup> Both the Poisson–Boltzmann bulk electrostatic and B3LYP/6-31+G(d,p) QM/MM coupling corrections were calculated from 100 MD snapshots (adjacent data points separated by  $\sim 3$  ps, which is similar to the appropriate block size determined from the reverse cumulative average analysis of the SCC-DFTB/MM trajectories) taken



**Figure 3.** Linear fits and convergence of the free energy derivatives (for  $\Delta G^{(1)}$ ) for the five small molecules in solution. (a) Linear fits of the free energy derivatives as a function of  $\lambda$ ; example of free energy derivative convergence for imidazole with  $\lambda = 0.5$  based on (b) forward cumulative average and (d) reverse cumulative average. In (c), the  $\Phi$  convergence behavior is shown.

from the SCC-DFTB/MM trajectories. While the QM/MM coupling corrections were calculated from  $\lambda = 0$  and  $\lambda = 1$  trajectories, the Poisson–Boltzmann bulk electrostatic corrections were determined from an average of all three  $\lambda$  values (integrated linear fit). In the B3LYP/MM calculations, the MM region included the entire sphere for spherical boundary simulations while only water molecules within 12 Å from the solute were included for periodic boundary simulations. For further improvements, gas-phase proton affinity calculations were carried out with CCSD/6-311+G(d,p) for both SCC-DFTB minimized structures in the gas phase (for glycine, which is not stable in the deprotonated state, vertical proton affinity was computed) and a set of 50 solute configurations collected from SCC-DFTB/MM trajectories.

**2.5. T4-Lysozyme.** Using the SCC-DFTB/MM based GSBP approach, the  $pK_a$  shifts for histidine 31 (H31) and lysine 102 (K102) in the T4-lysozyme M102K mutant<sup>88</sup> were calculated relative to imidazole and propyl-amine in solution, respectively. These pairs were chosen because the  $pK_a$  shifts were among the largest known experimentally due to a change in environment, and as a result, they have been regarded as stringent test cases for electrostatic models.<sup>43</sup>  $\Delta G^{(1)}$  (Scheme 1) was calculated using the GSBP protocol with SCC-DFTB/MM free energy

simulations. Five  $\lambda$  values ( $\lambda = 0, 0.2, 0.5, 0.8$ , and  $1.0$ ) were sampled for 3 to 12 ns, depending on the stability of the protein and the convergence of the free energy derivative; more  $\lambda$  values were simulated compared to the solution studies to consider the possibility of more complex reorganization of the protein upon a change in protonation state.<sup>44</sup> The convergence of free energy derivatives was analyzed using the reverse cumulative average protocol.<sup>55</sup> The  $pK_a$  shifts were calculated solely from the differences between  $\Delta G^{(1)}$ 's calculated in solution and in the T4-lysozyme. Since the ZPE and internal contributions are independent of environment, they do not contribute; in addition, the van der Waals contributions were assumed to be sufficiently small and were ignored. The QM/MM coupling corrections were not expected to be as critical for shifts due to the change in environment, although this assumption should be evaluated more carefully in the future.

Two setups were used because the H31 and K102 residues are in different regions of the protein and are  $\sim 20$  Å apart. For H31, the system was centered on the N $\epsilon$  of H31 (see Figure 2a); for K102, the system was centered on the amine group of K102 (see Figure 2b). In both cases, the protein was partitioned into a 20 Å inner region with the remaining portion of the system in the outer region. The static field due to the outer region atoms,

**TABLE 2:  $\Delta G^{(1)}$  (in kcal/mol) Component (Scheme 1) for the Five Small Molecules Studied<sup>a</sup>**

	Fshift	Ext	GSBP	Ewald
glycine		126.5	136.5	128.3
acetic acid	143.2	152.7	143.5	134.1
imidazole <sup>b</sup>	170.0	131.0	140.9	131.2
ethanethiol		135.4	127.0	118.1
methylamine		124.9	134.9	126.1
propylamine <sup>b</sup>			135.4	

<sup>a</sup> No bulk solvation contribution is included in the Fshift and Ext results (see Table 3). Based on the reverse cumulative average analysis, the statistical error in the computed  $\Delta G^{(1)}$  values is on the order of 0.2–0.3 kcal/mol (see Figure 4d). <sup>b</sup> Imidazole and propylamine were used as the reference for His 31 and Lys 102, respectively, in the pK<sub>a</sub> shift calculations for the M102K mutant of the T4 lysozyme.

**TABLE 3: Various Bulk Solvation Contributions (in kcal/mol) to the Deprotonation Free Energy Considered for the Small Molecules in Solution<sup>a</sup>**

	$\Delta G^{\text{Born}, a}$	$\Delta G^{\text{PB1}, b}$ Fshift	$\Delta G^{\text{PB2}, c}$ Ext	$\Delta G^{\text{PB1}, b}$
glycine	9.2		19	10
acetic acid	−9.2	−1	−8	−10
imidazole	9.2	−19	18	10
ethanethiol	−9.2		−13	−10
methylamine	9.2		23	10

<sup>a</sup> Born correction for the solvation of a sphere with the radius of 18 Å and  $\epsilon=1$ . The sign of the correction depends on whether the solute is charged in the protonated or deprotonated state. The radius of 18 Å was chosen for the 16 Å spherical boundary simulations because a 2 Å dielectric buffer was considered based on the oxygen–oxygen distribution in bulk water. <sup>b</sup> The Poisson–Boltzmann correction was calculated with the dielectric interface defined by the van der Waals radii in the CHARMM 22 force field. The PB calculations were calculated at all three  $\lambda$  values (100 configurations) and the final correction was obtained by integrating from  $\lambda=0$  to 1 based on a linear fit. The statistical error for those corrections is rather large, on the order of 10 kcal/mol. <sup>c</sup> Similar to the PB1 corrections, except that the dielectric interface was defined differently: the entire region within the 16 Å sphere was set to be  $\epsilon=1$ , and the radii for the water oxygens were set to 2.2 Å (see text). In contrast to PB1 corrections, these PB2 corrections have small statistical error of  $\sim 1$  kcal/mol.

$\phi_s^{(o)}$ , was evaluated with the LPBE using a focusing scheme that placed a 56 Å cube of fine grid (0.4 Å) into a larger 132 Å cube of coarse grid (1.2 Å). The reaction field matrix, **M**, was evaluated using 400 spherical harmonics. In the Poisson–Boltzmann calculations for the outer region, one set of calculations used  $\epsilon_p = 1$  and 0.0 M salt, while the second set used  $\epsilon_p = 4$  and 0.1 M salt.

During the MD simulations, the inner region was further partitioned into Newtonian and Langevin regions. All atoms beyond 20 Å were fixed, those between 16 and 20 Å were treated with Langevin MD, while the rest were treated with standard Newtonian MD. All protein atoms in the Langevin region were harmonically constrained with force constants determined directly from the B-factors in the PDB file. Langevin MD atoms were updated heuristically during the simulation to consistently treat protein groups and water molecules that may switch regions during dynamics. All bonds involving hydrogen were constrained with the SHAKE algorithm<sup>86</sup> to allow 2 fs time steps. The entire system was heated to 300 K and equilibrated for  $\sim 200$  ps prior to any free energy simulations.

For the QM/MM partitioning of the K102 case, the SCC-DFTB region included the propyl-amine of the lysine side chain, while all other atoms in the inner region were treated by the CHARMM 22 force field.<sup>89</sup> A link atom<sup>13</sup> was added between the C $\beta$  and C $\delta$  to fulfill the valence of the QM region, and the point charges of the link host group (C $\beta$ , H $\beta$ 1, and H $\beta$ 2) were

**TABLE 4: Bonded, Zero-Point-Energy (ZPE) Contributions to the Free Energy of Deprotonation and the QM/MM Free Energy Correction as Well as the QM Proton Affinity Correction (in kcal/mol)**

	$\Delta G^{\text{bond}, a}$	$\Delta \text{ZPE}^b$	$\Delta G_{\text{B3-SCC}}^c$	$\Delta G_{\text{B3-SCC}}^d$	$\Delta E_{\text{PA}}^e$	$\Delta E_{\text{PA}}^f$
glycine	−4.2	−7.9	−8.7	−8.4	6.5	7.2
acetic acid	−4.2	−8.8	−10.3	−10.0	6.5	6.8
imidazole	−5.6	−8.6	0.3	−0.1	1.7	1.9
ethanethiol	−3.7	−6.0	12.6	11.9	8.3	8.4
methylamine	−5.7	−9.2	13.0	14.1	2.6	2.5

<sup>a</sup> The force constants for the angle ( $k_\theta$ ) and dihedral ( $k_\tau$ ) used to calculate the bonded terms (eq 9), are the same as those in the CHARMM 22 force field. The other constants are as follows:  $V_0=136.22$  T/P (Å<sup>3</sup>),  $\Lambda=17.457(\text{mT})^{-1/2}$  (Å),  $T=300.0$  K,  $P=1.0$  atm,  $m=1.008$  g/mol. <sup>b</sup> The ZPE corrections were calculated based on gas-phase optimized structures at the SCC-DFTB level. The exception is glycine, for which the zwitterionic form is not stable in the gas phase; thus the ZPE was estimated based on structures collected from solution MD simulations (see text). <sup>c</sup> The corrections were calculated based on eq 12 using 100 snapshots (with a time interval of 3 ps) from both  $\lambda=0$  and  $\lambda=1$  Ext trajectories; all MM atoms were included in the correction calculations. <sup>d</sup> Same as c, but for the Ewald simulations, which included 100 snapshots (with a time interval of 2.7 ps) with both  $\lambda=0$  and  $\lambda=1$  trajectories; only MM water molecules within 12 Å were included. <sup>e</sup> The correction in gas-phase proton affinity was estimated based on CCSD/6-311+G(\*,\*) calculations at SCC-DFTB optimized geometries. <sup>f</sup> Similar to e, but the solute structures (25) were collected from SCC-DFTB/MM GSBP simulations at both  $\lambda=0$  and  $\lambda=1$ .

**TABLE 5: Effect of the van der Waals Parameters for the Acidic Proton on  $\Delta G^{(1)}$  and  $\Delta G^{(\text{vdW})}$  (Scheme 1) for Acetic Acid and Imidazole (in kcal/mol)**

	$\Delta G^{(1)}$	$\Delta G^{(\text{vdW})}$
acetic acid <sup>a</sup>	152.7	0.14
acetic acid <sup>b</sup>	153.8	−1.45
imidazole <sup>a</sup>	131.0	0.15
imidazole <sup>b</sup>	130.1	−0.21

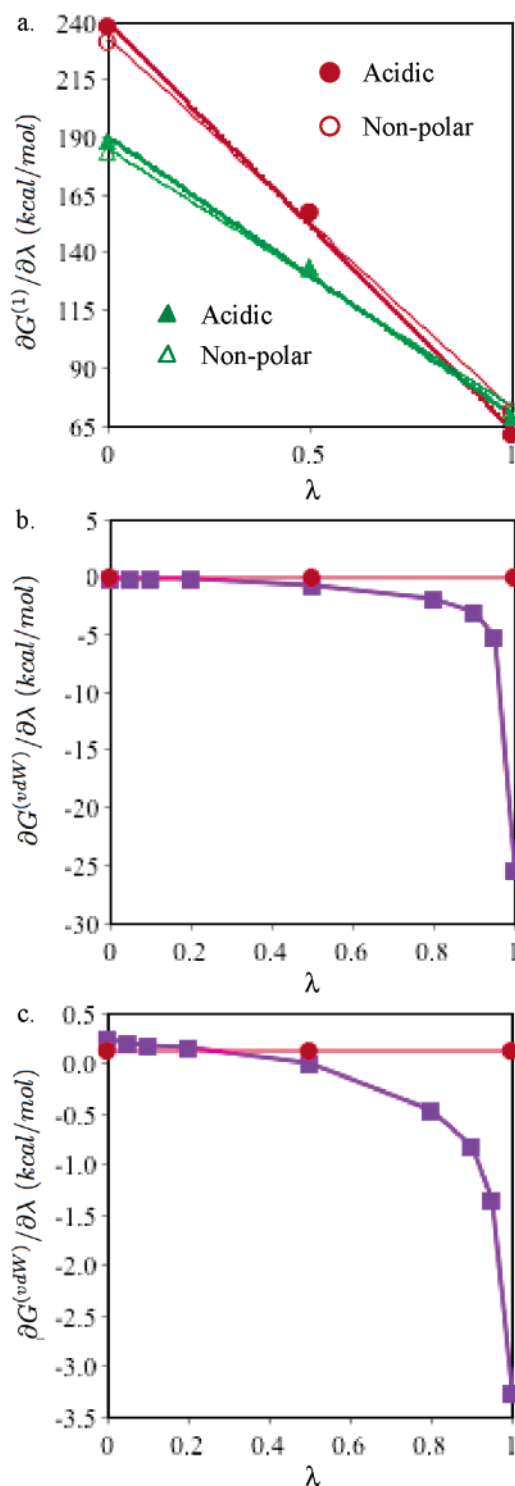
<sup>a</sup> The acidic proton (dummy atom) was treated with the van der Waals parameters for a polar hydrogen in the CHARMM 22 force field:  $\epsilon=-0.04$  kcal/mol;  $\sigma=0.28$  Å. <sup>b</sup> The acidic proton (dummy atom) was treated with the van der Waals parameters for a nonpolar hydrogen in the CHARMM 22 force field:  $\epsilon=-0.02$  kcal/mol;  $\sigma=1.32$  Å.

excluded from the QM calculation.<sup>21</sup> One of the three hydrogens on the amine group was selected randomly to be converted to a dummy atom during the free energy simulations. For the H31 case, the 4-methyl imidazole moiety was treated with SCC-DFTB and the rest with CHARMM 22 force field; the link atom was placed between the C $\alpha$  and C $\beta$ , and the point charges of the host group (C $\alpha$ , H $\alpha$ , and the carbonyl of the side chain) were excluded.

### 3. Results and Discussion

**3.1. Small Molecules in Solution.** The pK<sub>a</sub> shifts between a series of small molecules (see Figure 1) were calculated for spherical and periodic boundary conditions with the free energy perturbation approach as described above. The five small molecules were chosen here to represent titratable groups of amino acids (with a pK<sub>a</sub> ranging from 2.4 to 10.7); four of them represent titratable side chains, and Gly was chosen for studying the pK<sub>a</sub> of the C-terminus. These calculations are invaluable for rigorously evaluating the ability of QM/MM methods to predict appropriate shifts due to the different electronic structures associated with different molecules. In addition, these calculations are necessary to characterize the reference state for determining the pK<sub>a</sub> shift due a different environment; in this study, the pK<sub>a</sub> shifts of H31 and K102 are determined in a T4-

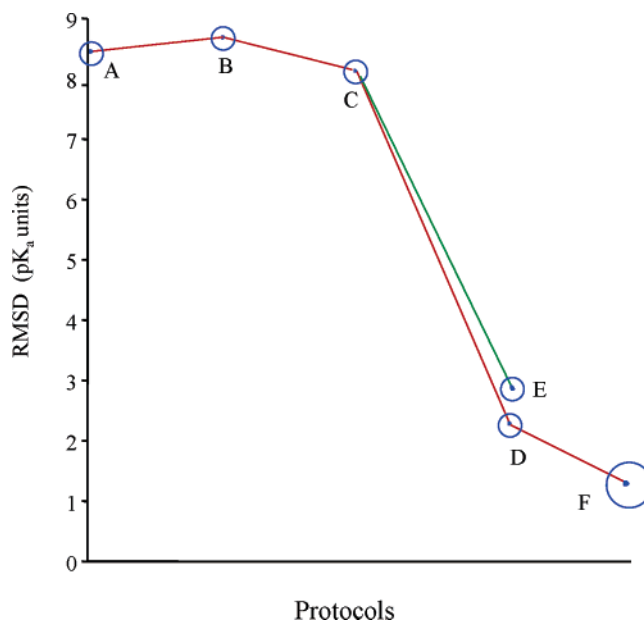




**Figure 4.** Effect of van der Waals parameters on the computed  $pK_a$  of imidazole (triangle) and acetic acid (circle). (a) Linear fit of the free energy derivatives of  $\Delta G^{(1)}$  with the polar (thick lines) and nonpolar (thin lines) hydrogen van der Waals parameters in the CHARMM 22 force field for the acidic proton (dummy atom). In (b) and (c), the free energy derivatives for annihilating the polar (circle) and nonpolar (box) dummy atoms are shown for acetic acid and imidazole, respectively.

lysozyme mutant relative to imidazole and propylamine in solution, respectively.

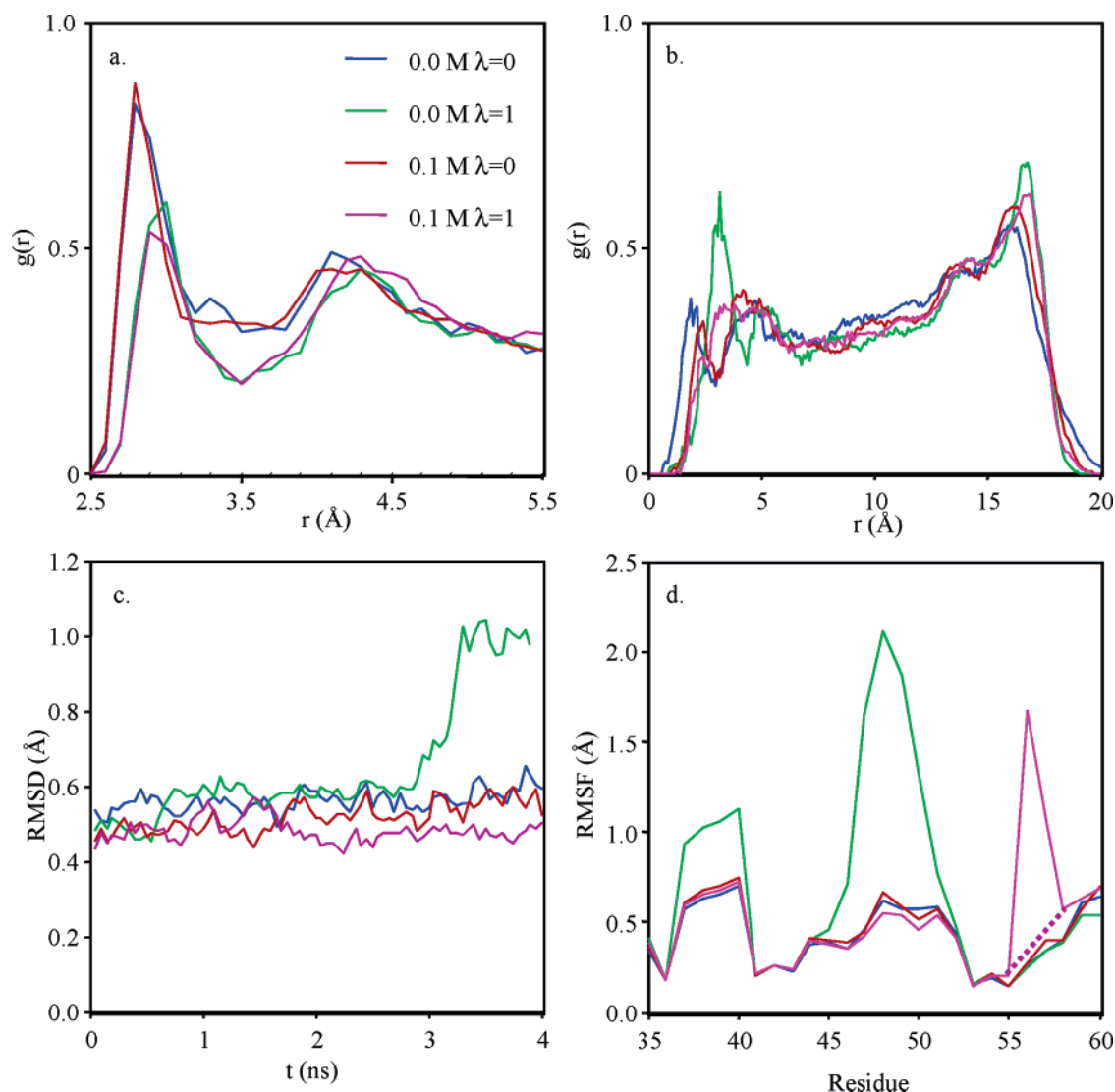
**3.1.1. Gas-Phase Proton Affinities.** Before presenting  $pK_a$  results in solution, it is useful to comment on the gas-phase proton affinities (PAs) at different levels of calculations. As shown in Table 1, SCC-DFTB overestimates the PAs for acetic acid and glycine while it underestimates the PAs for the other



**Figure 5.** Schematic plot of the root-mean-square differences (RMSD) for  $pK_a$  shifts (relative to  $-\text{COOH}$  in glycine) compared to experimental values as different protocols are used with the extended electrostatics (Ext) simulations, which include different corrections. (A)  $\Delta G^{(1)}$  for SCC-DFTB/MM without corrections. (B) Born correction for an 18 Å radius sphere added to (A). (C)  $\Delta ZPE$  added to (B). (D) B3LYP/6-31+G\*\*/MM correction added to (C). (E) Gas-phase proton affinity correction to SCC-DFTB based on CCSD/6-311+G\*\* calculations added to (C). (F) Gas-phase proton affinity CCSD/6-311+G\*\* correction to B3LYP/6-31+G\*\* added to (E). Clearly, both gas phase and QM/MM corrections were necessary to reproduce the experimental trend.

three systems, as compared to both B3LYP and CCSD calculations; the difference is as large as 20 kcal/mol in some cases. Significant errors in PAs have been observed for several semiempirical methods including AM1, PM3, and SCC-DFTB,<sup>90</sup> which is likely due to the minimal basis set character of these methods. Although systematic improvements can be made by including better treatment of polarization,<sup>91</sup> we instead pursue perturbative corrections at higher QM levels in this study. It is worth pointing out that B3LYP with a modest basis set (e.g., 6-31+G(d,p)) produces PAs that may significantly differ from CCSD/6-311+G(d,p) calculations, even when relative values are considered; e.g., the relative PA between imidazole and ethylthiol are 24.7 and 32.7 kcal/mol at the B3LYP/6-31+G(d,p) and CCSD/6-311+G(d,p) levels, respectively. Therefore, for quantitative  $pK_a$  calculations, we improve the SCC-DFTB/MM results with both B3LYP/MM and CCSD calculations (see below). Although even higher level of theories can be pursued, the CCSD/6-311+G(d,p) level of calculation is often found reliable for relative PAs in closed-shell molecules.<sup>92,93</sup>

**3.1.2. Free Energy Perturbation with SCC-DFTB/MM Potential and Different Electrostatic Models.** The free energy for converting the proton to a dummy atom,  $\Delta G^{(1)}$ , plays a dominant role in determining the  $pK_a$  in the DTSC scheme. As shown in Figure 3, the corresponding free energy derivatives converge rather quickly (on the order of 100 ps after the initial 80 ps of equilibration, Figure 3b) and fit nicely ( $R^2 \sim 0.99$ , Figure 3a) with the expected linear dielectric response of the solution. In addition, the reverse cumulative average analysis<sup>55</sup> successfully coarse grained the data set into statistically independent blocks for the entire trajectory after the initial 80 ps of equilibration. The  $\Phi$ -value<sup>55</sup> converged to between 5 and 7, with 5 ps blocks (Figure 3c); 7 ps blocks were used to ensure an appropriate



**Figure 6.** The structure of water and stability of the protein in the  $pK_a$  calculations for His 31 in the M102K mutant of T4 lysozyme using 0.0 and 0.1 M SCC-DFTB/CHARMM-GSBP simulations. The radial distribution of water oxygens about the acidic nitrogen in His 31 ( $N\epsilon 2$ ) and origin of the simulation sphere are shown in (a) and (b), respectively. The root-mean-square deviation (RMSD) for the protein backbone atoms relative to the X-ray structure and the root-mean-square fluctuation (RMSF) for  $\alpha$ -carbons are shown in (c) and (d), respectively. The structure of the protein was destabilized in the deprotonated simulations ( $\lambda = 1$ ): (c) the RMSD increased substantially after 3 ns, and (d) RMSF of residues about D47 (0.0 M) and N55 (0.1 M) are significantly increased relative to all other simulations.

**TABLE 6:  $pK_a$  Shifts (in  $pK_a$  Units), Relative to the  $-\text{COOH}$  Group in Glycine, and the RMS Difference from Experimental Values for the Five Small Molecules Studied<sup>a</sup>**

	expt <sup>b</sup>	Ext <sup>c</sup>	GSBP <sup>d</sup>	Ewald
glycine	0	0.0 (0.0) [0.0]	0.0 [0.0]	0.0
acetic acid	2.4	-2.9 (3.4) [2.2]	2.8 [1.6]	2.1
imidazole	4.6	3.6 (4.4) [4.4]	4.3 [4.3]	2.8
ethanethiol	8.2	0.9 (10.8) [9.6]	10.8 [9.6]	10.0
methylamine	8.3	11.2 (8.3) [8.3]	8.3 [8.3]	9.3
RMSD		4.3 (1.2) [0.6]	1.2 [0.7]	1.2

<sup>a</sup> Calculated  $pK_a$  shifts include  $\Delta G^{(1)}$  (Table 2), various solvation corrections (Table 3), bonded-term, ZPE, QM/MM, and QM corrections (Table 4). <sup>b</sup> The experimental values were obtained from CRC Handbook of Physical Chemistry. <sup>c</sup> The three columns differ in the bulk solvation correction used (Table 3): values without parentheses or brackets used the PB1 correction; values with parentheses were obtained with the 18 Å Born correction; values with brackets were obtained with the PB2 correction. <sup>d</sup> Values in the brackets were obtained by replacing the outer contribution, which corresponds to the Born correction with a radius of 18 Å, with the PB2 correction (Table 3).

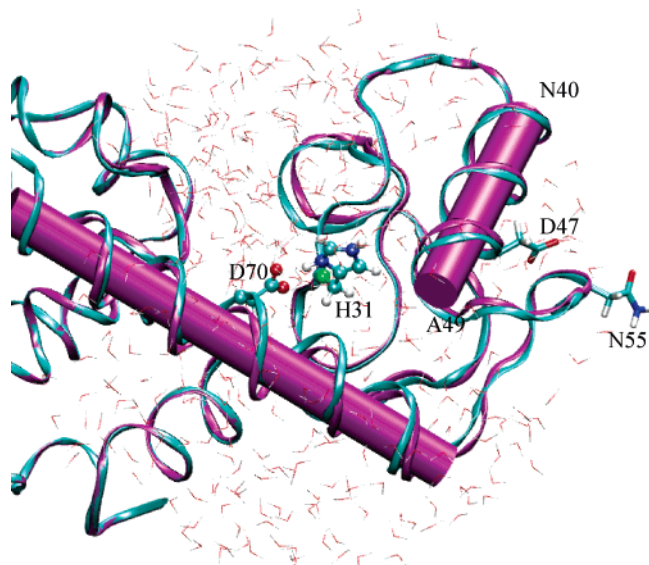
**TABLE 7:  $pK_a$  for His 31 and Lys 102 in the M102K Mutant of the T4-Lysozyme<sup>a</sup>**

		expt	DTSC
His 31	0.0 M		8.2
His 31	0.1 M	9.1 <sup>b</sup>	8.5
Lys 102	0.0 M		-5.1
Lys 102	0.1 M	6.5 <sup>c</sup>	-2.9
Lys 102	Linear <sup>d</sup>		6.3

<sup>a</sup>  $pK_a$  was calculated by adding the computed  $pK_a$  shifts (from solution to the enzyme) to the  $pK_a$  values of reference compounds in solution. The reference for His and Lys was taken as imidazole ( $pK_a=7.0$ ) and propylamine ( $pK_a=10.5$ ), respectively; note that only  $\Delta G^{(1)}$  was considered in computing the  $pK_a$  shift. <sup>b</sup> Determined for the wild-type T4-lysozyme by Anderson et al. (ref 27). <sup>c</sup> Determined for the M102K mutant by Dao-pin et al. (ref 28). <sup>d</sup> For the protonated state ( $\lambda=0$ ), the local structure surrounding Lys 102 is assumed to be unfolded, thus Lys 102 becomes fully solvent exposed. Based on the free energy derivative for propylamine in solution (197.1 kcal/mol) at  $\lambda=0$  and that from the 0.1 M enzyme simulation (62.0 kcal/mol) at  $\lambda=1$ , a linear fit was used to estimate  $\Delta G^{(1)}$ .

block size, which gave an estimate for the error in free energy derivatives of 0.2–0.3 kcal/mol (see Figure 3d).

The differences between  $\Delta G^{(1)}$  with different electrostatic models (Table 2) originated from inner region electrostatics



**Figure 7.** Average structure for the H31 protonated simulation (0.0 M salt) of the M102K mutant of T4 lysozyme (cyan) overlaid with the crystal structure (silver). For stable portions of all simulations, the average structure did not deviate significantly from the crystal structure for either 0.0 or 0.1 M GSBP setups.

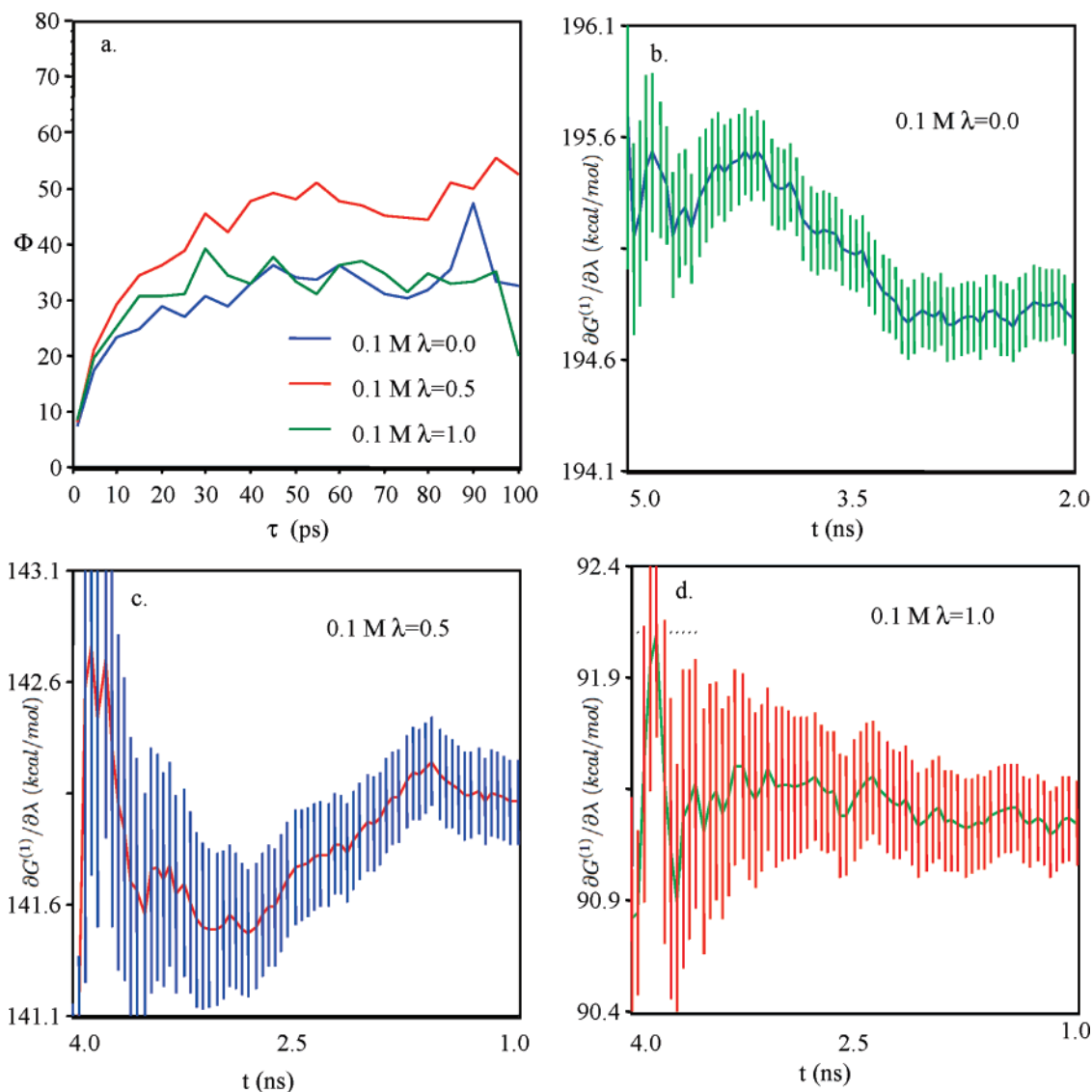
treatment and long-range contributions. The GSBP and Ewald results, both included long-range contributions, agreed very well with each other in terms of relative  $\Delta G^{(1)}$  values (root-mean-square difference of 0.9 kcal/mol); there was a nearly constant shift ( $\sim 8$ – $9$  kcal/mol) between the absolute values, due presumably to the neglect of correction of net-charge effects in the current QM/MM implementation of the Ewald summation and slight difference in the water density in the two sets of simulations (e.g.,  $\sim 0.0333$  mol/ $\text{\AA}^3$  in Ewald and  $\sim 0.0366$  mol/ $\text{\AA}^3$  in GSBP). The Ext and GSBP results for  $\Delta G^{(1)}$  differ substantially since the former did not include the contribution from the bulk solvent. The magnitude of the difference, however, was very similar among different solutes. Indeed, adding a Born correction (9.2 kcal/mol) for charging an ion in a 18  $\text{\AA}$  sphere ( $\epsilon = 1$ ) surrounded by an infinite bulk ( $\epsilon = 80$ ) brought the Ext results into close agreement with the GSBP values with the root-mean-square difference decreased from 9.5 to 0.7 kcal/mol. Therefore, the Ext and GSBP trajectories were not significantly different in terms of configuration distributions. In contrast, the Fshift results remained poor when the same Born correction was included. As discussed in previous work,<sup>10,94,95</sup> the unbalanced treatment of QM/MM (note that no cutoff is applied to QM/MM electrostatic interactions) and MM/MM interactions in Fshift calculations caused substantial overpolarization of the system, which could not be corrected with a simple Born formula. Significant improvement in the Fshift results was observed when the correction was estimated using the linearized Poisson-Boltzmann (PB) calculations with the dielectric surface defined using van der Waals radii based on the CHARMM 22 force field. For example, the difference between  $\Delta G^{(1)}$  from Fshift and Ext calculations for acetic acid and imidazole was 9.5 and 39 kcal/mol, respectively; with the PB correction (Table 3), these differences shrunk substantially to 2.5 and 2.0 kcal/mol, respectively. As often commented in the literature,<sup>39,96</sup> this definition of the dielectric surface allows the bulk dielectric to penetrate into the low-dielectric region (in this case the small molecule plus explicit water molecules), resulting in a heterogeneous mixture of  $\epsilon = 1$  and  $\epsilon = 80$  media, which cancels overpolarization of the system due to truncated

(Fshift) electrostatic treatment. Although the cancellation of over-polarization brought the Fshift and Ext results into better agreement, there were significant fluctuations (standard deviation of  $\sim 10$  kcal/mol) in the computed PB corrections among the set of 100 snapshots due to variations of the dielectric interface. This limited the statistical accuracy of applying such heterogeneous-dielectric PB corrections ( $\Delta G^{\text{PB1}}$  in Table 3), thus Fshift results will not be discussed further.

It is interesting to further comment on the bulk solvation contribution in the GSBP results. As mentioned above, the GSBP simulations explicitly included water molecules within 16  $\text{\AA}$  from the center of the sphere; the dielectric boundary between  $\epsilon = 1$  and  $\epsilon = 80$ , however, was set at the radius of 18  $\text{\AA}$ . This *smooth* dielectric boundary was established such that reaction field matrix (**M**) remains constant as the explicit solvent molecules fluctuate during the MD, and the 2  $\text{\AA}$  buffer was chosen based on the oxygen–oxygen radial distribution in bulk water. For the simple small molecule/water systems considered here, the bulk contribution in GSBP corresponds to an 18  $\text{\AA}$  Born correction (+9.2 kcal/mol for  $\Delta G^{(1)}$ ). This consideration, however, may not be entirely accurate as the explicit solvent sphere has a rough dielectric boundary with a fluctuating radius<sup>22</sup> during MD. To quantitatively estimate the bulk electrostatic contribution, linearized PB calculations were carried out for 100 configurations at each  $\lambda$  value. The region occupied by the small molecule and explicit solvent (within 16  $\text{\AA}$ ) was assigned to have  $\epsilon = 1$  (i.e., no high-dielectric penetration was allowed); in addition, the oxygen atoms in water were assigned a radius of 2.2  $\text{\AA}$ ,<sup>97</sup> which defined a rough dielectric boundary for each snapshot. Averaging over the configurations yielded a value of 10 kcal/mol for the bulk contribution to  $\Delta G^{(1)}$  (column  $\Delta G^{\text{PB2}}$  of Table 3), corresponding to an effective Born radius between 16 and 17  $\text{\AA}$ . In contrast to the heterogeneous-dielectric PB calculations ( $\Delta G^{\text{PB1}}$  in Table 3) mentioned above, the  $\Delta G^{\text{PB2}}$  calculations have a much simpler dielectric interface, thus much smaller fluctuations (standard deviation of  $\sim 1$  kcal/mol). In short, the GSBP setup based on the 18  $\text{\AA}$  sphere ( $\epsilon = 1$ ) has pushed the bulk dielectric slightly too far from the explicit solvent region, and this effect should be taken into account in the analysis of GSBP  $\text{pK}_a$  results (see below). Based on the similar argument, a Born correction to the Ext results should also use the value of 10 kcal/mol rather than 9.2 kcal/mol.

The combined contribution from zero point energy and the bonded-term corrections were between  $-9.7$  and  $-14.9$  kcal/mol (Table 4). Similar to previous work,<sup>49</sup> contribution from the van der Waals term of the dummy atom ( $\Delta G_{\text{D}}^{\text{vdW}}$ ) was found to be insignificant, with a value of  $\sim 0.15$  kcal/mol for both imidazole and acetic acid (Table 5). This was expected since the van der Waals radius of the dummy atom is small compared to the heavy atom that it is bound to, as parametrized for polar hydrogen atoms in the CHARMM 22 force field.<sup>89</sup> With different van der Waals parameters, the precise value of this contribution would change but the effect on relative  $\text{pK}_a$  should remain small, as long as the van der Waals parameters are within a physically reasonable range. To test this, additional calculations for acetic acid and imidazole were carried out with an extreme set of van der Waals parameters for the acidic proton based on the nonpolar hydrogen in the CHARMM 22 force field.<sup>89</sup> The “hydrophobic” proton led to changes in  $\Delta G^{(1)}$  of +1.1 and  $-0.9$  kcal/mol for acetic acid and imidazole, respectively (Table 5); the  $\Delta G^{\text{vdW}}$  changed from  $\sim 0.15$  to  $-1.45$  and  $-0.21$  kcal/mol, respectively. Combining both contributions yielded changes of  $-0.5$  kcal/mol for acetic acid and  $-1.3$  kcal/



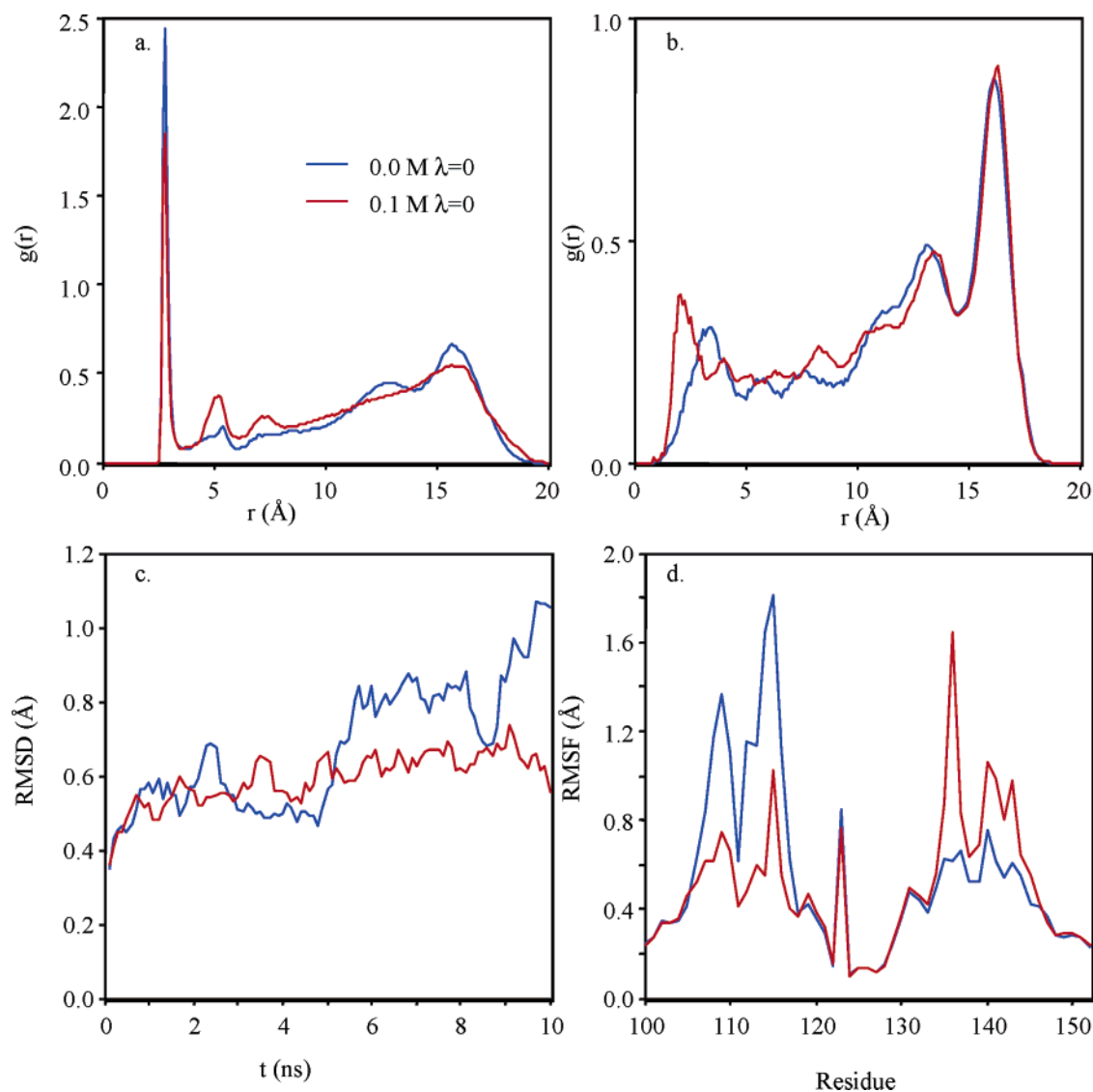


**Figure 8.** Convergence of the free energy derivatives for the H31  $pK_a$  simulation in the M102K T4 lysozyme mutant with the 0.1 M GSBP simulation. In (a), the convergence behavior for  $\Phi$  is shown. The reverse cumulative averages calculated with 40 ps blocks for  $\lambda = 0.0$ , 0.5, and 1.0 are shown in (b), (c), and (d), respectively. For  $\lambda = 0.0$ , the initial 2 ns of a 5 ns simulation were discarded, and for  $\lambda = 0.5$  and 1.0 the initial 1 ns of the 4 ns simulations were discarded.

mol for imidazole; these give the upper limits of change in  $\Delta G^{(1)} + \Delta G^{\text{vdW}}$  with different van der Waals parameters for the acidic proton (dummy atom). We note that the effect of the van der Waals parameters on  $\Delta G^{\text{vdW}}$  was larger for acetic acid than for imidazole. This was expected because  $\Delta G^{\text{vdW}}$  measures the contribution from the dummy atom when the solute is in the deprotonated state; for acetic acid, the deprotonated state is charged, in contrast to the neutral deprotonated imidazole, thus the influence of the van der Waals parameters on  $\partial G^{\text{vdW}}/\partial \lambda$  is more pronounced, as evident in Figure 4c and 4d.

**3.1.3. Improvements over SCC-DFTB/MM Results.** As discussed above, there were significant differences between PAs from SCC-DFTB and higher level QM (B3LYP, CCSD) calculations. In addition, the SCC-DFTB/MM electrostatic interaction was calculated as that between QM Mulliken charges and MM point charges, in contrast to the more rigorous treatment based on one-electron integrals. Moreover, SCC-DFTB is less polarizable than higher level QM methods due to the minimal basis set used. Therefore, it was expected that corrections for both gas-phase PA and QM/MM interactions are needed for improvements in the calculated  $pK_a$  values. As shown in Figure 5, which is based on the Ext simulations with

the finite sphere boundary, this is precisely the situation found here. Combining the contributions thus far ( $\Delta G^{(1)}$ ,  $\Delta ZPE$ , and  $\Delta G^{(2)}$ ) at only the SCC-DFTB/MM level, protocol C produces a root-mean-square difference (RMSD) for the  $pK_a$  shifts (relative to  $-\text{COOH}$  in glycine) of 8  $pK_a$  units compared to experimental values. The large error was dominated by ethanethiol and methylamine, for which SCC-DFTB had large errors in the gas-phase PA (Table 1). The QM/MM correction with B3LYP/6-31+G(d,p)/MM was rather significant, in the range of 10 kcal/mol (Table 4); this was larger than the typical difference between SCC-DFTB and B3LYP for gas-phase PAs (Table 1), suggesting that corrections for the electrostatic interaction between QM and MM atoms and solute polarization are sizable. Including such QM/MM correction with B3LYP/6-31+G(d,p)/MM perturbative calculations (protocol D) reduces the RMSD to 2.6  $pK_a$  units. Including only gas-phase PA corrections at the CCSD/6-311+G(d,p) level relative to SCC-DFTB (protocol E) had a similar effect and reduced the RMSD to  $\sim 2.4$   $pK_a$  units; the gas phase correction seemed insensitive to the solute geometry in the CCSD calculations (compare the last two columns in Table 4). Including both QM/MM corrections (B3LYP/6-31+G(d,p)/MM vs SCC-DFTB) and gas-phase



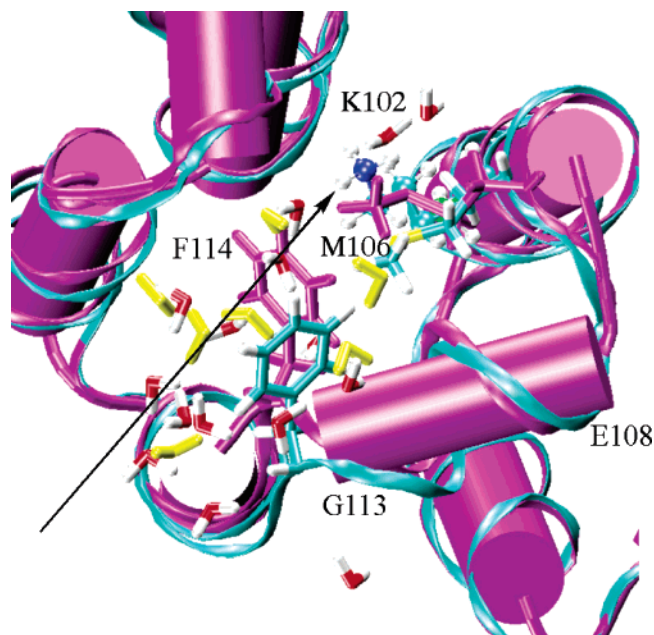
**Figure 9.** Structure of water and stability of the protein in the protonated state of Lys 102 ( $\lambda = 0.0$ ) for the M102K mutant of T4 lysozyme using 0.0 and 0.1 M SCC-DFTB/CHARMM-GSBP simulations. The radial distribution of water oxygens about the acidic nitrogen in Lys 102 ( $N\zeta$ ) and origin of the simulation sphere are shown in (a) and (b), respectively. The root-mean-square deviation (RMSD) for the protein backbone relative to the X-ray structure and root-mean-square fluctuation (RMSF) for  $\alpha$ -carbons are shown in (c) and (d), respectively.

PA corrections (CCSD/6-311+G(d,p)//SCC-DFTB vs B3LYP/6-31+G(d,p)//SCC-DFTB) (protocol F), the RMSD dropped to 1.2 pK<sub>a</sub> units (Table 6). Finally, as commented above, the bulk contribution in GSBP and Ext simulations was better described with the Born correction corresponding to a sphere of radius between 16 and 17 Å instead of 18 Å considering the fluctuating dielectric interface. Taking this into account, the RMSD in both Ext and GSBP results became about 0.7 pK<sub>a</sub> unit (Table 6). We emphasize that making this correction to the bulk contribution was not motivated by a better fit to experimental results but due to a more precise consideration of the dielectric boundary (see section 3.1.2).

For the Ewald-sum simulations, as discussed above, the relative  $\Delta G^{(1)}$ 's were very similar to GSBP results with a RMSD between them of 0.9 kcal/mol. The B3LYP/MM corrections, although calculated using a smaller number of MM atoms (12 Å from the QM solute vs the entire 16 Å sphere in the spherical model), were found to be rather close to those for the spherical simulations. This was expected because the QM/MM correction was presumably dominated by the first few solvation shells. Including all the components ( $\Delta G^{(1)}$ , bonded-term,  $\Delta ZPE$ , B3LYP/MM, and CCSD corrections), the RMSD for relative

pK<sub>a</sub> values with the Ewald simulations was 1.2 pK<sub>a</sub> units, which was very close to the spherical boundary simulations with carefully chosen bulk corrections.

In summary, the agreement between our simulations and experimental values is very encouraging. It seems that for the case of a small solute in pure solution, including the dielectric response of the bulk solvent during the simulation is not crucial, provided that a decent number of explicit solvent molecules are included. The long-range effects can be introduced after the simulation, although the most effective protocol depends on the electrostatic model (e.g., Fshift vs Ext). For accurate pK<sub>a</sub> shifts, corrections at the level of QM and QM/MM interactions are both important, and the protocol of combining efficient QM/MM sampling procedure with perturbative corrections at higher levels of theory was found to give satisfactory results. It is important to note that all corrections were calculated at higher levels of theory using the configurations sampled by SCC-DFTB (gas phase) or SCC-DFTB/MM (solution) simulations. The GSBP results are satisfactory, although the 2.0 Å buffer region between the bulk and the explicit solvent system appeared to be slightly too large. We note that the surface polarization term was not included in the current study, which may lead to



**Figure 10.** Comparison of the average structure for the 0.0 M salt protonated K102 simulation (cyan) to the crystal structure for the protein (purple) and water (yellow) atoms. The reorientation of Phe 114 plays a key role in the initial impouring of waters into the hydrophobic region containing the charged K102; this may have initiated some of the changes seen in the Glu 108 to Gly 113  $\alpha$ -helix. In addition to the Phe 114 and the Glu 108- Gly 113  $\alpha$ -helix, the orientation of Met 106 is also perturbed.

improvements when considering size dependence of the results.<sup>79,98</sup> Nevertheless, the spherical systems used here seem to be sufficiently large to reproduce the experimental trend.

**3.2. T4-Lysozyme.** The series of small molecules provided the opportunity to quantify the ability to compute  $pK_a$  for molecules with different electronic structures in similar environments. This section will investigate the ability to reproduce  $pK_a$  shifts due to different environments where the stability of the protonation state will depend on the ability of the system to stabilize charges via electrostatic interactions. An ideal protein system for such investigations is T4-lysozyme, which comes from the bacteriophage T4 and catalyzes the hydrolytic cleavage of polysaccharides. In the following, we discuss our results for the  $pK_a$  shift in H31 (Figure 2a) and K102 (Figure 2b) in the M102K mutant (Table 7). The experimentally measured  $pK_a$  shifts for those residues are among the largest known and therefore present a stringent test of the simulation protocol, as emphasized by Warshel and co-workers.<sup>43</sup>

**3.2.1. Histidine 31.** The H31 is involved in a salt-bridge interaction with D70, which is surface accessible from one side and close to the hydrophobic core on the other side. Both experimental and computational investigations found that salt bridges at the surface usually have little effect on the overall stability of proteins<sup>88,96</sup> due to the similar solvent accessibility in the folded and unfolded states. In contrast, positioning charged residues into the low-dielectric protein interior may be energetically destabilizing unless the charges can be neutralized or satisfied with electrostatic interactions (e.g., salt bridges). In fact, salt bridges that are buried in hydrophobic regions often contribute significantly to the stability; the H31-D70 salt bridge in T4-lysozyme has been shown to contribute 3–5 kcal/mol of stabilization<sup>27</sup> where the  $pK_a$  of H31 and D70 are shifted up to 9.1 and down to 0.5, respectively. These shifts helped explain why T4-lysozyme is most stable at pH 5.0 instead of its isoelectric point of 10. This behavior is interesting since most

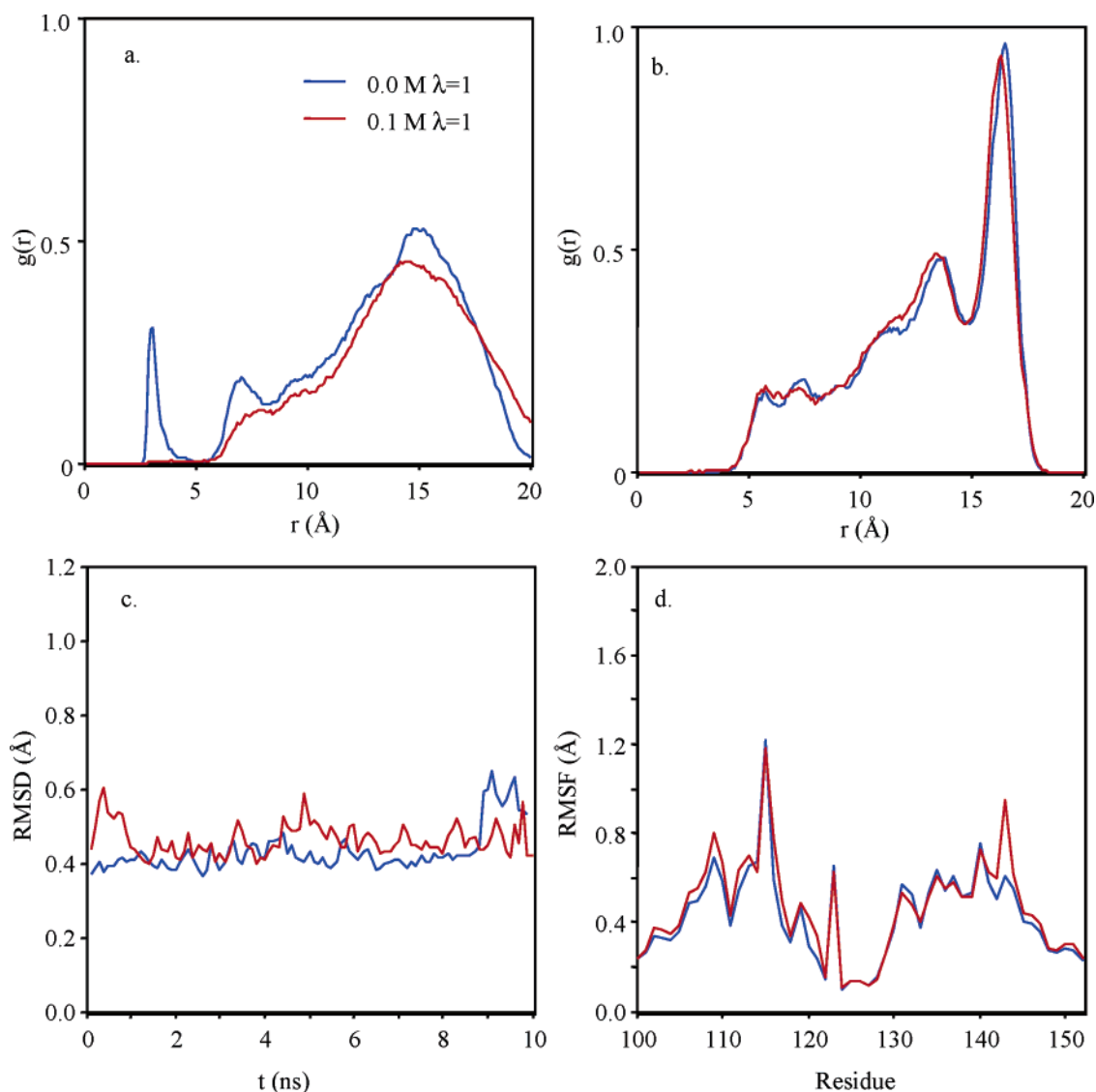
proteins are expected to be most stable at the isoelectric point due to the minimization of electrostatic repulsion among charged groups.<sup>99</sup> At pH 5, H31 ( $pK_a \sim 7$ ) is protonated and D70 is deprotonated in the unfolded state, thus little energy is required to form the salt bridge in the folded state.

As mentioned above,  $\Delta G^{(1)}$  was calculated with two GSBP setups (0.0 and 0.1 M salt concentration with protein dielectric in the outer region ( $\epsilon_p$ ) of 1 and 4, respectively). The protonated state ( $\lambda = 0$ ) was stable for both GSBP setups. The distributions of water molecules about the acidic nitrogen, N $\epsilon$  (Figure 6a), and about the origin of the sphere (Figure 6b) as well as the root-mean-square difference (RMSD) for backbone atoms relative to the X-ray structure (Figure 6c) and root-mean-square fluctuations (RMSF) of  $\alpha$ -carbons (Figure 6d) were very similar for both  $\lambda = 0$  simulations. For the partially ( $\lambda = 0.5$ ) and fully deprotonated ( $\lambda = 1.0$ ) states, the protein structure did not deviate significantly from the crystal structure during most interval of the simulations (Figure 7). However, signs of structural instability were observed after  $\approx 4$  ns in both 0.0 and 0.1 M setups, although the causes of the instability seemed different. For the 0.0 M set, D47 displaced significantly to the interior of the simulation sphere to become better solvated in both  $\lambda = 0.5$  and  $\lambda = 1.0$  simulations, causing significant change in both RMSD (relative to the X-ray structure) and RMSF in the nearby region (Figure 6c,d). In addition, while the radial distributions of water oxygens about N $\epsilon 2$  were similar between sets of simulations (Figure 6a), the distribution of water oxygens about the origin was very different in the  $\lambda = 1.0$ , 0.0 M simulation. As seen in Figure 6b, the first peak at  $\approx 4$  Å was significantly higher, due most likely to the solvation shell of the displaced D47 residue. For the 0.1 M set, D47 remained in place although the RMSD in the backbone increased from 0.5 to 0.7 Å after 4 ns (data not shown), due to increased mobility in the loop containing N55 (Figure 6d). These observations are consistent with the expectation that the deprotonated state would be less stable.

Despite the minor structural instability observed in the  $\lambda = 0.5$  and 1.0 simulations, the data set for  $\Delta G^{(1)}$  was successfully coarse grained into statistically independent blocks for all  $\lambda$  values in the reverse cumulative average after the equilibration portions of the trajectories ( $\approx 1$ –2 ns) were discarded (Figure 8). The  $\Phi$  converged at a higher value (40–50) (Figure 8a) when compared to the small molecules (Figure 3a), reflecting the need for larger block sizes; with 40 ps blocks, the estimated errors in the free energy derivatives are in the range of 0.2–0.4 kcal/mol (Figure 8b,c,d). Fitting the free energy derivatives linearly ( $R^2 \sim 0.99$ ) gives  $pK_a$  shifts (relative to imidazole in solution) of 1.2 and 1.5  $pK_a$  units for the 0.0 and 0.1 M simulations, respectively. Adding this  $pK_a$  shift to the experimental  $pK_a$  (7.0) of imidazole in solution, our calculated  $pK_a$  is 8.2 (0.0 M) and 8.5 (0.1 M), which are in decent agreement with the experimental value of 9.1  $pK_a$  units (Table 7).

**3.2.2. Lysine 102.** The M102K T4-lysozyme mutant was introduced to study the effect of positioning a charged residue into a hydrophobic pocket.<sup>88</sup> In this mutant, the  $pK_a$  of K102 (Figure 2b) was measured to have a large downward shift by four  $pK_a$  units to 6.5 with differential titrations and NMR. This corresponds to a significant destabilization of 2 to 9 kcal/mol for pH ranging from 10 to 3, respectively. The M102K mutant crystal structure (1.9 Å resolution at pH 6.8)<sup>88</sup> was found to be very similar to the wild type, where the only significant difference is in the increased mobility (as reflected by the B-factors) of the E108-G113  $\alpha$ -helix and the mobility of K102 compared to the methionine residue in the WT enzyme. A



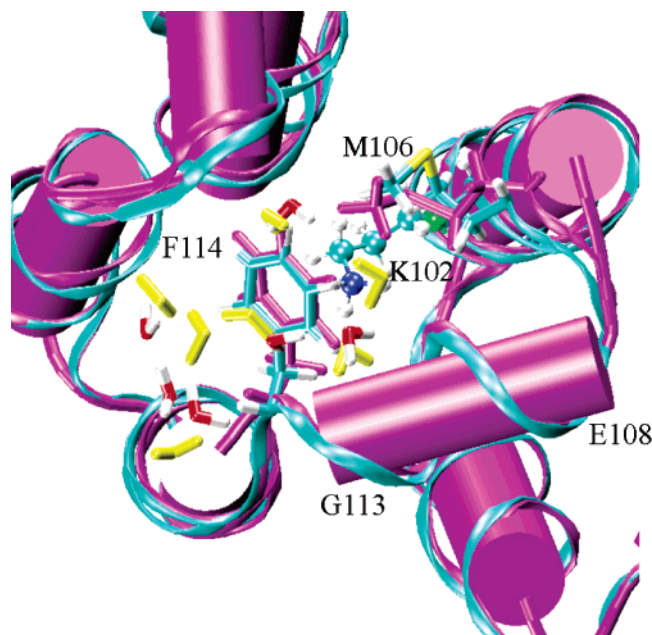


**Figure 11.** Structure of water and stability of the protein in the deprotonated state of Lys 102 ( $\lambda = 1.0$ ) for the M102K mutant of T4 lysozyme using 0.0 and 0.1 M SCC-DFTB/CHARMM-GSBP simulations. The radial distribution of water oxygens about the acidic nitrogen in Lys 102 ( $N\zeta$ ) and origin of the simulation sphere are shown in (a) and (b), respectively. The root-mean-square deviation (RMSD) for the protein backbone relative to the X-ray structure and root-mean-square fluctuation (RMSF) for  $\alpha$ -carbons are shown in (c) and (d), respectively. Both sets of simulations produced a stable deprotonated state.

possible explanation for the increased mobility in this  $\alpha$ -helix was suggested to result from a mixing of the two K102 protonation states at pH 6.8, where it is completely disordered for the protonated state and more structured (as in the WT enzyme) for the deprotonated state. Although this could not be verified at pH 6.8, the CD signal was found to yield a 43% reduction in helical content at pH 3.0;<sup>88</sup> since the 43% reduction is too high to account for by only the E108-G113  $\alpha$ -helix, other  $\alpha$ -helices must be disordered in the protonated state as well. Here, we report results from our calculations that support this interpretation.

In contrast to the H31 case, calculating the  $pK_a$  shift for K102 proved much more problematic. The protonated state of the 0.0 M simulation was found to be unstable with an RMSD eventually increasing from 0.5 to 1.0 Å (Figure 9), which is alarming considering that a significant part of the protein was fixed during the GSBP simulations. While not as dramatic, the 0.1 M simulation also yielded a higher RMSD ( $\approx 0.6$  Å) compared to the deprotonated state ( $\approx 0.4$  Å). The origin of the higher RMSDs in both GSBP simulations stems from the structural changes in two  $\alpha$ -helices in close proximity to K102.

In the 0.0 M simulation, the E108-G113  $\alpha$ -helix partially unwound due to a dramatic migration ( $\approx 12$  Å) of the nearby F114 to the surface of the protein, allowing water molecules to go into the interior of the protein to solvate the protonated K102 (Figure 10). In the 0.1 M simulation, F114 merely swung open to allow water access to the protonated lysine and the E108-G113  $\alpha$ -helix remained structured (Figure 10); the RMSF for the K135–R145 helix-turn-helix, on the other hand, increased compared to the 0.0 M protonated and both sets of deprotonated simulations (Figure 9d). Overall, the 0.0 and 0.1 M protonated-state simulations produced very different distributions of water molecules, RMSDs (relative to the X-ray structure), and RMSFs (Figure 9). The common trend of the structural instability of the protonated state (with increased  $\alpha$ -helix motions), however, was consistently reproduced by both simulations. The deprotonated state simulations were very similar with both GSBP setups (Figure 11). The only significant difference was reflected in the radial distribution of water about the acidic nitrogen ( $N\zeta$ ), where there was a shift toward the  $N\zeta$  in the 0.0 M simulation. The small peak at 3.5 Å in the 0.0 M simulation is due to a water molecule that remained hydrogen-bonded with K102



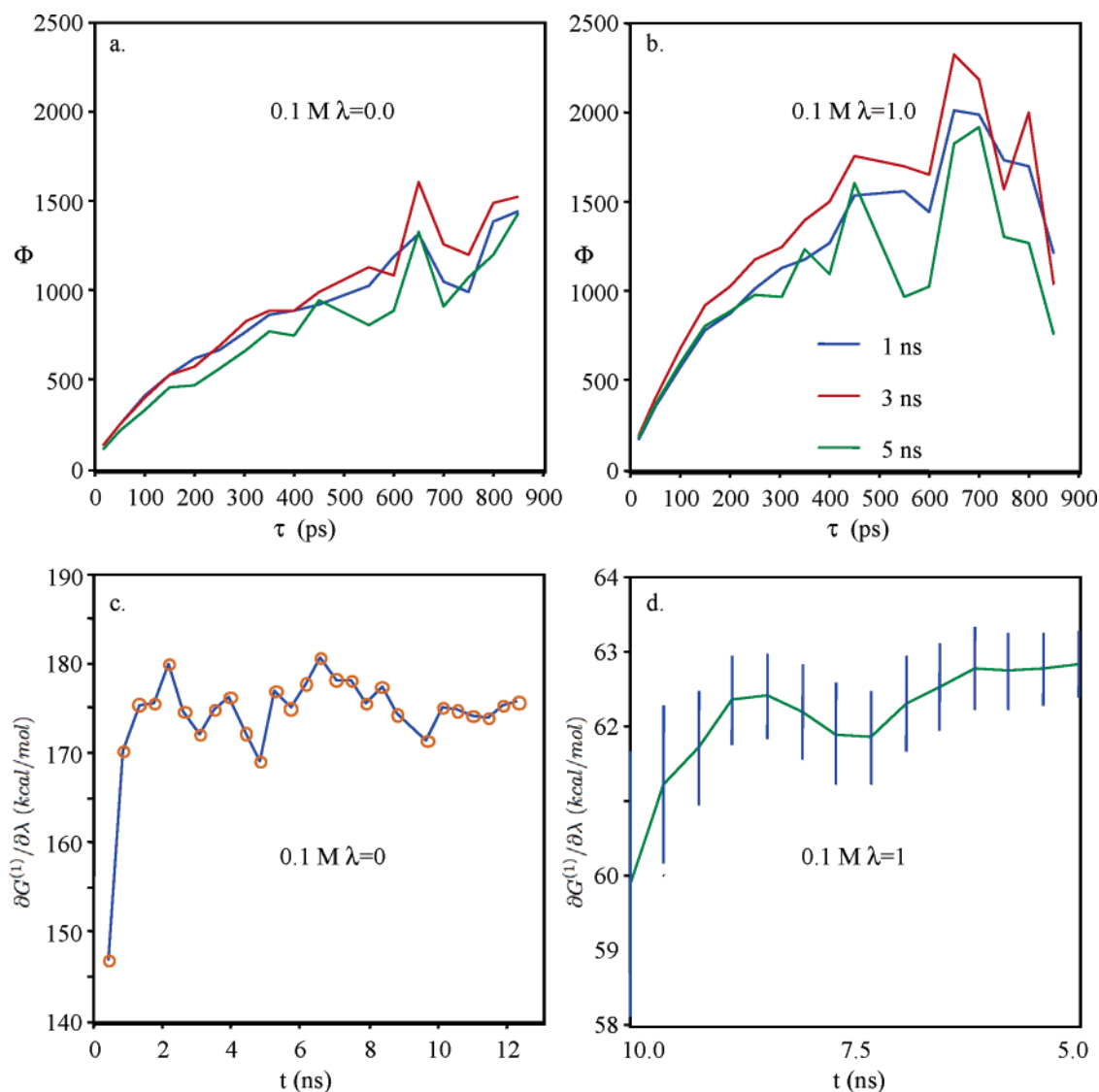
**Figure 12.** Comparison of the average structure for the 0.0 M salt deprotonated K102 simulation (cyan) to the crystal structure for the protein (purple) and water (yellow) atoms. Unlike the protonated case (Figure 10), the average simulation structure and the crystal data are not significantly different. The Phe 114 residue was clearly seen blocking the deprotonated K102 from the solvent molecules.

behind the F114 gate; this water molecule was placed during the initial solvation, and it escaped early in the 0.1 M simulation. The deprotonated structure was very similar to the crystal structure with the F114 acting as a hydrophobic barrier that prevents solvation of the K102 (Figure 12). The observations from the protonated and deprotonated simulations support the idea that K102 is deprotonated in the crystal structure and the protonated K102 disrupts the structure significantly. The convergence behavior of the free energy derivatives was found less satisfactory for K102 compared to the H31 case. The  $\Phi$  for  $\Delta G^{(1)}$  did not converge for any of the  $\lambda$  values after more than 10 ns for each  $\lambda$  (Figure 13). The  $\Phi$  value may be converging for the  $\lambda = 1.0$  (deprotonated) simulation with 400 ps blocks; 40 ns of simulation was estimated to be necessary in order to confirm this, which has not been carried out considering the high computational cost. Taking the forward cumulative average and ignoring obvious areas of equilibration (for example, see Figure 13c), the estimated  $pK_a$  shift (relative to propylamine in solution) is  $-15.6$  and  $-13.4$   $pK_a$  units for 0.0 and 0.1 M simulations, respectively. These are much lower than the experimental shift of  $-4.0$   $pK_a$  units. The origin of this discrepancy seems to be the structural instability of the protonated state. Inherent to GSBP, the system is constrained so that it cannot respond completely to the protonated K102. If one assumes that K102 in the protonated state is completely solvent exposed, the free energy derivative for  $\lambda = 0$  can be taken as that for propylamine in solution (197.1 kcal/mol). Using this value together with the free energy derivative for  $\lambda = 1$  in the realistic lysozyme simulations, a linear fit model would produce a  $pK_a$  shift of 4.2, which is in close agreement with the experimental result of 4.0. Although this agreement might be coincidental, considering the success we have had for all other cases studied here, this result strongly argues that protonated K102 significantly destabilizes the protein structure and is likely to be exposed to the solvent, in contrast to the crystal structure.

#### 4. Concluding Remarks

Due to the rapid technological developments in the area of structural biology, high-resolution structures of enzyme systems of increasing size and complexity have become available.<sup>100</sup> These structures offer the exciting opportunity of deciphering the structure–functional relationship<sup>101,102</sup> in enzymes with atomic detail. In such investigations, computational methods based on energetics considerations,<sup>12,11</sup> such as the hybrid QM/MM methods, hold great promise. A crucial step in making QM/MM methods a practical tool in quantitative studies of enzyme systems is the validation of their accuracy and the revelation of factors that determine the accuracy. In a series of recent studies, we have examined the importance of van der Waals interactions,<sup>14</sup> QM/MM frontier treatment,<sup>21</sup> and configuration sampling<sup>8</sup> in QM/MM simulations, primarily using SCC-DFTB as the QM method due to its computational efficiency. Besides the pioneering work of Warshel and co-workers in the context of empirical valence bond model,<sup>79,11</sup> the accurate treatment of electrostatics for QM/MM methods has only been addressed recently.<sup>5–10</sup> Considering the importance of electrostatics in enzyme catalysis<sup>26,11</sup> and extensive discussions in the context of classical simulations,<sup>3,4</sup> it is expected that electrostatic treatment plays a decisive role in determining the reliability of QM/MM simulations. Motivated by such considerations, we have recently implemented consistent electrostatic models for both spherical boundary<sup>23,10</sup> and periodic boundary QM/MM simulations. In the current study, we have chosen to use  $pK_a$  calculations to validate these electrostatic models as well as to explore ways to push forward the quantitative accuracy of QM/MM simulations. In addition to the importance of titratable residues to the structure and function of biomolecules,  $pK_a$  calculations were selected due to the sensitivity of results to electrostatics and the availability of accurate experimental results.

Relative  $pK_a$  calculations for five small molecules in solution and two residues in the interior of the T4 lysozyme have produced encouraging results that highlight components in determining the quantitative accuracy of QM/MM simulations. The small molecule studies demonstrated that inconsistent electrostatic model (force-shift based cutoff treatment for MM interactions combined with no-cutoff treatment for QM/MM interactions) produced results that were difficult to correct in a robust way. This is somewhat unfortunate as it is the default option for most QM/MM simulations in the program CHARMM. The extended electrostatics, generalized solvent boundary potential (GSBP), and Ewald treatments gave consistent results for the solution calculations, once an appropriate bulk contribution was carefully chosen. The calculations also demonstrated that the level of QM potential and the way that QM/MM interaction is evaluated both contribute significantly to the accuracy of the QM/MM results, even when relative quantities, such as  $pK_a$  shifts among different small molecules, are considered. In the present study, a set of simple corrections based on SCC-DFTB/MM trajectories and higher QM level and QM/MM single point energy calculations were found to work well; the general applicability of such perturbative treatment will be further evaluated in other quantitative studies, such as reduction potential calculations.<sup>25</sup> Careful benchmark calculations of different properties using several protein systems are useful for determining the quantitative importance of two factors not explored in detail here, which concern the electronic polarizability of the protein environment near the QM region and surface polarization effect in the spherical boundary condition.<sup>79,98</sup>



**Figure 13.** Convergence of the free energy derivatives for  $\lambda = 0.0$  and  $\lambda = 1.0$  for the K102  $pK_a$  simulation in the K102M mutant of the T4 lysozyme with the 0.1 M GSBP setup. The convergence behavior of  $\Phi$  in the reverse cumulative average with increasing block size is shown for  $\lambda = 0.0$  (a) and  $\lambda = 1.0$  (b); although  $\Phi$  may be converging for  $\lambda = 1.0$  with very large block size (400 ps), much longer simulations would be required to verify this. Overall, the  $\Phi$  (a,b) failed to converge due to large fluctuations in the free energy derivative, as shown in (c), where forward cumulative averages (containing e400 ps blocks) yield values ranging from 170 to 180 kcal/mol. The reverse cumulative average (with 400 ps blocks) of the  $\lambda = 1.0$  free energy derivative (after the initial 5 ns are discarded) is shown in (d), which gives a value between 61 and 63 kcal/mol.

Calculations for the residues in the interior of the T4 lysozyme, which were suggested as stringent tests of electrostatic treatments,<sup>42</sup> illustrate both the promise and limitation of the QM/MM protocol based on the GSBP approach. On one hand, the  $pK_a$  shift for the semi-buried His 31 was reproduced well with SCC-DFTB/CHARMM-GSBP simulations, provided that sufficient sampling (multiple nanoseconds) is carried out and convergence is carefully monitored.<sup>55</sup> No higher-level QM correction was made due to the expected error cancellation between solution and enzyme simulations. On the other hand, calculating the  $pK_a$  shift for Lys 102 proved to be a daunting task; initial stages of major conformational reorganization of the protein due to the change in the protonation state of Lys 102 was observed. Since only limited conformational flexibility of the protein is allowed in typical GSBP setups, the estimated  $pK_a$  after more than 10 ns of simulation for each  $\lambda$  window is still far below the experimental value, although the shift is in the correct direction. Such significant conformational response to a change in the protonation state is expected for most buried residues, for which microscopic models are, in principle, the

only robust approach as predictive tools. While macroscopic models<sup>39,103,96</sup> may be parametrized to obtain correct  $pK_a$  values for buried residues in specific proteins, it is difficult to employ these models to make novel prediction for systems that step outside empirical paradigms. More importantly, calculations based on macroscopic models do not generally offer any signature of “warning” for difficult cases while microscopic simulations do.

In addition to quantitatively evaluating the accuracy of QM/MM simulations, the ability to calculate reliable  $pK_a$  values in protein systems provides a powerful tool to lay out the framework of proton transfers, enabling the direct characterization of the beginning and end-points of such events. Often in proteins, the proton is transferred between two different residue types combining, in essence, the complexity of both our small molecule (with different electronic structures) and T4 lysozyme calculations (with the change in environments). Long-range proton transfers in carbonic anhydrase and proton pumping in cytochrome *c* oxidase are two examples where the electronic structure and electrostatic interactions combine to provide the



free energy surface that enables proteins to accomplish amazing tasks. Our group is actively investigating the thermodynamics and kinetics of proton transfers in these systems using QM/MM techniques that have been calibrated by pK<sub>a</sub> calculations.

**Acknowledgment.** The authors thank Peter Koenig for interesting discussions regarding the link atom treatment for PA calculations with SCC-DFTB. The authors also acknowledge support from the University of Wisconsin, Madison and the National Science Foundation (MCB-0314327, CHE-CAREER-0348649). Q.C. also acknowledges a Research Fellowship from the Alfred P. Sloan Foundation.

## References and Notes

- Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.
- Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.
- Davis, M. E.; Mccammon, J. A. *Chem. Rev.* **1990**, *90*, 509–521.
- Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.
- Gao, J.; Alhambra, C. *J. Chem. Phys.* **1997**, *107*, 1212–1217.
- Nam, K.; Gao, J.; York, D. M. *J. Chem. Theor. Comp.* **2005**, *1*, 2–13.
- Gregersen, B. A.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 536–556.
- Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643–8653.
- Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583–5595.
- Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905.
- Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.
- Field, M. J. *J. Comp. Chem.* **2002**, *23*, 48–58.
- Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*(6), 700–733.
- Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467–6478.
- Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43–52.
- Das, D.; Eurenus, K. P.; Billings, E. M.; Sherwood, P.; Chattfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534–10547.
- Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
- Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
- Zhang, Y.; Lee, T.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46–54.
- Konig, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082–9095.
- Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*(12), 9050–9063.
- Im, W.; Bernéche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*(7), 2924–2937.
- Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Academic Press: San Diego, 1996.
- Schaefer, P.; Sampson, N.; Vrielink, A.; Cui, Q. in preparation.
- Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W. H. Freeman and Company, 1999.
- Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 2403–2408.
- Dao-pin, S.; Anderson, D. E.; Baase, W. A.; Dahlquist, F. W.; Matthews, B. W. *Biochemistry* **1991**, *30*, 11521–11529.
- Ondrechen, M. J.; Clifton, J. G.; Ringe, D. *Proc. Natl. Acad. Sci.* **2001**, *22*, 12473–12478.
- Wikstrom, M. *Curr. Opin. Struct. Biol.* **1998**, *8*, 480.
- Ferguson-Miller, S.; Babcock, G. T. *Chem. Rev.* **1996**, *96*, 2889–2907.
- Brzezinski, P.; Larsson, G. *Biochim. Biophys. Acta* **2003**, *1605*, 1–13.
- Popovic, D. M.; Stuchebrukhov, A. A. *J. Am. Chem. Soc.* **2004**, *126*, 1858–1871.
- Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610–5620.
- Gordon, M.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106*, 3486–3494.
- Konecny, R.; Li, J.; Fisher, C. L.; Dillet, V.; Bashford, D.; Noodleman, D. A. *Inorg. Chem.* **1999**, *38*, 940.
- Yang, A.; Honig, B. *J. Mol. Biol.* **1993**, *231*, 459–474.
- Antosiewicz, J.; Mccammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.
- Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comp. Chem.* **1987**, *9*, 327–335.
- Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368–8372.
- Sham, Y. Y.; Chu, Z. T.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400–417.
- Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400–417.
- Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.
- Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 3572–3575.
- Jorgensen, W. L.; Briggs, J. M. *J. Am. Chem. Soc.* **1989**, *111*, 4190–4197.
- Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.
- Li, G.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 14521–14528.
- Grabowski, P.; Riccardi, D.; Gomez, M. A.; Asthagiri, D.; Pratt, L. R. *J. Phys. Chem. A* **2002**, *106*, 9145–9148.
- Tissandier, M.; Cowen, K.; Feng, W.; Gundlach, E.; Cohen, M.; Earhart, A.; Coe, J. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.
- Zhan, C.; Dixon, D. *J. Phys. Chem. A* **2002**, *105*, 11534–11540.
- McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1973.
- Štrajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333–13343.
- Yang, W.; Bitetti-Putzer, R.; Karplus, M. *J. Chem. Phys.* **2004**, *120*, 2618–2628.
- Herschbach, D. R. *J. Chem. Phys.* **1959**, *31*, 1652–1661.
- Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 103–118.
- Boresch, S. *Mol. Sim.* **2002**, *28*, 13–37.
- Elstner, M.; Porezag, D.; Jungnickel, G.; Elstner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*(11), 7260–7268.
- Elstner, M.; Frauenheim, T.; Suhai, S. *THEOCHEM* **2003**, *632*, 29.
- Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- Bartlett, R. J. *Annu. Rev. Phys. Chem.* **32**, 359–401.
- Haharan, P. C.; Pople, J. A. *Theor. Chim. Acta.* **1973**, *28*, 213–222.
- Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.
- Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*(2), 569–585.
- Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462–3471.
- Ma, S. K. *Statistical Mechanics*; World Scientific: Singapore, 1985.
- Straatsma, T. P.; Mccammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
- Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- Bennett, C. H. *J. Chem. Phys.* **1976**, *22*, 245–268.
- Kuwajima, S.; Warshel, A. *J. Chem. Phys.* **1988**, *89*, 3751.
- Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Chem. Phys.* **1996**, *100*, 1206–1215.
- Bogusz, S., III; T. E. C.; Brooks, B. R. *J. Chem. Phys.* **1998**, *108*, 7070–7084.
- Hünenberger, P. H.; Mccammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.
- Weber, W.; Hünenberger, P. H.; Mccammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.
- Kastenholz, M. A.; Hünenberger, P. H. *J. Phys. Chem. B* **2004**, *108*, 774–788.
- Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124–129.
- Steinbach, P. J.; Brooks, B. R. *J. Comp. Chem.* **1994**, *15*, 667–683.
- Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*(2), 187–217.
- Essmann, U.; Perera, L.; Berkowitz, M. L. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- Brooks, C. L., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159–181.
- Brooks, C. L., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159.
- Rychaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.
- Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

- (88) Dao-pin, S.; Söderlind, E.; Baase, W. A.; Wozniak, J. A.; Sauer, U.; Matthews, B. W. *J. Mol. Biol.* **1991**, *221*, 873–887.
- (89) Jr., A. D. M.; Bashford, D.; Bellot, M.; Jr., R. L. D.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; III, W. E. R.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (90) Range, K.; Riccardi, D.; Elstner, M.; Cui, Q.; York, D., in preparation.
- (91) Elstner, M.; Cui, Q., work in progress.
- (92) Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frisch, M. J. *J. Chem. Phys.* **1998**, *109*, 10570–10579.
- (93) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.
- (94) Wood, R. H. *J. Chem. Phys.* **1995**, *103*, 6177–6187.
- (95) Ashbaugh, H. S.; Wood, R. H. *J. Chem. Phys.* **1997**, *106*, 8135–8139.
- (96) Dong, F.; Zhou, H. X. *Biophys. J.* **2002**, *83*, 1341–1347.
- (97) Nina, M.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239–5248.
- (98) Darden, T.; Pearlman, D.; Pedersen, L. G. *J. Chem. Phys.* **1998**, *109*, 10921–10935.
- (99) Acampora, G.; J. Hermans, J. *J. Am. Soc.* **1967**, *89*, 1543–1547.
- (100) Burley, S. K.; Almo, S. C.; Bonanno, J. B.; Capel, M.; Chance, M. R.; Gaasterland, T.; Lin, D.; Sali, A.; Studier, F. W.; Swaminathan, S. *Nature Genet.* **1999**, *23*, 151–157.
- (101) Whisstock, J. C.; Lesk, A. M. *Q. Rev. Biophys.* **2003**, *36*, 307–340.
- (102) Jones, S.; Thornton, J. M. *Curr. Opin. Chem. Biol.* **2004**, *8*, 3–7.
- (103) Yang, A.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 252–265.