

# Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules

Patricia Schaefer, Demian Riccardi, and Qiang Cui

Citation: *The Journal of Chemical Physics* **123**, 014905 (2005); doi: 10.1063/1.1940047

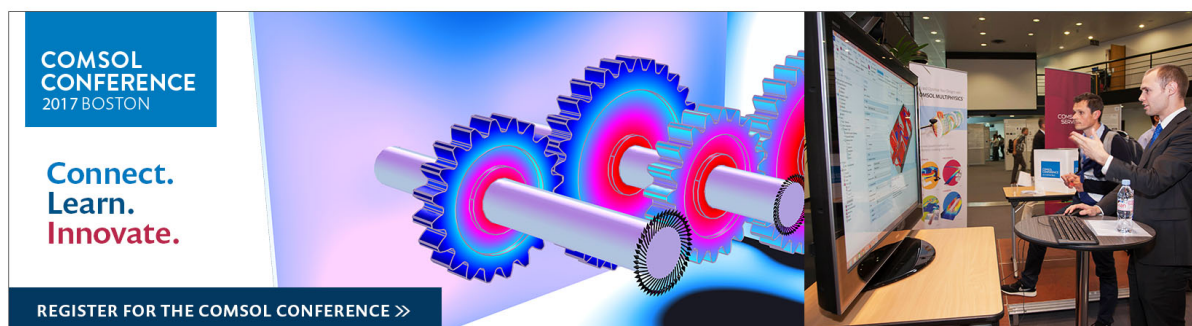
View online: <http://dx.doi.org/10.1063/1.1940047>

View Table of Contents: <http://aip.scitation.org/toc/jcp/123/1>

Published by the *American Institute of Physics*

---

---



COMSOL  
CONFERENCE  
2017 BOSTON

Connect.  
Learn.  
Innovate.

REGISTER FOR THE COMSOL CONFERENCE >>

The banner features a central graphic of three interlocking gears with a blue and red color scheme. To the right, a photograph shows two men in business attire at a conference booth, with one man pointing at a large computer monitor displaying a software interface. The background of the booth is white with blue and red accents.

# Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules

Patricia Schaefer, Demian Riccardi, and Qiang Cui<sup>a)</sup>

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

(Received 14 February 2005; accepted 29 April 2005; published online 12 July 2005)

A robust approach for dealing with electrostatic interactions for spherical boundary conditions has been implemented in the QM/MM framework. The development was based on the generalized solvent boundary potential (GSBP) method proposed by Im *et al.* [*J. Chem. Phys.* **114**, 2924 (2001)], and the specific implementation was applied to the self-consistent-charge density-functional tight-binding approach as the quantum mechanics (QM) level, although extension to other QM methods is straightforward. Compared to the popular stochastic boundary-condition scheme, the new protocol offers a balanced treatment between quantum mechanics/molecular mechanics (QM/MM) and MM/MM interactions; it also includes the effect of the bulk solvent and macromolecule atoms outside of the microscopic region at the Poisson–Boltzmann level. The new method was illustrated with application to the enzyme human carbonic anhydrase II and compared to stochastic boundary-condition simulations using different electrostatic treatments. The GSBP-based QM/MM simulations were most consistent with available experimental data, while conventional stochastic boundary simulations yielded various artifacts depending on different electrostatic models. The results highlight the importance of carefully treating electrostatics in QM/MM simulations of biomolecules and suggest that the commonly used truncation schemes should be avoided in QM/MM simulations, especially in simulations that involve extensive conformational samplings. The development of the GSBP-based QM/MM protocol has opened up the exciting possibility of studying chemical events in very complex biomolecular systems in a multiscale framework. © 2005 American Institute of Physics. [DOI: 10.1063/1.1940047]

## I. INTRODUCTION

Electrostatic interactions play an important role in determining the structure and function of biomolecules.<sup>1–5</sup> This is particularly true for enzymes,<sup>6</sup> in which electrostatic interactions are believed to be fine tuned in the active site, thus providing a significant rate enhancement compared to solution. Therefore, it is imperative that electrostatic interactions are treated in a reliable and efficient manner for simulating enzyme systems. Although a significant amount of work has been done along this line in the context of classical simulations,<sup>7,8</sup> much less effort has been made in the area of quantum mechanics/molecular mechanics (QM/MM) simulations<sup>9–15</sup> due to technical complications introduced by the QM atoms. Considering the great potential of QM/MM methods in the mechanistic analysis of enzymes,<sup>6,16–18</sup> it is important that robust and efficient algorithms for computing electrostatics are introduced into QM/MM simulations; the impact due to different treatments of electrostatics on the simulation result should also be investigated systematically.

The Ewald summation method, combined with periodic boundary conditions, is a widely utilized technique for the accurate calculation of electrostatic interactions in biomolecular systems.<sup>19,20</sup> The method has been implemented in the QM/MM framework by several groups, mostly with

semiempirical level of QM approaches.<sup>11,12</sup> The Ewald method includes a complete representation of electrostatic energy and forces from an infinite periodic array of systems. Although it avoids artifacts associated with truncating electrostatic interactions, a new set of concerns arises due to the imposed artificial periodicity.<sup>21–23</sup> As a result, a large number of solvent molecules is needed to ensure that these artifacts are minimal,<sup>24</sup> and this results in a high computational cost that severely limits the size of systems that can be studied using the Ewald method with either classical or QM/MM potential functions.

In contrast to the all-atom Ewald technique, macroscopic continuum models are among the simplest methods used for modeling a biomolecule/solvent system. The macromolecule is treated as a single low-dielectric medium with embedded fixed charges surrounded by a high-dielectric medium representing the solvent.<sup>3,7,25,26</sup> The continuum model is widely used in both the quantum chemistry community<sup>27,28</sup> and the classical simulations of macromolecules.<sup>29,30</sup> Carefully parametrized models have been successful in reproducing the electrostatic contribution to the solvation free energy of small solutes and can provide useful estimates for macromolecular systems while remaining computationally efficient.<sup>31</sup> By ignoring the atomic and structural details of the solvent, continuum models suffer from a number of problems.<sup>4,29,32</sup> For example, it has become increasingly clear that specific

<sup>a)</sup>Electronic mail: cui@chem.wisc.edu

water molecules play a crucial role in determining the strength and specificity of macromolecular binding interactions.<sup>33–37</sup>

Various proposals for creating a compromise between the two extremes of the all-atom representation in the Ewald technique and the continuum models have been presented by numerous authors in both contexts of small-molecule reactivity analysis<sup>27,28,38–40</sup> and macromolecular simulations.<sup>41–48</sup> For example, the spherical solvent boundary potential (SSBP) model<sup>49</sup> was developed by Beglov and Roux based on suggestions laid out in earlier models.<sup>42</sup> In SSBP, a small number of explicit solvent molecules (the first few solvation shells) are included explicitly along with the solute, while the influence of the remaining bulk solvent is represented with an effective boundary potential. The system is split into regions of explicit and implicit treatments based on a separation of the multidimensional solute–solvent configurational integral in terms of the  $n$  “inner” solvent molecules and the remaining “outer” bulk solvent molecules; Kirkwood’s classical multipolar expansion<sup>50</sup> is used to approximate the electrostatic reaction-field free energy. It was found that thermodynamic properties, such as the solvation free energy of metal ions and the potential of mean force associated with salt bridge interactions between amino acid side-chain analogs, have only a very weak dependence on the size of explicit solvent sphere, suggesting that the SSBP model is rather robust.<sup>51</sup>

As the complexity of biophysical and biochemical problems probed by simulations increases, it is often preferable for the simulation region to include only a portion of the biomolecular system explicitly. In such applications, the irregular shape of the macromolecule/solvent dielectric interface outside the simulation region needs to be considered. The Kirkwood multipolar expansion used in SSBP, which is valid only for an arbitrary charge distribution inside a spherical cavity enclosed in a dielectric continuum, can no longer be used. Thus, Im *et al.* extended the SSBP methodology to macromolecule/solvent systems with arbitrary geometries. This treatment incorporates both the influence of the macromolecular charge distribution outside the simulation region and the electrostatic response from the irregular dielectric interfaces. The method, referred to as the generalized solvent boundary potential (GSBP) method,<sup>52</sup> allows for an accurate treatment of electrostatics for a small subset of atoms embedded in a large solvated macromolecule of arbitrary geometry. Within the framework of GSBP, all atoms in the inner region belonging to a macromolecule, cofactor, ligand, or solvent undergo explicit dynamics, while any remaining macromolecule and solvent molecules lying outside of this inner region are included implicitly in terms of a bulk solvent-shielded static field and a reaction field due to the irregular macromolecule/solvent dielectric boundary.

In this work, we develop a robust and efficient approach for treating long-range electrostatics in QM/MM simulations by adapting the GSBP method to the QM/MM framework. Specifically, the implementation was made with the self-consistent-charge density-functional tight-binding (SCC-DFTB) method<sup>53,54</sup> as the QM level in the CHARMM program,<sup>55</sup> the details of the implementation and preliminary

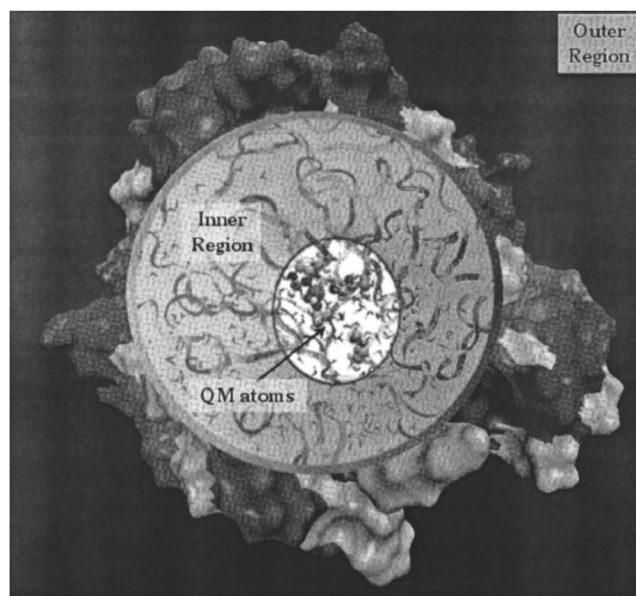


FIG. 1. Schematic representation of the GSBP partition of a solvated biomolecule in the QM/MM framework. Atoms in the inner region are represented explicitly, including a portion of the biomolecule, any ligand or cofactor, and solvent molecules; some atoms are treated with a QM potential. In the outer region, the remaining biomolecule, ligand, or cofactor atoms are represented explicitly but their positions are fixed; the solvent in this region is replaced by a dielectric continuum. The contributions from all MM atoms and the dielectric solvent are included in the QM/MM calculations in a self-consistent manner.

benchmark calculations that illustrate the power and limitation of the new approach are presented here. In Sec. II we review the theoretical details of the method<sup>52</sup> and present its implementation. In Sec. III, the accuracy of the implementation is tested using a series of calculations with simple point-charge models and small molecular systems. This is followed by discussions of the calculations involving active site dynamics of human carbonic anhydrase II to illustrate the importance of electrostatics in biomolecular simulations and to assess the performance of the new method. Finally, we end with a brief summary and conclusions in Sec. IV.

## II. THEORY

In this section we first review the theory of the GSBP method and then discuss the necessary modifications required to extend the theory to a QM/MM framework. The current work deals with the SCC-DFTB method<sup>53,54</sup> as the QM level, though extension to other QM methods is straightforward.

### A. GSBP

Consider a system comprised of a macromolecule in a particular configuration,  $\mathbf{R}$ , immersed in a bulk liquid of  $N$  solvent molecules with a total potential energy of  $U_{\text{tot}}(\mathbf{R}, 1, 2, \dots, N)$ . The GSBP treatment separates the  $N$  solvent molecules into two groups, the  $n$  “inner” solvent molecules within the region of interest and the remaining  $N-n$  “outer” solvent molecules [shown schematically in Fig. 1]. This partition makes it possible to integrate out the contribution of the outer solvent molecules such that their influence

on the inner region is taken into account implicitly. Specifically, equilibrium quantities involving only the degrees of freedom of the reduced system can be expressed rigorously in terms of the potential of mean force (PMF)<sup>56</sup>  $\mathcal{W}(\mathbf{R}, 1, 2, \dots, n)$ ,

$$\mathcal{W}(\mathbf{R}, 1, 2, \dots, n) = \frac{1}{C} \int e^{-\beta U_{\text{tot}}(\mathbf{R}, 1, 2, \dots, N)} d(n+1) \cdots dN, \quad (1)$$

where  $C$  is a normalization constant. The PMF can be obtained by calculating the reversible thermodynamical work necessary to assemble the reduced system,

$$\mathcal{W}(\mathbf{X}) = U(\mathbf{X}) + \Delta\mathcal{W}_{\text{elec}}(\mathbf{X}) + \Delta\mathcal{W}_{\text{CR}}(\mathbf{X}) + \Delta\mathcal{W}_{\text{NP}}(\mathbf{X}), \quad (2)$$

where  $\mathbf{X}$  has been used to represent Cartesian coordinates of atoms in the reduced system,  $\mathbf{X} = (\mathbf{R}, 1, 2, \dots, n)$ . The term  $U(\mathbf{X})$  is the microscopic potential-energy function that is comprised of bonded (bonds, angles, and torsions) and non-bonded contributions (Lennard-Jones and Coulombic interactions);  $\Delta\mathcal{W}_{\text{elec}}$ ,  $\Delta\mathcal{W}_{\text{CR}}$ , and  $\Delta\mathcal{W}_{\text{NP}}$  represent the electrostatic solvation, configurational restriction, and nonpolar contributions to the PMF, respectively.<sup>52</sup> The goal of the GSBP method is to provide a closed-form expression for the electrostatic solvation contribution to the PMF that is valid for any arbitrary geometry and charge distribution.

To make the problem more manageable, the calculation is separated into inner- and outer-region contributions. The charge density is expressed as the sum of outer and inner regions,  $\rho(\mathbf{X}) = \rho^{(i)}(\mathbf{X}_i) + \rho^{(o)}(\mathbf{X}_o)$ , where  $\mathbf{X}_i$  and  $\mathbf{X}_o$  represent the coordinates of atoms in the inner and outer regions, respectively. The coordinates for the outer-region atoms are held fixed during the simulation; in most simulations of macromolecular systems, the coordinates of outer-region atoms are based on x-ray structures and are not chosen arbitrarily. Likewise, the potential energy becomes the sum of three terms,  $U(\mathbf{X}) = U^{(ii)}(\mathbf{X}_i) + U^{(oo)}(\mathbf{X}_o) + U^{(io)}(\mathbf{X})$ .

In general the electrostatic solvation free energy  $\Delta\mathcal{W}_{\text{elec}}(\mathbf{X})$  for a given set of point charges,  $\rho(\mathbf{r}) = \sum_{\alpha} q_{\alpha} \delta(\mathbf{r} - \mathbf{r}_{\alpha})$ , is given by

$$\Delta\mathcal{W}_{\text{elec}}(\mathbf{X}) = \frac{1}{2} \sum_{\alpha} q_{\alpha} \phi_{\text{rf}}(\mathbf{r}_{\alpha}), \quad (3)$$

where the reaction-field potential  $\phi_{\text{rf}}(\mathbf{r})$  is obtained by subtracting a reference electrostatic potential computed in vacuum  $\phi_v(\mathbf{r})$  from the electrostatic potential computed in the dielectric solvent environment  $\phi_s(\mathbf{r})$ . The electrostatic potentials are determined as solutions of the Poisson-Boltzmann (PB) equation<sup>3,56</sup> with the appropriate dielectric boundary and charge distributions

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r}) \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}). \quad (4)$$

Specifically in the GSBP method, as atoms (charges) in the “inner” region move around throughout a simulation, one would need to solve the PB equation at every new configuration, which is prohibitively expensive for macromolecules. To make the method computationally more tractable, we note

that the electrostatic term can be separated into three components that express outer-outer, inner-outer, and inner-inner interactions, respectively,

$$\Delta\mathcal{W}_{\text{elec}} = \Delta\mathcal{W}_{\text{elec}}^{(oo)} + \Delta\mathcal{W}_{\text{elec}}^{(io)} + \Delta\mathcal{W}_{\text{elec}}^{(ii)}. \quad (5)$$

The quantity  $\Delta\mathcal{W}_{\text{elec}}^{(oo)}$  arises from the reaction-field interactions of the charges in the outer region and is constant during the simulation since the outer region is fixed. The second term,  $\Delta\mathcal{W}_{\text{elec}}^{(io)}$ , accounts for the electrostatic interaction between charges in the inner region and the reaction field due to charges in the outer region. In combination with the direct Coulombic interaction between the inner- and outer-region atoms, the total inner-outer electrostatic interaction can be expressed as

$$\begin{aligned} \Delta\mathcal{W}_{\text{elec}}^{(io)} + U_{\text{elec}}^{(io)} &= \sum_{\alpha \in \text{inner}} q_{\alpha} \phi_{\text{rf}}^{(o)}(\mathbf{r}_{\alpha}) + U_{\text{elec}}^{(io)} \\ &= \sum_{\alpha \in \text{inner}} q_{\alpha} \phi_s^{(o)}(\mathbf{r}_{\alpha}), \end{aligned} \quad (6)$$

where  $\phi_s^{(o)}(\mathbf{r}_{\alpha})$  is the electrostatic potential at  $\mathbf{r}_{\alpha}$  due purely to explicit atoms in the outer region with their dielectric response and  $\phi_{\text{rf}}^{(o)}(\mathbf{r}_{\alpha})$  is the corresponding reaction-field potential (i.e., without the vacuum Coulombic component). The  $\phi_s^{(o)}(\mathbf{r}_{\alpha})$  can be calculated using values saved on a regular grid, which needs to be computed only once, making the calculation of the inner-outer term efficient. Finally, the inner-inner contribution is expressed in terms of the inner-region reaction field,

$$\Delta\mathcal{W}_{\text{elec}}^{(ii)} = \frac{1}{2} \sum_{\alpha \in \text{inner}} q_{\alpha} \phi_{\text{rf}}^{(i)}(\mathbf{r}_{\alpha}). \quad (7)$$

This term is troublesome since the value of  $\phi_{\text{rf}}^{(i)}$  is dependent on the instantaneous configuration of the inner region. To make the computation efficient, the charge density of the inner region is expressed in the form of a basis set expansion,

$$\rho^{(i)}(\mathbf{r}) = \sum_m c_m b_m(\mathbf{r}), \quad (8)$$

where the basis functions are dependent on the shape of the inner region (e.g., Legendre polynomials are used for orthorhombic inner regions while spherical harmonics are used for spherical inner regions). The coefficients  $c_m$  of the expansion can be written as

$$c_m = \sum_n S_{mn}^{-1} Q_n, \quad (9)$$

with the elements of the overlap matrix  $S_{mn}$  given by

$$S_{mn} = \int d\mathbf{r} b_m(\mathbf{r}) b_n(\mathbf{r}). \quad (10)$$

The quantities  $Q_n$  are the generalized multipole moments,

$$Q_n = \sum_{\alpha \in \text{inner}} q_{\alpha} b_n(\mathbf{r}_{\alpha}). \quad (11)$$

The basis set representation of the inner-inner contribution to the electrostatic solvation free energy is then given in a matrix form,



$$\Delta\mathcal{W}_{\text{elec}}^{(ii)} = \frac{1}{2} \sum_{mn} Q_m M_{mn} Q_n, \quad (12)$$

where  $M$  is the generalized reaction-field matrix for the system [we assumed that an orthonormal basis set is used; for more general cases, see (Ref. 52)],

$$M_{mn} = \int d\mathbf{r} b_m(\mathbf{r}) \phi_{\text{rf}}(\mathbf{r}; b_n(\mathbf{r})), \quad (13)$$

where  $\phi_{\text{rf}}(\mathbf{r}; b_n(\mathbf{r}))$  is the reaction field at position  $\mathbf{r}$  due to basis function  $b_n(\mathbf{r})$ . Since the dielectric boundary between the inner and outer regions is assumed to be smooth on average,  $M$  is independent of the instantaneous configuration (and charge distribution) of the inner region. Thus, it needs to be computed only once and stored for computational efficiency.

Putting all the terms together, Eq. (2) becomes

$$\begin{aligned} \mathcal{W}(\mathbf{X}) = & U^{(oo)}(\mathbf{X}_o) + \Delta\mathcal{W}_{\text{elec}}^{(oo)}(\mathbf{X}_o) + U^{(ii)}(\mathbf{X}_i) + U_{\text{bonded}}^{(io)}(\mathbf{X}) \\ & + U_{\text{LJ}}^{(io)}(\mathbf{X}) + \Delta\mathcal{W}_{\text{CR}}(\mathbf{X}) + \Delta\mathcal{W}_{\text{NP}}(\mathbf{X}) \\ & + \sum_{\alpha \in \text{inner}} q_\alpha \phi_s^{(o)}(\mathbf{r}_\alpha) + \frac{1}{2} \sum_{mn} Q_m M_{mn} Q_n, \end{aligned} \quad (14)$$

where the final two terms are the core of the GSBP method; both  $U^{(oo)}(\mathbf{X}_o)$  and  $U^{(ii)}(\mathbf{X}_i)$  include Coulombic, LJ, and bonded terms.

## B. QM/MM implementation of GSBP

Choosing a small subset of inner-region atoms to be treated with quantum mechanics (typically fewer than 100 atoms), we extend the GSBP approach to a combined QM/MM framework (shown schematically in Fig. 1).

In the most popular form,<sup>57</sup> the QM atoms of the system interact with the MM atoms through electrostatic and van der Waals terms. Accordingly, the total energy of the system is

$$E^{\text{tot}} = \langle \Psi | \hat{H}^{\text{QM}} + \hat{H}_{\text{elec}}^{\text{QM/MM}} | \Psi \rangle + E_{\text{vdW}}^{\text{QM/MM}} + E^{\text{MM}}, \quad (15)$$

where  $\langle \Psi | \hat{H}^{\text{QM}} + \hat{H}_{\text{elec}}^{\text{QM/MM}} | \Psi \rangle$  gives the electronic energy of the QM atoms in the presence of MM atoms,  $E_{\text{vdW}}^{\text{QM/MM}}$  is the van der Waals interaction between the QM and MM atoms, and  $E^{\text{MM}}$  is the energy of the MM atoms. In the framework of GSBP, the QM/MM-related terms in the potential of mean force  $\mathcal{W}(\mathbf{X})$  can be shown to adopt the form of

$$\langle \Psi | \hat{H}^{\text{QM}} + \hat{H}_{\text{elec}}^{\text{QM/i}} + \hat{H}_{\text{elec}}^{\text{QM/o}} | \Psi \rangle + \Delta\mathcal{W}_{\text{elec}}(\mathbf{X}_{\text{QM}}, \Psi; \mathbf{X}_{\text{MM}}). \quad (16)$$

The first term represents the direct Coloumbic interaction between the QM and MM atoms (in both the inner and outer regions) and  $\Delta\mathcal{W}_{\text{elec}}(\mathbf{X}_{\text{QM}}, \Psi; \mathbf{X}_{\text{MM}})$  is the electrostatic solvation free energy of the QM atoms in the presence of the MM atoms and bulk solvent; we note that, although written as separate terms, the electrostatic solvation free energy contributes to the self-consistent solution of the QM wave function  $\Psi$ . Similar to the above decomposition of electrostatic solvation free energy in the classical framework,  $\Delta\mathcal{W}_{\text{elec}}(\mathbf{X}_{\text{QM}}, \Psi; \mathbf{X}_{\text{MM}})$  contains three components,

$$\begin{aligned} \Delta\mathcal{W}_{\text{elec}}(\mathbf{X}_{\text{QM}}, \Psi; \mathbf{X}_{\text{MM}}) = & \frac{1}{2} \sum_{mn} Q_m^{\text{QM}} M_{mn} Q_n^{\text{QM}} \\ & + \sum_{mn} Q_m^{\text{QM}} M_{mn} Q_n^{\text{MM}} \\ & + \int d\mathbf{r} \rho^{\text{QM}}(\mathbf{r}) \phi_{\text{rf}}^{(o)}(\mathbf{r}), \end{aligned} \quad (17)$$

where  $\rho^{\text{QM}}(\mathbf{r})$  is the charge density of the QM atoms. The first term in Eq. (17) gives the reaction-field interaction concerning the QM atoms, while the second term gives the reaction-field interaction between the QM atoms and the inner-region MM atoms. The last term gives the interaction between the QM atoms and the reaction field due to outer-region atoms.

When the “link atom” approach is used for connecting QM and MM regions, some MM atoms need to be excluded from the QM/MM interaction terms in the calculation, using either the exclude-link-host atom or the exclude-link-host group approach.<sup>57</sup> We account for this in the GSBP context by simply subtracting the contributions of the excluded atoms,

$$\begin{aligned} \Delta\mathcal{W}_{\text{elec}}(\mathbf{X}_{\text{QM}}, \Psi; \mathbf{X}_{\text{MM}}) = & \frac{1}{2} \sum_{mn} Q_m^{\text{QM}} M_{mn} Q_n^{\text{QM}} \\ & + \sum_{mn} Q_m^{\text{QM}} M_{mn} (Q_n^{\text{MM}} - Q_n^{\text{EX}}) \\ & + \int d\mathbf{r} \rho^{\text{QM}}(\mathbf{r}) \phi_{\text{rf}}^{(c)}(\mathbf{r}), \end{aligned} \quad (18)$$

where  $Q_n^{\text{EX}}$  is defined as

$$Q_n^{\text{EX}} = \sum_{\alpha \in \text{EX}} q_\alpha b_n(\mathbf{r}_\alpha). \quad (19)$$

For a general QM approach, both  $Q_n^{\text{QM}}$  and the last term in Eq. (17) depend on the electron density of the QM region. In the SCC-DFTB approach, the situation is simplified by the fact that the electron density is represented by a collection of Mulliken charges,<sup>53,54</sup>

$$\rho^{\text{QM}}(\mathbf{r}) = \sum_{A \in \text{QM}} \Delta q^A \delta(\mathbf{r} - \mathbf{R}_A), \quad (20)$$

where  $\Delta q^A$  is the Mulliken charge of QM atom  $A$ . Thus, calculating  $Q_m^{\text{QM}}$  is a straightforward extension of the classical expression,

$$\begin{aligned} Q_m^{\text{QM}} = & \int d\mathbf{r} \rho^{\text{QM}}(\mathbf{r}) b_m(\mathbf{r}) = \int d\mathbf{r} \Delta q^A \delta(\mathbf{r} - \mathbf{R}_A) b_m(\mathbf{r}) \\ = & \sum_{A \in \text{QM}} \Delta q^A b_m(\mathbf{R}_A). \end{aligned} \quad (21)$$

Likewise, the interaction between the QM atoms and the reaction field due to the outer region has the form

$$\sum_A \Delta q^A \phi_{\text{rf}}^{(o)}(\mathbf{R}_A). \quad (22)$$

Using the variational principle, the GSBP contributions are included in the SCC-DFTB matrix elements during the self-consistent field (SCF) iteration, with the addition of the following terms:

$$\frac{1}{2}S_{\mu\nu}\sum_{A\in\text{QM}}[\Gamma_{CA}(\mathbf{R}_C,\mathbf{R}_A)+\Gamma_{DA}(\mathbf{R}_D,\mathbf{R}_A)]\Delta q^A + \frac{1}{2}S_{\mu\nu}[\Omega(\mathbf{R}_C)+\Omega(\mathbf{R}_D)] \quad \mu\in C,\nu\in D. \quad (23)$$

Here,  $\mu$  and  $\nu$  run over a minimal set of localized pseudo-atomic Slater orbitals located on QM atoms  $C$  and  $D$ , respectively, and  $S_{\mu\nu}$  is the overlap integral associated with the two basis functions. The terms  $\Gamma_{AB}(\mathbf{R}_A,\mathbf{R}_B)$  and  $\Omega(\mathbf{R}_A)$  are defined according to the equations

$$\Gamma_{AB}(\mathbf{R}_A,\mathbf{R}_B) \equiv \sum_{mn} b_m(\mathbf{R}_A)M_{mn}b_n(\mathbf{R}_B),$$

$$\Omega(\mathbf{R}_A) \equiv \sum_{mn} b_m(\mathbf{R}_A)M_{mn}(Q_n^{\text{MM}} - Q_n^{\text{EX}}) + \phi_s^{(o)}(\mathbf{R}_A).$$

Note that we have combined the direct Coloumbic and reaction-field contributions to the interaction between the QM and the outer regions in  $\Omega(\mathbf{R}_A)$  because the total electrostatic potential  $\phi_s^{(o)}(\mathbf{R}_A)$  is used.

Additionally, new terms are needed in the force calculations. The total force on an atom in the inner region is obtained by taking the first derivative of the PMF with respect to the position of that atom. The new terms added to the force in the SCC-DFTB/MM implementation of GSBP are

$$-\Delta q^A \sum_{mn} Q_m^{\text{QM}} M_{mn} \frac{\partial b_n(\mathbf{R}_A)}{\partial \mathbf{R}_A} - \Delta q^A \sum_{mn} \frac{\partial b_m(\mathbf{R}_A)}{\partial \mathbf{R}_A} M_{mn} (Q_n^{\text{MM}} - Q_n^{\text{EX}}) - \phi_s^{(o)}(\mathbf{R}_A) \frac{\partial \Delta q^A}{\partial \mathbf{R}_A} \quad (24)$$

for QM atom  $A$  located at position  $\mathbf{R}_A$  and

$$-\Delta q_\alpha \sum_{mn} Q_m^{\text{QM}} M_{mn} \frac{\partial b_n(\mathbf{r}_\alpha)}{\partial \mathbf{r}_\alpha} \quad (25)$$

for MM atom  $\alpha$  at position  $\mathbf{r}_\alpha$ . Calculations for the derivative of the basis function<sup>52</sup> and Mulliken charges<sup>53</sup> have been discussed in previous studies.

### III. TEST CALCULATIONS AND APPLICATIONS

In this section, we first compare GSBP results using QM/MM and classical potential functions for a simple point-charge system as a way to validate our implementation. The convergence of the method with respect to the size of the basis set was briefly explored using a more realistic example of a solvated QM imidazole. We will then illustrate the QM/MM-GSBP method and the importance of carefully treating electrostatics in bimolecular simulations, with preliminary application to the enzyme human carbonic anhydrase II.

#### A. Simple point-charge model

A QM/MM model was constructed with a  $N^+$  ion, treated quantum mechanically, and a classical point charge ( $q_2 = +1.0 e$ ) inside a spherical cavity with a radius of 12 Å. The

cavity was centered at the origin and enclosed in a dielectric continuum with  $\epsilon=80$ . Im *et al.*<sup>52</sup> used a classical version of this model (comprised of two point charges) as an important comparison between the basis set approach and finite difference PB calculations. This model is ideal for validating our implementation of the SCC-DFTB/MM-GSBP approach because SCC-DFTB/MM is expected to give *exactly* the same result as a pure classical model for a QM region that includes only a  $N^+$  ion due to the Mulliken charge description of the SCC-DFTB electron distribution.<sup>53</sup>

All PB calculations for calculating the electrostatic potential and the reaction-field matrix were performed using a 65<sup>3</sup> cubic grid centered at the origin, with a grid spacing of 0.8 Å (coarse grid), followed by a focusing procedure to a grid spacing of 0.4 Å (fine grid).<sup>58</sup> The inner-region charge density was expressed using a normalized set of 361 spherical harmonics.

Throughout the calculations,  $q_1$  (the  $N^+$  ion) was fixed near the center of the cavity while  $q_2$  (a MM point charge) was gradually moved away along the  $x$  axis from the center towards the dielectric boundary. Excellent agreement was found (not shown) between the QM/MM and classical model results for both the electrostatic solvation free energy and the corresponding derivative with respect to the  $x$  coordinate of  $q_1$ , which indicates that the current QM/MM implementation of GSBP is reliable. The electrostatic solvation free energy diverged (i.e., showed a dramatic increase in magnitude) as  $q_2$  approached the dielectric boundary; e.g., the electrostatic solvation free energy was  $-54.86$  and  $-115.53$  kcal/mol for both QM/MM-GSBP and MM-GSBP when  $q_2$  was at 0.5 and 11.5 Å from the center, respectively. This divergence phenomena has been well noted in the literature for reaction-field-based methods;<sup>49,50,59,60</sup> we note that such behavior is independent of the potential function used for the inner region.

#### B. Tests with imidazole

Small molecule models were also used as tests of the QM/MM-GSBP approach. An imidazole molecule was solvated in an 18-Å water sphere around its center of mass to carve out a cavity in a uniform dielectric continuum,  $\epsilon=80$ . The imidazole molecule was treated with SCC-DFTB while the solvent molecules were treated classically. **To account for the water exclusion radius, a 2-Å layer of the water sphere was then deleted for all subsequent simulations.**<sup>52</sup> The resulting system contained a total of 591 TIP3P water molecules.<sup>61</sup> To calculate  $\mathbf{M}$  and  $\phi_s^{(o)}$ , PB equations were solved with a grid identical to the one used with the point-charge model. The basis functions used to express the charge density of the inner region were generated from spherical harmonics up through the 20th order (400 basis functions).

It is interesting to consider the convergence of the free-energy (potential of mean force) and force values with respect to the size of the basis set. As shown in Fig. 2(a), the solvation free energy (of the entire solute-solvent system) converges rather slowly as a function of the basis size, which is expected based on the observation for the point-charge model; i.e., the presence of water molecules close to the

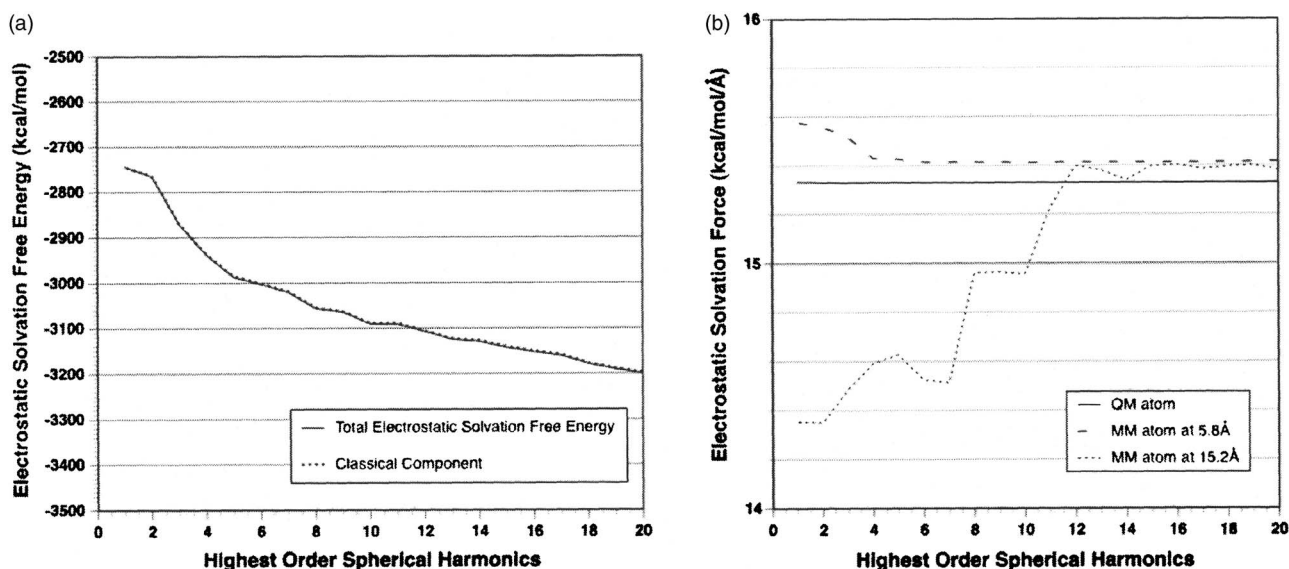


FIG. 2. Convergence of the GSBP results vs the size of the basis set for a solvated imidazole. (a) Electrostatic solvation free energy of the entire system; the solid line is the full QM/MM result and the dashed line is the classical contribution. (b) Electrostatic solvation contribution to the  $x$  direction of the force for three atoms in different regions. The solid line is for a QM atom, the dashed lines are for two MM atoms at 5.8 and 15.2 Å from the center of the cavity, respectively.

dielectric boundary requires a large number of basis functions for the multipolar expansion around the center of the sphere to converge. The QM atoms make only a small contribution to the electrostatic solvation free energy, which shows a very weak dependence on the size of the basis; this behavior is expected since the QM region includes only a small neutral solute (imidazole) in the center of the water sphere.

Based on similar arguments, the convergence of the force (due to GSBP) with respect to the basis set size depends on the location of the atom [Fig. 2(b)]. For the QM atoms and MM atoms close to the center of the sphere, the convergence is fast and requires spherical harmonics no higher than the order sixth. For a water molecule close to the dielectric boundary, however, the convergence is rather poor, and spherical harmonics up to the 18th order are needed. Evidently, a delicate balance must be considered in designing realistic simulations to keep the region of interest far enough from the boundary while keeping the simulation region small to minimize computational costs. Quantitative tests such as  $pK_a$  calculations are useful for validating the system size and the order of spherical harmonics in realistic applications.<sup>62</sup>

### C. Carbonic anhydrase II

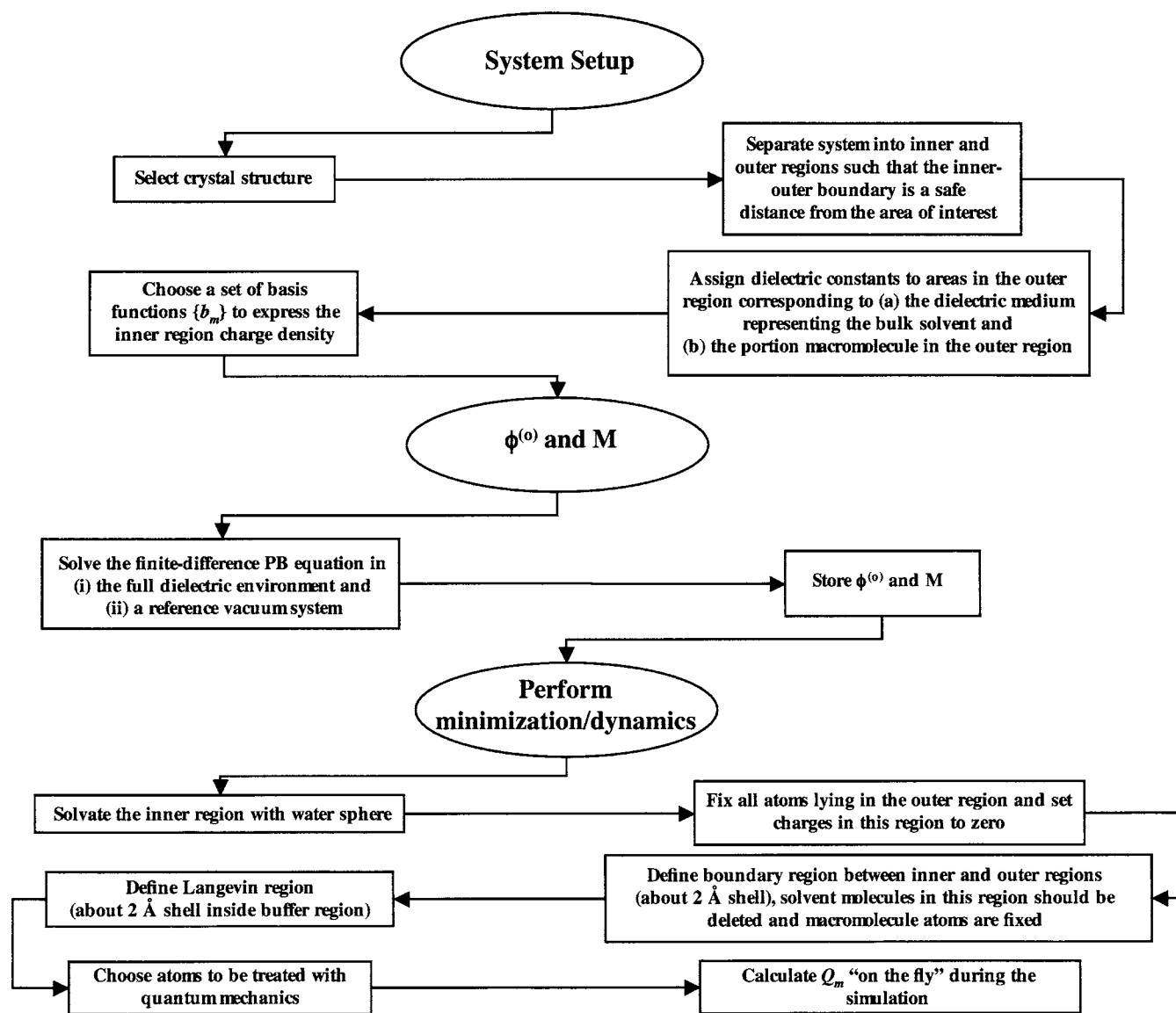
To illustrate the setup and performance of the QM/MM-GSBP method for a realistic system as well as demonstrate the importance of a reliable treatment of electrostatics in enzyme simulations, we present preliminary results from our study of the human carbonic anhydrase II (CA). CA is a Zn(II)-containing metalloenzyme comprised of a single polypeptide chain of 260 residues that moderates respiration by catalyzing the conversion between  $\text{CO}_2$  and  $\text{HCO}_3^-$ ; this has a direct effect on the release of  $\text{O}_2$  to the tissues through hemoglobin.<sup>63</sup> An important step in the functional cycle of CA involves a proton transfer (PT) between the zinc-bound water and His 64 close to the surface of the enzyme. The PT

spans about 8 Å and therefore is expected to be mediated by water molecules in the active site.<sup>64–66</sup> The significant charge reorganization associated with such an intermediate-range PT implies that a careful treatment of electrostatics is important, which makes CA an ideal system for illustrating the QM/MM-GSBP approach. Here we briefly compare QM/MM-GSBP with more traditional QM/MM simulations using the stochastic boundary condition with different approximate treatments for electrostatics.<sup>67,68</sup> A more thorough analysis and comparison of QM/MM simulations using GSBP and other boundary conditions as well as previous theoretical studies<sup>69–72</sup> will be reported separately.

### 1. Computational setup

The procedure for setting up an enzyme system for QM/MM-GSBP simulations is summarized in scheme 1I which involves partitioning of the systems into inner and outer regions, calculating and storing of the GSBP-related quantities ( $\phi_s^{(o)}$  and the reaction field matrix  $\mathbf{M}$ ), and initial minimization/equilibration.

As the starting atomic configuration, the x-ray structure for CA from human erythrocytes with 1.54-Å resolution (PDB code 2CBA)<sup>73</sup> was used. The active site was solvated in a 25-Å water sphere centered on the zinc ion. Successive overlays of the water sphere were performed with all solvent molecules within 2.8 Å of any nonhydrogen protein atom removed in each iteration. The added solvent molecules were relaxed at 300 K for 20 ps with all protein and crystallographic water molecules held fixed. The solvation layers beyond 18 Å were then deleted, providing the final structure used as the starting point for subsequent simulations. The outer (fixed) region contained 1229 out of the total 4525 atoms, and the remaining 3296 atoms (2846 protein atoms together with 150 TIP3P water molecules<sup>61</sup>) belonged to the inner region [Fig. 3(a)]. Simulations of two protonation states were explored here; one set involved zinc-bound water



SCHEME 1.

and neutral His 64, while the other involved a zinc-bound hydroxide and a protonated His 64. Consistent with earlier literature,<sup>69</sup> these states are referred to as CHOH and COHH, respectively.

In the PB calculations for  $\phi_s^{(o)}$  and the reaction field matrix  $M$ , the dielectric constant for the bulk solvent region was set to 80 while that for the protein in the outer region was set to one in most cases and four in others. The protein-solvent boundary was set using the optimized atomic Born radii by Nina *et al.*<sup>74</sup> No salt was included and the ionic screening constant  $\kappa^2(\mathbf{r})$  was set to 0 everywhere. All PB calculations were performed using a  $111^3$  cubic grid centered on the zinc ion, with a coarse grid spacing of 1.2 Å, followed by a focusing procedure to a grid spacing of 0.4 Å. The inner-region charge density was expressed using the first 20th-order spherical harmonics with a total of 400 basis functions. Quantitative tests using  $pK_a$  calculations in both solution and protein systems have shown that the system size and order of spherical harmonics are appropriate.<sup>62</sup>

In the molecular-dynamics (MD) calculations, all atoms

between 18 and 20 Å were held fixed to provide a buffer region while allowing for a smooth spherical dielectric boundary at 20 Å. Atoms in the spherical shell from 16 to 18 Å were treated with Langevin dynamics, while atoms within 16 Å were treated with Newtonian dynamics [Fig. 3(a)] similar to the deformable stochastic boundary method.<sup>75</sup> A time step of 1.0 fs was used and all bonds to hydrogen atoms were constrained with SHAKE<sup>76</sup> Figure 3(b) shows a detailed view of the active site. For the qualitative purposes of studying equilibrium properties, only a small subset of atoms were treated with quantum mechanics (SCC-DFTB<sup>53,54</sup>), which included the zinc ion, the three coordinating histidine residues (His 94, His 96, and His 119), and the zinc-bound water molecule (or hydroxide ion in the COHH state). In the study of proton transfer energetics, the side chain of His 64 and bridging water molecules were also treated with SCC-DFTB. Other protein atoms plus explicit solvent molecules were described with the CHARMM22 force field.<sup>77</sup> Link atoms<sup>57</sup> were introduced between  $C_\alpha$  and  $C_\beta$  atoms of the histidine residues to saturate the valence of the



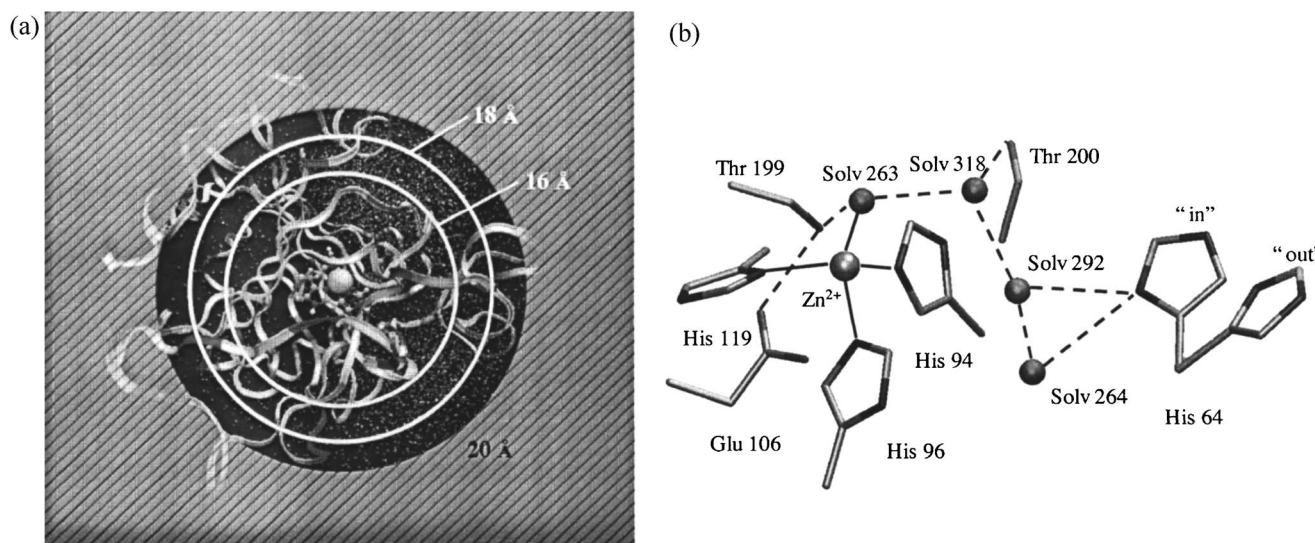


FIG. 3. Carbonic anhydrase II system based on the x-ray structure from human erythrocytes at 1.54-Å resolution (PDB code 2CBA). (a) The simulation setup of 20-Å water sphere centered on the zinc ion in the active site. The zinc ion, shown as a van der Waals sphere, several histidine residues (His 94, His 96, His 119, and His 64), and the zinc-bound water in the active site are shown in CPK format, the protein backbone is shown in ribbon and the water molecules are shown in dots. In the GSBP framework, all outer-region atoms (atoms beyond 20 Å from Zn) are fixed, and the entire system is embedded in an isotropic dielectric continuum ( $\epsilon=80$ ) representing the bulk solvent. All water molecules in the spherical shell from 18 to 20 Å are deleted to account for the water exclusion radius.<sup>52</sup> In the boundary region, all atoms from 18 to 20 Å are held fixed while atoms in the region from 16 to 18 Å are treated with Langevin dynamics. All atoms within 16 Å are fully flexible throughout the simulations. (b) The active site of CA with several important residues (colored according to atom type). Hydrogen atoms have been neglected for clarity.

boundary QM atoms. The link atoms interact with the MM atoms, except those atoms belonging to the group of the “link host” MM atom (the  $C_{\alpha}$ ,  $H_{\alpha}$ , N, and HN atoms in this case), through electrostatic terms; no van der Waals interactions were included for the link atoms.

To compare with the QM/MM-GSBP simulations, conventional stochastic boundary condition simulations<sup>47,75</sup> were also carried out. The x-ray structure was solvated in a 25-Å water sphere centered on the zinc ion, and atoms outside of 25 Å were deleted. Atoms in the spherical shell from 16 to 25 Å were treated with Langevin dynamics in an analogous way as in the GSBP setup; the same QM/MM partition scheme was used. All charges beyond 10 Å of the zinc ion were screened according to the Poisson–Boltzmann charge-scaling scheme<sup>78</sup> to approximately account for solvent shielding beyond that provided by the explicit solvent molecules included in the model. Two popular approximate schemes for treating electrostatics in CHARMM, force-shift and extended electrostatics, were tested. In the former, interactions between atoms with greater than 12-Å separation were ignored; thus this scheme will be referred to as “cutoff” in figures and discussions. In the extended electrostatics model, which will be referred to as “ext,” interactions beyond 12 Å were treated using a multipolar expansion including dipolar and quadrupolar terms.<sup>55</sup> To make the comparison between different protocols meaningful, several independent nanosecond simulations with different initial velocity distributions were carried out for each protocol.

Before moving on to the discussion of simulation results, it is useful to emphasize the fundamental differences between GSBP, cutoff, and extended electrostatics simulations in the QM/MM framework. In the CHARMM implementation of SCC-DFTB/MM, QM/MM electrostatic interactions are

treated without any truncation; as a result, the cutoff simulations treat QM/MM and MM/MM interactions in a very unbalanced way, which may induce unphysical features (*vide infra*) into QM/MM simulations. The extended electrostatics model treats the MM/MM interactions with essentially no cutoff and therefore handles QM/MM and MM/MM interactions in a balanced manner; however, the effects of the bulk solvent and macromolecule atoms beyond the simulated system are missing, which may also cause unexpected artifacts. The GSBP approach implemented here alleviates these troubles and therefore is expected to be most reliable.

## 2. Results and discussion

The catalytic cycle of CA follows a ping-pong kinetic scheme. The first half cycle in the hydration direction involves the nucleophilic attack of the zinc-bound hydroxide on the carbon in  $CO_2$ . The resulting  $HCO_3^-$  is then displaced by a water molecule and released into a solution.<sup>64</sup> In the second half cycle, which is the step of interest here, the zinc-bound water transfers its proton to the His 64 near the surface of CA, thus regenerating the catalytic species of zinc-bound hydroxide; the transferred proton is released into a solution as the doubly protonated His 64 side-chain flips from a buried (“in”) conformer to a solvent-exposed (“out”) conformer. Since the distance between the zinc-bound water and His 64 in the x-ray structure<sup>66</sup> was observed to be too long (7.5 Å) for a direct transfer, water molecules in the active site were assumed to act as bridges that relay the proton transfer (PT).<sup>64–66</sup> Although this mechanism has been widely accepted, the precise number of water molecules involved in such bridges and the nature of events that limit the PT rate have been under heated debate.<sup>65,69–72,79–82</sup> Issues of

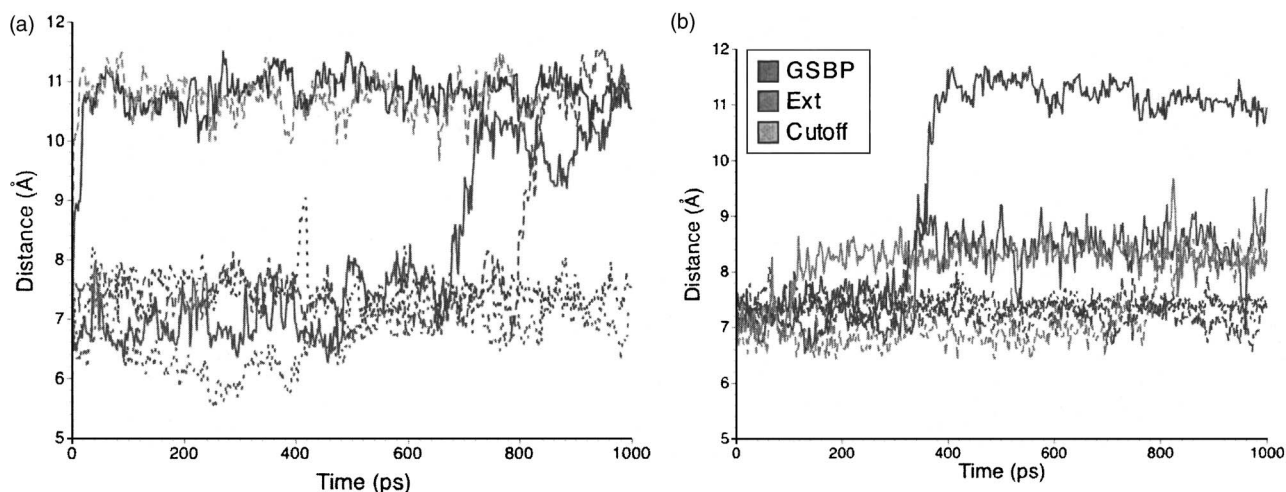


FIG. 4. Time evolution of the distance between the zinc ion and the ND1 atom of His 64 in independent QM/MM simulations of CA. (a) COHH state (zinc-bound hydroxide ion and protonated His 64 residue). (b) CHOH state (zinc-bound water and neutral His 64). Results are colored according to the electrostatic treatment in the simulation. GSBP results are shown in blue, ext results are shown in green, and cutoff results are shown in orange. Multiple independent simulations of each treatment are indicated by solid, dashed, and dotted lines of a single color.

critical importance include the flexibility of the proton acceptor (His 64), the distribution of water molecules in the active site, and the type of “water wires” formed between the zinc-bound water and His 64, as well as the energetics of proton transfer associated with a specific type of water bridge. In the following, we briefly compare results from GSBP- and stochastic-boundary-condition-based SCC-DFTB/MM simulations concerning those issues; more detailed discussions and comparison with previous studies will be given elsewhere.

*a. Flexibility of His 64.* To function as an effective proton shuttle between the CA active site and bulk solution, the side chain of His 64 is expected to be flexible. Indeed, both in and out conformations have been observed in the x-ray structures of CA<sup>73</sup> [Fig. 3(b)]. For the COHH state, in all cutoff simulations performed, the His 64 side chain consistently flipped to the “out” position in the very early stage of the trajectory (results from only one MD simulation are shown for clarity in Fig. 4(a)). This behavior is likely due to the unbalanced QM/MM and MM/MM treatments in the cutoff simulations, which overestimate the repulsion between the protonated His 64 and zinc-hydroxide (both groups bear a charge of +1). For example, the favorable interaction between Glu 106 in the active site [Fig. 3(b)] and the protonated His 64, which counteracts the repulsion between His 64 and the zinc site, is truncated in the cutoff scheme. As expected, such artifact was not seen in neither GSBP nor ext simulations, and both in and out conformations were sampled. Of the two GSBP simulations presented here, one flipped out very rapidly in less than 50 ps, while the other stayed “in” for the entire 1 ns of the simulation. The ext simulations also sampled both in and out conformations, although His 64 did not flip out till 650 ps of the simulation had passed in one case and 800 ps in the other [Fig. 4(a)]. The third ext simulation only sampled the in conformation for the full 1 ns of the production run. However, unlike the GSBP result, the His 64 side chain reached a region very close ( $\sim 5.0$  Å compared to the value  $\sim 8.0$  Å in the x-ray data) to the zinc atom, as shown in

Fig. 5(a). Such a distinct deviation from the x-ray structure is likely an artifact due to the neglect of bulk solvation in the ext simulations, which may cause an overestimation of attraction between the protonated His 64 and active site residues (i.e., opposite to the cutoff simulations). In the CHOH case, which involves a neutral His 64, the results seem to be less sensitive to the treatment of electrostatics. Indeed, the problems of His 64 consistently flipping out in the cutoff simulations and that of an unusually close distance between the zinc ion and His 64 in the ext simulations were no longer visible [Fig. 4(b)]. However, the agreement with experiment in terms of the comparison between the average structure from simulations and the x-ray data depends on the electrostatic model. As shown in Figs. 5(b) and 5(c), GSBP results agree very well with the x-ray structure for both the in and out conformations in the respective trajectories. Average results from cutoff and ext simulations, however, are intermediate between the in and out conformers in the x-ray data; we emphasize that the average structures were calculated using only the last 500 ps of trajectories to ensure that the observed result is not an artifact due to the mixture of in and out conformers. These observations clearly demonstrated the importance of electrostatic interactions in determining configurational sampling of an enzyme system with complex and heterogeneous charge distribution. Based on the comparison with available x-ray data, GSBP appears to be robust while both ext and cutoff treatments have problems generating consistently reliable results.

*b. Water distribution.* Radial distribution functions (RDFs) provide insights into the ordering of water molecules around the zinc ion in the active site. The RDFs between the zinc ion and the oxygen atom of the water molecules (excluding the oxygen of the zinc-bound water/hydroxide) were normalized to bulk concentration for water,  $0.0334$  molecules/Å<sup>3</sup>. Since the water distribution is expected to be different with His 64 adopting the in and out conformers, we discuss the corresponding results separately. The radial distribution function of the COHH state showed

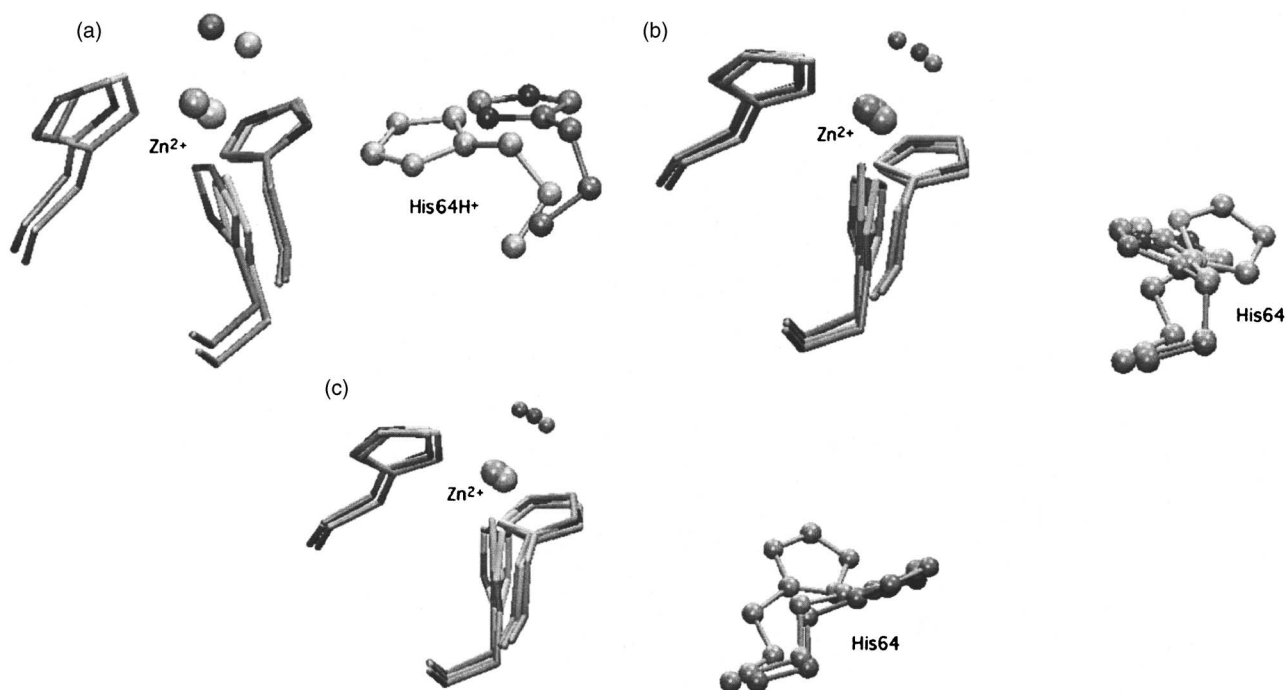


FIG. 5. Superposition of the averaged structures of several active site residues, the zinc ion, and the zinc-bound water molecule (hydroxide ion) with the x-ray structure (colored by atom type). Green, orange, and yellow structures are produced from GSBP, cutoff and ext simulations, respectively. (a) The x-ray structure with His 64 in the in conformation superposed with a snapshot from an ext simulation [dotted green line in Fig. 4(a)] where the His 64 side chain sampled a region very close to the zinc ion ( $\sim 5.0$  Å compared to  $\sim 8.0$  Å in the x-ray data). (b) The x-ray structure with His 64 in the in conformation superposed with structures representing the time average over the last 500 ps of a GSBP [blue dashed line in Fig. 4(b)] and a cutoff simulation [solid orange line in Fig. 4(b)]. (c) The x-ray structure with His 64 in the out conformation superposed with structures representing the time average over the last 500 ps of a GSBP [solid blue line in Fig. 4(b)] and a cutoff simulation [solid orange line in Fig. 4(b)]. The orange structure here is the same as that shown in (b).

three peaks for GSBP simulations when His 64 is “in” while ext simulations where His 64 is “in” showed four peaks [Fig. 6(a)]. The three peaks are located at 3.0 to 3.5, 5.6, and 8.0 Å from the zinc ion [dashed blue line in Fig. 6(a)] in the GSBP results; these three solvation shells were also observed in the RDFs of the ext simulations with another peak at 4.6 Å from the zinc. In addition to a difference in the number of peaks in the RDFs, simulations differed in relative sizes of the individual peaks. This is made clear by integrating RDFs over a given distance from the zinc ion; as shown in Table I, the GSBP simulation consistently gives fewer water molecules close to the zinc compared to the ext simulations. For simulations with His 64 in the out conformation, there is better agreement between GSP, ext, and cutoff results in terms of the feature of the RDF [Fig. 6(b)] and integrated number of water molecules (Table I). Comparing in and out results, the major trend is that fewer number of water molecules are close to the active site, and the difference is most significant for GSBP simulations. The RDFs for the CHO state showed a different picture [Fig. 6(b)]. There are no noticeable difference in the number of peaks for in and out conformations. However, the RDFs corresponding to the ext and cutoff simulations resulting in His 64 positioned intermediate between the in and out conformations [Figs. 5(b) and 5(c)] are somewhat different from the others. This orientation of His 64 seems to cause a tighter packing of water molecules in the first solvation shell than what is otherwise observed. For example, the first cutoff simulation [solid orange line in Fig. 4(b)] reached this unusual conformation in the first 200 ps of the trajectory, and though the second cut-

off simulation [dashed orange line in Fig. 4(b)] did not reach this conformation until the last 200 ps of the 1-ns simulation, the RDFs for these two simulations are remarkably similar. They both have a large peak near 3.0 Å as well as a rather large peak around 6.0 Å, with a less significant peak located in the middle of these two and a final peak near 8.0 Å. By contrast, GSBP simulations showed four peaks located at 3.2, 4.6, 5.8, and 8.0 Å from the zinc ion; somewhat similar results were observed from the second ext simulation (dashed green line). These differences are further shown by the integrated RDFs in Table I, and the most obvious trend is that cutoff simulations yielded a substantially larger number of water molecules in the active site compared to ext and GSBP simulations. Apparently, due to the unbalanced QM/MM and MM/MM treatments, the cutoff simulations resulted in an overpolarization of the system and the +2 charge of the QM region pulled water molecules in. These results further emphasized the negative effects caused by the inconsistent treatment of electrostatics in the cutoff QM/MM simulations.

*c. Water wire formation.* The role of water wires in CA and other systems involving long-range PT has been extensively discussed.<sup>69,70,72</sup> We have analyzed the current CA simulations for the presence of water bridges containing one to six-water molecule bridges. We consider a water bridge to be present in a trajectory frame if both distance ( $O-O < 3.5$  Å) and angle ( $O-H-O \geq 140^\circ$ ) criteria are satisfied simultaneously for all the water molecules connecting the zinc-bound water and His 64. The water bridge is considered to be productive if a proton can successfully be passed from



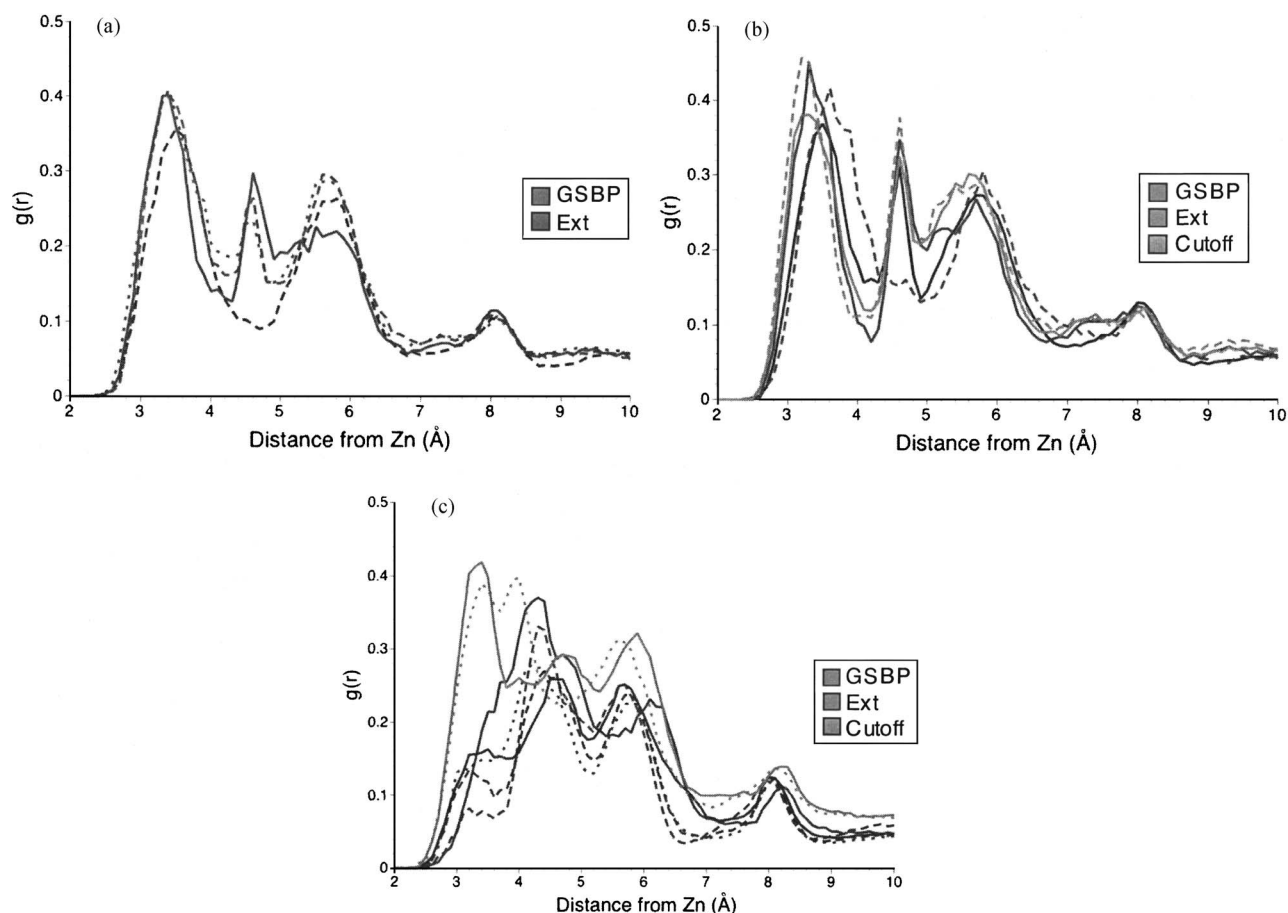


FIG. 6. Radial distribution functions for the zinc center to the oxygen in the surrounding water molecules, normalized to bulk water density. The zinc-bound water/hydroxide is excluded from the analysis. (a) COHH state with His 64 in the in conformation; (b) COHH state with His 64 in the out conformation; (c) CHO state. Results are colored according to the electrostatic treatment in the simulation. GSBP results are shown in blue, ext results are shown in green, and cutoff results are shown in orange. Multiple independent simulations of each treatment are shown in some cases and indicated by solid, dashed, and dotted lines of a single color.

donor to acceptor through the bridge, and unproductive otherwise (e.g., one water molecule serves as a double donor or double acceptor). The results are categorized by the population of the water bridge, defined by the number of frames

TABLE I. Integrated radial functions of COHH and CHO states of CA<sup>a,b</sup>

Distance from Zn <sup>2+</sup> (Å)	COHH			CHOH		
	GSBP	Ext	Cutoff	GSBP	Ext	Cutoff
5.0	3.0 (3.8)	4.0 (3.9)	(4.0)	3.7 (3.3)	3.2 [4.5]	5.0 [5.1]
7.0	8.5 (10.1)	9.9 (10.2)	(10.7)	9.3 (8.1)	8.0 [10.7]	12.0 [12.3]
10.0	15.1 (17.8)	17.5 (18.4)	(19.6)	16.2 (14.4)	15.1 [17.7]	23.1 [22.1]
17.0	111.9 (111.2)	117.0 (116.9)	(120.1)	105.5 (104.9)	107.1 [111.2]	127.7 [130.6]
Full model <sup>a</sup>	148	999	999	148	993	993

<sup>a</sup>The radial distribution function is normalized to bulk water concentration, 0.0334 molecule/Å<sup>3</sup>, and are shown in Fig. 6.

<sup>b</sup>Results corresponding to “in” and “out” conformations of residue His 64 are shown without and with parentheses, respectively; the Ext and Cutoff simulations with His 64 positioned in an unexpected intermediate positioning between the “in” and “out” conformations, results are shown in brackets.

<sup>c</sup>Total number of water molecules in the setup.

with a productive water bridge of a specific length relative to the total frames of the analyzed trajectory (Fig. 7). When more than one productive water bridge is present in a frame, only the shortest one is counted. In the COHH state, GSBP, ext, and cutoff simulations showed good agreement when His 64 is in the out conformation [solid bars in Fig. 7(a)]. These simulations favored four-water molecule bridges, which is expected since a longer water bridge would be required to accommodate the longer Zn-H64 distance. By the same token, the preference for two-water molecule bridges exhibited in ext and GSBP simulations with His 64 “in” is also expected, given the shorter Zn-H64 distance [striped bars in Fig. 5(a)]; we note, however, that the GSBP simulations gave much lower water wire populations than do the ext simulations. For the CHO state, only the GSBP simulation yielded out His 64 conformation that is consistent with the x-ray result; water wires containing four, five, and six water molecules showed very similar populations [striped bars in Fig. 5(b)]. With the His 64 “in,” GSBP simulations also showed a nearly equal distribution over water bridges ranging from two to six water molecules [solid bars in Fig. 5(b)]. In one of the two ext simulations, the His 64 sampled the unusual position intermediate between the in and out configurations [Figs. 5(b) and 5(c)], while it stayed in the



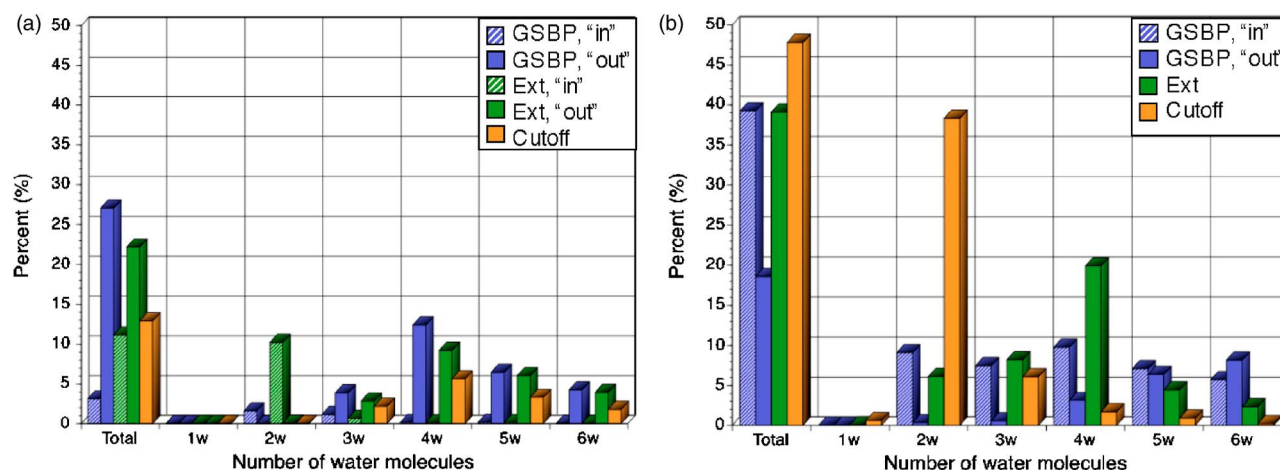


FIG. 7. Percentage of trajectory frames containing productive water bridges between zinc-bound water/hydroxide and ND1 atom of His 64 from all simulations, shown as all sizes of water bridges as well as a distribution by the number of water molecules participating in the bridge. Results for in and out conformations of His 64 are shown separately with solid bars and striped bars, respectively. (a) COHH state. (b) CHOH state.

active site in the other simulation. Thus, it is somewhat surprising for ext simulations to favor a four-water molecule bridge, but not unreasonable. In both of the cutoff simulations, His 64 was positioned in that unusual configuration as well. However, cutoff results favored a shorter, two-water molecule bridge in contrast to the ext results. This difference may be caused by the overpolarization of the system in the cutoff simulations, in which the active site was flooded with water. It is important to note that while our analysis of water bridge formation is in good agreement with our Zn-H64 data, further analysis needs to be performed to confirm this point. A more detailed analysis in relation to the configuration of His 64 is necessary. Furthermore, we have only considered water bridge formation to the ND1 atom of His 64, but the NE2 proton may also be considered the donor in the COHH state. These analyses, which have important functional implications, will be reported separately.

*d. Proton transfer energetics.* The ultimate quantities of interest in the study of CA include the dominant PT pathway and energetics. To explore the effect of electrostatic treatment on the PT energetics, adiabatic mapping was used to determine the approximate barrier height and endothermicity of PT involving three bridging water molecules (x-ray water numbers: 2, 30, and 56).<sup>73</sup> In the adiabatic mapping, the four

transferring protons were assumed to be moving in a fully concerted manner, and their transfer was controlled by the antisymmetric stretch involving the relevant proton donor, the proton, and the proton acceptor atoms. Although this protocol is not expected to give the precise barrier height and endothermicity, it is sufficient for comparing different electrostatic treatments. The possibility of slipping into different regions in the conformational space during different sets of adiabatic mapping calculations was minimized by using one coordinate set, with GSBP on or off; with GSBP off, either the cutoff or ext treatment was used. As shown in Table II, the electrostatic treatment had only a relatively small effect on the barrier height but a much more pronounced effect on the endothermicity, as much as 7 kcal/mol between GSBP and ext calculations. This is expected because charge separation is more significant after the proton transfer. Although the effect on the barrier height seems small in those adiabatic mapping calculations, we argue that the effect on free-energy barrier could be more significant since the electrostatic treatment may perturb the conformation of important residues (e.g., His 64) and the distribution of water molecules in the active site, as described above. This was indeed confirmed by the potential of mean force simulations, which will be reported separately. We note that the energetics seem to be rather insensitive to the dielectric constant used for the outer region;  $\epsilon=1,4$  gave essentially the same result within 0.5 kcal/mol. Finally, we note that both barrier and exothermicity from the adiabatic mapping calculations are sufficiently different from experiments, which indicated a barrier about 8–10 kcal/mol and a PT being nearly thermoneutral.<sup>83</sup> Indeed, the potential of mean force simulations (not included) found much better agreement with these experimental data, which highlighted the importance of configuration sampling in the study of long-range PT processes.

TABLE II. Energetics of proton transfer (kcal/mol) from representative adiabatic mapping calculations in carbonic anhydrase with the fSCC-DFTB/CHARMM potential and different electrostatic treatments.<sup>a</sup>

Model	Barrier	Endothermicity
GSBP	18.9 (18.8) <sup>b</sup>	6.9 (6.5) <sup>b</sup>
Ext	19.2	13.0
Cutoff	20.4	10.7

<sup>a</sup>The crystal structure was used. Three bridging water molecules (crystal water 2, 30, and 56 in Ref. 65) were assumed to be involved in a fully concerted proton transfer between the zinc-bound water and His 64. The protons were pushed in the adiabatic mapping calculations using antisymmetric stretch involving the relevant proton donor, the proton and proton acceptor atoms.

<sup>b</sup>The dielectric constant for the outer region is set to be 1 and 4, respectively, for values with and without parentheses.

## IV. CONCLUSIONS

As the computational hardware improves, combined QM/MM simulations have become increasingly powerful in the study of condensed phase problems, ranging from reac-

tivity analysis in solution, materials, and biomolecules to molecular spectra assignment in complex environments. The quantitative accuracy of such simulations clearly depends on many factors, such as the level of QM method and the reliability of the MM force field for the problem at hand. With a given choice of QM and MM potential functions, other technical details of the simulation may also make an important contribution.<sup>84–88</sup> For biomolecules, which are characterized by complex and heterogeneous charge distributions, the treatment of electrostatics is expected to make an important impact on the simulation results. Although this issue has been emphasized repeatedly in classical simulations, less systematic work has been done in the QM/MM framework.<sup>9–15</sup> We recently stressed the importance of careful treatment of electrostatics in QM/MM studies of processes involving long-range charge separation or net change in the total charge (e.g., change in the oxidation<sup>14</sup> or protonation state<sup>62,89</sup>), and we adopted a charge-scaling approach<sup>78</sup> in these studies. In the current work, we implemented a more systematic and numerically robust way of treating QM/MM electrostatics by adapting the generalized solvent boundary potential (GSBP) approach of Im *et al.*<sup>52</sup> in the QM/MM framework. The resulting new protocol treats the QM/MM and MM/MM interactions in a balanced manner and includes long-range electrostatics at the Poisson–Boltzmann level due to the bulk solvent and macromolecule atoms outside of the region being explicitly simulated.

The implementation of the method was verified based on studies of simple systems such as point charges and a solvated imidazole. The performance of the method was illustrated with the application to the enzyme human carbonic anhydrase II (CA), and a comparison was made to the more standard stochastic boundary condition protocol with different electrostatic treatments. Properties of interest included orientation of the proton acceptor (His 64), distribution of water molecules in the active site, formation of water wires connecting the proton donor and acceptor groups, and energetics of proton transport. It was found that results from GSBP-based QM/MM simulations were more consistent with available experimental data concerning the active site. Both extended electrostatics and cutoff treatments resulted in unusual configurations of the His 64 residue and different distribution of water molecules in the active site, which have important implications in the function of the enzyme. Although only a single reaction path involving three bridging water molecules was explored for the proton transport, it is likely that the electrostatic treatment has a major impact on the computed energetics, especially the free energy of the process.

In the cutoff simulations, QM/MM interactions have no cutoff while MM/MM interactions are truncated (e.g., with a force-shift scheme here) beyond a given radius. This causes the imbalance of QM/MM and MM/MM interactions, which overpolarizes the system so that the active site is flooded with additional water molecules; the orientation of certain functionally important residues (e.g., protonated His 64 in the COHH state) may also be affected. Therefore, we strongly suggest that truncation-based cutoff schemes should

be avoided in QM/MM simulations. In stochastic boundary simulations with extended electrostatics, the imbalance between QM/MM and MM/MM interactions is avoided although the bulk solvation and contribution from the macromolecule outside of the stochastic region are neglected. This may also cause a significant deviation of the simulation from reality.

In conclusion, the GSBP-based QM/MM approach represents a significant improvement over the popular stochastic boundary condition and is computationally more effective than methods based on the periodic boundary condition, especially for very large macromolecules. The advantage of GSBP is particularly important for simulations that involve extensive conformational samplings, such as free-energy calculations, because configuration distribution may depend sensitively on the treatment of electrostatics. The development of this protocol has opened up the exciting possibility of theoretically studying chemical events in complex biomolecular systems such as peptide synthesis in the ribosome, proton pumping in membrane proteins, and transcription in RNA polymerase, which are being pursued in our group. There is, however, plenty of space for further developments. For example, a uniform dielectric constant was applied to macromolecule atoms in the outer region, which may not be sufficiently accurate for highly charged systems such as nucleic acids. More sophisticated dielectric-response models<sup>90</sup> may be considered, which can be particularly important for high-salt conditions. In addition, as mentioned in the example of solvated imidazole, the reaction-field term diverges as the inner atoms approach the dielectric boundary; this is a well-known phenomena and does not cause a serious concern as long as the region of interest is sufficiently far from the boundary. Nevertheless, it would be of interest to explore multipolar expansions around multiple centers or boundary element formulations<sup>13</sup> to enhance the convergence of the calculation with respect to the size of the basis set. When the region of interest is sufficiently close to the bulk water, effects other than electrostatics should also be treated carefully.<sup>42</sup> Finally, in addition to all those technical developments, establishing stringent benchmark systems to quantitatively evaluate and adjust the accuracy of the boundary potential methods is of paramount importance.<sup>32</sup> In this regard, comparing  $pK_a$ 's, redox potentials as well as various spectra to experimental measurements will be very useful. Results from studies along this line will be reported in the near future.<sup>62</sup>

## ACKNOWLEDGMENTS

P.S. acknowledges a KTI fellowship from the NSF (KTI-0139335) and would like to thank P. Koenig for helpful discussions related to the current work. Q.C. acknowledges discussions with Professor B. Roux and Professor D. York on boundary condition methods. This work was partially supported by grants from the National Science Foundation (MCB-0314327, CHE-CAREER-0348649) and the Research Corporation. QC is a Alfred P. Sloan Research Fellow.

- <sup>1</sup> A. Fersht, *Enzyme Structure, Mechanism and Protein Foldings* 3rd ed. (Freeman, San Francisco, 1998).
- <sup>2</sup> C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Advances in Chemical Physics Vol. LXXI (Wiley, New York, 1988).
- <sup>3</sup> B. Honig and A. Nicholls, *Science* **268**, 1144 (1995).
- <sup>4</sup> A. Warshel and A. Papazyan, *Curr. Opin. Struct. Biol.* **8**, 211 (1998).
- <sup>5</sup> A. Warshel and S. T. Russell, *Q. Rev. Biophys.* **17**, 283 (1984).
- <sup>6</sup> A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.* **32**, 425 (2003).
- <sup>7</sup> M. E. Davis and J. A. McCammon, *Chem. Rev. (Washington, D.C.)* **90**, 509 (1990).
- <sup>8</sup> C. Sagui and T. A. Darden, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 155 (1999).
- <sup>9</sup> V. Luzhkov and A. Warshel, *J. Comput. Chem.* **13**, 199 (1992).
- <sup>10</sup> J. Florian and A. Warshel, *J. Phys. Chem. B* **101**, 5583 (1997).
- <sup>11</sup> J. Gao and C. Alhambra, *J. Chem. Phys.* **107**, 1212 (1997).
- <sup>12</sup> K. Nam, J. Gao, and D. M. York, *J. Chem. Theory Comput.* **1**, 2 (2005).
- <sup>13</sup> B. A. Gregersen and D. M. York, *J. Phys. Chem. B* **109**, 536 (2005).
- <sup>14</sup> G. Li, X. Zhang, and Q. Cui, *J. Phys. Chem. B* **107**, 8643 (2003).
- <sup>15</sup> A. R. Dinner, X. Lopez, and M. Karplus, *Theor. Chem. Acc.* **109**, 118 (2003).
- <sup>16</sup> J. Gao, *Acc. Chem. Res.* **29**, 298 (1996).
- <sup>17</sup> Q. Cui and M. Karplus, *Adv. Protein Chem.* **66**, 315 (2003).
- <sup>18</sup> J. Gao and D. G. Truhlar, *Annu. Rev. Phys. Chem.* **53**, 467 (2002).
- <sup>19</sup> S. Bogusz, T. E. Cheatham III, and B. R. Brooks, *J. Chem. Phys.* **108**, 7070 (1998).
- <sup>20</sup> T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
- <sup>21</sup> S. Kuwajima and A. Warshel, *J. Chem. Phys.* **89**, 3751 (1988).
- <sup>22</sup> P. H. Hünenberger and J. A. McCammon, *Biophys. Chem.* **78**, 69 (1999).
- <sup>23</sup> P. H. Hünenberger and J. A. McCammon, *J. Chem. Phys.* **110**, 1856 (1999).
- <sup>24</sup> W. Weber, P. H. Hünenberger, and J. A. McCammon, *J. Phys. Chem. B* **104**, 3668 (2000).
- <sup>25</sup> C. Tanford and J. Kirkwood, *J. Am. Chem. Soc.* **79**, 5333 (1957).
- <sup>26</sup> J. Warwicker and H. Watson, *J. Mol. Biol.* **157**, 671 (1982).
- <sup>27</sup> J. Tomasi and M. Persico, *Chem. Rev. (Washington, D.C.)* **94**, 2027 (1994).
- <sup>28</sup> C. J. Cramer and D. G. Truhlar, *Chem. Rev. (Washington, D.C.)* **99**, 2161 (1999).
- <sup>29</sup> T. Simonson, *Rep. Prog. Phys.* **66**, 737 (2003).
- <sup>30</sup> M. Feig and C. L. Brooks, *Curr. Opin. Struct. Biol.* **14**, 217 (2004).
- <sup>31</sup> A. Jean-Charles, A. Nicholls, K. Sharp, B. Honig, A. Tempczyk, T. Hendrickson, and W. C. Still, *J. Am. Chem. Soc.* **113**, 1454 (1991).
- <sup>32</sup> C. N. Schutz and A. Warshel, *Proteins: Struct., Funct., Genet.* **44**, 400 (2001).
- <sup>33</sup> G. A. Papoian, J. Ulander, M. P. Ewastwood, Z. Luthey-Schulten, and P. G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3352 (2004).
- <sup>34</sup> G. A. Papoian, J. Ulander, and P. G. Wolynes, *J. Am. Chem. Soc.* **125**, 9170 (2003).
- <sup>35</sup> C. K. Reddy, A. Das, and B. Jayaram, *J. Mol. Biol.* **314**, 619 (2001).
- <sup>36</sup> S. Bergqvist, M. A. Williams, R. O'Brien, and J. E. Ladbury, *J. Mol. Biol.* **336**, 829 (2004).
- <sup>37</sup> A. Baerga-Ortiz, S. Bergqvist, J. G. Mandell, and E. A. Komives, *Protein Sci.* **13**, 166 (2004).
- <sup>38</sup> L. Onsager, *J. Am. Chem. Soc.* **58**, 1486 (1936).
- <sup>39</sup> M. M. Karelson and M. C. Zerner, *J. Phys. Chem.* **96**, 6949 (1992).
- <sup>40</sup> P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webber, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, D. Cohen, *J. Chem. Phys.* **105**, 1968 (1996).
- <sup>41</sup> A. Warshel, *Chem. Phys. Lett.* **55**, 454 (1978).
- <sup>42</sup> G. King and A. Warshel, *J. Chem. Phys.* **91**, 3647 (1989).
- <sup>43</sup> S. A. Adelman, *J. Chem. Phys.* **73**, 3145 (1980).
- <sup>44</sup> S. A. Adelman and C. L. Brooks III, *J. Phys. Chem.* **86**, 1511 (1982).
- <sup>45</sup> A. Warshel, *J. Phys. Chem.* **86**, 2218 (1982).
- <sup>46</sup> M. Berkowitz and J. A. McCammon, *Chem. Phys. Lett.* **90**, 215 (1982).
- <sup>47</sup> C. L. Brooks III and M. Karplus, *J. Chem. Phys.* **79**, 6312 (1983).
- <sup>48</sup> D. Borgis, N. Levy, and M. Marchi, *J. Chem. Phys.* **119**, 3516 (2003).
- <sup>49</sup> D. Beglov and B. Roux, *J. Chem. Phys.* **100**, 9050 (1994).
- <sup>50</sup> J. G. Kirkwood, *J. Chem. Phys.* **2**, 351 (1934).
- <sup>51</sup> A. Masunov and T. Lazaridis, *J. Am. Chem. Soc.* **125**, 1722 (2003).
- <sup>52</sup> W. Im, S. Bernèche, and B. Roux, *J. Chem. Phys.* **114**, 2924 (2001).
- <sup>53</sup> M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* **58**, 7260 (1998).
- <sup>54</sup> Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, *J. Phys. Chem. B* **105**, 569 (2001).
- <sup>55</sup> B. Brooks, R. Brucoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- <sup>56</sup> D. A. McQuarrie, *Statistical Mechanics* (Harper & Row, New York, 1973).
- <sup>57</sup> M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700 (1990).
- <sup>58</sup> M. Gilson, K. Sharp, and B. Honig, *J. Comput. Chem.* **9**, 327 (1988).
- <sup>59</sup> H. L. Friedman, *Mol. Phys.* **29**, 1533 (1975).
- <sup>60</sup> J. A. C. Rullmann and P. T. van Duijnen, *Mol. Phys.* **61**, 293 (1987).
- <sup>61</sup> W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- <sup>62</sup> D. Riccardi, P. S. Schaefer, and Q. Cui, *J. Phys. Chem. B* (in press).
- <sup>63</sup> L. Backman, *Eur. J. Biochem.* **120**, 257 (1981).
- <sup>64</sup> D. W. Christianson and C. A. Fierke, *Acc. Chem. Res.* **29**, 331 (1996).
- <sup>65</sup> C. Tu, D. N. Silverman, C. Forsman, B.-H. Jonsson, and S. Lindskog, *Biochemistry* **28**, 7913 (1989).
- <sup>66</sup> A. E. Eriksson, T. A. Jones, and A. Liljas, *Proteins: Struct., Funct., Genet.* **4**, 274 (1988).
- <sup>67</sup> B. R. Brooks, R. E. Brucoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- <sup>68</sup> R. H. Stote, D. J. States, and M. Karplus, *J. Chim. Phys. Phys.-Chim. Biol.* **88**, 2419 (1991).
- <sup>69</sup> S. Toba, G. Colombo, and K. M. J. Merz, *J. Am. Chem. Soc.* **121**, 2290 (1999).
- <sup>70</sup> D. Lu and G. A. Voth, *Proteins: Struct., Funct., Genet.* **33**, 119 (1998).
- <sup>71</sup> S. Braun-Sand, M. Strajbl, and A. Warshel, *Biophys. J.* **87**, 2221 (2004).
- <sup>72</sup> C. N. Schutz and A. Warshel, *J. Phys. Chem. B* **108**, 2066 (2004).
- <sup>73</sup> K. Häkansson, M. Carlsson, L. A. Svensson, and A. Liljas, *J. Mol. Biol.* **227**, 1192 (1992).
- <sup>74</sup> M. Nina, D. Beglov, and B. Roux, *J. Phys. Chem. B* **101**, 5239 (1997).
- <sup>75</sup> C. L. Brooks III and M. Karplus, *J. Am. Chem. Soc.* **79**, 6312 (1983).
- <sup>76</sup> J. P. Ryckaert, G. Ciccotti, and H. J. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- <sup>77</sup> A. D. MacKerell Jr., D. Bashford, M. Bellott *et al.*, *J. Phys. Chem. B* **102**, 3586 (1998).
- <sup>78</sup> T. Simonson, G. Archontis, and M. Karplus, *J. Phys. Chem. B* **101**, 8349 (1997).
- <sup>79</sup> C. Forsman, G. Behravan, B.-H. Jonsson, Z. Liang, S. Lindskog, X. Ren, J. Sandström, and K. Wallgren, *FEBS Lett.* **229**, 360 (1988).
- <sup>80</sup> C. K. Tu and D. N. Silverman, *Biochemistry* **24**, 5881 (1985).
- <sup>81</sup> D. N. Silverman, C. K. Tu, S. Lindskog, and G. C. Wynns, *J. Am. Chem. Soc.* **101**, 6734 (1979).
- <sup>82</sup> J. Aqvist and A. Warshel, *J. Mol. Biol.* **224**, 7 (1992).
- <sup>83</sup> D. N. Silverman, *Methods Enzymol.* **249**, 479 (1995).
- <sup>84</sup> D. Riccardi, G. Li, and Q. Cui, *J. Phys. Chem. B* **108**, 6467 (2004).
- <sup>85</sup> Y. Zhang, T. Lee, and W. Yang, *J. Chem. Phys.* **110**, 46 (1999).
- <sup>86</sup> A. Altun and W. Thiel, *J. Phys. Chem. B* **109**, 1268 (2005).
- <sup>87</sup> H. Y. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hermans, and W. T. Yang, *Proteins: Struct., Funct., Genet.* **44**, 484 (2001).
- <sup>88</sup> P. H. Konig, M. Hoffmann, T. Frauenheim, and Q. Cui, *J. Phys. Chem. B* **109**, 9082 (2005).
- <sup>89</sup> G. Li and Q. Cui, *J. Phys. Chem. B* **107**, 14521 (2003).
- <sup>90</sup> X. Y. Song, *J. Chem. Phys.* **116**, 9359 (2002).