Combining Self-Consistent-Charge Density-Functional Tight-Binding (SCC-DFTB) with Molecular Mechanics by the Generalized Hybrid Orbital (GHO) Method

Jingzhi Pu, Jiali Gao,* and Donald G. Truhlar*

Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant Street Southeast, Minneapolis, Minnesota 55455-0431

Received: February 2, 2004; In Final Form: March 30, 2004

In combined quantum mechanical and molecular mechanical (QM/MM) calculations with the QM/MM boundary at a covalent bond, the generalized hybrid orbital (GHO) method has been shown to provide a well balanced and stable connection between the QM and MM regions. The GHO method has previously been developed for semiempirical molecular orbital methods based on neglect of diatomic differential overlap (GHO-NDDO) and for the ab initio Hartree−Fock level (GHO-AIHF). In the present work, we formulate the GHO algorithm and its analytical gradients for treating the QM subsystem by the self-consistent-charge density-functional tight-binding (SCC-DFTB) theory. To obtain a good description of the bond length at the QM/MM boundary, a parametrized empirical correction term involving the GHO boundary atom and its QM frontier neighbor is added. Geometries and Mulliken charges obtained from GHO-SCC-DFTB calculations are compared to the fully QM results for a set of 18 molecules and ions with various functional groups close to the boundary, and we verified that we reproduced the full C−C stretch potential in ethane and in propanoate. The torsion barrier of *n*-butane around the central C−C bond is studied with the GHO boundary atom placed at different locations. Finally, the energetics of the method are tested for the proton affinities of a series of 15 alcohols, amines, thiols, and acids. The results indicate that the GHO treatment for combining SCC-DFTB with molecular mechanics is both theoretically robust and satisfactory for practical use. In Supporting Information we present parameters for boundaries that cut through O−C and S−C bonds.

1. Introduction

Combined quantum mechanical and molecular mechanical (QM/MM) methods are powerful tools for studying chemical reactivity in large systems.[1−4] Although various schemes have been proposed and widely applied, they differ considerably in the treatment of the QM/MM boundary when it runs through a covalent bond. Under such circumstances, special care must be taken to saturate the free valence caused by cutting the QM frontier bond.

A number of methods have been proposed to truncate the wave function for the QM fragment at the boundary region.[1−43] One popular way is adding an additional atom (usually a hydrogen atom is used) to cap the QM fragment.[5−8] Although this so-called "link atom" approach is widely used, the QM frontier bond (the bond formed between the link atom and its QM frontier neighbor) may experience unphysical polarization if the electrostatic interactions involving the link atom are not properly handled.[9−12] Recently, delocalized Gaussian functions have been used to correct the strong polarization near the QM/MM boundary region.[13,14] A more fundamental approach is to use effective potentials,[15,16] and a third, even more justifiable family of methods is based on the self-consistent-field approximation, such as the local self-consistent field (LSCF) algorithm[17,18] proposed by Rivail and co-workers. In the LSCF method, the frontier bond connecting the QM and MM fragments is represented by a set of strictly localized bond orbitals (SLBOs). Because these bond orbitals are strictly localized, the authors assumed that they are transferable from one system to another. Therefore, one can construct the bond orbitals from calculations on model compounds calculations. The self-consistent-field (SCF) optimization of the QM wave functions is carried out over an active subspace that excludes SLBOs. The LSCF method was originally[17,18] formulated for semiempirical wave functions with the neglect of diatomic differential overlap[44] (NDDO) approximation; then it was generalized to Hartree−Fock (HF) and post-HF wave functions and to density functional theory (DFT), and analytical gradients were formulated.[19,20] Recently, specific force field parameters have also been developed for LSCF.[21] A disadvantage of the LSCF approach is that the bond orbital must be reconstructed for systems with different bonding situations. A similar approach called the frozen orbital method has been proposed by Friesner and co-workers, where a library of frozen densities representing the frozen orbitals was parametrized for side chains of amino acids.[22−24] An important feature of the QM/MM methods mentioned so far is that they all include the electrostatic effect of the MM subsystem on the QM subsystem; other methods[29−31] (IMOMM) circumvent some of the difficulties of treating this polarization effect by neglecting it, but that can be a very serious approximation.[14,45]

The present article is concerned with the generalized hybrid orbital (GHO) method,[25−28] which is similar to the LSCF technique. In the GHO treatment, the QM/MM boundary atom is represented by a set of sp^3 hybrid orbitals, where one of these hybrid orbitals participates in the SCF procedure to optimize the QM wave function, and the other three hybrid orbitals are frozen. The major improvement of the GHO method compared to the LSCF method is that the construction of the bond orbitals results from a hybridization scheme that is completely determined by the local geometry of the QM/MM boundary region,

10.1021/jp049529z CCC: $27.50
Published on Web 06/02/2004

and therefore the model calculations for parametrizing the bond orbital in each new situation are avoided. The GHO method was originally[25−27] proposed for NDDO wave functions, in which form it has been used successfully in recent enzyme dynamics studies.[3,46−53] Recently, the GHO formalism was extended to the ab initio Hartree−Fock (AIHF) level,[28] including unrestricted Hartree−Fock (UHF). The key step in the GHO-AIHF extension is the satisfaction of orthogonality constraints involving auxiliary hybrid basis functions, which were not needed when the GHO algorithm was applied with the NDDO approximation. The extension of the GHO method to ab initio HF wave functions established a solid theoretical framework for formulating the GHO method as a fundamental quantum mechanical approach to quantal−classical boundary problems. However, AIHF theory is generally inaccurate due to the lack of electron correlation. Therefore it is desirable to treat the QM subsystem more accurately, and this may be accomplished by developing a method for hybrid density functional theory[54] and for post-Hartree−Fock correlated methods[55] based on a Hartree−Fock reference.

An alternative strategy is to try to improve the accuracy while retaining the low cost of NDDO methods. The recently developed self-consistent-charge density-functional tight-binding (SCC-DFTB) method is particularly encouraging for its efficiency and transferability to various scales of systems.[56−66] Starting from a second-order expansion of the Kohn−Sham total energy with respect to density fluctuations, the total energy is self-consistently minimized to incorporate the relaxation of the initial charge density at the Mulliken population level.[56,57] The SCC-DFTB method has been shown to provide reaction energies, geometries, and vibrational frequencies for a set of small organic molecules that are of comparable accuracy to full DFT results.[56] Parameters for sulfur[58] and zinc[59] have been developed recently. With corrections for long-range dispersion forces incorporated, the method can be applied to study hydrogen bonding and stacking interactions in biological systems.[60−63] This method has been implemented in the CHARMM program[67] for link atom QM/MM calculations.[64−66] A generalized-Born-plus-atomic-surface-tension solvation model[68] and a class IV charge model[69] have also been developed on the basis of the SCC-DFTB method.

With the increasing number of applications of this method, it will be useful to have a more fundamental QM/MM boundary treatment, and we provide such a treatment in the present paper. In particular, we formulate the GHO algorithm for combined QM/MM calculations in which the QM part is described by the SCC-DFTB theory. The presentation of the method is organized as follows. Following a brief review of the theory, the key modifications of the formulas for combining SCC-DFTB and GHO as well as a set of practical procedures for implementation are given in section 2. Section 3 summarizes the determination of parameters for the case in which the boundary atom is an sp^3 methylene carbon. In section 4, we present and discuss tests of the proposed GHO-SCC-DFTB method. Section 5 contains concluding remarks.

2. Theory

In this section, we describe the theoretical background for combining the SCC-DFTB method with molecular mechanics using the GHO method. The presentation is organized as follows. First we briefly review the SCC-DFTB theory and the GHO method. Next we describe our GHO implementation on the basis of the SCC-DFTB framework and derive the analytical gradient expression that takes account of the GHO basis transformation. At the end of this section, the requirement of orbital orthogonalization is discussed.

2.1. SCC-DFTB. The method was derived from a second-order expansion of the Kohn−Sham energy with respect to a charge density fluctuation $\delta\rho$ from a given reference density ρ_0:[56,57,70]

$$E = \sum_i^{\mathrm{occ}} n_i \langle \phi_i | H^0 | \phi_i \rangle + \frac{1}{2}\int\int \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \left.\frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho\delta\rho'}\right|_{\rho 0} \right) \delta\rho(\mathbf{r})\, \delta\rho'(\mathbf{r}')\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' + \left\{ -\frac{1}{2}\int\int \frac{\rho_0(\mathbf{r}')\, \rho_0(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' + E_{\mathrm{xc}}[\rho_0(\mathbf{r})] - \int V_{\mathrm{xc}}[\rho_0]\, \rho_0(\mathbf{r})\, \mathrm{d}\mathbf{r} \right\} + E_{\mathrm{core}} \quad (1)$$

where n_i is the occupation number of the Kohn−Sham orbital ϕ_i, E_{core} is the effective core−core repulsion energy, and the first term is the expectation value of the Kohn−Sham Hamiltonian based on the reference density ρ_0:

$$H^0 = -\frac{1}{2}\nabla^2 - \sum_\alpha \frac{Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{r}|} + \int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, \mathrm{d}\mathbf{r}' + V_{\mathrm{xc}}[\rho_0(\mathbf{r})] \quad (2)$$

where $\mathbf{R}_\alpha$ denotes the atomic coordinates of the atom α. In these equations, V_{xc} and E_{xc} are the exchange-correlation potential and exchange-correlation energy, respectively. The Kohn−Sham orbitals ϕ_i in eq 1 are expanded in a minimal basis set of localized pseudoatomic Slater orbitals with a frozen core,

$$\phi_i = \sum_\mu c_\mu^i \chi_\mu \quad (3)$$

where the pseudoatomic orbitals χ_μ are obtained by solving an atomic Kohn−Sham equation for neutral atoms. Under the two-center approximation, the zero-order Hamiltonian matrix elements are expressed as

$$H_{\mu\nu}{}^0 = \begin{cases} \epsilon_\mu & \mu = \nu \\ \langle \chi_\mu | T + V_{\mathrm{eff}}[\rho_0^\alpha + \rho_0^\beta] | \chi_\nu \rangle & \mu \in \alpha,\ \nu \in \beta \\ 0 & \mu,\nu \in \alpha, \text{ and } \mu \neq \nu \end{cases} \quad (4)$$

where ϵ_μ are the orbital energies obtained from solving the Kohn−Sham equations for neutral atoms, the effective one-electron operator V_{eff} incorporates both the Columbic and the exchange-correlation contribution in eq 2, and ρ_0^α is the reference density on atom α. The second term in eq 1 takes into account the second-order energy due to the charge fluctuation. With a multipole expansion of the density fluctuation truncated at the monopole term, the second-order energy can be approximated by[56]

$$E^{2\mathrm{nd}} = \frac{1}{2}\int\int \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \left.\frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho\delta\rho'}\right|_{\rho_0} \right) \delta\rho(\mathbf{r})\, \delta\rho'(\mathbf{r}')\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' = \frac{1}{2}\sum_{\alpha,\beta} \gamma_{\alpha\beta} q_\alpha q_\beta \quad (5)$$

where q_α is the partial charge of atom α obtained from Mulliken population analysis,[71] $\gamma_{\alpha\alpha}$ is related to the chemical hardness

of α, and $\gamma_{\alpha\beta}$ is an effective Coulomb operator that tends to $R_{\alpha\beta}^{-1}$ at large interatomic distances, where $R_{\alpha\beta}$ is $|\mathbf{R}_\alpha - \mathbf{R}_\beta|$. The sum of the third and fourth terms in eq 1 is assumed to be short ranged and pairwise; their sum is called the repulsive energy E_{rep}. With these approximations, the SCC-DFTB total energy is finally written as[56]

$$E^{\text{SCC-DFTB}} = E_{\text{bs}} + E_{\text{rep}} \tag{6a}$$

where the first term is the band structure energy defined by

$$E_{\text{bs}} = \sum_i^{occ} n_i \langle \phi_i | H^0 | \phi_i \rangle + \frac{1}{2}\sum_{\alpha,\beta} \gamma_{\alpha\beta} q_\alpha q_\beta \tag{6b}$$

The repulsive energy for each pair of atoms in the system is obtained from the difference of E_{bs} and the DFT energy for diatomic species or small model systems. Because the second term in E_{bs} depends on the Mulliken charges, which in turn depend on the coefficients in eq 3, the application of the variational principle to eq 6a leads to a secular equation that must be solved self-consistently:

$$\mathbf{HC} = \mathbf{SC}\epsilon \tag{7a}$$

where

$$\mathbf{H} = \mathbf{H}^0 + \mathbf{H}^1 \tag{7b}$$

$$C_{\nu i} = c_\nu^{\,i} \tag{7c}$$

$$S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle \tag{7d}$$

and

$$H^1_{\mu\nu} = \frac{1}{2} S_{\mu\nu} \sum_\xi (\gamma_{\alpha_\mu \xi} + \gamma_{\alpha_\nu \xi}) q_\xi \tag{7e}$$

where α_μ denotes the atom on which basis function μ is centered and ϵ is a diagonal matrix of orbital energies.

The gradient of the SCC-DFTB energy is

$$\frac{\partial E^{\text{SCC-DFTB}}}{\partial \mathbf{R}_\alpha} = \sum_i n_i \sum_{\mu\nu} c_\mu^i c_\nu^i \left[\frac{\partial H^0_{\mu\nu}}{\partial \mathbf{R}_\alpha} - \left(\epsilon_i - \frac{1}{2}\sum_\xi (\gamma_{\alpha\mu\xi} + \gamma_{\alpha\nu\xi}) q_\xi \right) \frac{\partial S_{\mu\nu}}{\partial \mathbf{R}_\alpha} \right] + q_\alpha \sum_\xi \frac{\partial \gamma_{\alpha\xi}}{\partial \mathbf{R}_\alpha} q_\xi + \frac{\partial E_{\text{rep}}}{\partial \mathbf{R}_\alpha} \tag{8}$$

For the QM/MM implementation of the SCC-DFTB method in CHARMM, the interaction of the MM point charges with the QM fragment is approximated by the electrostatic interaction between MM point charges and Mulliken charges of the QM atoms:[64]

$$H^{\text{el}}_{\text{QM/MM}} = \sum_{\alpha \in \text{QM}} \sum_{\beta \in \text{MM}} \frac{Q_\alpha q_\beta}{R_{\alpha\beta}} \tag{9}$$

and the gradient of this term must be added to eq 8.

2.2. GHO-SCC-DFTB/MM. The GHO boundary carbon (denoted by B) is taken in this work to be an sp^3 carbon. Note that although the GHO boundary B can in principle be in other hybridization states or even be any other type of atom, the present choice of the QM/MM boundary is sufficient for most situations in practical studies for enzymes. The QM atom directly bonded to the GHO boundary atom is called the frontier atom, and it is denoted as A. Having defined the A (QM frontier) and B (GHO boundary) atoms, we can partition the entire system into a QM subsystem and an MM subsystem (Figure 1). We note that although the GHO boundary atom B is included in the QM subsystem, it is actually both a QM atom and an MM atom, as we will discuss below. The QM atoms other than B in the QM subsystem are fully treated by quantum mechanics and are defined as "fully QM atoms" subsequently. We will denote fully QM atoms as Q throughout the paper. Because we select an sp^3 hybridized carbon as the GHO boundary B, there are three molecular mechanics atoms bonded to atom B. In Figure 1, these MM neighbors of atom B are denoted by X, Y, and Z.

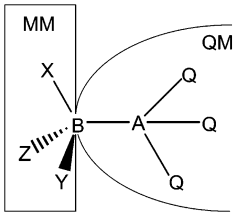

Figure 1. Schematic representation of the QM/MM partition in the GHO method.

In the SCC-DFTB theory, a minimum set of pseudoatomic orbitals is used for the basis set. For a GHO-SCC-DFTB/MM treatment, only the atoms in the fully QM subsystem will be represented by these basis functions; the basis functions on the GHO boundary atom are treated differently. The basis functions on fully QM atoms are denoted by χ_μ, with $\mu = 1, 2, ..., N$, where N denotes the number of such basis functions. For the GHO boundary atom B, one can construct a set of hybrid orbitals $\{\eta_B, \eta_X, \eta_Y, \eta_Z\}$ based on the hybridization of the atomic s and p basis functions on B. Note that SCC-DFTB uses a minimum basis set; therefore only a single set of valence s and p orbitals must to be considered in the above hybridization for a carbon boundary atom. The orbital η_B is called the active hybrid orbital, and it points toward A. The other three hybrid orbitals $\{\eta_X, \eta_Y, \eta_Z\}$ are called auxiliary hybrid orbitals. The hybridization scheme is described by a basis transformation matrix $\mathbf{T}_b$, defined in detail previously.[26] One can construct the hybrid orbitals from the s and p atomic orbtials on B by applying the basis transformation:

$$\begin{pmatrix} \eta_B \\ \eta_x \\ \eta_y \\ \eta_z \end{pmatrix} = \mathbf{T}_b^{\dagger} \begin{pmatrix} s \\ p_X \\ p_Y \\ p_Z \end{pmatrix} \tag{10}$$

By construction,[24] the hybrid orbitals $\{\eta_B, \eta_X, \eta_Y, \eta_Z\}$ form an orthonormal set:

$$S_{cd} = \langle \eta_c | \eta_d \rangle = \delta_{cd} \qquad (c, d = B, X, Y, Z) \tag{11}$$

The matrix $\mathbf{T}$ that transforms the atomic orbital (AO) basis to the hybrid (H) basis over the entire QM subsystem is

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_b \end{pmatrix} \tag{12}$$

where $\mathbf{I}_N$ is an N dimensional unit matrix, and $\mathbf{T}_b$ is the 4×4 hybridization matrix that is used to construct the local hybrid orbitals on GHO boundary B as in eq 10. By this convention, we have placed the boundary atom at the end of the QM atom list.

The basis functions χ_μ on fully QM atoms plus the active hybrid basis function η_B form the $(N+1)$-dimensional active basis space for the SCF calculation; functions in this active space are denoted by χ_a. The $N+1$ occupied and virtual Kohn−Sham orbitals (KSO) resulting from solving the secular equation of the SCC-DFTB theory are linear combinations of these active basis functions:

$$\phi_i = \sum_{\mu=1}^{N} c_{\mu i}\chi_\mu + c_{Bi}\eta_B = \sum_{a=1}^{N+1} c_{ai}\chi_a \qquad (i = 1, 2, ..., N+1) \tag{13}$$

where $\chi_{N+1} = \eta_B$. Each of the three auxiliary basis functions forms an auxiliary KSO by itself and is frozen in the SCF procedure:

$$\begin{aligned}\phi_{N+2} &= \eta_X \\ \phi_{N+3} &= \eta_Y \\ \phi_{N+4} &= \eta_Z\end{aligned} \tag{14}$$

Each frozen auxiliary KSO provides an effective charge distribution to mimic the bond formed by atom B and one of its MM neighbors X, Y, or Z. The charge density P^H_{bb} assigned to each auxiliary orbital is chosen to be 1- $q_B/3$ where q_B is the MM point charge of B. This represents a uniform partition of the MM point charge on B over the auxiliary orbitals. In particular, q_B equals −0.27 and −0.18 for a methyl carbon and a methylene carbon, respectively, in the CHARMM force field. Thus the auxiliary orbital occupancy for these two key cases is 1.09 and 1.06, respectively. As mentioned above, the GHO boundary atom B exists in two regions in the GHO method. As a QM atom, its orbitals contribute to the construction of QM wave functions. As an MM atom, the fractional densities in frozen orbitals on B are determined by its MM point charge, and the chemical bonding with X, Y, and Z is modeled by the MM force field that is used. Therefore, atom B is both a QM and an MM atom in the GHO treatment.

The total energy is the sum of the QM energy, the MM energy, and the QM/MM interaction energy:

$$E^{tot} = E_{QM} + E_{MM} + E_{QM/MM} \tag{15}$$

where E_{QM} and E_{MM} denote the energies of the QM and MM subsystems, respectively, and $E_{QM/MM}$ is the interaction energy. For E_{MM}, the contributions of terms involving only QM atoms (fully QM atoms Q and the boundary atom B) are removed. Note that the electrons in GHO auxiliary orbitals experience the MM charges and therefore make a contribution in the QM/MM interaction energy. In addition, $E_{QM/MM}$ contains the nuclear−nuclear repulsion between the QM nuclei and MM charges, and the nonbonded van der Waals interactions between QM and MM atoms.

As stated above, the GHO-SCC-DFTB wave functions are optimized over an active QM basis set space with reduced dimensionality and the frozen auxiliary hybrid orbitals only provide electron fields for the active QM orbitals. To enforce the constraint, the self-consistent-field (SCF) procedure must be modified. A practical way to do this is as follows.

i. Form the basis transformation matrix **T** from the AO basis to the hybrid (H) basis according to eqs 10 and 12.

ii. Transform the overlap matrix **S** from the AO basis to the H basis:

$$\mathbf{S}^H_{N+4} = \mathbf{T}^\dagger \mathbf{S}^{AO}_{N+4} \mathbf{T}$$

iii. Drop columns and rows corresponding to the auxiliary orbitals to obtain the reduced overlap matrix for the active orbitals: $\mathbf{S}^H_{N+1}$

$$\mathbf{S}^H_{N+4} \xrightarrow{\text{drop auxiliary}} \mathbf{S}^H_{N+1}$$

iv. Form the total Hamiltonian matrix in the AO basis: $\mathbf{H}^{AO}_{N+4}$, which is a sum of the charge-independent part $(\mathbf{H}^0)^{AO}_{N+4}$ and the charge-dependent part $(\mathbf{H}^1)^{AO}_{N+4}$. (Note that for the first iteration, the Hamiltonian matrix only includes the charge-independent part.)

v. Transform the total Hamiltonian matrix from the AO basis to the H basis, and obtain the auxiliary orbtial energies:

$$\mathbf{H}^H_{N+4} = \mathbf{T}^\dagger \mathbf{H}^{AO}_{N+4} \mathbf{T}$$

vi. Drop columns and rows corresponding to auxiliary orbitals to obtain the reduced Hamiltonian matrix for the active orbitals: $\mathbf{H}^H_{N+1}$

$$\mathbf{H}^H_{N+4} \xrightarrow{\text{drop auxiliary}} \mathbf{H}^H_{N+1}$$

vii. Solve the secular equation in the space of the active hybrid basis, to obtain the active KSOs in the hybrid basis $\mathbf{C}^H_{N+1}$:

$$\mathbf{H}^H_{N+1}\,\mathbf{C}^H_{N+1} = \mathbf{S}^H_{N+1}\,\mathbf{C}^H_{N+1}\epsilon$$

viii. Append the auxiliary orbitals to form the total KSO matrix.

$$\mathbf{C}^H_{N+1} \xrightarrow{\text{append auxiliary}} \mathbf{C}^H_{N+4}$$

ix. Transform the KSO coefficient matrix from the H basis to the AO basis.

$$\mathbf{C}^H_{N+4} = \mathbf{T}\mathbf{C}^{AO}_{N+4}$$

x. Carry out a Mulliken population analysis to obtain partial charges q_α, and form the new Hamiltonian matrix $\mathbf{H}^{AO}_{N+4}$.

xi. Calculate the band structure energy by summing over orbital energies, and check convergence. If not converged, go back to step iv.

xii. After the SCF is converged, evaluate the GHO-SCC-DFTB total energy. For a parametrized version (see Section 3), add the pairwise empirical correction (E_{cor}) between a GHO boundary atom B and its QM frontier neighbor A.

2.3. GHO-SCC-DFTB Gradients. The basis transformation matrix contributes a nonvanishing force on the GHO boundary atom B and on MM atoms bonded to B (denoted X, Y, and Z). This introduces additional terms into the GHO-SCC-DFTB gradients. The GHO-SCC-DFTB analytical gradient can be derived by starting with the same expression used for GHO-AIHF:[28]

$$\frac{\partial E^{GHO-SCC-DFTB}}{\partial \mathbf{R}_\alpha} = \frac{\partial E^{SCC-DFTB}}{\partial \mathbf{R}_\alpha} + \sum_{\mu\nu}^{N+4} \frac{\partial P^{AO}_{\mu\nu}}{\partial \mathbf{R}_\alpha} H^{AO}_{\mu\nu} - \sum_{\mu\nu}^{N+4} \frac{\partial W^{AO}_{\mu\nu}}{\partial \mathbf{R}_\alpha} S^{AO}_{\mu\nu} \tag{16}$$

where the **P** and **W** are the density matrix and the energy-weighted density matrix, respectively, in the AO basis, **H** and **S** are the total Hamiltonian matrix and overlap matrix in SCC-

DFTB, and $\mathbf{R}_\alpha$ denotes the nuclear coordinates of atoms B, X, Y, and Z. The derivative of the density matrix and energy-weighted density matrix can be further expanded as

$$\frac{\partial \mathbf{P}^{\mathrm{AO}}}{\partial \mathbf{R}_\alpha} = \frac{\partial (\mathbf{T}\mathbf{P}^{\mathrm{H}}\mathbf{T}^\dagger)}{\partial \mathbf{R}_\alpha} = \frac{\partial \mathbf{T}}{\partial \mathbf{R}_\alpha}\mathbf{P}^{\mathrm{H}}\mathbf{T}^\dagger + \mathbf{T}\mathbf{P}^{\mathrm{H}}\frac{\partial \mathbf{T}^\dagger}{\partial \mathbf{R}_\alpha} + \mathbf{T}\frac{\partial \mathbf{P}^{\mathrm{H}}}{\partial \mathbf{R}_\alpha}\mathbf{T}^\dagger \quad (17)$$

$$\frac{\partial \mathbf{W}^{\mathrm{AO}}}{\partial \mathbf{R}_\alpha} = \frac{\partial (\mathbf{T}\mathbf{W}^{\mathrm{H}}\mathbf{T}^\dagger)}{\partial \mathbf{R}_\alpha} = \frac{\partial \mathbf{T}}{\partial \mathbf{R}_\alpha}\mathbf{W}^{\mathrm{H}}\mathbf{T}^\dagger + \mathbf{T}\mathbf{W}^{\mathrm{H}}\frac{\partial \mathbf{T}^\dagger}{\partial \mathbf{R}_\alpha} + \mathbf{T}\frac{\partial \mathbf{W}^{\mathrm{H}}}{\partial \mathbf{R}_\alpha}\mathbf{T}^\dagger \quad (18)$$

where the last terms are not included in the gradient terms because the corresponding density in the hybrid basis is truly variational, so they are zeroes.[72]

To evaluate the gradient analytically according to eq 16, the density matrix and energy-weighted density matrix $\mathbf{W}^{\mathrm{AO}}$ are required. The density matrix elements $P^{\mathrm{AO}}_{\mu\nu}$ are defined as

$$P^{\mathrm{AO}}_{\mu\nu} = \sum_{i=1}^{N+4} n_i c_{\mu i} c_{\nu i} \quad (19)$$

where $c_{\mu i}$ and $c_{\nu i}$ are the orbital coefficients for atomic basis functions μ and ν in KSO i; the MOs with indexes of $i = N + 1$, $N + 2$, $N + 3$ correspond to the auxiliary KSOs constructed from each of the auxiliary hybrid basis functions. Note that these auxiliary KSOs are fractionally occupied with an occupation number n_i ($i = N + 1$, $N + 2$, $N + 3$) of $1 - q_{\mathrm{B}}/3.0$, where q_{B} is the MM point charge on the GHO boundary B. Similarly, the auxiliary KSOs has been included in formulating the energy weighted density matrix element $W^{\mathrm{AO}}_{\mu\nu}$:

$$W^{\mathrm{AO}}_{\mu\nu} = \sum_{i=1}^{N+4} n_i \epsilon_i c_{\mu i} c_{\nu i} \quad (20)$$

The orbital energies for the $N + 1$ active KSOs can be obtained in the conventional way[56] by solving the secular equation (eq 7a) in the active space. The auxiliary orbital energies are not directly available, because the auxiliary KSOs are excluded from the active SCF space. However, we evaluate the auxiliary orbital energies explicitly as expectation values of the GHO-SCC-DFTB Hamiltonian operator, which yields

$$\epsilon_i = \sum_{\mu\nu=1}^{N+4} c_{\mu i} c_{\nu i} H_{u\nu} \quad (i = N + 2, N + 3, N + 4) \quad (21)$$

In particular, we evaluate this expression using the hybrid basis, in which only diagonal Hamiltonian elements survive:

$$\epsilon_i = H^{\mathrm{H}}_{bb} \qquad (i = N + 2, N + 3, N + 4;\ b = x, y, z) \quad (22)$$

The occupation number n_i for an auxiliary orbital is equal to $(1 - q_{\mathrm{B}}/3.0)$, as we stated above.

Notice that no special orthogonalization treatments have been imposed for GHO-SCC-DFTB. Because the SCC-DFTB method is intrinsically a nonorthogonal tight-binding theory, the overlap matrix **S** for solving the secular equation (eq 7a) and carrying out the population analysis is not a unit matrix. Therefore, the auxiliary hybrid basis functions $\{\eta_{\mathrm{X}}, \eta_{\mathrm{Y}}, \eta_{\mathrm{Z}}\}$ have nonvanishing overlaps with other fully QM basis functions. Note that the orthogonality between the active hybrid basis function η_{B} and the three auxiliary basis functions has been satisfied in eq 11. In the current implementation, we simply neglect the other overlap elements involving the auxiliary hybrid basis functions in the hybrid basis:

$$S^{\mathrm{H}}_{\mu b} = 0 \qquad (\mu = 1, 2, ..., N;\ b = x, y, z) \quad (23)$$

This approximation can be simply accomplished by dropping entries of the overlap matrix and Hamiltonian matrix corresponding to the auxiliary basis functions. The consequence of the approximation in eq 23 is that all Kohn−Sham orbitals including the auxiliary orbitals are mutual orthogonal in the GHO-SCC-DFTB treatment:

$$\langle \phi_i | \phi_j \rangle = \delta_{ij} \qquad (i, j = 1, 2, ..., N + 4) \quad (24)$$

Although our experience[28] showed that explicit orbital orthogonalization is indispensable for obtaining even qualitatively reasonable results at the ab initio HF level (GHO-AIHF), this does not seem to be required for GHO-SCC-DFTB. This phenomenon might be largely related to the semiempirical character of the SCC-DFTB method, in which the construction of the Hamiltonian matrix only takes account of diatomic interactions (eqs 4 and 7e).

3. Parametrization

In this section, the GHO-SCC-DFTB method is parametrized for use with the CHARMM22 force field.[73] Note that CHARMM27[74] protein parameters are essentially identical[75] to those in CHARMM22, so the current parametrization should be useful with CHARMM27 as well.

Optimization of molecular geometries by the unparametrized GHO-SCC-DFTB method underestimates the bond distance between the GHO boundary atom B and the QM frontier atom A. Similar behavior was observed for unparametrized semiempirical GHO-AM1[25] and GHO-PM3.[27] The underestimated A−B bond distances across the QM/MM boundary can be related to the difficulty of treating the boundary consistently. In GHO-AIHF, we found that the unbalance across the boundary could be remediated by introducing a set of integral scaling factors for one-electron kinetic energy integrals involving the boundary orbitals.[28] By analogy, in GHO-SCC-DFTB, the diatomic Hamiltonian matrix elements involving the boundary orbital could be adjusted to improve the results. In fact, it would be very easy to scale these matrix elements as a whole because in the present implementation of the SCC-DFTB method in CHARMM, all **H** and **S** (overlap matrix) elements are precalculated and tabulated as simple functions of interatomic distances. However, decomposition of these matrix elements into individual contributions, i.e., kinetic energies, nuclear attraction potential energies, repulsion energies related to a given density, and exchange-correlation terms, is not well defined. One could also adjust the chemical hardness parameters in γ_{BB} or the parameters in the pairwise repulsive terms involving the boundary atom. Note that E_{rep} directly affects the geometries, but it does not affect the wave functions, except indirectly through the geometry, and therefore the atomic charges are insensitive to these changes.

We found that the most serious problem with the unparametrized method is with the geometries, not the charges. Therefore, keeping the above analysis in mind, we decided that the adjustment of the A−B pairwise repulsive term is the best way to parametrize the GHO-SCC-DFTB method. We therefore added an empirical correction E_{cor} to E_{rep}:

$$E_{\mathrm{rep}} \rightarrow E_{\mathrm{rep}} + E_{\mathrm{cor}} \quad (25)$$

TABLE 1: Empirical Correction Parameters for GHO-SCC-DFTB with the QM/MM Boundary Cutting a C−C Bond

E_{cor} coefficient	C−C
c_1[a]	6.422
c_2[b]	−55.715
c_4[c]	7.040

[a] In kcal/mol/Å [b] In kcal/mol/Å^2 [c] In kcal/mol/Å^4

where E_{cor} is composed of a harmonic polynomial plus a quartic term:

$$E_{cor} = c_1 R_{AB} + c_2 {R_{AB}}^2 + c_4 {R_{AB}}^4 \quad (26)$$

where R_{AB} denotes the distance between a GHO boundary atom B and the QM frontier atom A connected to it; and coefficients c_1, c_2, and c_4 are parameters to be empirically determined. We note that the polynomial correction term in eq 26 effectively does not allow the A−B bond to dissociate. This is not a problem in QM/MM calculations because the boundary region is chosen away from the chemical process, and although the A−B bond can fluctuate from its equilibrium geometry, there should not be enough energy to break it.

The total fitness function used for determining the E_{cor} parameters is an equally weighted combination of a geometry fitness function and a charge fitness function:

$$F_{total} = -({F_{geom}}^2 + {F_{charge}}^2)^{1/2} \quad (27)$$

where the geometry and charge fitness functions are

$$F_{geom} = -\left[\frac{1}{K}\sum_m \left(\sum_r \left(\frac{r^{GHO} - r^{SCC-DFTB}}{r_0}\right)^2 + \sum_\theta \left(\frac{\theta^{GHO} - \theta^{SCC-DFTB}}{\theta_0}\right)^2\right)\right]^{1/2} \quad (28)$$

$$F_{charge} = -\left[\left(\frac{q_B^{GHO} - q_B^{MM}}{q_0}\right)_{m=1}^2\right]^{1/2} \quad (29)$$

In eq 28, the training set for geometries contains 10 molecules (m = 1,2, ..., 10); in particular, the training set molecules and the QM/MM division can be found in Table 2 where the training set molecules are listed in bold; K is the number of unique bond distances (r) and angles (θ) in the training set molecules (note that if two bond distances or angles are equal by symmetry, we include them only once); q_B^{GHO} is the Mulliken charge on the GHO boundary atom B in propane (which is the first molecule in the training set); q_B^{MM} is the MM point charge for the methylene carbon in propane (this equals −0.18 in CHARMM); and r_0 (0.02 Å), θ_0 (2°), and q_0 (0.05 atomic charge) are scale units for distances, angles, and charges, respectively. We take the reference geometry (denoted SCC-DFTB in eq 28) to be the geometry optimized at the fully QM level, i.e., SCC-DFTB. Note the GHO boundary atom B is a carbon atom in a CH_2 group and the frontier atom is also chosen as a carbon atom. The parameters are optimized by maximizing the total fitness function using a genetic algorithm.[76] Table 1 gives the optimized coefficients of eq 26 for the cases where the QM/MM boundary cuts a C−C bond.

In the optimization, although the QM frontier atom A is always a carbon atom, we did not restrict the type of hybridization on the frontier carbon. However, one may need more flexibility in choosing the QM frontier atom for some situations.

TABLE 2: Key A−B Bond Distance (Å) for Unparametrized and the Parametrized GHO-SCC-DFTB Compared to the SCC-DFTB Fully QM Results[a]

system	GHO-SCC-DFTB (unparametrized)	GHO-SCC-DFTB (with E_{cor})	SCC-DFTB
CH_3BH_2−AH_3	1.427	1.489	1.509
CH_3BH_2−AH_2CH_3	1.432	1.496	1.518
CH_3BH_2−$AH_2C(O)OH$	1.430	1.495	1.519
CH_3BH_2−AH_2NH_2	1.441	1.512	1.528
CH_3BH_2−$AH_2NH_3^+$	1.429	1.491	1.518
CH_3BH_2−AH_2NH^-	1.458	1.543	1.540
CH_3BH_2−AH_2OH	1.435	1.500	1.513
CH_3BH_2−AH_2O^-	1.477	1.619	1.606
CH_3BH_2−AH_2SH	1.432	1.496	1.515
CH_3BH_2−AH_2S^-	1.442	1.515	1.520
CH_3BH_2−AH=CH_2	1.424	1.490	1.494
CH_3BH_2−$A(O)NH_2$	1.450	1.526	1.521
CH_3BH_2−$A(O)OH$	1.444	1.514	1.506
CH_3BH_2−$A(O)O^-$	1.495	1.643	1.621
CH_3BH_2−$A(O)OCH_3$	1.446	1.519	1.507
ethyl benzene[b]	1.429	1.494	1.502
histidine[b]	1.432	1.502	1.492
alanine dipeptide[c]	1.432	1.503	1.526

[a] The training set is in bold. [b] B is the methylene carbon, and A is the ipso ring carbon. [c] B is C_α, and A is a carbonyl carbon.

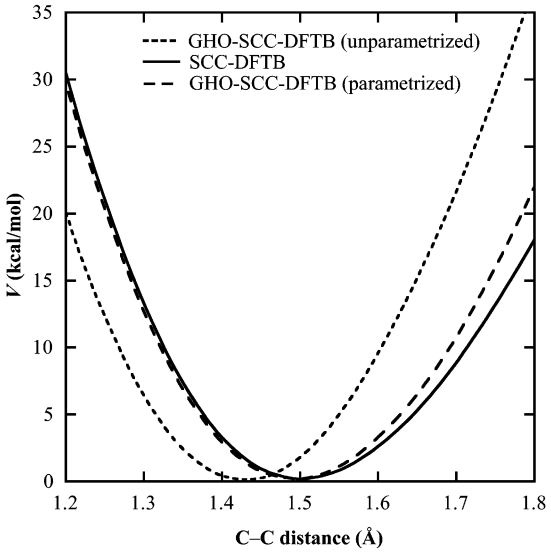

Figure 2. C−C stretching potential in ethane, treated by fully QM (SCC-DFTB) and QM/MM (GHO-SCC-DFTB). Results of both the unparametrized version and the parametrized version with an empirical correction E_{cor} for GHO-SCC-DFTB are given.

For example, to study phosphate hydrolysis processes in ribonucleic acids, one may need to put the QM/MM boundary along an O−C bond instead of a C−C bond. The parameters for an A atom other than a carbon can be obtained by a procedure similar to the one we used for C−C bonds. Parameters for O−C and S−C boundaries are presented in the Supporting Information, with recommended E_{cor} coefficients and test results; however, we restrict the main text to boundaries that pass through C−C bonds.

4. Results and Discussion

4.1. Stretch Potential. To test the new method, we study the C−C stretch potential for ethane and propanoate. The former has a typical C−C bond distance, and the latter is an extreme case with a very large C−C equilibrium distance. Figures 2 and 3 show the stretch potential curves obtained by using unparametrized and parametrized GHO-SCC-DFTB and com-

TABLE 3: Mean Unsigned Errors in Bond Lengths (Å) and Bond Angles (deg) with GHO-SCC-DFTB[a]

	GHO-SCC-DFTB (unparametrized)					GHO-SCC-DFTB (with E_{cor})				
system	A−B	Q−A	B−M	Q−A−B	A−B−M	A−B	Q−A	B−M	Q−A−B	A−B−M
$CH_3BH_2AH_3$	0.083	0.006	0.018	1.9	1.4	0.020	0.006	0.015	2.3	1.6
$CH_3BH_2AH_2CH_3$	0.086	0.005	0.018	1.2	0.7	0.021	0.006	0.015	1.7	0.6
$CH_3BH_2AH_2C(O)OH$	0.089	0.004	0.020	1.3	0.9	0.024	0.005	0.017	0.9	0.7
$CH_3BH_2AH_2NH_2$	0.087	0.003	0.019	1.3	0.8	0.016	0.005	0.016	1.4	0.8
$CH_3BH_2AH_2NH_3^+$	0.089	0.008	0.022	1.6	0.8	0.027	0.009	0.020	1.4	0.5
$CH_3BH_2AH_2NH^-$	0.082	0.007	0.026	2.9	2.7	0.003	0.004	0.028	2.7	2.8
$CH_3BH_2AH_2OH$	0.078	0.005	0.019	1.9	1.0	0.013	0.007	0.016	1.8	0.9
$CH_3BH_2AH_2O^-$	0.129	0.010	0.027	2.0	2.5	0.012	0.003	0.031	2.5	2.9
$CH_3BH_2AH_2SH$	0.084	0.008	0.019	1.3	0.9	0.019	0.010	0.016	1.7	0.6
$CH_3BH_2AH_2S^-$	0.078	0.001	0.026	1.2	1.1	0.005	0.004	0.027	1.7	1.2
$CH_3BH_2AH{=}CH_2$	0.070	0.004	0.017	0.7	0.5	0.005	0.005	0.017	1.0	0.6
$CH_3BH_2A(O)NH_2$	0.061	0.007	0.018	0.6	2.2	0.012	0.013	0.015	1.1	1.9
$CH_3BH_2A(O)OH$	0.062	0.004	0.016	0.5	3.3	0.008	0.009	0.015	0.9	2.9
$CH_3BH_2A(O)O^-$	0.126	0.012	0.025	1.0	1.2	0.021	0.006	0.032	0.6	1.1
$CH_3BH_2A(O)OCH_3$	0.061	0.007	0.018	0.6	2.2	0.012	0.013	0.015	1.1	1.9
ethyl benzene	0.094	0.004	0.026	0.7	2.3	0.024	0.002	0.027	1.0	2.5
histidine	0.060	0.002	0.010	1.2	1.0	0.009	0.002	0.009	1.5	1.0
alanine dipeptide	0.094	0.004	0.026	0.7	2.3	0.024	0.002	0.027	1.0	2.5
training set	0.088	0.006	0.020	1.3	1.4	0.015	0.006	0.019	1.4	1.3
whole set	0.083	0.005	0.020	1.2	1.5	0.014	0.006	0.019	1.4	1.4

[a] The training set is in bold.

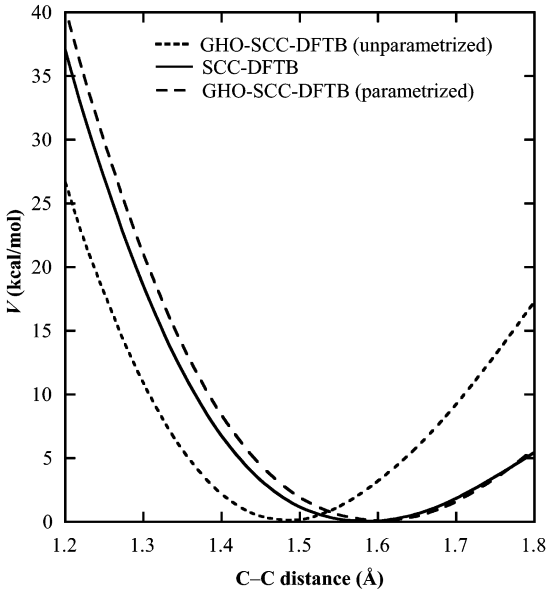

Figure 3. C−C stretching potential in propanoate, treated by fully QM (SCC-DFTB) and QM/MM (GHO-SCC-DFTB). Results of both the unparametrized version and the parametrized version with an empirical correction E_{cor} for GHO-SCC-DFTB are given.

pare them to the fully SCC-DFTB results for ethane and propanoate, respectively. The unparametrized method gives C−C potential energy curves with a shape in good agreement with the fully SCC-DFTB results. However, the locations of the minima of the potential wells are underestimated by 0.08 and 0.12 Å for ethane and propanoate, respectively. With the aid of the E_{cor} term, the parametrized GHO-SCC-DFTB method gives accurate equilibrium C−C distances for these systems, and the C−C stretch potential energy curves obtained by the parametrized GHO-SCC-DFTB method agree well with the results of fully QM calculations over a wide range from 1.2 to 1.8 Å. Note that we do not modulate the E_{cor} term with any cutoff or decay functions. This simplification is justified because one should not place the QM/MM boundary at a bond that dissociates during a simulation. The functional form of the E_{cor} term developed here should be suitable for practical applications.

TABLE 4: Mulliken Atomic Charges (au) Determined at the GHO-SCC-DFTB Level for the CAH3−CB Fragment in Propane

atom	GHO-SCC-DFTB (unparametrized)	GHO-SCC-DFTB (with E_{cor})	SCC-DFTB
C_A	−0.21	−0.21	−0.20
H_1	0.06	0.07	0.06
H_2	0.07	0.07	0.06
H_3	0.07	0.07	0.06
C_B	−0.17	−0.18	−0.08
C_AH_3	−0.01	−0.01	−0.01

4.2. Geometry. The bond distances between the GHO boundary atom B and the QM frontier atom A are listed in Table 2, for both the unparametrized and the parametrized model. Table 2 shows that the unparametrized GHO-SCC-DFTB underestimates the A−B bond distance by 0.06−0.13 Å. However, the parametrized GHO-SCC-DFTB yields A−B bond lengths that deviate by only 0.004−0.02 Å from the fully QM results. In Table 3, we report the mean-unsigned-errors (MUEs) of bond distances and bond angles involving the GHO boundary atom of the GHO-SCC-DFTB results as compared to the fully QM results. The A−B bond distance given by GHO-SCC-DFTB only deviates from the QM results by 0.015 Å on average. The bond angles near the QM/MM boundary reproduce the QM results within 1.5°. The MUEs for the whole test set and the training set are similar, which may indicate that the training set is diverse enough and is representative of various bonding situations near the QM/MM boundary.

4.3. Mulliken Charges. Although we do adjust the charges in our parametrization, we find that the charges obtained from Mulliken population analysis are not sensitive to the empirical corrections, because the E_{cor} term does not change the electronic wave function directly. We report the Mulliken partial charge determined by GHO-SCC-DFTB/CHARMM for propane and acetic acid in Tables 4 and 5. The QM methyl group in propane as calculated by GHO-SCC-DFTB is within 0.01 au of neutral for both the unparamtrized and the parametrized calculations. The results show that the atomic charge is not changed very much by the paramtrization. This is partly due to the fact that the E_{cor} term does not alter the QM wave function for a given molecular geometry.

TABLE 5: Mulliken Atomic Charges (au) Determined at the GHO-SCC-DFTB Level for Acetic Acid

atom	GHO-SCC-DFTB (unparametrized)	GHO-SCC-DFTB (with E_{cor})	SCC-DFTB
O(=CA)	−0.48	−0.42	−0.50
O(−C_A)	−0.43	−0.47	−0.45
H(O)	0.34	0.34	0.33
C_A(=O)	0.65	0.66	0.63
C_B	−0.34	−0.37	−0.28

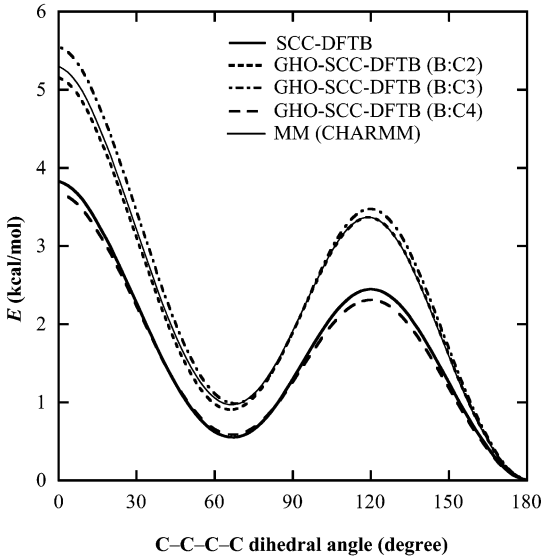

Figure 4. Potential energy curve for the internal rotation around the C2−C3 bond in *n*-butane using parametrized GHO-SCC-DFTB, pure QM (SCC-DFTB), and MM (CHARMM22). For the GHO-SCC-DFTB method, three cases are studied, where the GHO boundary atom is placed at C2, C3, and C4, respectively.

4.4. Rotational Barriers. The potential energy profile for internal rotation about the central C−C bond in *n*-butane has been frequently used to test QM/MM boundary algorithms.[25−28] As we mentioned in our previous paper,[28] the X−A−B−X molecular mechanical torsion energy should be included in the GHO-SCC-DFTB total energy, because both the QM and MM contributions are important for the torsion barrier. Figure 4 plots the internal rotational barrier for cases where the GHO boundary atom is placed at different locations with respect to the central C−C bond. The choice of placing the GHO boundary atom at the second, third, and fourth carbon corresponds a small-, a medium-, and a large-sized QM fragment. For the case with the GHO boundary at C4, both of the central atoms involved in the rotation are described by full QM; therefore the rotational barrier faithfully reproduces the QM results. If the central C−C bond whose torsion is being considered coincides with the A−B bond for the QM/MM division (C3 boundary case), the rotation barrier is described by both the QM and MM terms. In this case, because MM predicts higher internal rotation barriers than SCC-DFTB, the combination of the QM/MM barrier heights is similar to the MM one. With the size of the QM fragment reduced (C2 boundary case), the description of the rotational barrier is eventually dominated by MM.

4.5. Proton Affinities. In combined QM/MM calculations, the QM energy is of central importance for evaluating the QM/MM boundary treatment. Proton affinities are a sensitive test of how well the boundary regions are described. With the energy of the proton defined as zero, the proton affinity for a base X can be calculated as the energy difference between a base species (denoted X or X^-) and its protonated form (denoted XH^+ or XH). Table 6 compares the proton affinities calculated by GHO-SCC-DFTB with the results obtained from fully QM calculations by SCC-DFTB. The proton affinity test suite listed in Table 6 has been used previously in testing the GHO-AIHF algorithm.[28] The average absolute deviation of the proton affinities from the fully QM results for 30 cases in Table 6 is only 1.7 kcal/mol, which is only about 1% of the typical proton affinity. Note that 15 out of the 30 cases are designed to push the method to its limit by putting the GHO boundary only one bond away from the X−H bond being dissociated. Even for these extremely hard cases, the proposed GHO-SCC-DFTB method gives reasonable agreement with the fully QM calculations with a mean unsigned error of 2.4 kcal/mol. It is encouraging that these deviations are reduced to only 1.2 and 0.6 kcal/mol, respectively, if the GHO boundary is moved one bond and two bonds further away from the protonated/deprotonated center, respectively. These results suggest that one obtains more accurate results if one places the GHO boundary atom at least one carbon atom away from the reactive center.

TABLE 6: Proton Affinities (kcal/mol) Using GHO-SCC-DFTB/CHARMM Compared to the SCC-DFTB Fully QM Results

system	GHO-SCC-DFTB (with E_{cor})	SCC-DFTB	QM/MM vs QM
BH_3−AH_2O^-	251.5	255.2	−3.7
CH_3BH_2−AH_2O^-	256.8	253.7	3.0
BH_3−$AH_2CH_2O^-$	253.9	253.7	0.1
$CH_3CH_2BH_2$−AH_2O^-	256.3	253.1	3.2
CH_3BH_2−$AH_2CH_2O^-$	255.5	253.1	2.4
BH_3−$AH_2CH_2CH_2O^-$	252.7	253.1	−0.4
BH_3−AH_2NH_2	65.2	68.5	−3.3
CH_3BH_2−AH_2NH_2	67.1	69.3	−2.2
BH_3−$AH_2CH_2NH_2$	67.6	69.3	−1.7
$CH_3CH_2BH_2$−AH_2NH_2	66.9	69.9	−3.0
CH_3BH_2−$AH_2CH_2NH_2$	69.3	69.9	−0.5
BH_3−$AH_2CH_2CH_2NH_2$	68.8	69.9	−1.0
BH_3−AH_2NH^-	273.2	274.0	−0.8
CH_3BH_2−AH_2NH^-	276.2	272.6	3.6
BH_3−$AH_2CH_2NH^-$	272.3	272.6	−0.3
$CH_3CH_2BH_2$−AH_2NH^-	275.8	272.0	3.8
CH_3BH_2−$AH_2CH_2NH^-$	274.0	272.0	2.0
BH_3−$AH_2CH_2CH_2NH^-$	271.6	272.0	−0.4
BH_3−AH_2S^-	212.7	213.7	−1.0
CH_3BH_2−AH_2S^-	215.0	212.8	2.2
BH_3−$AH_2CH_2S^-$	212.1	212.8	−0.7
$CH_3CH_2BH_2$−AH_2S^-	214.7	212.4	2.3
CH_3BH_2−$AH_2CH_2S^-$	213.7	212.4	1.3
BH_3−$AH_2CH_2CH_2S^-$	212.0	212.4	−0.5
BH_3−$A(O)O^-$	222.4	224.6	−2.1
CH_3BH_2−$A(O)O^-$	222.8	223.9	−1.1
BH_3−$AH_2C(O)O^-$	223.6	223.9	−0.3
$CH_3CH_2BH_2$−$A(O)O^-$	222.1	223.2	−1.2
$CH_3BH_2AH_2C(O)O^-$	225.4	223.2	2.2
BH_3−$AH_2CH_2C(O)O^-$	222.7	223.2	−0.5

Another way to put the results in Table 6 into perspective is to compare to a previous study. Twenty-one of the cases in Table 6 (the first 18 rows, $BH_3AH_2CO_2^-$, and the last two rows) were also studied by Amara and Field with QM/MM methods based on Hartree−Fock theory for QM, OPLS-AA for MM, and a link atom approach with smear charge distributions.[14] Their two best methods are denoted $LA_g(4,0)$ and $LA_g(4,3)$, and for these 21 cases, their mean unsigned deviations of QM/MM from QM in proton affinities are 3.3 and 2.3 kcal/mol, respectively. Our mean unsigned deviation for these 21 cases is 1.8 kcal/mol. The mean SCC-DFTB proton affinity in these 21 cases is 190.4 kcal/mol, so a 1.8 kcal/mol mean unsigned error is less than 1%. This accuracy is satisfactory enough for many applications.

5. Concluding Remarks

In this paper, we used the GHO approach to combine a QM fragment described by the SCC-DFTB method with a fragment described by molecular mechanics, and we present formulas for analytical gradients. By adding a parametrized empirical term, we obtain an improved description of the bond length at the QM/MM boundary. The proposed method is robust for geometry optimizations with various functional groups present near the boundary. The electronic and energetic properties of the GHO-SCC-DFTB method are further tested against the full QM results. We conclude that the GHO-SCC-DFTB method provides an electrostatically stable representation of the QM and MM boundary, where the SCC-DFTB theory is used to describe the QM part. This work also has implications for applying the GHO-SCC-DFTB/MM methodology to solid-state and other condensed-phase systems, such as enzymes.

Acknowledgment. We thank Marcus Elstner for providing SCC-DFTB integration tables and parameter files. This work has been supported in part by grant no. CHE00-92019 from the National Science Foundation and by a grant (GM46736) from the National Institutes of Health.

Supporting Information Available: Tables of E_{cor} in GHO-SCC-DFTB parametrized for cutting C−O and C−S bonds at the QM/MM boundary, Mulliken charges, and geometries for such test cases. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Gao, J., Thompson, M., Eds. *Combined Quantum Mechanical and Molecular Mechanical Methods*; ACS Symposium Series 712; American Chemical Society: Washington, DC, 1998.
(2) Gao, J. *Rev. Comp. Chem.* **1996**, *7*, 119.
(3) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
(4) Monard, G.; Prat-Resina, X.; Gonzalez-Lafont, A.; Lluch, J. M. *Int. J. Quantum Chem.* **2003**, *93*, 229.
(5) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(6) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(7) Gao, J.; Xia, X. F. *Science* **1992**, *258*, 631.
(8) *Hyperchem User Manual*, *Computational Chemistry*; Hypercube, Inc.: Ontario, Canada, 1994.
(9) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
(10) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
(11) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
(12) Ferre, N.; Olivucci, M. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 71.
(13) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.
(14) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.
(15) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
(16) Dilabio, G.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.
(17) Ferenczy, G. G.; Rivail, J.-L.; Surjan, P. R.; Naray-Szabo, G. *J. Comput. Chem.* **1992**, *13*, 830.
(18) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.
(19) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.
(20) Monard, G.; Loos, M.; Théry V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153.
(21) Ferre, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610.
(22) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.
(23) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *Chem. Phys. Lett.* **2000**, *321*, 113.
(24) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442.
(25) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
(26) Amara, P.; Field. M. J.; Alhambra, C.; Gao, J.; *Theor. Chem. Acc.* **2000**, *104*, 336.
(27) Garcia-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2004**, *111*, 280.
(28) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.
(29) Dapprich, S.; Komiromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. (THEOCHEM)* **1999**, *461*, 1.
(30) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.
(31) Derat, E.; Bouquant, J.; Humbel, S. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 61.
(32) Hillier, I. H. *J. Mol. Struct. (THEOCHEM)* **1999**, *463*, 45.
(33) Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462.
(34) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452.
(35) Turner, A. J.; Moliner, V.; Williams, I. H. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323.
(36) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. *Proteins Struct. Funct. Genet.* **1999**, *37*, 157.
(37) Aida, M.; Yamataka, H.; Dupuis, M. *Int. J. Quantum Chem.* **2000**, *77*, 199.
(38) Cui, Q.; Karplus, M. *J. Chem. Phys.* **2000**, *112*, 1133.
(39) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656.
(40) Hurley, M. M.; Wright, J. B.; Lushington, G. H.; White, W. E. *Theor. Chem. Acc.* **2003**, *109*, 160.
(41) Swart, M. *Int. J. Quantum Chem.* **2003**, *91*, 177.
(42) Basiuk, V. A. *J. Phys. Chem. B* **2003**, *107*, 8890.
(43) Kerdcharoen, T.; Birkenheuer, U.; Kruger, S.; Woiterski, A.; Rosch, N. *Theor. Chem. Acc.* **2003**, *109*, 285.
(44) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902. Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1.
(45) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549.
(46) Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M. L.; Villa, J. *Acc. Chem. Res.* **2002**, *35*, 341.
(47) Alhambra, C.; Gao, J.; Corchado, J. C.; Villa, J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1999**, *121*, 2253.
(48) Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Gao, J.; Truhlar, D. G. *J. Am. Chem. Soc.* **2000**, *122*, 8197.
(49) Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326.
(50) Alhambra, C.; Sanchez, M. L.; Corchado, J. C.; Gao, J.; Truhlar, D. G. *Chem. Phys. Lett.* **2001**, *347*, 512, **2002**, *355*, 388 (E).
(51) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Comput. Chem.* **2003**, *24*, 177.
(52) Poulsen, T. D.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2003**, *107*, 9567.
(53) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558.
(54) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.
(55) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213.
(56) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B.* **1998**, *58*, 7260.
(57) Frauenheim, T.; Seifert, G.; Elstner, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. *Phys. Stat. Solidi (b)* **2000**, *217*, 41.
(58) Niehaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct. (THEOCHEM)* **2001**, *541*, 185.
(59) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565.
(60) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.
(61) Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366.
(62) Kumar, A.; Elstner, M.; Suhai, S. *Int. J. Quantum Chem.* **2003**, *95*, 44.
(63) Hu, H.; Elstner, M.; Hermans, J. *Proteins*: *Struct., Funct., Genet* **2003**, *50*, 451.
(64) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
(65) Cui, Q.; Elstner, M.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 2721.
(66) Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct.* (*THEOCHEM*) **2003**, *632*, 29.
(67) *Chemistry at HARvard Macromolecular Mechanics* (CHARMM) computer program, as described in: Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(68) Xie, L.; Liu, H. Y. *J. Comput. Chem.* **2002**, *23*, 1404.
(69) Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2545.

(70) Foulkes, W.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520.
(71) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(72) Pulay, P. *Mol. Phys.* **1969**, *17*, 197. Pulay, P. In *Modern Theoretical Chemistry*; Schaefer, H. F., Ed.; Plenum Press: New York, 1977; p 153, Vol. 4.
(73) Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; 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, B.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(74) Mackerell A. D., Jr. In *Computational Biochemistry and Biophysics*; Becker, O., Mackerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Dekker: New York, 2001; p 7. Mackerell, A. D., Jr.; Nilsson, L. In *Computational Biochemistry and Biophysics*; Becker, O., Mackerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Dekker: New York, 2001; p 441.
(75) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
(76) Carroll, D. L. *FORTRAN Genetic Algorithm (GA) Driver*. See http://cuaerospace.com/carroll/ga.html.