Generalized Hybrid Orbital (GHO) Method for Combining Ab Initio Hartree−Fock Wave Functions with Molecular Mechanics

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

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

Received: September 15, 2003; In Final Form: November 17, 2003

The generalized hybrid orbital (GHO) method provides a way to combine quantum mechanical (QM) and molecular mechanical (MM) calculations on a single molecular system or supramolecular assembly by providing an electrostatically stable connection between the QM portion and the MM portion. The GHO method has previously been developed for semiempirical molecular orbital calculations, on the basis of neglect of diatomic differential overlap (GHO−NDDO); in the present work, it is extended to the ab initio Hartree−Fock (HF) level (GHO−AIHF). First, the theoretical foundation for the GHO−AIHF extension is discussed, and four different approaches are proposed to overcome the nonorthogonality between active molecular orbitals (MOs) and auxiliary MOs. In the first scheme, the auxiliary hybrid basis functions are projected out of the active QM basis. The second scheme neglects the diatomic differential overlap between the auxiliary basis and the active QM basis. In the third scheme, hybrid orbitals are constructed from Löwdin-type symmetric orthogonalized atomic orbitals on the basis of global Löwdin orthogonalization. The fourth procedure involves local Löwdin orthogonalization. The procedures for implementing the four GHO−AIHF schemes are described, and analytical gradient expressions are derived. The unparametrized GHO−AIHF method is tested for hydrocarbons with various basis sets, in particular, the geometries and charges are compared with pure QM calculations for ethane, ethyl radical, and *n*-octane, and the method is tested for the torsion potential around the central bond in *n*-butane. Furthermore, a parametrization of the GHO−AIHF method for the MIDI! basis is presented and tested for 16 molecules and ions with various functional groups near the QM/MM boundary. The results show the robustness of the algorithm and illustrate the significant improvement made by introducing several one-electron integral-scaling parameters. Finally, the energetic performance of the method is tested by comparing the proton affinities for a set of small model compounds (alcohols, amines, thiols, and acids) to results obtained from fully QM calculations. We conclude that the GHO−AIHF scheme provides a reasonable fundamental solution to the problem of combining an ab initio quantum mechanical electronic structure calculation with molecular mechanics.

I. Introduction

Modeling the energetics and dynamics of macromolecular systems and large complexes presents a major challenge for modern theoretical chemistry, because of the size and intricacy of the systems. A powerful tool for meeting this challenge is the combined use of quantum mechanics and molecular mechanics (QM/MM).[1−41] The motivation for this combined approach is that processes involving bond breaking, bond forming, and electronic excitation should be described by QM, whereas much of the remainder of the molecule (or system) may be adequately treated by classical force fields, i.e., MM. Thus, combined QM/MM methods synthesize the computational accuracy of QM with the computational efficiency of MM for large systems. In the type of QM/MM synthesis considered here, the system is partitioned into a small subsystem that is treated by QM and a large subsystem that is treated by MM. A special case is the treatment of solvation for which the boundary between the QM and MM subsystems can be placed between solute and solvent molecules that are not covalently bonded to each other. When using QM and MM within a single molecule, such as an organometallic complex or a macromolecule, the boundary may pass through one or more covalent bonds. For example, in enzymatic reactions, some protein residues (as well as the substrate) participate in the chemical reactions, and, therefore, they must be included in the QM region; thus, the QM/MM boundary must pass through covalent bonds of the protein. The treatment of the QM and MM boundary is far from straightforward, which represents a major concern in the accuracy of combined QM/MM methods. Several different methods have been developed to truncate the QM electronic wave function gracefully at the QM/MM boundary when it passes through a bond. This paper describes new approaches to this problem.

The most straightforward strategy to treat QM/MM boundaries is the so-called "link atom" approach, which uses a H atom to cap the free valence of the QM fragment.[2,3] Because of its simplicity, the link atom approach has been widely used in all types of QM/MM applications, both at the semiempirical and ab initio levels. A drawback of the "link atom" method is that it introduces additional degrees of freedom that are not present in the original molecular system. This introduces complications in the energy definition, the optimization of geometries, and the polarization of the bond between the QM boundary atom

* Authors to whom correspondence should be addressed. E-mail: truhlar@umn.edu, gao@chem.umn.edu.

10.1021/jp036755k CCC: $27.50
Published on Web 01/03/2004

and the link atom; these problems have been thoroughly discussed,[17,29,35,41] and methods to circumvent the problem have been proposed.[29,42] In a so-called "double link atom" approach,[39] two link atoms are added at each QM/MM boundary to alleviate electrostatic unbalance introduced by a single link atom, and delocalized Gaussian functions are used to correct the strong polarization near the QM/MM boundary region. Although the question of correctly balancing the polarization at the QM/MM boundary when one uses link atoms continues to be troublesome, meaningful results can be obtained, especially when one introduces refinements (for example, parametrized treatments of the electrostatic potentials or point charges that are smeared out near the boundary).[17,29,35,42] However, it is reasonable to ask whether there is a more fundamental way to join the quantum mechanical and classical mechanical regions without introducing the extra degrees of degrees of freedom that are associated with nonphysical atoms.

An example of a more fundamental approach that has been developed to provide a quantum mechanical description of bonds at the QM/MM boundaries is to use localized orbitals. The first application of this approach was described by Rivail and co-workers, using the local self-consistent field (LSCF) algorithm.[6−9,25] In the LSCF framework, the chemical bonds that connect the QM and MM fragments are called frontier bonds, and they are represented by a set of strictly localized bond orbitals (SLBOs), which are determined by calculations for small model compounds. The strictly localized character of these orbitals helps to ensure that they are transferable from the model system to the large molecule. The SLBOs are excluded from the self-consistent field (SCF) optimization of the large molecule, to prevent their mixture with other QM basis functions. The LSCF method was first developed for semiempirical Hamiltonians, and it was then generalized to ab initio Hartree−Fock (HF), post-HF, and density functional theory (DFT) methods with analytical gradients.[7,8] Recently, specific force-field parameters have also been developed for the LSCF method.[9] Although the LSCF framework is theoretically more robust than using link atoms, the unavoidable need for model studies for each individual system is a drawback. To solve this problem, Friesner and co-workers parametrized a library of frozen densities for all side chains of amino acids.[37] A detailed quantitative test of the LSCF approach showed that, similar to the link atom method, it can be satisfactory if used with special care.[29]

The link-atom and LSCF methods continue to be used in various forms. For example, Friesner and co-workers called their LSCF library a "frozen orbital" approach,[37] and Zhang et al.[36] developed a "pseudo-bond" method, which is a link-atom implementation, in which the link atom is a pseudo-halogen atom whose lone-pair orbitals and electrons serve as an effective core potential (ECP). The ECP parameters are adjusted to mimic the properties of the original bond. Recently, Swart proposed a version of the link atom approach called AddRemove.[41]

Another approach with some similarity to the LSCF method is the generalized hybrid orbital (GHO) method.[26] In this approach, a set of four hybrid orbitals is assigned to each MM atom at a QM/MM boundary; such atoms are called boundary atoms and are denoted by the symbol B. One of the hybrid orbitals, called the active orbital, is directed toward the QM atom to which the B atom is bonded; this atom is denoted by the symbol A. In the initial application, the boundary atom has been chosen to be an sp^3 C atom, although other atoms can also be used. The hybridization scheme is completely determined by the local geometry of four atoms (the B atom and the three MM atoms to which it is bonded), and one hopes that this will make the parametrization transferable so that one might not need to perform calculations on new model systems each time the GHO method is applied. Furthermore, the hybridization of these orbitals varies dynamically during molecular dynamics simulations, whereas, in the LSCF and the frozen orbital methods, the localized orbitals are treated as independent of geometrical variations. The active hybrid orbital is allowed to participate in the SCF optimizations, and it mixes freely with other QM basis functions. The other three hybrid orbitals, which are called auxiliary orbitals, do not mix and are excluded from the SCF procedure; however, they provide a quantum mechanical representation of the charge distribution in the bonds of the boundary atom to its MM neighbors, and this delocalized representation is more robust than using point charges at the boundary. In the semiempirical molecular orbital[43] framework for which the method was developed, these treatments neglect diatomic differential overlap (NDDO).

The GHO boundary treatment has been used successfully in recent enzyme dynamics studies that have been based on combined QM/MM potential energy surfaces (PESs).[27,44−47] These applications used semiempirical QM models,[43] which often have large errors in quantitative energy calculations; thus, specific parametrizations[48] are typically needed. To overcome this limitation, the present article presents an extension of the GHO formalism to the ab initio Hartree−Fock (AIHF) level. In addition to being of interest in its own right, a formulation based on AIHF is a first step toward developing methods for hybrid DFT and for post-HF correlated methods based on a HF reference.

The organization of the paper is as follows. The GHO−AIHF algorithm is presented in Section II. Details of orthogonalization procedures are described in Section III. For simplicity, the main text is restricted to closed-shell singlets; the changes required to treat open-shell systems are presented in an appendix. The implementation of GHO−AIHF is discussed in Section IV. Section V presents tests of the algorithm with no new parameters. Section VI discusses further parametrization by scaling integrals that involve orbitals at the QM/MM boundary. Section VII contains concluding remarks.

II. Theory

The presentation of the GHO−AIHF method is organized as follows. In Subsection IIA, a brief review of the principal elements of the GHO method in given, and some useful terms and notation used throughout the later discussions are defined. In Subsection IIB, the major theoretical concerns are discussed, and a general overlap constraint is formulated. In Section III, we propose four different strategies for enforcing this constraint.

A. The Generalized Hybrid Orbital (GHO) Method with Neglect of Overlap. Because the GHO algorithm has been presented at the semiempirical level elsewhere,[26] we provide only a brief summary of the major elements relevant to the further development in the present article. In this section, we describe the case in which the QM system is approximated by neglect of diatomic differential overlap[43] (NDDO) and the QM/MM partition of the molecular system is placed at an sp^3 C atom, which is called a GHO boundary atom B; in principle, B can be any other type of atom or have other hybridizations, although the present choice is sufficient for treating a large number of systems, including most enzyme systems. The B atom is both a QM and an MM atom in the GHO method, and the QM atom bonded to the B atom is called the frontier atom A. The three MM atoms directly bonded to the B atom are denoted

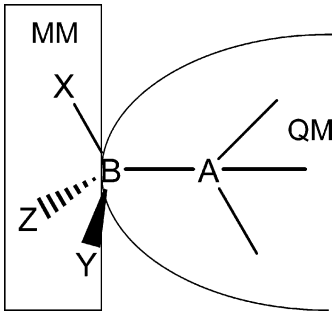

Figure 1. Schematic representation of the QM/MM partition in the GHO method.

by the symbols X, Y, and Z (see Figure 1). We define atoms other than B in the QM subsystem as "fully QM atoms", which are denoted by the symbol Q, because they are fully treated by quantum mechanics. Let the total number of basis functions on fully QM atoms be N, and denote these by χ_u, where $u = 1, 2, ..., N$. For the GHO boundary atom B, a set of hybrid orbitals $\{\eta_B, \eta_x, \eta_y, \eta_z\}$ is constructed by hybridization of the atomic s and p valence basis functions on the B atom. The orbital η_B is the active hybrid orbital. The other three hybrid orbitals are the auxiliary orbitals, denoted by η_b (where $b = x, y, z$). A basis transformation matrix $\mathbf{T}_b$, which has been defined in detail previously,[26] relates these hybrid orbitals to the atomic s and p orbitals on the B atom:

$$\begin{pmatrix} \eta_B \\ \eta_x \\ \eta_y \\ \eta_z \end{pmatrix} = \mathbf{T}_b^{\dagger} \begin{pmatrix} s \\ p_x \\ p_y \\ p_z \end{pmatrix} \tag{1}$$

As a result of the hybridization, $\{\eta_B,\eta_x,\eta_y,\eta_z\}$ is an orthonormal set:

$$S_{cd} = \langle \eta_c | \eta_d \rangle = \delta_{cd} \qquad (\text{where } c,d = \text{B}, x, y, z) \tag{2}$$

The fully QM basis functions (χ_u) 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 molecular orbitals that result from diagonalization of the Fock matrix[43] of the SCF calculation are linear combinations of these active basis functions:

$$\varphi_i = \sum_{u=1}^{N} c_{ui}\chi_u + c_{Bi}\eta_B = \sum_{a=1}^{N+1} c_{ai}\chi_a \qquad (i = 1, 2, .., N + 1) \tag{3}$$

where $\chi_{N+1} = \eta_B$. Each of the three auxiliary basis functions forms an auxiliary MO by itself, which, for each geometry, is to be frozen in the SCF procedure:

$$\begin{aligned} \varphi_{N+2} &= \eta_x \\ \varphi_{N+3} &= \eta_y \\ \varphi_{N+4} &= \eta_z \end{aligned} \tag{4}$$

The frozen auxiliary MOs provide an effective charge distribution to mimic the fractional charges present in the three bonds that are formed by the B atom and its MM neighbors {X, Y, Z}. The occupation P^H_{bb} assigned to each auxiliary orbital is chosen to be $1 - q_B/3$, where q_B is the MM partial charge of the B atom, so that the MM partial charge on the B atom is evenly distributed over three auxiliary orbitals. For example, the CHARMM parametrization of the MM force field has $q_B = -0.27$ for a methyl carbon and $q_B = -0.18$ for a methylene carbon. Thus, the auxiliary orbital occupancy for these two key cases is 1.09 and 1.06, respectively. (It should be emphasized that all partial atomic charges on MM atoms, including the boundary atoms, are retained without modification, and they all interact with QM electrons and nuclei.) Many of the notations defined in this paragraph will be used throughout the remainder of the article; we summarize them in Table 1 for further reference.

TABLE 1: Notation

symbol	description
B	the GHO boundary atom
A	the frontier atom, i.e., the QM atom to which the B atom is directly bonded
Q	fully QM atoms, including the A atom but not the B atom
X, Y, Z	the three MM atoms to which the B atom is directly bonded
χ_u	fully QM basis functions on Q; index $u = 1-N$
η_B	active hybrid basis function on the B atom
η_b	auxiliary hybrid basis on the B atom; the index $b = 1-3$ (or takes the values x, y, and z)
χ_a	active basis functions (χ_u plus η_B); index $a = 1-(N + 1)$. Note that $\chi_{N+1} = \eta_B$.
q_B	the MM partial charge on the B atom

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

$$E^{tot} = E_{QM} + E_{MM} + E_{QM/MM} \tag{5}$$

where E_{QM} and E_{MM} are the internal energies of the QM and MM subsystems, respectively, with the MM terms that involve only QM atoms (i.e., that involve only fully QM atoms and the boundary atom) removed, and $E_{QM/MM}$ contains (i) the interaction energy of delocalized electrons (including those in auxiliary orbitals) with MM partial charges, (ii) the interaction energy of QM nuclei with MM partial charges, and (iii) nonbonded van der Waals interactions between QM and MM atoms. Because the NDDO method includes only valence electrons, all QM nuclear charges (including the nuclear charge on the B atom) are reduced by the number of core electrons. It is useful to rewrite eq 5 as

$$E^{tot} = E^{orb} + E_{MM} + E^{nuc}_{QM,QM/MM} \tag{6}$$

where E^{orb} is the electronic component of the sum of E_{QM} and component (i) of $E_{QM/MM}$ and $E^{nuc}_{QM,QM/MM}$ is the sum of components (ii) and (iii) of $E_{QM/MM}$ and the nucleus repulsion terms in the E_{QM} term.

In the GHO formalism, there are two classes of MOs: active (act) and auxiliary (aux). Consequently, we can write

$$E^{orb} = E^{act} + E^{aux} \tag{7}$$

These energy terms can be further expressed as

$$E^{act} = \sum_{u\upsilon=1}^{N+1} P^H_{u\upsilon}\left[(H^H_{u\upsilon} + I^H_{u\upsilon}) + \frac{1}{2}\sum_{st=1}^{N+1} P^H_{st}(st||u\upsilon) + \sum_{b=1}^{3} P^H_{bb}(bb||u\upsilon)\right] \tag{8}$$

$$E^{aux} = \sum_{b=1}^{3} P^H_{bb}\left[(H^H_{bb} + I^H_{bb}) + \frac{1}{2}\sum_{c=1}^{3} P^H_{cc}(cc||bb)\right] \tag{9}$$

where P_{ij} and H_{ij} are elements of the density matrix **P** and the

conventional one-electron matrix **H**,[42,43] which contains the kinetic energy of electrons and the attraction from QM nuclei; and **I** represents the energy due to interaction with MM partial charges. The superscript "H" is added to indicate that these quantities are evaluated in the hybrid (H) basis. The shorthand notation of the two-electron integral is defined as

$$(uv||\lambda\sigma) = (uv|\lambda\sigma) - \frac{1}{2}(u\lambda|v\sigma) \qquad (10a)$$

where

$$(uv|\lambda\sigma) = \int\int d\tau_1 d\tau_2 u^*(1)v(1)\lambda^*(2)\sigma(2) \qquad (10b)$$

and τ_i denotes the spatial coordinates of electron i. The active basis indices $\{u,v,s,t\}$ all run from 1 to $N+1$, and the auxiliary basis indices $\{b,c\}$ run from 1 to 3. The SCF procedure only optimizes E^{act} over the active basis space; the energy E^{aux}, which is due to auxiliary orbitals, is a fixed constant for a given geometry, although its derivatives, with respect to nuclear coordinates, are nonzero because of the transformation matrix $\mathbf{T}_b$. After further basis transformation, the total energy can be conveniently evaluated in the atomic orbital (AO) basis as

$$E^{tot} = \frac{1}{2}\sum_{uv=1}^{N+4} P_{uv}^{AO}(H_{uv}^{AO} + F_{uv}^{AO}) + E_{MM} + E_{QM,QM/MM}^{nuc} \qquad (11)$$

B. Generalized Hybrid Orbitals for a Nonorthogonal Basis. The formulation[26] of the GHO method that we have just reviewed was simplified in several ways by the approximations made in the semiempirical NDDO Hamiltonians. An especially critical issue is that, because all basis functions are assumed to be orthogonal to each other in these methods, all auxiliary MOs are automatically orthogonal to active MOs, even though they are not eigenvectors of the Fock matrix. Moreover, the quantum mechanical calculation is semiempirically parametrized in terms of a small number of quantities such as β, U_{ss}, and U_{pp},[43] and many types of integrals are set equal to zero,[42] so that only a small set of integral types must be considered at the GHO boundary.[26] Finally, the density force term[49,50] vanishes in the GHO energy derivative expression (see below), because the overlap matrix is unity under the NDDO approximation. For ab initio HF theory, all the aforementioned simple features become more complicated.

There are four main aspects that must be considered in the treatment of the QM/MM boundary in the generalized hybrid orbital method, to extend the method to ab initio molecular orbital theory. First is the representation of the boundary atom. Second is the generality with respect to using arbitrary basis sets for the QM subsystem. Third is the question of whether the hybrid orbitals must be explicitly orthogonalized to the orbitals of the fully QM atoms. Finally, we must mitigate systematic errors that may occur due to the boundary treatment and learn if they can be removed by scaling electronic integrals that involve the boundary orbitals.

In deciding which representation to use for the boundary atom, we note that all combined QM/MM approaches contain empirical parameters, because these calculations ultimately must balance the treatment of a QM fragment and a highly parametrized MM component. Nevertheless, we want to gain the advantages of parameter-free HF calculations in regions removed from the boundary, and, thus, we want to retain flexibility in basis set selection for the QM fragment in regions far from the boundary. Because reasonably accurate ab initio HF calculations usually require at least a polarized split valence basis set, but polarization functions on the boundary atom would unnecessarily complicate the treatment, we want to allow for different basis sets on the boundary atom and the fully quantum atoms. Regardless of whether the same basis set is used for the boundary atom, there will be an imbalance between the QM fragment and the boundary atom, which may ultimately need be resolved by parametrizing certain electronic integrals. Nevertheless, one criterion for judging whether the QM and MM regions have been joined in a physically reasonable way is that any reparametrization should yield good agreement with full QM calculations. Because parameters are necessarily involved anyway, it is reasonable to use a representation of the boundary atom that is simple and straightforward to implement. Following this line of reasoning, we decided to forego explicit consideration of the 1*s* core electrons on B, and we have assigned an STO-3G[51] valence-only (STO-3Gv) minimal basis set to the B atom. To balance the neglect of core electrons, the nuclear charge of the boundary atom B, for carbon, is reduced from 6 to 4. With this choice of basis functions on the B atom, the formulation of the hybrid basis orbitals in the GHO−AIHF method is greatly simplified. It should be emphasized that we impose no restrictions on the basis functions for atoms in the QM region. The basis imbalance between the fully QM atoms and the boundary atom B may be compensated (if necessary) by parametrizations of the integrals that involve orbitals on the B atom, as we will discuss in Section VI.

The next major issue to be addressed in extending the GHO method to the ab initio HF level is the treatment of orthogonality constraints of MOs. This concern is present not only in the GHO method but also in the LSCF-type methods. The ab initio HF equations[52] are derived from the variational principle for the energy under MO orthonormality constraints:

$$\mathbf{C}^\dagger\mathbf{S}\mathbf{C} = \mathbf{I} \qquad (12)$$

where **C** is a matrix whose columns are the MO coefficients and the dagger symbol (†) denotes a transpose. The MOs that satisfy orthonormality constraints and diagonalize the Lagrange multiplier matrix are called canonical MOs. The canonical MOs are unique, except for degeneracies and phase functions.[52,53] In GHO calculations, only a subset of the MOs are canonical MOs. The auxiliary orbitals in the GHO method are not orthogonal to the active MOs because the auxiliary orbitals are not eigenvectors of the Fock operator. To retain the orthogonality constraint of eq 12 for active MOs and auxiliary orbitals, the following conditions must be imposed, in addition to the SCF procedure in the optimization of the $N+1$ active MOs:

$$(\varphi_a|\varphi_b) = 0 \qquad (a = 1, 2, ..., N+1; b = N+2, N+3, N+4) \qquad (13)$$

where φ_a and φ_b represent active and auxiliary MOs, respectively. Equation 13 can be formulated in terms of the mixed atomic and hybrid basis orbitals:

$$\langle\chi_a|\eta_b\rangle = 0 \qquad (a = 1, 2, ..., N+1; b = x, y, z) \qquad (14)$$

Taking into consideration that the active hybrid basis function η_B (χ_{N+1}) is constructed as being orthogonal to the other auxiliary basis functions (see eq 2), the portion of eq 14 that must be enforced separately is

$$S_{ub} = \langle\chi_u|\eta_b\rangle = 0 \qquad (u = 1, 2, ..., N; b = x, y, z) \qquad (15)$$

In the present work, four different approaches for enforcing eq 15 are proposed and tested: (i) basis set projection, (ii) the

neglect of diatomic differential auxiliary overlap, (iii) hybridization based on global Löwdin orthogonalized atomic orbitals, and (iv) hybridization based on local Löwdin orthogonalization (LLO) near the boundary. The detailed orthogonalization procedures will be discussed in Section III.

The analytical energy gradients in the GHO method are evaluated as the sum of the conventional HF gradient[49,50] and a correction term that is due to the basis transformation:

$$\frac{\partial E^{\mathrm{orb}}}{\partial \boldsymbol{q}} = \frac{\partial E^{\mathrm{HF}}}{\partial \boldsymbol{q}} + \sum_{u\nu}^{N+4}\left(\frac{\partial P_{u\nu}^{\mathrm{AO}}}{\partial \boldsymbol{q}}\right)F_{u\nu}^{\mathrm{AO}} - \sum_{u\nu}^{N+4}\left(\frac{\partial W_{u\nu}^{\mathrm{AO}}}{\partial \boldsymbol{q}}\right)S_{u\nu}^{\mathrm{AO}} \quad (16)$$

where $\mathbf{W}^{\mathrm{AO}}$and $\mathbf{S}^{\mathrm{AO}}$ are the energy-weighted density matrix[49] and overlap matrix, respectively, in the AO basis; and $\boldsymbol{q}$ denotes only the nuclear coordinates of the boundary atom and its MM neighbors X, Y, and Z. The expressions for energy-gradient components for other atoms are unaffected by the GHO procedure. Note that the $\partial\mathbf{W}^{\mathrm{AO}}/\partial\boldsymbol{q}$ term comes from the GHO basis transformation on the density forces, and it is not present in the semiemprical GHO gradient formula, because, in semiempirical methods, the overlap matrix is assumed to be a unit matrix. In ab initio calculations, the matrixes **P** and **W** are further transformed until the differentiated function is a true variational density:

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

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

where **T** is the total basis transformation matrix from the hybrid basis to the AO basis:

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix} \quad (19)$$

We do not need to evaluate the last terms (derivative of a variational density) in eqs 17 and 18, because they have already been included in the conventional HF gradient calculations.[49,50] The gradients ($\partial\mathbf{T}/\partial\boldsymbol{q}$) of the basis transformation matrix was derived elsewhere.[26]

To evaluate the GHO−AIHF gradient analytically, according to eq 16, the energy-weighted density matrix $\mathbf{W}^{\mathrm{AO}}$ is required. The energy-weighted density matrix element $W_{u\nu}^{\mathrm{AO}}$ can be written as

$$W_{u\nu}^{\mathrm{AO}} = \sum_{i=1}^{N+4} n_i^{\mathrm{occ}}\epsilon_i c_{ui}c_{\nu i} \quad (20)$$

where the sum includes the auxiliary MOs; c_{ui} and $c_{\nu i}$ are the respective orbital coefficients for atomic basis functions u and ν in molecular orbital i; and n_i^{occ} and ϵ_i denote the occupation number and the orbital energy for molecular orbital i, respectively. The orbital energies for the $N+1$ active MOs can be obtained in the customary way by solving Roothaan's equation[52] in the active space. The auxiliary orbital energies are not directly available, because the auxiliary MOs are excluded from the active SCF space. However, we evaluate the auxiliary orbital energies explicitly as expectation values of the Fock operator, which yields

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

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

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

The occupation number n_i^{occ} for an auxiliary orbital is equal to $(1 - q_{\mathrm{B}}/3.0)$, as in Subsection IIA.

III. Orthogonalization Procedures

In this section, we present four different approaches for enforcing the orthogonality constraints in eq 15. The first one is to project the auxiliary basis out of the active basis and develop the MOs by an expansion over the projected active basis. The projection operation makes the active basis orthogonal to the auxiliary basis, while retaining the strictly localized character of the auxiliary orbitals. The second method involves the NDDO approximation at the boundary. Specifically, we directly neglect the diatomic differential overlaps between auxiliary hybrid orbitals and basis functions on fully QM atoms; this is called the neglect of diatomic differential auxiliary overlap (NDDAO) approximation. For consistency, the corresponding two-electron integrals that involve the two-center charge distributions between auxiliary orbitals and active QM basis functions also are neglected. In the third and fourth approaches, the hybrid orbitals are developed by hybridization of Löwdin[54,55] symmetrically orthogonalized atomic orbitals (OAOs). This hybridization scheme is justified by the fact that the symmetric OAOs maximally resemble the original AOs in the least-squares sense. The hybrid orbitals from OAOs consequently resemble hybrid orbitals that are based on the original AOs. In the third method, all basis functions are orthogonal to each other; however, in the fourth method, orthogonalization is performed only locally near each boundary.

A. Projected Basis Method. The essential step of this approach is to construct a set of basis functions in the QM region, orthogonal to the auxiliary orbitals, by the following projection:

$$|\tilde{\chi}_u\rangle = (1 - \sum_{b=1}^{3} S_{ub}^2)^{-1/2}(|\chi_u\rangle - \sum_{b=1}^{3}|\eta_b\rangle\langle\eta_b|\chi_u\rangle) \qquad (u = 1, 2, ..., N;\; b = x, y, z) \quad (23)$$

where S_{ub} is the overlap integral between χ_u and η_b, and the fully QM basis functions after the projection are denoted as $\tilde{\chi}_u$. The normalization factor is obtained based on the fact that the hybrid basis functions are constructed to be orthonormal to each other (eq 2). Instead of using eq 3, the $N+1$ active MOs are expanded over the N projected fully QM basis functions ($\tilde{\chi}_u$) and one active hybrid basis function (η_{B}):

$$\varphi_i = \sum_{u=1}^{N}\tilde{c}_{ui}\tilde{\chi}_u + c_{\mathrm{B}i}\eta_{\mathrm{B}} = \sum_{a=1}^{N+1}\tilde{c}_{ai}\tilde{\chi}_a \quad (i = 1, 2, ..., N+1) \quad (24)$$

where $\tilde{\chi}_{N+1} = \chi_{N+1} = \eta_{\mathrm{B}}$. The auxiliary MOs are still the auxiliary hybrid basis functions as expressed in eq 4. The orthogonality between active MOs and auxiliary MOs is satisfied, because of the orthogonality between $\tilde{\chi}_u$ and η_b, similar to eq 15:

$$\tilde{S}_{ub} = \langle\tilde{\chi}_u|\eta_b\rangle = 0 \qquad (u = 1, 2, ..., N;\; b = x, y, z) \quad (25)$$

Thus, the overall transformation from the $N + 4$ atomic orbitals into the projected orthogonal hybrid atomic orbitals can be expressed as follows:

$$\tilde{\chi} = \mathbf{T}\chi \tag{26}$$

where $\tilde{\chi}$ is a column matrix of the projected orbitals and χ is the column matrix of the atomic orbitals. The transformation matrix is given below:

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_b \end{pmatrix}\mathbf{M} \tag{27}$$

where $\mathbf{T}_b$ is the 4×4 hybridization transformation matrix and $\mathbf{M}$ relates the hybrid (H) basis to the projected hybrid (PH) basis:

$$\mathbf{M} = \begin{pmatrix} c_1 & 0 & .. & 0 & 0 & 0 & 0 & 0 \\ 0 & c_2 & .. & 0 & 0 & 0 & 0 & 0 \\ : & : & : & : & : & : & : & : \\ 0 & 0 & .. & c_N & 0 & 0 & 0 & 0 \\ 0 & 0 & .. & 0 & 1 & 0 & 0 & 0 \\ -c_1S_{1x} & -c_2S_{2x} & .. & -c_NS_{Nx} & 0 & 1 & 0 & 0 \\ -c_1S_{1y} & -c_2S_{2y} & .. & -c_NS_{Ny} & 0 & 0 & 1 & 0 \\ -c_1S_{1z} & -c_2S_{2z} & .. & -c_NS_{Nz} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{28}$$

where c_u ($u = 1, 2, ..., N$) in eq 28 is the normalization factor for the projected basis $\tilde{\chi}_u$:

$$c_u = (1 - \sum_{b=1}^{3} S_{ub}^2)^{-1/2} \qquad (u = 1, 2, ..., N) \tag{29}$$

Because the hybrid basis functions are not changed by the projection operation, their transformation matrix is a unit submatrix in $\mathbf{M}$. The projection operation makes the active basis orthogonal to the auxiliary basis but retains the strictly localized character of the auxiliary orbitals.

The algorithm for the SCF procedure using the projected hybrid basis function is given below:

(1) Form the projection matrix $\mathbf{M}$ according to eqs 28 and 29.

(2) Form the total transformation matrix $\mathbf{T}$ from the AO basis to the PH basis using eq 27.

(3) Transform the overlap matrix to the PH basis:

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

(4) Drop columns and rows corresponding to the auxiliary orbitals to obtain the reduced overlap matrix for the active orbitals $\mathbf{S}_{N+1}^{PH}$:

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

(5) Form the Löwdin transformation matrix for the active PH basis: $(\mathbf{S}_{N+1}^{PH})^{-1/2}$.

(6) Guess the total density matrix in the AO basis: $\mathbf{P}_{N+4}^{AO}$.

(7) Form the total Fock matrix in the AO basis: $\mathbf{F}_{N+4}^{AO}$.

(8) Transform the Fock matrix from the AO basis to the PH basis:

$$\mathbf{F}_{N+4}^{PH} = \mathbf{T}^{\dagger}\mathbf{F}_{N+4}^{AO}\mathbf{T}$$

(9) Drop columns and rows corresponding to auxiliary orbitals to obtain the reduced Fock matrix for the active orbitals $\mathbf{F}_{N+1}^{PH}$:

$$\mathbf{F}_{N+4}^{PH} \xrightarrow{\text{drop auxiliary}} \mathbf{F}_{N+1}^{PH}$$

(10) Löwdin transform the Fock matrix from the PH basis to the orthogonalized hybrid (OH) basis:

$$\mathbf{F}_{N+1}^{OH} = (\mathbf{S}_{N+1}^{PH})^{-1/2}\mathbf{F}_{N+1}^{PH}(\mathbf{S}_{N+1}^{PH})^{-1/2}$$

(11) Diagonalize the Fock matrix in the OH basis to obtain a new set of active MOs by solving Roothaan's equation in the $(N + 1)$-dimensional active space:

$$\mathbf{F}_{N+1}^{OH}\mathbf{C}_{N+1}^{OH} = \mathbf{C}_{N+1}^{OH}\epsilon$$

(12) Back Löwdin transform the MOs to the PH basis:

$$\mathbf{C}_{N+1}^{PH} = (\mathbf{S}_{N+1}^{PH})^{-1/2}\mathbf{C}_{N+1}^{OH}$$

(13) Form the active density matrix in the PH basis: $\mathbf{P}_{N+1}^{PH}$.

(14) Add the diagonal auxiliary density P_{bb} to the active density matrix, i.e., form the total density matrix in the PH basis: $\mathbf{P}_{N+4}^{PH}$.

(15) Transform the density matrix back to the AO basis:

$$\mathbf{P}_{N+4}^{AO} = \mathbf{T}\mathbf{P}_{N+4}^{PH}\mathbf{T}^{\dagger}$$

(16) Compute the total energy and test for SCF convergence. If not yet converged, go to step (7).

To derive the analytical gradient expression for the GHO−AIHF method based on the projected basis, we start from eq 16, and derivatives on $\mathbf{P}^{AO}$and $\mathbf{W}^{AO}$are expressed as

$$\frac{\partial \mathbf{P}^{AO}}{\partial \boldsymbol{q}} = \frac{\partial(\mathbf{T}\mathbf{P}^{PH}\mathbf{T}^{\dagger})}{\partial \boldsymbol{q}} = \frac{\partial \mathbf{T}}{\partial \boldsymbol{q}}\mathbf{P}^{PH}\mathbf{T}^{\dagger} + \mathbf{T}\frac{\partial \mathbf{P}^{PH}}{\partial \boldsymbol{q}}\mathbf{T}^{\dagger} + \mathbf{T}\mathbf{P}^{PH}\frac{\partial \mathbf{T}^{\dagger}}{\partial \boldsymbol{q}} \tag{30}$$

$$\frac{\partial \mathbf{W}^{AO}}{\partial \boldsymbol{q}} = \frac{\partial(\mathbf{T}\mathbf{W}^{PH}\mathbf{T}^{\dagger})}{\partial \boldsymbol{q}} = \frac{\partial \mathbf{T}}{\partial \boldsymbol{q}}\mathbf{W}^{PH}\mathbf{T}^{\dagger} + \mathbf{T}\frac{\partial \mathbf{W}^{PH}}{\partial \boldsymbol{q}}\mathbf{T}^{\dagger} + \mathbf{T}\mathbf{W}^{PH}\frac{\partial \mathbf{T}^{\dagger}}{\partial \boldsymbol{q}} \tag{31}$$

where basis transformation matrix $\mathbf{T}$ is defined in eq 27. The derivatives on the transformation matrix $\mathbf{T}$ can be written as

$$\frac{\partial \mathbf{T}}{\partial \boldsymbol{q}} = \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_b \end{pmatrix}\frac{\partial \mathbf{M}}{\partial \boldsymbol{q}} + \begin{pmatrix} 0 & 0 \\ 0 & \partial \mathbf{T}_b/\partial \boldsymbol{q} \end{pmatrix}\mathbf{M} \tag{32}$$

where $\mathbf{M}$ is the projection transformation matrix defined by eqs 28 and 29. The derivatives of elements in $\mathbf{M}$ can be formulated as

$$\frac{\partial c_u}{\partial \boldsymbol{q}} = \frac{\partial(1 - \sum_{b=1}^{3} S_{ub}^2)^{-1/2}}{\partial \boldsymbol{q}} = c_u^3 \sum_{b=1}^{3} S_{ub}\frac{\partial S_{ub}}{\partial \boldsymbol{q}} \tag{33}$$

$$\frac{\partial(-c_u S_{ub})}{\partial \boldsymbol{q}} = -\frac{\partial c_u}{\partial \boldsymbol{q}}S_{ub} - c_u\frac{\partial S_{ub}}{\partial \boldsymbol{q}} \tag{34}$$

Recall that S_{ub} is the overlap integral between χ_u and η_b in the hybrid basis; thus,

$$\frac{\partial \mathbf{S}^{\mathrm{H}}}{\partial \boldsymbol{q}} = \frac{\partial \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix}^{\dagger} \mathbf{S}^{\mathrm{AO}} \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix}}{\partial \boldsymbol{q}}$$

$$= \frac{\partial \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix}^{\dagger}}{\partial \boldsymbol{q}} \mathbf{S}^{\mathrm{AO}} \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix}^{\dagger} \frac{\partial \mathbf{S}^{\mathrm{AO}}}{\partial \boldsymbol{q}} \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix}^{\dagger} \mathbf{S}^{\mathrm{AO}} \frac{\partial \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix}}{\partial \boldsymbol{q}} \quad (35)$$

The partial derivatives in the first and third terms only operate on the lower right-hand element of the matrices.

B. NDDAO Approximation. If the MO expansion strictly follow eqs 3 and 4, as we discussed in Section IIB, eq 15 is a necessary and sufficient condition to ensure that the auxiliary MOs are orthogonal to the active MOs. The simplest approach is to follow semiempirical approximations, in which the orbital orthogonality of eq 15 is ensured by neglecting the diatomic differential overlap between the auxiliary orbitals η_b and basis functions χ_u on QM atoms. This is called the neglect of diatomic differential auxiliary overlap (NDDAO) approximation. For consistency, two-electron integrals that involve the differential overlap $\chi_u\eta_b$ also are neglected, because such a charge distribution vanishes under the NDDAO approximation:

$$(\chi_u\eta_b|XY) = 0 \quad (36)$$

where X and Y denote any basis functions. To elucidate the effects of the NDDAO approximation on the SCF procedure and the GHO energy, we first write the Fock matrix in the hybrid basis, without any approximation:

$$F^{\mathrm{H}}_{u\upsilon} = H^{\mathrm{H}}_{u\upsilon} + \sum_{st=1}^{N+1} P^{\mathrm{H}}_{st}\left[(st|u\upsilon) - \frac{1}{2}(su|t\upsilon)\right] + \sum_{b=1}^{3} P^{\mathrm{H}}_{bb}\left[(bb|u\upsilon) - \frac{1}{2}(bu|b\upsilon)\right] \quad (37)$$

$$F^{\mathrm{H}}_{bb} = H^{\mathrm{H}}_{bb} + \sum_{u\upsilon=1}^{N} P^{\mathrm{H}}_{u\upsilon}\left[(u\upsilon|bb) - \frac{1}{2}(ub|\upsilon b)\right] + 2\sum_{u=1}^{N} P^{\mathrm{H}}_{u\mathrm{B}}\left[(u\mathrm{B}|bb) - \frac{1}{2}(ub|\mathrm{B}b)\right] + P^{\mathrm{H}}_{\mathrm{BB}}\left[(\mathrm{BB}|bb) - \frac{1}{2}(\mathrm{B}b|\mathrm{B}b)\right] + \sum_{c=1}^{3} P^{\mathrm{H}}_{cc}\left[(cc|bb) - \frac{1}{2}(cb|cb)\right] \quad (38)$$

$$F^{\mathrm{H}}_{u\mathrm{B}} = H^{\mathrm{H}}_{u\mathrm{B}} + \sum_{st=1}^{N+1} P^{\mathrm{H}}_{st}\left[(st|u\mathrm{B}) - \frac{1}{2}(su|t\mathrm{B})\right] + \sum_{b=1}^{3} P^{\mathrm{H}}_{bb}\left[(bb|u\mathrm{B}) - \frac{1}{2}(bu|b\mathrm{B})\right] \quad (39)$$

$$F^{\mathrm{H}}_{\mathrm{BB}} = H^{\mathrm{H}}_{\mathrm{BB}} + \sum_{st=1}^{N+1} P^{\mathrm{H}}_{st}\left[(st|\mathrm{BB}) - \frac{1}{2}(s\mathrm{B}|t\mathrm{B})\right] + \sum_{b=1}^{3} P^{\mathrm{H}}_{bb}\left[(bb|\mathrm{BB}) - \frac{1}{2}(b\mathrm{B}|b\mathrm{B})\right] \quad (40)$$

where the indices $\{u,\upsilon\}$ for fully QM basis function run from 1 to N in eqs 37−39. Note that other off-diagonal terms of the Fock matrix that involve η_b (F^{H}_{ub}, $F^{\mathrm{H}}_{\mathrm{B}b}$, and F^{H}_{bc}) are not presented explicitly here, because they neither participate in the SCF iteration (dropped as auxiliary entries) nor enter the energy evaluation (weighted by zero densities P^{H}_{ub}). Because there is no $\chi_u\eta_b$ differential overlap in eq 40, the $F^{\mathrm{H}}_{\mathrm{BB}}$ term remains unchanged. According to eq 36, the exchange integrals $(bu|b\upsilon)$ in eq 37, $(ub|\upsilon b)$ and $(ub|\mathrm{B}b)$ in eq 38, and $(bu|b\mathrm{B})$ in eq 39 should be neglected, as a consequence of the NDDAO approximation. However, we do not neglect the $(ub|\mathrm{B}b)$ and $(bu|b\mathrm{B})$ terms, because it is more complicated to code this because of the occurrence of three hybrid orbitals. Therefore, only eqs 37 and 38 are affected by the NDDAO approximation as implemented here. The modified Fock matrix then can be written as the sum of the original Fock matrix and a correction term:

$$(F^{\mathrm{H}}_{u\upsilon})^{\mathrm{NDDAO}} = F^{\mathrm{H}}_{u\upsilon} + \frac{1}{2}\sum_{b=1}^{3} P^{\mathrm{H}}_{bb}(bu|b\upsilon) \quad (41)$$

$$(F^{\mathrm{H}}_{bb})^{\mathrm{NDDAO}} = F^{\mathrm{H}}_{bb} + \frac{1}{2}\sum_{u\upsilon=1}^{N} P^{\mathrm{H}}_{u\upsilon}(ub|\upsilon b) \quad (42)$$

Although integrals $(ub|\upsilon b)$ in eqs 41 and 42 are expressed in the hybrid basis, one can easily rewrite them in terms of the readily available AO integrals:

$$(ub|\upsilon b) = \sum_{\lambda,\sigma}^{s,p\ \text{on B}} (T_{\mathrm{b}})_{\lambda b}(T_{\mathrm{b}})_{\sigma b}(\lambda u|\sigma\upsilon) \quad (43)$$

Combining eqs 41−43, the modified Fock matrix, using the NDDAO approximation, can be expressed as

$$(F^{\mathrm{H}}_{u\upsilon})^{\mathrm{NDDAO}} = F^{\mathrm{H}}_{u\upsilon} + \frac{1}{2}\sum_{b=1}^{3} P^{\mathrm{H}}_{bb} \sum_{\lambda,\sigma}^{s,p\ \text{on B}} (T_{\mathrm{b}})_{\lambda b}(T_{\mathrm{b}})_{\sigma b}(\lambda u|\sigma\upsilon) \quad (44)$$

$$(F^{\mathrm{H}}_{bb})^{\mathrm{NDDAO}} = F^{\mathrm{H}}_{bb} + \frac{1}{2}\sum_{u\upsilon=1}^{N} P^{\mathrm{H}}_{u\upsilon} \sum_{\lambda,\sigma}^{s,p\ \text{on B}} (T_{\mathrm{b}})_{\lambda b}(T_{\mathrm{b}})_{\sigma b}(u\lambda|\upsilon\sigma) \quad (45)$$

With the NDDAO correction terms added in, the electron−electron repulsion between electrons present in auxiliary orbitals and the orbitals on fully QM atoms is increased, and exchange between such orbitals is eliminated under the NDDAO assumption. The modified electrostatic fields due to the NDDAO approximation are consistently reflected in the modified Fock matrix. Before finalizing the algorithm, we note that the one-electron integrals that involve η_b (H^{H}_{ub}, $H^{\mathrm{H}}_{\mathrm{B}b}$, and H^{H}_{bc}) are not necessarily kept unchanged if one neglects the $\chi_u\eta_b$ differential overlap; however, no further modifications on them are needed in GHO, because of the same reasons as those for the F^{H}_{ub} terms.

Next, we present a practical procedure to perform a GHO−AIHF calculation using the NDDAO approximation. We define the total transformation matrix $\mathbf{T}$ between the AO basis and the hybrid basis as

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & \mathbf{T}_{\mathrm{b}} \end{pmatrix} \quad (46)$$

The SCF procedure for the NDDAO method is modified correspondingly as follows:

(1) Transform the overlap matrix to the hybrid basis:

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

(2) Drop columns and rows for auxiliary entries of overlap matrix in the hybrid basis:

$$\mathbf{S}^{\mathrm{H}}_{N+4} \xrightarrow{\text{drop auxiliary}} \mathbf{S}^{\mathrm{H}}_{N+1}$$

(3) Form the Löwdin transformation matrix for the active hybrid basis: $(\mathbf{S}^{\mathrm{H}}_{N+1})^{-1/2}$.

(4) Guess the total density matrix in the AO basis: $\mathbf{P}^{\mathrm{AO}}_{N+4}$.

(5) Form the total Fock matrix in the AO basis: $\mathbf{F}^{\mathrm{AO}}_{N+4}$.

(6) Transform the Fock matrix in the AO basis to the hybrid basis:

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

(7) Modify the total Fock matrix in the hybrid basis according to eqs 44 and 45, to take into consideration the NDDAO approximation.

(8) Drop columns and rows corresponding to the auxiliary orbitals to obtain the reduced Fock matrix for active orbitals $\mathbf{F}^{\mathrm{H}}_{N+1}$:

$$\mathbf{F}^{\mathrm{H}}_{N+4} \xrightarrow{\text{drop auxiliary}} \mathbf{F}^{\mathrm{H}}_{N+1}$$

(9) Löwdin transform the Fock matrix from the hybrid basis to the OH basis:

$$\mathbf{F}^{\mathrm{OH}}_{N+1} = (\mathbf{S}^{\mathrm{H}}_{N+1})^{-1/2}\mathbf{F}^{\mathrm{H}}_{N+1}(\mathbf{S}^{\mathrm{H}}_{N+1})^{-1/2}$$

(10) Diagonalize the Fock matrix in the OH basis to obtain a new set of active MOs by solving Roothaan's equation in the active space:

$$\mathbf{F}^{\mathrm{OH}}_{N+1}\mathbf{C}^{\mathrm{OH}}_{N+1} = \mathbf{C}^{\mathrm{OH}}_{N+1}\epsilon$$

(11) Back Löwdin transform the MOs to the nonorthogonal hybrid basis:

$$\mathbf{C}^{\mathrm{H}}_{N+1} = (\mathbf{S}^{\mathrm{H}}_{N+1})^{-1/2}\mathbf{C}^{\mathrm{OH}}_{N+1}$$

(12) Form the active density matrix in the hybrid basis $\mathbf{P}^{\mathrm{H}}_{N+1}$.

(13) Add the diagonal auxiliary density P_{bb} to the active density matrix, i.e., form the total density matrix in the hybrid basis $\mathbf{P}^{\mathrm{H}}_{N+4}$.

(14) Transform the density matrix back to the AO basis:

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

(15) Compute the total energy and test for convergence. If not yet converged, go back to step (5).

The analytical gradient in the NDDAO approximation involves extra terms for the neglected exchange energies:

$$\left(\frac{\partial E}{\partial \boldsymbol{q}}\right)^{\mathrm{NDDAO}} = \left(\frac{\partial E}{\partial \boldsymbol{q}}\right) + \frac{1}{2}\sum_{b=1}^{3} P^{\mathrm{H}}_{bb} \sum_{u\nu=1}^{N} P^{\mathrm{H}}_{u\nu} \sum_{\lambda,\sigma}^{s,p\ \text{on B}} \frac{\partial}{\partial \boldsymbol{q}}[(T_{\mathrm{b}})_{\lambda b}(T_{\mathrm{b}})_{\sigma b}(u\lambda|\nu\sigma)] \quad (47)$$

where the first term ($\partial E/\partial \boldsymbol{q}$) represents the gradients computed from eq 16, using derivative routines available for a Fock matrix without NDDAO correction terms. The second correction term involves derivatives of the two-electron integral $\partial(u\lambda|\nu\sigma)/\partial \boldsymbol{q}$, and the basis transformation matrix $\partial \mathbf{T}/\partial \boldsymbol{q}$.

C. Global Löwdin Orthogonalization (GLO) Method. The Löwdin symmetric orthogonalization method[54,55] can be used to construct a full set of orthogonalized atomic orbitals (OAOs),[54,55] which can be constructed from the AO basis for all QM atoms (including the GHO boundary atom B) by the following transformation:

$$\begin{pmatrix} \chi^{\mathrm{L}}_1 \\ \vdots \\ \chi^{\mathrm{L}}_N \\ s^{\mathrm{L}} \\ p^{\mathrm{L}}_x \\ p^{\mathrm{L}}_y \\ p^{\mathrm{L}}_z \end{pmatrix} = \mathbf{T}^{\mathrm{LO}} \begin{pmatrix} \chi_1 \\ \vdots \\ \chi_N \\ s \\ p_x \\ p_y \\ p_z \end{pmatrix} \quad (48)$$

where $\{s,p_x,p_y,p_z\}$ denotes the AO basis on B, and χ_u ($u = 1, 2, ..., N$) are basis functions on fully QM atoms; the corresponding OAO basis functions are labeled by superscript "L", and $\mathbf{T}^{\mathrm{LO}}$ is the Löwdin orthogonalization transformation matrix, which is given by the relation[54,55]

$$\mathbf{T}^{\mathrm{LO}} = (\mathbf{S}^{\mathrm{AO}}_{N+4})^{-1/2} \quad (49)$$

where $\mathbf{S}^{\mathrm{AO}}_{N+4}$ is the overlap matrix for all $N + 4$ AO basis functions. The OAOs form an orthonormal set, i.e.,

$$\mathbf{S}^{\mathrm{OAO}}_{N+4} = \mathbf{I} \quad (50)$$

One can show that the OAOs resemble the original AOs maximally in the least-squares sense.[54,55] Specifically, $\{s^{\mathrm{L}},p^{\mathrm{L}}_x,p^{\mathrm{L}}_y,p^{\mathrm{L}}_z\}$ resembles $\{s,p_x,p_y,p_z\}$ on the boundary atom B. If we construct the hybrid basis by a hybridization of $\{s^{\mathrm{L}},p^{\mathrm{L}}_x,p^{\mathrm{L}}_y,p^{\mathrm{L}}_z\}$, the orbitals of the resulting orthogonalized hybrid (OH) basis also maximally resemble those obtained from direct hybridization of the original AOs:

$$\begin{pmatrix} \eta_{\mathrm{B}} \\ \eta_x \\ \eta_y \\ \eta_z \end{pmatrix} = \mathbf{T}_{\mathrm{b}}{}^{\dagger} \begin{pmatrix} s^{\mathrm{L}} \\ p^{\mathrm{L}}_x \\ p^{\mathrm{L}}_y \\ p^{\mathrm{L}}_z \end{pmatrix} \quad (51)$$

In this spirit, the OH basis retains the major chemical characteristics of the original hybrid basis; however, the functions of the OH basis are orthogonal to the remaining N basis functions χ^{L}_u ($u = 1, 2, ..., N$), which are not hybridized:

$$\mathbf{S}^{\mathrm{OH}}_{N+4} = \mathbf{I} \quad (52)$$

Because the Löwdin OAO basis contains "orthogonalization tails" from other atom centers,[56] the hybrid basis defined by eq 51 also contains these tails. Although it might be argued that the delocalized nature of the hybrid Löwdin OAOs makes them inappropriate for use in the GHO method, we have determined that this does not present a problem. Strictly localized orbitals were strongly favored in the LSCF and in the frozen-orbital treatments, because, in that type of method, the localized orbitals represent the frontier QM/MM bonding orbitals. The major motivation of using a localized orbital in those methods is to promote transferability of the bonding orbital. In the GHO approach, the QM frontier bonding orbital is actually described as a delocalized orbital, because the active hybrid basis function η_{B} is allowed to mix with other active basis functions in forming active MOs. Thus, localization of hybrid orbitals is not a requirement for the GHO method, at least for the active hybrid

orbital η_B. Furthermore, in the AO basis, all *s* and *p* orbitals on the B atom have non-negligible overlaps with basis functions from other QM atoms. The largest overlaps are between the basis on the B atom and its QM neighbor the A atom. The basis functions on the B atom become delocalized to other centers in the SCF step, regardless of whether we orthogonalize them to the fully QM basis. Actually, Löwdin OAOs are more localized than AOs in some aspects. For example, the nodal regions of Löwdin OAOs around other centers effectively prevent electrons from penetrating to the neighborhood of other atoms.[56,57] For these reasons, we think the inclusion of orthogonalization tails in the hybrid basis is physically justified. Our further investigations show that the orthogonalization can be localized to orbitals on atoms close to the GHO boundary, because only those orbitals have significant overlaps with the auxiliary orbitals, and the smaller overlaps with atoms far away can be neglected, as in the NDDAO method. The local orthogonalization scheme is considered in Subsection IIID. In the remainder of this section, we present the algorithm using Löwdin orthogonalization over the entire molecule.

The total transformation matrix **T** that relates the AO basis to the OH basis is

$$\mathbf{T} = \mathbf{T}^{\mathrm{LO}}\begin{pmatrix}\mathbf{I}_N & 0\\ 0 & \mathbf{T}_b\end{pmatrix} \tag{53}$$

The modified SCF procedure for the hybrid Löwdin method then can be described as given below:

(1) Guess the total density matrix in the AO basis $\mathbf{P}^{\mathrm{AO}}_{N+4}$.

(2) Form the total Fock matrix in the AO basis $\mathbf{F}^{\mathrm{AO}}_{N+4}$.

(3) Transform the Fock matrix from the AO basis to the OH basis:

$$\mathbf{F}^{\mathrm{OH}}_{N+4} = \mathbf{T}^{\dagger}\mathbf{F}^{\mathrm{AO}}_{N+4}\mathbf{T}$$

(4) Drop columns and rows corresponding to auxiliary orbitals to obtain the reduced Fock matrix for active orbitals $\mathbf{F}^{\mathrm{OH}}_{N+1}$:

$$\mathbf{F}^{\mathrm{OH}}_{N+4} \xrightarrow{\text{drop auxiliary}} \mathbf{F}^{\mathrm{OH}}_{N+1}$$

(5) Diagonalize the Fock matrix in the $(N+1)$-dimensional OH basis to obtain a new set of active MOs by solving Roothaan's equation in the active space:

$$\mathbf{F}^{\mathrm{OH}}_{N+1}\mathbf{C}^{\mathrm{OH}}_{N+1} = \mathbf{C}^{\mathrm{OH}}_{N+1}\epsilon$$

(6) Form the active density matrix in OH active basis $\mathbf{P}^{\mathrm{OH}}_{N+1}$.

(7) Add the diagonal auxiliary density P_{bb} to the active density matrix, i.e., form the total density matrix in the OH basis $\mathbf{P}^{\mathrm{OH}}_{N+4}$.

(8) Transform the density matrix to the AO basis:

$$\mathbf{P}^{\mathrm{AO}}_{N+4} = \mathbf{T}\mathbf{P}^{\mathrm{OH}}_{N+4}\mathbf{T}^{\dagger}$$

(9) Compute the total energy and test for SCF convergence. If not yet converged, go back to step (2).

Finally, we formulate the analytical gradient for GHO−AIHF using hybrid Löwdin OAOs; eq 16 is the starting point. Density derivatives on $\mathbf{P}^{\mathrm{AO}}$and $\mathbf{W}^{\mathrm{AO}}$, similar to eqs 17 and 18, can be obtained by

$$\frac{\partial \mathbf{P}^{\mathrm{AO}}}{\partial \boldsymbol{q}} = \frac{\partial(\mathbf{T}\mathbf{P}^{\mathrm{OH}}\mathbf{T}^{\dagger})}{\partial \boldsymbol{q}} = \frac{\partial \mathbf{T}}{\partial \boldsymbol{q}}\mathbf{P}^{\mathrm{OH}}\mathbf{T}^{\dagger} + \mathbf{T}\mathbf{P}^{\mathrm{OH}}\frac{\partial \mathbf{T}^{\dagger}}{\partial \boldsymbol{q}} + \mathbf{T}\frac{\partial \mathbf{P}^{\mathrm{OH}}}{\partial \boldsymbol{q}}\mathbf{T}^{\dagger} \tag{54}$$

$$\frac{\partial \mathbf{W}^{\mathrm{AO}}}{\partial \boldsymbol{q}} = \frac{\partial(\mathbf{T}\mathbf{W}^{\mathrm{OH}}\mathbf{T}^{\dagger})}{\partial \boldsymbol{q}} = \frac{\partial \mathbf{T}}{\partial \boldsymbol{q}}\mathbf{W}^{\mathrm{OH}}\mathbf{T}^{\dagger} + \mathbf{T}\mathbf{W}^{\mathrm{OH}}\frac{\partial \mathbf{T}^{\dagger}}{\partial \boldsymbol{q}} + \mathbf{T}\frac{\partial \mathbf{W}^{\mathrm{OH}}}{\partial \boldsymbol{q}}\mathbf{T}^{\dagger} \tag{55}$$

where the basis transformation matrix **T** is defined according eq 53; again, the last terms of eqs 54 and 55 have already been collected in conventional HF gradient calculations. The derivatives on the transformation matrix **T** can be written as

$$\frac{\partial \mathbf{T}}{\partial \boldsymbol{q}} = \mathbf{T}^{\mathrm{LO}}\begin{pmatrix}0 & 0\\ 0 & \partial \mathbf{T}_b/\partial \boldsymbol{q}\end{pmatrix} + \frac{\partial \mathbf{T}^{\mathrm{LO}}}{\partial \boldsymbol{q}}\begin{pmatrix}\mathbf{I} & 0\\ 0 & \mathbf{T}_b\end{pmatrix} \tag{56}$$

The derivatives of the Löwdin transformation matrix have been formulated in the development of a previous algorithm,[58] and they are directly available in the form of $\partial(\mathbf{S}^{\mathrm{AO}}_{N+4})^{+1/2}/\partial \boldsymbol{q}$. Utilizing this available result, one can obtain $\partial \mathbf{T}^{\mathrm{LO}}/\partial \boldsymbol{q}$, as required for eq 56, by the following transformation, which is derived by combining eq 49 with the result in Appendix A:

$$\frac{\partial \mathbf{T}^{\mathrm{LO}}}{\partial \boldsymbol{q}} = -(\mathbf{S}^{\mathrm{AO}}_{N+4})^{-1/2}\frac{\partial(\mathbf{S}^{\mathrm{AO}}_{N+4})^{+1/2}}{\partial \boldsymbol{q}}(\mathbf{S}^{\mathrm{AO}}_{N+4})^{-1/2} \tag{57}$$

As an alternative to the aforementioned procedure, one can first hybridize AOs on the B atom to create a set of hybrid basis functions; the orthogonalized hybrid (OH) basis then can be created by Löwdin orthogonalization of the hybrid basis to the AOs on other QM atoms. However, it can be shown that the hybridization and Löwdin orthogonalization operators commute with each other,[56] and we also performed numerical tests showing that the total energy is invariant, with respect to the order of hybridization and orthogonalization. Because this variant is identical to the previous one, in terms of energy and gradients, we do not include any detailed discussion of it.

D. Local Löwdin Orthogonalization (LLO) Method. Finally, we consider a variation of the GHO−AIHF method using Löwdin OAOs. Instead of doing a Löwdin orthogonalization over the entire molecule, we restrict the orthogonalization to orbitals near the GHO boundary. This scheme should be particularly attractive, at least from a conceptual viewpoint, for a large QM subsystem or for QM subsystems with two or more boundary atoms, because the effect of each boundary is entirely localized to the region near that boundary. If one optimizes integral scale factors or new MM parameters to correct imbalances near the transition state empirically, those parameters might be expected to be more transferable if one uses local Löwdin orthogonalization.

The Löwdin orthogonalization (LLO) method is identical to the global Löwdin orthogonalization (GLO) method in all respects, except the definition of $\mathbf{T}^{\mathrm{LO}}$. In the LLO method, one retains only overlap integrals that involve boundary atoms and a subset of fully QM atoms. For example, in Section V, we will consider including only overlap integrals that involve (LLO:F) boundary atoms and frontier atoms, (LLO:FG) boundary atoms, frontier atoms, and geminal atoms (i.e., A atoms and other atoms bonded to A atoms), or (LLO:FGV) boundary atoms, atoms bonded to, geminal to, and vicinal to boundary atoms.

Although the local orthogonalization method is preferred, because it makes the minimal changes in the quantum mechanical treatment in regions far from the boundary, one must be careful to avoid conformations where the distant regions of a chain of QM atoms bend back toward the boundary. Such problems do not occur for any of the test cases considered in

TABLE 2: Optimized Geometries and Partial Charges for Ethane at the Pure HF Level with Conventional Basis Sets[a]

	Bond Distance (Å)			Angle (deg)		Mulliken Charge		Löwdin Charge	
basis	C_A-C_B	C_A-C_H	C_B-C_H	$H-C_A-C_B$	C_A-C_B-H	q_A	q_B	q_A	q_B
STO-3G	1.538	1.086	1.086	110.7	110.7	−0.17	−0.17	−0.08	−0.08
STO-4G	1.535	1.082	1.082	110.7	110.7	−0.18	−0.18	−0.09	−0.09
3-21G	1.542	1.084	1.084	110.8	110.8	−0.60	−0.60	−0.26	−0.26
6-31G	1.530	1.084	1.084	111.2	111.2	−0.45	−0.45	−0.31	−0.31
6-31G(d)	1.527	1.086	1.086	111.2	111.2	−0.48	−0.48	−0.44	−0.44
6-31G(d,p)	1.527	1.086	1.086	111.2	111.2	−0.33	−0.33	−0.29	−0.29
6-31+G(d)	1.528	1.086	1.086	111.2	111.2	−0.55	−0.55	−0.58	−0.58
6-31+G(d,p)	1.527	1.086	1.086	111.2	111.2	−0.37	−0.37	−0.43	−0.43
MIDI!	1.541	1.087	1.087	111.0	111.0	−0.55	−0.55	−0.27	−0.27

[a] Symmetry is not imposed in these reference calculations. For comparison to GHO−AIHF results, we refer to the two carbons as the A and B atoms, even though there is actually no GHO boundary atom in a pure QM calculation. Note that $q_{AH_3} = q_{BH_3} = 0$ in these calculations.

TABLE 3: Optimized Geometries and Partial Charges for Ethane at the Pure HF Level with Mixed Basis Sets[a] and at the Pure MM Level (Last Row)

	Bond Distance (Å)			Angle (deg)		Mulliken Charge			Löwdin Charge		
basis	C_A-C_B	C_A-C_H	C_B-C_H	$H-C_A-C_B$	C_A-C_B-H	q_A	q_B	q_{AH_3}[b]	q_A	q_B	q_{AH_3}[b]
STO-3G	1.512	1.088	1.077	109.1	111.2	−0.02	−0.75	0.48	0.01	−0.40	0.13
STO-4G	1.509	1.084	1.078	109.0	111.2	−0.04	−0.75	0.48	0.00	−0.39	0.12
3-21G	1.534	1.085	1.075	109.6	110.4	−0.29	−0.85	0.58	−0.30	−0.26	−0.01
6-31G	1.515	1.085	1.078	109.5	111.5	−0.06	−0.88	0.61	−0.36	−0.23	−0.04
6-31G(d)	1.511	1.085	1.078	109.5	111.5	−0.06	−0.88	0.61	−0.48	−0.20	−0.07
6-31G(d,p)	1.509	1.088	1.077	110.1	111.2	0.06	−0.91	0.64	−0.43	−0.17	−0.10
6-31+G(d)	1.512	1.085	1.085	108.9	111.3	−0.20	−0.87	0.60	−0.81	−0.11	−0.16
6-31+G(d,p)	1.511	1.086	1.085	108.9	111.4	−0.03	−0.87	0.60	−0.66	−0.11	−0.16
MIDI!	1.531	1.088	1.074	109.1	111.1	−0.12	−0.90	0.63	−0.31	−0.26	−0.01
MM[c]	1.529	1.111	1.111	110.3	110.3	−0.27[d]	−0.27[d]	0.00[d]	−0.27[d]	−0.27[d]	0.00[d]

[a] The basis on one methyl group is the STO-3Gv basis, and the basis on the other methyl group is indicated in column 1. We refer to the C atom described by an STO-3Gv basis as C_B, and the C atom described by the larger basis as C_A. For the C atom described by STO-3Gv, we omit the core electrons and decrease the nuclear charge by two. [b] $q_{BH_3} = -q_{AH_3}$. [c] All MM results in this paper are for the CHARMM force field of ref 61. [d] MM partial charges.

the present paper; however, they could occur in other applications. In such cases, one should use the GLO method.

IV. Implementation

We have implemented all four of the proposed GHO−AIHF algorithms in the GAMESSPLUS package,[59,60] which works with the CHARMM/GAMESS interface.[28] The MM force field is the CHARMM22 parametrization.[61] For the MM term in the energy expression, we follow the original QM/MM paper by Field et al.,[3] i.e., we remove the MM terms that involve exclusively QM atoms (the GHO boundary atom B is also considered as a QM atom for this purpose). Following this principle, the torsion terms to be removed are different from the QM/MM implementation in the previous[28] CHARMM/GAMESS interface. In that interface, any *X*−QM−QM−*X* terms were removed, regardless of whether the 1,4 centers are QM atoms or MM atoms. This rule may be reasonable for QM/MM calculations when a link atom is used, but it is not valid for a GHO treatment. Removing all *X*−QM−QM−*X* torsion terms with GHO seriously underestimates the internal rotation barrier for the QM−QM bond if one of the central atoms is the frontier atom A and the other is the GHO boundary atom B. For that case, the QM Hamiltonian in GHO provides only a partial description of the torsion barrier (mainly due to the presence of electrons in auxiliary hybrid orbitals), and the major portion of the torsion is determined by the MM force field, which should not be removed from the total energy.

V. Tests of Unparametrized Method

In previous work, a wide variety of test data has been used for validation of combined QM/MM methods, including, for example, atomic charges,[6a,7,26,29,36,40] conformational energies,[35] diamides,[36] di-, tri-, and tetra-peptides,[6b,29,35] dipole moments,[35,39] geometries,[6b,26a,29a,36,39] HOMO and LUMO energies,[6a] potential energy curves for bond stretching,[36,39,40] protein simulations,[26b,44−48] proton affinities,[26b,29,35,36,39,40] proton transfer,[6a,7] size effects,[6b] torsion potentials,[6a,26,29,35,39] vibrational frequencies,[35] hydrogen bonding energies,[4] and zwitterions.[6b] In the present paper, we will consider atomic charges, dipeptides, geometries, proton affinities, size effects, torsion potentials, and zwitterions. However, to begin, we will concentrate on atomic charges and geometries, because we believe that these provide the most direct test of charge and interaction balance at the QM−MM interface.

In this section, we will present QM/MM results for several different basis sets in the fully QM region and with the STO-3Gv basis on the boundary atom. In this section, no integrals are scaled and no MM terms are altered. Thus, there are no new parameters, and we can test the criterion proposed in Section II, namely that the method should give qualitatively reasonable results without parametrization. Before considering these tests, we consider results from pure QM calculations with full basis sets and mixed basis sets. The QM basis sets include single-ζ valence basis set (STO-3G, STO-4G),[51] doubly split valence basis sets (3-21G, 6-31G),[62−64] basis sets that include polarization functions and diffuse functions [6-31G(d), 6-31G(d,p), 6-31+G(d), 6-31+G(d,p)],[64,65] and finally the MIDI! basis,[66] which was proposed to give accurate geometry and partial charges at the HF level. The results obtained with these basis sets, and with mixed basis sets where one methyl group is treated by STO-3Gv, are given in Tables 2 and 3 for comparison. (Table 3 also shows pure MM results.) Tables 2 and 3 show not only geometrical parameters but also partial

TABLE 4: QM/MM Optimized Geometries and Charges for Ethane by Unparametrized GHO−AIHF Level Using a Projected QM Basis

	Bond Distance (Å)			Angle (deg)		Mulliken Charge			Löwdin Charge		
basis	C_A-C_B	C_A-C_H	C_B-C_H	$H-C_A-C_B$	C_A-C_B-H	q_A	q_B	q_{AH_3}[a]	q_A	q_B	q_{AH_3}[a]
STO-3G	1.542	1.091	1.118	110.0	119.6	0.02	−0.47	0.20	0.10	−0.45	0.18
STO-4G	1.533	1.088	1.119	110.0	119.6	0.00	−0.47	0.20	0.08	−0.44	0.17
3-21G	1.514	1.083	1.119	112.1	120.3	−0.10	−0.54	0.27	−0.09	−0.30	0.03
6-31G	1.475	1.084	1.121	112.8	120.9	0.20	−0.61	0.34	−0.13	−0.28	0.01
6-31G(d)	1.473	1.085	1.121	113.2	121.0	0.15	−0.62	0.35	−0.36	−0.20	−0.07
6-31G(d,p)	1.470	1.086	1.121	113.4	121.0	0.31	−0.62	0.35	−0.23	−0.20	−0.07
6-31+G(d)	1.477	1.083	1.121	111.9	121.1	0.15	−0.63	0.36	−0.52	−0.17	−0.10
6-31+G(d,p)	1.475	1.085	1.121	112.0	121.1	0.32	−0.63	0.36	−0.38	−0.16	−0.11
MIDI!	1.497	1.086	1.119	112.2	120.7	0.05	−0.61	0.34	−0.10	−0.30	0.03

[a] $q_{BH_3} = -q_{AH_3}$.

TABLE 5: QM/MM Optimized Geometries and Charges for Ethane by Unparametrized GHO−AIHF with the NDDAO Approximation

	Bond Distance (Å)			Angle (deg)		Mulliken Charge			Löwdin Charge		
basis	C_A-C_B	C_A-C_H	C_B-C_H	$H-C_A-C_B$	C_A-C_B-H	q_A	q_B	q_{AH_3}[a]	q_A	q_B	q_{AH_3}[a]
STO-3G	1.288	1.086	1.134	112.0	111.2	−0.14	−0.30	0.03	−0.09	−0.28	0.01
STO-4G	1.277	1.082	1.135	112.0	111.1	−0.16	−0.29	0.02	−0.11	−0.27	0.00
3-21G	1.312	1.077	1.134	110.7	108.3	−0.60	−0.19	−0.08	−0.53	0.01	−0.28
6-31G	1.310	1.073	1.134	109.2	108.4	−0.44	−0.18	−0.09	−0.63	0.06	−0.33
6-31G(d)	1.258	1.077	1.135	109.2	108.7	−0.49	−0.15	−0.12	−0.89	0.20	−0.47
6-31G(d,p)	1.252	1.076	1.136	109.4	108.9	−0.34	−0.15	−0.12	−0.74	0.21	−0.48
6-31+G(d)	1.263	1.074	1.136	107.3	109.1	−0.53	−0.16	−0.11	−1.04	−0.24	−0.51
6-31+G(d,p)	1.257	1.074	1.136	107.6	109.2	−0.37	−0.16	−0.11	−0.89	0.25	−0.52
MIDI!	1.327	1.074	1.133	107.7	108.4	−0.53	−0.19	−0.08	−0.53	0.01	−0.28

[a] $q_{BH_3} = -q_{AH_3}$.

atomic charges obtained by Mulliken[67] population analysis and Löwdin[54,68] population analysis, which are denoted MPA and LPA, respectively.

First, we consider Table 2, which shows conventional full QM calculations with a range of basis sets. We see that geometries are relatively invariant to the basis set choice, but partial charges are dependent strongly on the basis set. Table 3 then shows the effect of describing one methyl group with the small STO-3Gv basis, while increasing the quality of the basis set on the other methyl group. Again, we see that geometries are relatively invariant to the lack of balance in the basis set, although using the mixed basis has a tendency to underestimate the C−C bond distances by ∼0.2 Å. Löwdin charges on the C atom with the small basis are similar to Löwdin charges when the entire molecule is treated with the small basis set, whereas Löwdin charges on the C atom with the large basis set are similar to Löwdin charges when the entire molecule is treated with the large basis set, which is encouraging, given the previously validated relatively good performance[69,70] of Löwdin charges for balanced basis sets. Mulliken charges, in contrast, are sometimes unstable.[69] In the present examples, Mulliken charges are uniformly more negative on the boundary atom than on the fully QM C atom. In contrast, Löwdin charges are less negative on the boundary atom when large basis sets are used for the fully QM atoms. We note the neither Mulliken nor Löwdin partial charges are used in any way in the QM/MM calculations; they are considered here as one possible test of whether the QM/MM boundary treatment has acceptably small charge transfer across nonpolar bonds. The importance of Table 3 is that it can be considered an upper bound on the quality of the performance that can be expected from GHO−AIHF. In other words, the most we should expect from unparametrized GHO−AIHF near the boundary (in ethane, all atoms are near the boundary) is the quality of the results in Table 3, although far from the boundary (in large molecules), one might hope for the quality of the results of Table 2.

We tested the unparametrized GHO−AIHF method for ethane, *n*-butane, *n*-octane, ethanol, ethylamine, and ethyl radical. First, we consider ethane. The QM/MM boundary is along the C−C bond in ethane. The GHO boundary C atom is labeled as C_B; and its QM neighbor C atom is labeled as C_A. Including the QM methyl group and C_B, there are five atoms to be treated quantum mechanically. An STO-3Gv basis is always assigned to the boundary carbon C_B; however, basis sets for fully QM atoms are not restricted, and we shall present results for several basis sets. We investigate the GHO−AIHF method with all four approaches to resolve the MO orthogonality constraint as described in Section III. The QM basis sets are the same as those used in Tables 2 and 3. To provide a comprehensive test of the methods, we consider geometries (obtained by optimization with analytic gradients), total energies (obtained from eq 11), and partial atomic charges. Partial atomic charges on QM atoms are calculated by MPA and LPA.

The GHO−AIHF optimized geometry and atomic charges obtained from population analysis with the projected basis, the NDDAO approximation, the global Löwdin orthogonalization (GLO) method, and the local Löwdin orthogonalization to frontier orbitals (LLO:F) treatments are listed in Tables 4−7. These tables show that reasonable geometries can be obtained for all the QM basis sets studied by GHO−AIHF QM/MM optimizations, even without any parametrization, although the trends are different than with the projection method. Because auxiliary orbitals are not SCF-optimized in GHO, these systematic deviations can be considered to be a normal consequence of using frozen orbitals. In comparison to the aforementioned two methods, the GLO method does not show a systematic error in this critical bond distance across the QM/MM boundary. Table 6 is encouraging, in that the C_A-C_B bond distance predicted by the third approach varies from 1.43 Å to 1.68 Å, only slightly deviating from the pure HF/full basis result of Table 2. Recall that the distinguishing feature of this method is that orthogonality tails on all other QM atoms are explicitly

TABLE 6: QM/MM Optimized Geometries and Charges for Ethane by Unparametrized GHO−AIHF, Based on Global Löwdin Orthogonalization (GLO)[a]

	Bond Distance (Å)			Angle (deg)		Mulliken Charge			Löwdin Charge		
basis	C_A-C_B	C_A-C_H	C_B-C_H	$H-C_A-C_B$	C_A-C_B-H	q_A	q_B	q_{AH_3}[b]	q_A	q_B	q_{AH_3}[b]
STO-3G	1.434	1.083	1.125	109.5	115.7	−0.04	−0.41	0.14	0.03	−0.39	0.12
STO-4G	1.426	1.079	1.125	109.4	115.8	−0.06	−0.40	0.13	0.01	−0.38	0.11
3-21G	1.548	1.077	1.120	107.2	113.7	−0.37	−0.46	0.19	−0.28	−0.24	−0.03
6-31G	1.521	1.077	1.121	107.0	114.0	−0.14	−0.52	0.25	−0.35	−0.22	−0.05
6-31G(d)	1.643	1.078	1.117	107.6	111.3	−0.31	−0.43	0.16	−0.56	−0.16	−0.11
6-31G(d,p)	1.668	1.080	1.116	107.1	111.2	−0.18	−0.43	0.16	−0.41	−0.17	−0.10
6-31+G(d)	1.652	1.076	1.117	106.9	110.8	−0.33	−0.45	0.18	−0.71	−0.13	−0.14
6-31+G(d,p)	1.678	1.077	1.116	106.5	110.7	−0.18	−0.45	0.18	−0.56	−0.14	−0.13
MIDI!	1.536	1.081	1.121	107.9	114.0	−0.21	−0.52	0.25	−0.29	−0.23	−0.04

[a] For this small molecule, the GLO method is identical to the LLO:FG method. [b] $q_{BH_3} = -q_{AH_3}$.

TABLE 7: QM/MM Optimized Geometries and Charges for Ethane by Unparametrized GHO−AIHF, Using Local Löwdin Orthogonalization to Frontier Orbitals (LLO:F) Only

	Bond Distance (Å)			Angle (deg)		Mulliken Charge			Löwdin Charge		
basis	C_A-C_B	C_A-C_H	C_B-C_H	$H-C_A-C_B$	C_A-C_B-H	q_A	q_B	q_{AH_3}[a]	q_A	q_B	q_{AH_3}[a]
STO-3G	1.413	1.082	1.126	108.7	116.5	−0.02	−0.42	0.15	0.03	−0.39	0.12
STO-4G	1.404	1.077	1.125	108.6	116.6	−0.04	−0.41	0.14	0.01	−0.38	0.11
3-21G	1.485	1.071	1.123	109.2	114.4	−0.28	−0.51	0.24	−0.26	−0.23	−0.04
6-31G	1.388	1.036	1.128	110.1	112.5	0.17	−0.67	0.40	−0.33	−0.19	−0.08
6-31G(d)	1.583	1.055	1.120	108.0	110.8	−0.15	−0.50	0.23	−0.57	−0.15	−0.12
6-31G(d,p)	1.581	1.056	1.120	108.0	110.8	−0.01	−0.50	0.23	−0.40	−0.15	−0.12
6-31+G(d)	1.593	1.066	1.119	108.9	110.9	−0.24	−0.48	0.21	−0.70	−0.13	−0.14
6-31+G(d,p)	1.590	1.066	1.119	109.1	110.8	−0.07	−0.49	0.22	−0.54	−0.13	−0.14
MIDI!	1.734	0.860	1.163	60.0	89.4	2.54	−1.29	1.02	0.27	−0.69	0.42

[a] $q_{BH_3} = -q_{AH_3}$.

included in the frozen auxiliary orbitals. The good performance of the third method probably comes from this more-realistic description of the auxiliary orbitals. Table 7 shows that, although the LLO:F method gives qualitatively correct results (except for the MIDI! basis), it is less satisfactory than the GLO method, in that the $H-C_A$ bond distance is generally underestimated. In terms of the bond angles across the QM/MM boundary, the last three methods show similar performance; there are deviations from the accurate results of 3°−5°. These deviations most likely are caused by basis imbalance near the QM/MM boundary, because the effect is exacerbated when the QM basis is changed from the STO-3G basis to a split valence basis. Another consideration is that these angles are closely coupled to the bond distance between C_A and C_B. For example, a C_A-C_B bond that is too short will favor greater C_A-C_B-H and $H-C_A-C_B$ angles, because stronger electron−electron repulsions are experienced by electron clouds that are present in C_A-H and C_B-H bonds.

Next, we comment on the atomic charges that are obtained from population analysis based on the GHO-type wave function. Although comparing charges from different population analysis schemes is sometimes misleading, which may be a consequence of the fact that atomic charge is not a well-defined observable parameter in quantum mechanics, the population analyses in Table 3 show that population-analysis charges show significant charge-transfer effects, as a result of basis set unbalance. A well-defined QM/MM boundary treatment should have little charge transfer between the two methyl groups in ethane when one of the C atoms is the GHO boundary atom. Therefore, a possible target in the parametrization process is to yield a net charge of zero for the purely QM methyl group, when one C atom is the boundary atom and the other is the frontier atom. As discussed in Section VI, other ways to test the charge balance across the boundary include (i) examination of the partial charges in propane when the center C atom is a boundary atom and (ii) examination of the optimized bond lengths and bond angles when charged groups are located near the boundary.

To test the ability of the GHO−AIHF method to reproduce the shape of the potential energy surface as well as the equilibrium geometries, we also investigated the internal rotation around the C2−C3 bond in *n*-butane. Figure 2 shows the energy profiles at various values of the dihedral angles for both QM/MM calculations for the GHO−AIHF treatment of the boundary, for pure QM calculations and for pure MM calculations. In each case, these results are based on geometries obtained by constrained optimization with the C1−C2−C3−C4 dihedral angle fixed at various values in the range of 0°−180°. Note that, for the GHO−AIHF case, we put the GHO boundary at C3; therefore, the torsion potential involves comparable contributions from QM and MM; this choice of boundary atom provides the most demanding possible test. The pure QM level produces a higher barrier than pure MM calculations. The most promising GHO−AIHF methods are those that use Löwdin OAOs; curves obtained from GLO and LLO:FG are both observed between the QM and MM results for angles of 0°−80° and very close to the MM results for angles of 80°−180°. The GHO−AIHF calculations using the projected basis significantly underestimate the torsion barrier, and when the NDDAO approximation is used, the torsion barrier height is overestimated. In principle, one could use different MM parameters at the boundary to compensate for any QM/MM imbalance; for example, one could use special dihedral MM parameters for torsions that involve frontier and boundary atoms. That is beyond the scope of the present article. Here, we just emphasize that the internal rotation barrier is qualitatively correct, even without new parameters, especially for the GLO and LLO:FG methods.

To study the difference between the GLO and LLO:FG methods further, we made a systematic comparison for *n*-octane in the fully extended conformation, where we treat the two methyl groups as MM and $(CH_2)_6$ as QM and test how many atoms must be included in each Löwdin orthogonalization. For clarity, we number the C atoms 1−8, and the H atoms that are

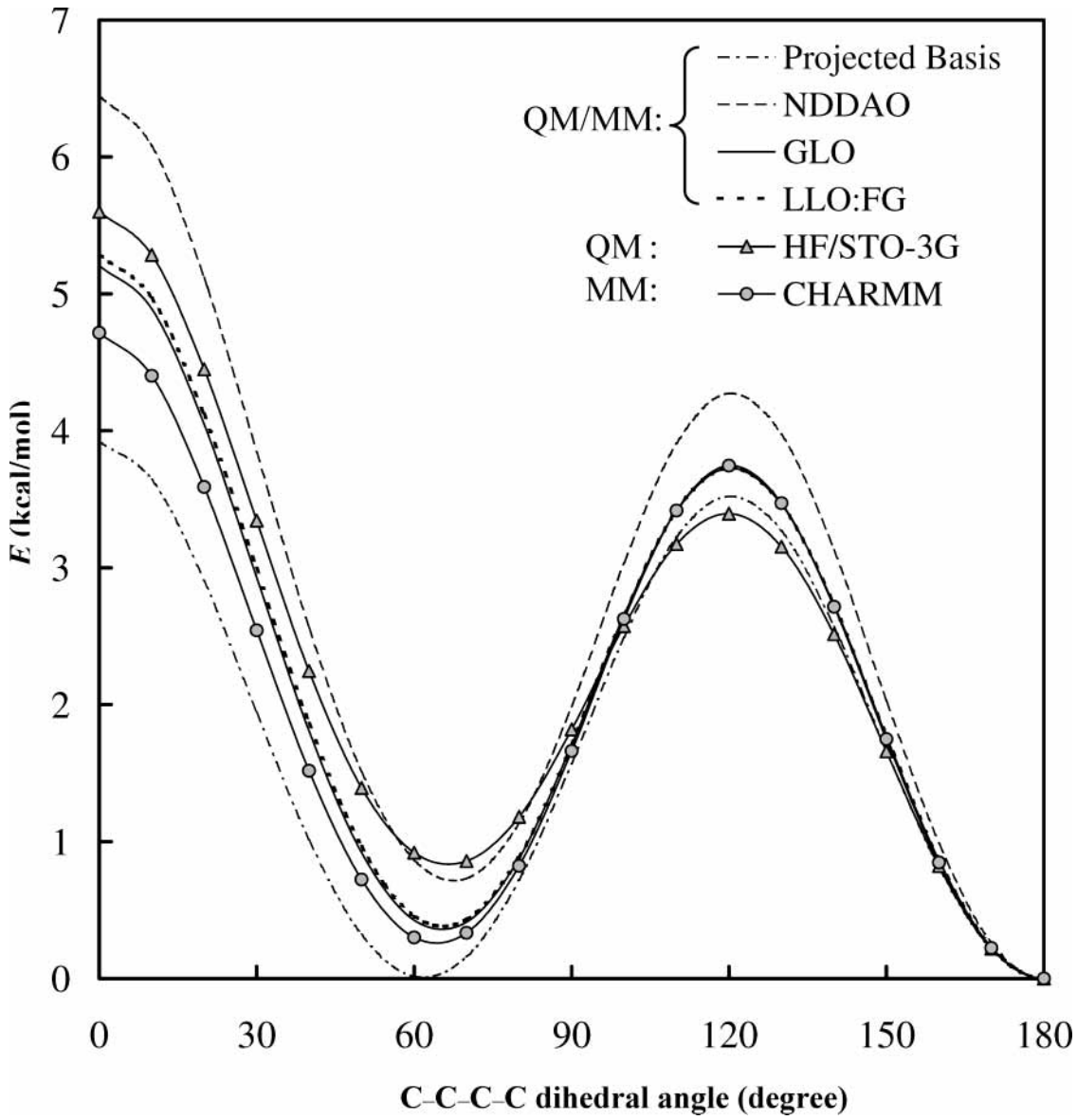

Figure 2. Potential energy curve for the internal rotation around the C2−C3 bond in *n*-butane, using unparametrized GHO−AIHF/STO-3G, pure QM (HF/STO-3G), and MM (CHARMM22). In all GHO calculations in this paper, we use the STO-3Gv basis set on the boundary atom, which, in the present case, is C3; the C1C2 ethyl group is the fully QM subsystem.

TABLE 8: Various Localized Löwdin Orthogonalization (LLO) Schemes in GHO Treatment for *n*-Octane[a]

	QM Atoms Included in Orthogonalization	
scheme	near GHO boundary atom C1	near GHO boundary atom C8
LLO:F	C2	C7
LLO:FG	C2, H2, C3	C7, H7, C6
LLO:FGV	C2, H2, C3, H3, C4	C7, H7, C6, H6, C5
GLO	all atoms over the entire molecule	all atoms over the entire molecule

[a] We number the C atoms C1−C8; the H atoms attached to C1 are labeled H1, those attached to C2 are labeled H2, etc. C1 and C8 are the GHO boundary atoms.

attached to C1 would be H1, those attached to C2 would be H2, etc. We then compare four QM/MM treatments as described in Table 8. The optimized geometries and charges for each treatment are listed in Tables 9 and 10. The geometry can be fully optimized when only the QM frontier atoms are included in the orthogonalization (whereas, in our experience, the geometry optimization often does not converge for methods that totally ignore the orthogonality concerns raised in Section IIB); however, the bond angles that involve the boundary atom (C1), the frontier atom (C2), and its QM neighbors (H2, C3), especially the C1−C2−C3 angle, are quantitatively underestimated. Including H2 into the orthogonalization slightly improves the bond distance at the QM/MM boundary (from 1.40 Å to 1.42 Å); however, we do not show full results of this calculation in which only two of the three atoms geminal to B are included. The C1−C2−C3 angle is greatly improved when C3 is included in the orthogonalization in the LLO:FG method. In fact, the local orthogonalization is well-converged after the QM frontier atoms and all its QM neighbors are included in the orthogonalization, i.e., adding more atoms that are further from the boundary (such as the LLO:FGV method) does not change the result significantly. Interestingly, the charges are not as sensitive

TABLE 9: GHO−AIHF/STO-3G Optimized Geometry Using Various Local Löwdin Orthogonalization (LLO) Schemes for *n*-Octane

scheme[a]	H−C1 (Å)	H−C1−C2 (deg)	C1−C2 (Å)	C−C2−C (deg)	C1−C2−H (deg)	C2−H (Å)	C2−C3 (Å)	C2−C3−H (deg)	C3−H (Å)	C−C3−C (deg)	C3−C4 (Å)	C−C4−C (deg)	C3−C4−H (deg)	C4−H (Å)
LLO:F	1.125	117.9	1.408	103.1	108.9	1.084	1.531	107.9	1.089	114.2	1.543	112.7	109.2	1.089
LLO:FG	1.125	116.0	1.437	111.2	108.2	1.085	1.540	108.9	1.088	112.5	1.543	112.4	109.3	1.088
LLO:FGV	1.125	116.0	1.437	111.8	108.1	1.085	1.540	109.2	1.088	112.3	1.544	112.4	109.3	1.088
GLO	1.125	116.0	1.437	111.8	108.1	1.085	1.540	109.2	1.088	112.3	1.544	112.4	109.3	1.088
pure QM[b]	1.086	110.5	1.541	112.6	109.3	1.088	1.545	109.3	1.086	112.3	1.545	112.5	109.2	1.088

[a] Schemes according to Table 8. [b] HF/STO-3G for the entire molecule.

TABLE 10: Charges for *n*-Octane at the GHO−AIHF/STO-3G Optimized Geometry Using Various Local Löwdin Orthogonalization (LLO) Schemes

	Mulliken Charge							Löwdin Charge						
scheme[a]	C1	C2	H2	C3	H3	C4	H4	C1	C2	H2	C3	H3	C4	H4
LLO:F	−0.43	0.06	0.05	−0.10	0.05	−0.10	0.05	−0.40	0.08	0.02	−0.04	0.02	−0.04	0.02
LLO:FG	−0.42	0.04	0.05	−0.10	0.05	−0.10	0.05	−0.40	0.08	0.02	−0.04	0.02	−0.04	0.02
LLO:FGV	−0.42	0.04	0.05	−0.10	0.05	−0.10	0.05	−0.40	0.08	0.02	−0.04	0.02	−0.04	0.02
GLO	−0.42	0.04	0.05	−0.10	0.05	−0.10	0.05	−0.40	0.08	0.02	−0.04	0.02	−0.04	0.02
pure QM[b]	−0.18	−0.09	0.05	−0.10	0.05	−0.10	0.05	−0.09	−0.03	0.02	−0.04	0.02	−0.04	0.02

[a] Schemes according to Table 8. [b] HF/STO-3G for the entire molecule.

to the orthogonalization extent as the geometries, as indicated in Table 10. This test shows that it is crucial to conduct explicit orthogonalization among orbitals near the boundary, because the overlaps between these orbitals are non-negligible. The orbitals on atoms far away from the boundary do not need to be included in the orthogonalization. We can either explicitly exclude them by localizing the orthogonalization, or we can do the orthogonalization over the entire molecule. These two options should give similar results, because the tails on centers far away from the boundary in the hybrid orbitals will be very small.

VI. Further Parametrization

Although the methods presented in the previous section give reasonable results even without any parametrization, one can obtain even better results if some parameters related to the boundary atom are added or readjusted. The objectives of the parametrization are to improve the electronegativity balance and bonding properties across the A−B interface. One way to monitor the electronegativity balance is to calculate the Mulliken or Löwdin charge on the quantal methyl group in propane when the middle C atom is treated as a GHO boundary. This should be close to zero, although it does not need to be exactly zero, because propane has a nonzero dipole moment and the A atom has a larger basis set than that of the B atom. The MM methyl group is neutral; therefore, the QM methyl group should also be neutral (by symmetry), in which case the partial charge on the B atom would be exclusively contributed from the extra charge densities in its auxiliary orbitals, i.e., matching the point charge of the B atom in MM. Therefore, one of the goals of our parametrization is that the auxiliary electrons should introduce minimal artificial polarizations of the A−B bond in propane. In propane, as well as in other molecules, the bonding properties at the QM−MM interface can also be monitored by calculating the bond distances, bond angles, and torsion potentials that involve QM atoms near the boundary. (Another possible choice of the electronegativity optimization target would have been ethane, which was used for the parameterization of the semiempirical GHO method.[26] In this case, the charge on the quantal methyl group is exactly zero. One reason for selecting propane for the present work is that, in the ethane test case, the B atom is a methyl carbon, but in propane it is a methylene carbon, which is more typical of real applications.)

The GHO treatment deviates from a pure QM calculation primarily for the following reasons. First, only a portion of the system is represented by a quantum mechanical wave function. Second, instead of a full SCF optimization over all basis functions, active MOs in the GHO method are only expanded over a reduced space, and auxiliary MOs are kept frozen. Third, the basis set balance across the boundary is sacrificed to maintain the simplicity of the algorithm. Finally, the screening effect of 1*s* core electrons on the GHO boundary atom is only crudely mimicked by a reduction of its nuclear charge by 2. All three of these aspects can be related to an inaccurate description of the interaction across the boundary. Two of the most straightforward ways to remedy these shortcomings are (i) to adjust the molecular mechanical parameters of the boundary atom and (ii) to scale the integrals that involve the orbitals on the boundary atom. The former corresponds to an MM fix, whereas the latter can be considered as a QM fix. In the following, we present a parametrization for the MIDI! basis,[66] in which we combine both strategies. The MIDI! basis (which is also sometimes called MIDIX) is an ideal choice for applying the GHO−AIHF method, in that it is designed to provide accurate geometries and partial charges at the HF level[66] and it also yields reasonable HF relative energies.[71] Furthermore, MIDI! does not include polarization functions on carbon; therefore, for large organic and biological systems (such as enzymatic systems), it is very efficient, in terms of computational cost. The LLO:FG version of the GHO−AIHF will be used as a starting point for adding parameters, because it gives the best geometries for MIDI! among the four orthogonalization schemes; at the same time, the orthogonalization is strictly localized at the boundary region that only involves orbitals on boundary, frontier, and geminal atoms.

First, we note that the A−B−M angles are always overestimated by ∼5° in the unparametrized GHO−AIHF geometries, where M represents an MM atom bonded to the GHO boundary atom B. This systematic error may come from several sources. For example, if the charge densities in the auxiliary orbitals are too small, the repulsions between auxiliary orbitals will be relatively weak, compared to the repulsion among the A−B bonding orbital and the auxiliary orbital, resulting in A−B−M angles that are too large. However, in combined QM/MM calculations, A−B−M angles are described both by QM

TABLE 11: Optimized Scaling Factors for Integrals Involving Boundary Orbitals for GHO−AIHF/MIDI! Calculations with the LLO:FG Method[a]

parameter	integral type	optimized value
c_1	$(s_A\|T\|s_B)$	0.9078
c_2	$(s_A\|T\|p_B)$	1.0257
c_3	$(p_A\|T\|s_B)$	1.0806
c_4	$(p_A\|T\|p_B)$	1.0283
c_5	$(s_B\|T\|s_B)$	0.9733
c_6	$(p_B\|T\|p_B)$	0.9858
c_7	$(s_B\|T\|p_B)$	0.9665

[a] MM equilibrium angles for A−B−M bends are decreased by 8°.

and MM; therefore, the predicted large bond angles could be a result of a poor balance between the QM and MM. Regardless of the precise source of error, we found that we can remove most of the systematic error by simply decreasing the equilibrium angles of all A−B−M molecular mechanics bending terms by 8°. (For example, when A and B are sp^3 C atoms and M is a H atom, we decrease the MM bond angle parameter from 110° to 102°.)

Next, accepting this change in the A−B−M bending potential, we determined a set of parameters for scaling integrals that involve the boundary orbitals. Because the integral scaling is used only to correct small inaccuracies, we prefer to make the scaling scheme as simple as possible. To avoid the complexity that is introduced by multicenter two-electron integrals, we chose to scale one-electron integrals only. For the same reason, even one-electron potential energy integrals are not considered, because they can involve up to three centers. It seems that scaling two-center one-electron potential energy integrals is less stable than scaling the kinetic energy integrals. Furthermore, if the scaling is localized to the AB pair, the scaling factors are expected to be more transferable from one system to another. Therefore, we restrict scaling to a subset of the one-electron kinetic energy integrals related the boundary: in particular, the scaling is restricted to integrals of the form $(v_A|T|v_B)$ and $(v_B|T|v_B)$, where v_B denotes a valence orbital on the boundary atom B, and v_A denotes a valence orbital on the frontier atom A. To take account of the different behavior of *s* and *p* orbitals, we also treat them separately, to add more flexibility. As a result of these considerations, we decided to scale the $(v_A|T|v'_B)$ and $(v_B|T|v'_B)$ ($v = s, p$; $v' = s, p$) types of one-electron integrals in our parametrization; the corresponding scaling factors are denoted by $c_1 - c_7$ as indicated in the first two columns of Table 11. (The integrals are scaled in the nonorthogonal nonhybridized atomic basis before the transformations to orthogonal, hybrid, and molecular orbitals.)

The optimal values of these scaling factors were determined using a microgenetic algorithm,[72] to maximize a fitness function *F* over a training set of molecules, in which the geometry fitness and the charge fitness are equally weighted:

$$F = -\left\{\frac{1}{K}\sum_m\left[\sum_r\left(\frac{r^{\mathrm{GHO}} - r^{\mathrm{HF}}}{r_0}\right)^2 + \sum_\theta\left(\frac{\theta^{\mathrm{GHO}} - \theta^{\mathrm{HF}}}{\theta_0}\right)^2\right] + \left(\frac{q_{\mathrm{B}}^{\mathrm{L,GHO}} - q_{\mathrm{B}}^{\mathrm{MM}}}{q_0}\right)_{m=1}^2\right\}^{1/2} \quad (58)$$

In eq 58, the training set for geometries contains five molecules ($m = 1, 2, ..., 5$): propane, 1-propanol, propanoic acid, 1-butene, and *n*-butane. *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), $q_{\mathrm{B}}^{\mathrm{L,GHO}}$ is the Löwdin charge on the GHO boundary atom B in propane, and $q_{\mathrm{B}}^{\mathrm{MM}}$ is the MM point charge for B (equal to −0.18 in CHARMM); r_0 (0.02 Å), θ_0 (2°), and q_0 (atomic charge of 0.05) are scale units for distance, angle, and charge, respectively. We take the reference geometry (denoted HF in eq 58) to be the geometry optimized at the pure QM level with the full basis set, i.e., HF/MIDI!. Note that the C atom in a CH_2 group is chosen as the GHO boundary atom B, and the C atom attached to a functional group is the frontier atom A, i.e., the QM/MM partitions are $CH_3BH_2AH_3$, $CH_3BH_2AH_2OH$, CH_3BH_2AOOH, $CH_3BH_2AH{=}CH_2$, and $CH_3BH_2AH_2CH_3$. The goal of the optimization is to find a set of scaling factors such that GHO−AIHF gives geometries that are in good agreement with those from pure QM calculations, whereas the charge transfer across QM/MM boundary is minimized. The last term in eq 58 accomplishes this minimization of the charge transfer, because one obtains a neutral quantal methyl group when $q_{\mathrm{B}}^{\mathrm{L,GHO}}$ matches $q_{\mathrm{B}}^{\mathrm{MM}}$ in propane. The purpose of including various functional groups near the boundary is to make the parameters more robust, especially when electronegative atoms (O) and unsaturated C atoms are close to the boundary.

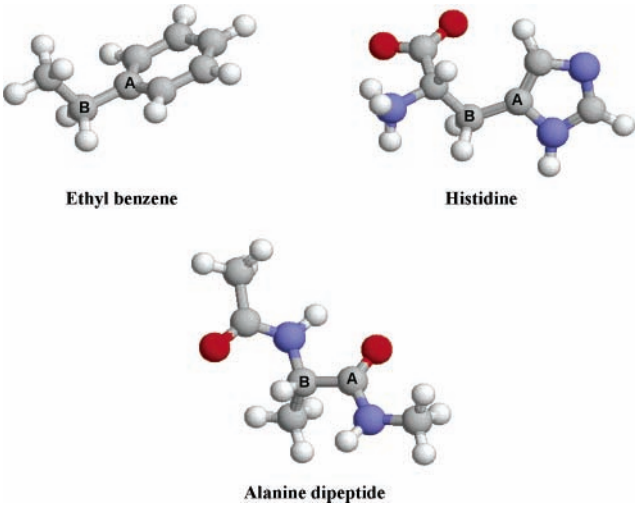

Figure 3. QM/MM partitions in ethyl benzene, histidine zwitterion, and alanine dipeptide. The GHO boundary atom is labeled as B, and the frontier QM atom is labeled as A.

The fitness function *F* was maximized, with respect to the scaling factors c_n, and the optimized scaling factors are tabulated in Table 11. We then tested the parametrized method against a wide variety of molecules that were not present in the parametrization training set. The test set contains systems with various functional groups (−SH, $-NH_2$, $-CONH_2$, $-C_6H_5$) near the boundary, as well as molecules with ionic charges near the boundary. We also include histidine zwitterion and alanine dipeptide as two realistic cases to test the reliability of applying the GHO−AIHF method to amino acids and proteins. (Figure 3 shows depictions of ethyl benzene, histidine, and alanine dipeptide molecules.)

First, we note that the parametrized model yields $q_{\mathrm{B}}^{\mathrm{L,GHO}} = -0.16$ in propane, which matches the desired value better than the unparametrized model (the unparametrized model gives $q_{\mathrm{B}}^{\mathrm{L,GHO}} = -0.15$). Furthermore, in *n*-butane, with C2 as the boundary atom and C3 as the frontier atom, for the parametrized model, $q_{\mathrm{B}}^{\mathrm{L,GHO}} = -0.17$, which is encouraging, because the charges in *n*-butane were not represented in the training set. As a result, for propane and *n*-butane, in the parametrized model, all methylene and methyl groups are neutral to within a partial charge of 0.02.

TABLE 12: A−B Bond Distance for the Unparametrized and the Parametrized[a] GHO−AIHF(LLO:FG)/MIDI! Results, Compared to All-Electron ab Initio QM Results

	Bond Distance (Å)		
system	unparametrized	parametrized	HF/MIDI![b]
$CH_3BH_2-AH_3$	**1.530**	**1.514**	**1.540**
$CH_3BH_2-AH_2CH_3$	**1.549**	**1.529**	**1.539**
$CH_3BH_2-AH_2C(O)OH$	1.546	1.518	1.539
$CH_3BH_2-AH_2NH_2$	1.568	1.548	1.545
$CH_3BH_2-AH_2NH_3^+$	1.554	1.536	1.533
$CH_3BH_2-AH_2OH$	**1.556**	**1.535**	**1.528**
$CH_3BH_2-AH_2O^-$	1.609	1.579	1.594
$CH_3BH_2-AH_2SH$	1.554	1.531	1.541
$CH_3BH_2-AH=CH_2$	**1.554**	**1.515**	**1.506**
$CH_3BH_2-A(O)NH_2$	1.628	1.569	1.523
$CH_3BH_2-A(O)OH$	**1.622**	**1.563**	**1.508**
$CH_3BH_2-A(O)O^-$	1.663	1.611	1.592
$CH_3BH_2-A(O)OCH_3$	1.609	1.547	1.509
ethyl benzene[c]	1.576	1.518	1.513
histidine[c,d]	1.578	1.488	1.500
alanine dipeptide[c]	1.622	1.555	1.555

[a] Seven scaling factors and one change in the MM parameters are listed in Table 11. Training set is indicated in boldface type. [b] Conventional quantum mechanics. [c] See Figure 3. [d] Zwitterion.

To further illustrate how the results can be improved by parametrization, we consider the key A−B bond distance. The values obtained for this bond distance for all molecules in the parametrization set and the test set are listed in Table 12 for both unparametrized and parametrized GHO−AIHF/MIDI!, and the mean unsigned errors (MUEs) are given in Table 13.

As shown in Tables 12 and 13, the parametrized version of GHO−AIHF significantly reduces the average errors in the A−B bond distances and A−B−M angles, although it is important to note that there are some troublesome cases. For example, when a carbonyl group is chosen as the frontier atom, the A−B bond has a tendency to be overestimated. Thus, if possible, one should avoid using a carbonyl C atom as the frontier atom. Nevertheless, the robustness of the algorithm and the scaling factors is impressive, considering the variety of the test cases. In particular, even including the difficult and troublesome cases, the MUE in all bonds that involve A or B is <0.02 Å, and the MUE in all bond angles that involve the A−B bond is only ∼2°. Furthermore, the average errors are only slightly larger for the entire test set than for the training set. The choice of molecules in Tables 12 and 13 does not mean that we recommend putting charged or highly functionalized groups close to the boundary, if there are other, more-suitable locations for a boundary; rather, these tables are designed to provide very difficult tests to validate the method.

Appendix B presents the extension of the present method to unrestricted Hartree−Fock[73] (UHF) for the QM/MM treatment (GHO−AIUHF) of open-shell systems. Table 14 shows the optimized geometries and charges obtained at GHO−AIUHF/MIDI! level for the ethyl radical, compared to pure QM calculations. The results show that, although the integral scaling factors are parametrized for closed-shell systems, they work equally well for open-shell system. Furthermore, as for closed-shell systems, the method is already qualitatively reasonable, even without parametrization.

Clearly, we could obtain better balance across the boundary by adding more parameters; however, this is not our goal in this paper. We prefer to emphasize that the results are qualitatively correct, even without any scaling, and that they can easily be improved by a very small amount of scaling.

TABLE 13: Mean Unsigned Errors in Bond Lengths and Bond Angles with GHO−AIHF (LLO:FG)/MIDI![a]

	Unparametrized					Parametrized				
	Bond Distance (Å)			Angle (deg)		Bond Distance (Å)			Angle (deg)	
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
$CH_3BH_2AH_3$	**0.010**	**0.006**	**0.019**	**3.2**	**3.9**	**0.026**	**0.004**	**0.020**	**2.6**	**1.6**
$CH_3BH_2AH_2CH_3$	**0.010**	**0.010**	**0.018**	**1.8**	**4.2**	**0.019**	**0.010**	**0.019**	**1.2**	**2.1**
$CH_3BH_2AH_2C(O)OH$	0.007	0.010	0.021	1.4	4.3	0.028	0.010	0.022	1.4	2.1
$CH_3BH_2AH_2NH_2$	0.023	0.009	0.017	2.5	4.7	0.003	0.008	0.017	2.0	2.8
$CH_3BH_2AH_2NH_3^+$	0.021	0.007	0.024	2.7	3.5	0.004	0.007	0.025	2.1	3.5
$CH_3BH_2AH_2OH$	**0.028**	**0.005**	**0.018**	**1.7**	**4.8**	**0.007**	**0.005**	**0.018**	**1.6**	**2.8**
$CH_3BH_2AH_2O^-$	0.011	0.002	0.016	2.5	7.0	0.014	0.004	0.016	2.6	5.4
$CH_3BH_2AH_2SH$	0.013	0.008	0.019	1.9	4.1	0.009	0.006	0.020	1.4	2.8
$CH_3BH_2AH=CH_2$	**0.047**	**0.004**	**0.023**	**2.1**	**4.1**	**0.009**	**0.003**	**0.024**	**0.7**	**1.6**
$CH_3BH_2A(O)NH_2$	0.105	0.008	0.022	2.4	3.4	0.046	0.009	0.024	2.1	3.7
$CH_3BH_2A(O)OH$	**0.114**	**0.008**	**0.021**	**1.4**	**2.5**	**0.055**	**0.008**	**0.023**	**1.6**	**2.9**
$CH_3BH_2A(O)O^-$	0.072	0.003	0.017	0.7	4.9	0.019	0.002	0.017	0.3	3.5
$CH_3BH_2A(O)OCH_3$	0.096	0.012	0.026	2.0	4.8	0.038	0.012	0.027	1.0	3.4
ethyl benzene[b]	0.064	0.001	0.027	1.1	1.6	0.005	0.002	0.028	0.4	1.4
histidine[b,c]	0.078	0.004	0.034	1.7	4.8	0.013	0.005	0.031	4.2	3.8
alanine dipeptide[b]	0.093	0.010	0.019	2.2	4.4	0.025	0.011	0.020	0.9	2.3
training set	0.042	0.006	0.021	2.0	3.9	0.021	0.006	0.021	1.5	2.2
entire set	0.048	0.007	0.020	2.0	4.1	0.019	0.007	0.021	1.5	2.8

[a] Seven scaling factors and one change in the MM parameters are listed in Table 11. Training set is in bold. [b] See Figure 3. [c] Zwitterion.

TABLE 14: Optimized Geometries and Partial Charges for Ethyl by GHO−AIUHF Based on Local Löwdin Orthogonalization (LLO), Compared to a Pure HF Calculation at the Same QM Level[a]

	Bond Distance (Å)			Angle (deg)		Mulliken charge		Löwdin charge	
basis	C_A-C_B	C_A-C_H	C_B-C_H	$H-C_A-C_B$	C_A-C_B-H	q_A	q_B	q_A	q_B
GHO−AIUHF/MIDI![b]	1.529	1.075	1.121	114.3	113.9	−0.07	−0.50	−0.21	−0.23
GHO−AIUHF/MIDI![c]	1.509	1.076	1.123	116.1	111.6	−0.03	−0.54	−0.19	−0.25
HF/MIDI!	1.503	1.076	1.088	120.7	111.4	−0.33	−0.60	−0.20	−0.28

[a] The boundary carbon (C_B) is the C atom in the methyl group, and the frontier C atom (C_A) is the radical center. We use the LLO:FG method in this table. [b] Unparametrized version. [c] Parametrized version (same parameters as used for Tables 11−13).

TABLE 15: Proton Affinities Calculated by Parametrized[a] GHO−AIHF/MIDI!, Based on Local Löwdin Orthogonalization (LLO:FG), Compared to Pure HF Calculations at the Same QM Level

	Proton Affinity (kcal/mol)		
species	GHO−AIHF/MIDI!	HF/MIDI!	QM/MM vs QM
$BH_3-AH_2O^-$	421.3	416.3	4.5
$CH_3BH-AH_2O^-$	422.1	415.1	7.1
$BH_3-AH_2CH_2O^-$	417.7	415.1	2.7
$CH_3CH_2BH_2-AH_2O^-$	421.6	414.8	6.8
$CH_3BH_2-AH_2CH_2O^-$	418.1	414.8	3.3
$BH_3-AH_2CH_2CH_2O^-$	416.2	414.8	1.4
$BH_3-AH_2NH_2$	229.2	232.6	−3.4
$CH_3BH_2-AH_2NH_2$	229.5	233.9	−4.4
$BH_3-AH_2CH_2NH_2$	232.4	233.9	−1.5
$CH_3CH_2BH_2-AH_2NH_2$	229.3	234.7	−5.4
$CH_3BH_2-AH_2CH_2NH_2$	232.7	234.7	−2.0
$BH_3-AH_2CH_2CH_2NH_2$	234.1	234.7	−0.6
$BH_3-AH_2NH^-$	443.0	439.9	3.2
$CH_3BH_2-AH_2NH^-$	443.6	438.7	4.9
$BH_3-AH_2CH_2NH^-$	440.4	438.7	1.7
$CH_3CH_2BH_2-AH_2NH^-$	443.2	438.3	4.9
$CH_3BH_2-AH_2CH_2NH^-$	440.7	438.3	2.4
$BH_3-AH_2CH_2CH_2NH^-$	439.3	438.3	1.0
$BH_3-AH_2S^-$	383.1	381.5	1.6
$CH_3BH_2-AH_2S^-$	383.4	380.6	2.8
$BH_3-AH_2CH_2S^-$	382.1	380.6	1.5
$CH_3CH_2BH_2-AH_2S^-$	383.2	380.5	2.7
$CH_3BH_2-AH_2CH_2S^-$	382.5	380.5	2.0
$BH_3-AH_2CH_2CH_2S^-$	381.2	380.5	0.7
$BH_3-A(O)O^-$	377.1	377.3	−0.2
$CH_3BH_2-A(O)O^-$	375.9	375.3	0.6
$BH_3-AH_2C(O)O^-$	377.3	375.3	1.9
$CH_3CH_2BH_2-A(O)O^-$	375.0	374.7	0.2
$CH_3BH_2AH_2C(O)O^-$	377.6	374.7	2.8
$BH_3-AH_2CH_2C(O)O^-$	376.1	374.7	1.3

[a] Seven scaling factors and one change in the MM parameters are listed in Table 11.

Finally, we present the results of using the parametrized GHO−AIHF method for proton affinities, which, because the charge state changes, present very stringent tests of the effect of the QM/MM boundary treatment on calculated energies. In the present article, we define the proton affinity as the zero-point-exclusive energy difference between a chemical species (denoted X or X^-) and its protonated form (denoted XH^+ or XH). Table 15 gives the proton affinities calculated using GHO−AIHF/MIDI! (LLO:FG, with scaled integrals) compared with fully QM (HF/MIDI!) results, where a set of compounds that consists of alcohols, amines, thiols, and acids is investigated. The QM/MM division of each system is specified in the first column of Table 15, where A denotes the QM frontier atom and B denotes the GHO boundary atom, as usual. The average absolute errors of the proton affinities for 30 cases in Table 15 are 2.6 kcal/mol. The most significant deviations from fully QM results were observed when the GHO boundary is only one bond away from the X−H bond that is being dissociated, where the MUE for 15 such cases is 3.5 kcal/mol. One would not recommend putting the boundary this close to a reaction center, especially for a reaction that involves a change in charge state, if any other choice is possible; rather, these tests are included to show what happens if one pushes the method to its limits. The MUE is reduced to 2.2 kcal/mol if one places the GHO boundary atom two bonds away from the protonated/depronated center. The results are comparable to the typical errors on proton affinities in QM/MM calculations by various other boundary treatments. For example, Amara et al. obtained an error of ∼3 kcal/mol on protonation energies for similar systems with two of their link-atom models.[35] As another example, the pseudobond approach, using various basis sets, yields mean absolute deviations in protonation energies of 2.9−7.7 kcal/mol.[36] On the basis of these comparisons, as well as what one can reasonably expect at the HF level, we conclude that the performance of the GHO−AIHF method for the energetics is satisfactory. We especially note that no proton affinities were used at any stage of the parametrization. If one were especially interested in proton affinities, one could probably reduce the errors by further parametrization; however, our goal was to present a more general parametrization that was based on geometries and charge balance as indicators of the charge polarization at the boundary.

VII. Concluding Remarks

Although great progress has been made in combining quantum mechanical electronic structure methods with classical molecular mechanics in a single algorithm, some of the fundamental problems of nonorthogonality between the explicit orbitals of the quantum mechanical subsystem and the implicit wave function of the classical mechanical subsystem are avoided by treatments that use artificial link atoms or consistently neglect all differential overlap (and compensate for this by the parameters of semiempirical molecular orbital theory). In the present study, a theoretical framework has been established that provides a more fundamental solution to this problem. We accomplish this by extending the generalized hybrid orbital (GHO) method to the ab initio Hartree−Fock (HF) level, in combination with quantum mechanical and molecular mechanical (QM/MM) calculations. Three important issues are investigated for this extension: (a) the orthogonality constraint that involves auxiliary orbitals, (b) the adequacy of a small basis set on the GHO boundary atom, and (c) the formulation of analytical gradients. In our treatment, a minimum valence basis set is used on the boundary atom to maintain the simplicity of the original hybridization scheme. Four methods are proposed to remove the nonorthogonality between active MOs and auxiliary MOs in GHO. To compute the GHO gradient analytically, the derivatives of the density and energy-weighted density matrices that are due to basis transformations must be included. With the aforementioned three questions answered satisfactorily, the GHO method can be applied at the ab initio HF level with a solid theoretical foundation. This provides a fundamental solution to the problem of orthogonality at a quantum mechanical−classical mechanical boundary.

Tests of the resulting algorithm showed that reasonable geometries and charges can be obtained even without any parametrization. Finally, we show that scaling some of the integrals that involve basis functions at the GHO boundary can improve the results, and this provides a simple way to parametrize the generalized hybrid orbital ab initio Hartree−Fock (GHO−AIHF) method for applications to practical problems. The scaled method is tested extensively for 30 proton affinities of neutral and negatively charged species; the average value of the 30 proton affinities is 369 kcal/mol, and the mean unsigned error of the GHO results from the fully QM results is only 2.6 kcal/mol.

Acknowledgment. This work has been supported in part by the National Science Foundation (Grant No. CHE00-92019) and by National Institutes of Health (Grant No. GM46736).

Appendix A

Here, we derive the expression for obtaining $\partial \mathbf{S}^{-1/2}/\partial q$ from $\partial \mathbf{S}^{+1/2}/\partial q$. We begin with

$$\mathbf{S}^{-1/2}\mathbf{S}^{+1/2} = \mathbf{I} \tag{A1}$$

Therefore,

$$\frac{\partial \mathbf{S}^{-1/2}}{\partial \boldsymbol{q}}\mathbf{S}^{+1/2} = -\mathbf{S}^{-1/2}\frac{\partial \mathbf{S}^{+1/2}}{\partial \boldsymbol{q}} \tag{A2}$$

and right multiplying by $\mathbf{S}^{-1/2}$ on both sides yields

$$\frac{\partial \mathbf{S}^{-1/2}}{\partial \boldsymbol{q}} = -\mathbf{S}^{-1/2}\frac{\partial \mathbf{S}^{+1/2}}{\partial \boldsymbol{q}}\mathbf{S}^{-1/2} \tag{A3}$$

Appendix B

This appendix presents the changes required to treat an open-shell system by unrestricted Hartree−Fock[73] (UHF) theory. The resulting formalism for ab initio UHF may be called GHO−AIUHF. We present the unrestricted GHO theory, using the hybrid Löwdin OAO method as an example. One can follow the procedures described in Section IIIC, except that two Fock matrices (one for α-spin electrons and one for β-spin electrons) are formed separately. Correspondingly, two sets of SCF equations are solved in the basis space of $N + 1$ dimensions to obtain active molecular orbitals (MOs). After the active α and β density matrices are constructed, one must append $0.5P^{\mathrm{H}}_{bb}$ as the charge density in α and β auxiliary MOs, respectively. For the gradient calculation, eq 16 is changed to

$$\frac{\partial E^{\mathrm{orb}}}{\partial \boldsymbol{q}} = \frac{\partial E^{\mathrm{HF}}}{\partial \boldsymbol{q}} + \sum_{\mu\nu}^{N+4}\left(\frac{\partial P^{\mathrm{AO},\alpha}_{\mu\nu}}{\partial \boldsymbol{q}}F^{\mathrm{AO},\alpha}_{\mu\nu} + \frac{\partial P^{\mathrm{AO},\beta}_{\mu\nu}}{\partial \boldsymbol{q}}F^{\mathrm{AO},\beta}_{\mu\nu}\right) - \sum_{\mu\nu}^{N+4}\frac{\partial W^{\mathrm{AO}}_{\mu\nu}}{\partial \boldsymbol{q}}S^{\mathrm{AO}}_{\mu\nu} \tag{B1}$$

where the middle terms are evaluated for α- and β-spin electrons separately.

References and Notes

(1) Warshel, A.; Levitt, M. *J. Mol. Biol*. **1976**, *103*, 227.
(2) Singh, U. C.; Kollman, P. A. *J. Comput. Chem*. **1986**, *7*, 718.
(3) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem*. **1990**, *11*, 700.
(4) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(5) Gao, J. *Rev. Comput. Chem*. **1996**, *7*, 119.
(6) (a) Ferenczy, G. G.; Rivail, J.-L.; Surjan, P. R.; Naray-Szabo, G. *J. Comput. Chem.* **1992**, *13*, 830. (b) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem*. **1994**, *15*, 269.
(7) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett*. **1996**, *263*, 100.
(8) Monard, G.; Loos, M.; Théry V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem*. **1996**, *58*, 153.
(9) Ferre, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem*. **2002**, *23*, 610.
(10) Maseras, F.; Morokuma, K. *J. Comput. Chem*. **1995**, *16*, 1170.
(11) (a) Vreven, T.; Morokuma, K. *J. Comput. Chem*. **2000**, *21*, 1419. (b) Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J. *J. Chem. Phys*. **2001**, *115*, 62.
(12) (a) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett*. **2002**, *355*, 257. (b) Morokuma, K. *Philos. Trans. R. Soc. London, Ser. A* **2002**, *360*, 1149.
(13) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Comput. Chem*. **1994**, *16*, 113.
(14) Monard, G.; Merz, K. M. *Acc. Chem. Res*. **1999**, *32*, 904.
(15) Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr. *J. Chem. Phys*. **2000**, *113*, 5604.
(16) (a) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem*. **1995**, *99*, 6374. (b) Thompson, M. A. *J. Phys. Chem*. **1996**, *100*, 14492.
(17) Bakowies, D.; Thiel, W. *J. Phys. Chem*. **1996**, *100*, 10580.
(18) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
(19) Lennartz, A.; Schafer, A.; Terstegen, F.; Thiel, W. *J. Phys. Chem. B* **2002**, *106*, 1758.
(20) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem*. **1996**, *60*, 1189.
(21) (a) Cummins, P. L.; Gready, J. E. *J. Comput. Chem*. **1997**, *18*, 1496. (b) Titmuss, J. J.; Cummins, P. L.; Rendall, A. P.; Bliznyuk, A. A.; Gready, J. E. *J. Comput. Chem*. **2002**, *23*, 1314.
(22) Bersuker, I. B.; Leong, M. K.; Boggs, J. E.; Pearlman, R. S. *Int. J. Quantum Chem*. **1997**, *63*, 1051.
(23) Tongraar, A.; Leidl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1997**, *101*, 6299.
(24) (a) Woo, T. K.; Cavallo, L.; Ziegler, T. *Theor. Chem. Acc*. **1998**, *100*, 307. (b) Woo, T. K.; Blochl, P. E.; Ziegler, T. *THEOCHEM* **2000**, *506*, 313.
(25) Gao, J., Thompson, M., Eds. *Combined Quantum Mechanical and Molecular Mechanical Methods*, ACS Symposium Series 712; American Chemical Society: Washington, DC, 1998.
(26) (a) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714. (b) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc*. **2000**, *104*, 336.
(27) (a) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem*. **2002**, *53*, 467. (b) Truhlar, D. G.; Gao, J.; Alhambra, C.; Mireia, G.-V.; Corchado, J.; Sanchez, M. L.; Villa, J. *Acc. Chem. Res*. **2002**, *35*, 341.
(28) Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462.
(29) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
(30) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
(31) (a) Hillier, I. H. *THEOCHEM* **1999**, *463*, 45. (b) Hall, R. J.; Hindle, S. A.; Burton, N. A.; Hillier, I. H. *J. Comput. Chem*. **2000**, *21*, 1433. (c) Nicoll, R. M.; Hindle, S. A. MacKenzie, G.; Hillier, I. H.; Burton, N. A. *Theor. Chem. Acc*. **2001**, *106*, 105.
(32) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656.
(33) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys*. **1999**, *110*, 10452.
(34) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. *J. Comput. Chem*. **2000**, *21*, 1088.
(35) Amara, P.; Field, M. J. *Theor. Chem. Acc*. **2003**, *109*, 43.
(36) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys*. **1999**, *110*, 46.
(37) (a) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem*. **1999**, *20*, 1468. (b) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *Chem. Phys. Lett*. **2000**, *321*, 113. (c) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem*. **2000**, *21*, 1442.
(38) (a) Rothlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. *J. Biol. Inorg. Chem*. **2000**, *5*, 236. (b) Laoio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys*. **2002**, *116*, 6941.
(39) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys*. **2002**, *117*, 10534.
(40) Dilabio, G.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys*. **2002**, *116*, 9578.
(41) Swart, M. *Int. J. Quantum Chem*. **2003**, *91*, 177.
(42) Alhambra, C.; Wu, L.; Zhang, Z.-Y.; Gao, J. *J. Am. Chem. Soc*. **1998**, *120*, 3858.
(43) (a) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys*. **1965**, *43*, S129. (b) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc*. **1985**, *107*, 3902. (c) Stewart, J. J. P. *J. Comput.-Aided Mol. Des*. **1990**, *4*, 1.
(44) (a) Alhambra, C.; Gao, J.; Corchado, J. C.; Villa, J.; Truhlar, D. G.; *J. Am. Chem. Soc*. **1999**, *121*, 2253. (b) Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Gao, J.; Truhlar, D. G. *J. Am. Chem. Soc*. **2000**, *122*, 8197.
(45) Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326.
(46) Alhambra, C.; Sanchez, M. L.; Corchado, J. C.; Gao, J.; Truhlar, D. G. *Chem. Phys. Lett*. **2001**, *347*, 512; **2002**, *355*, 388 (erratum).
(47) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Comput. Chem*. **2003**, *24*, 177.
(48) González-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem*. **1991**, *95*, 4618.
(49) Pulay, P. *Mol. Phys*. **1969**, *17*, 197.
(50) Pulay, P. In *Modern Theoretical Chemistry*; Schaefer, H. F., Ed.; Plenum Press: New York, 1977; Vol. 4, p 153.
(51) (a) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys*. **1969**, *51*, 2657. (b) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys*. **1970**, *52*, 2769.
(52) Roothaan, C. C. J. *Rev. Mod. Phys*. **1951**, *23*, 69.
(53) Gineitye, V. *Int. J. Quantum Chem*. **1999**, *72*, 559.
(54) Löwdin, P. O. *Adv. Quantum Chem*. **1970**, *5*, 185.
(55) Löwdin, P. O. *THEOCHEM* **1991**, *229*, 1.
(56) Weinhold, F.; Carpenter, J. E. *THEOCHEM* **1988**, *165*, 189.

(57) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495.

(58) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, J. C.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 5503.

(59) Pu, J.; Thompson, J. D.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Rinaldi, D. L.; Gao, J.; Cramer, C. J.; Truhlar, D. G. GAMESSPLUS-Version 4.1, University of Minnesota, Minneapolis, MN, 2003. (See http://comp.chem.umn.edu/gamessplus.)

(60) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(61) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J. E.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 2586.

(62) (a) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939. (b) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *104*, 2797.

(63) (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 72. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(64) (a) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654. (b) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta.* **1973**, *28*, 213.

(65) (a) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. *J. Comput. Chem.* **1983**, *4*, 294. (b) Spitznagel, G. W. Diplomarbeit; Universität Erlangen, Germany, 1982.

(66) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281.

(67) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(68) Baker, J. *Theor. Chim. Acta* **1985**, *68*, 221.

(69) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.

(70) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002**, *5*, 117.

(71) Lynch, B. J.; Truhlar, D. G. *Theor. Chem. Acc.*, in press.

(72) Carroll, D. L. FORTRAN Genetic Algorithm (GA) Driver, http://cuaerospace.com/carroll/ga.html.

(73) Pople, J. A.; Nesbet, R. K. *J. Chem. Phys.* **1954**, *22*, 571.