A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations

Jiali Gao,*,† Patricia Amara,‡ Cristobal Alhambra,† and Martin J. Field*,‡

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260, and Institut de Biologie Structurale Jean-Pierre Ebel 41, Avenue des Martyrs 38027 Grenoble, Cedex 1, France

Received: February 3, 1998; In Final Form: April 8, 1998

A generalized hybrid orbital (GHO) method has been developed at the semiempirical level in combined quantum mechanical and molecular mechanical (QM/MM) calculations. In this method, a set of hybrid orbitals is placed on the boundary atom between the QM and MM fragments, and one of the hybrid orbitals participates in the SCF calculation for the atoms in the QM region. The GHO method provides a well-defined potential energy surface for a hybrid QM/MM system and is a significant improvement over the "link-atom" approach by saturating the QM valencies with hydrogen atoms. The method has been tested on small molecules and yields reasonable structural, energetic, and electronic results in comparison with the results of the corresponding QM and MM approximations. The GHO method will greatly increase the applicability of combined QM/MM methods to systems comprising large molecules, such as proteins.

I. Introduction

Combined quantum mechanical and molecular mechanical (QM/MM) methods offer a powerful tool for the study of chemical reactions in solutions and in enzymes.[1] In this approach, a large molecular system is partitioned into a QM and an MM region. The QM region typically consists of the reactant molecule and participating amino acids, whereas the rest of the protein−solvent system is included in the MM region.[2−5] Consequently, bond-forming and -breaking processes can be modeled by quantum chemical methods in the presence of the explicit protein−solvent environment. For solution-phase systems, the QM and MM boundary division is straightforward since there are no covalent bonds connecting the solute and solvent molecules. For biopolymers, however, the QM/MM separation typically occurs within a large molecule, where one or more covalent bonds must be divided between the two regions. In this case, it is far less evident that the molecular fragments in both the QM and MM regions can be adequately treated. As a result, the QM/MM boundary division presents a major drawback for applications of combined QM/MM methods in biomolecular systems.

A number of approaches have been proposed to circumvent this problem. A simple solution is to saturate the valencies of the QM fragment with hydrogen atoms ("link atoms").[2,3,6] The link "atom" may also be methyl groups or pseudo-halogen atoms, which can be parametrized to mimic the properties of the neighboring MM fragment. The link-atom method has been widely used in combined QM/MM studies of proteins.[7] However, the method lacks a unique definition of the total energy for the hybrid QM/MM system since arbitrary corrections must be made to remove the effect of the added link atoms.[8] For example, Coulombic interactions between a link atom and MM atoms are excluded from the QM Hamiltonian in most implementations.[2,3,8] However, this results in arbitrary charge polarization in the QM system and unrealistically large partial charges on the hydrogen link atom. Furthermore, recent studies indicate that QM/MM results strongly depend on the placement of the link atom and on the number of MM atoms excluded from the classical electrostatic field that interacts with the QM region.[6]

The second approach, which is more appealing, is the local self-consistent field (LSCF) algorithm developed by Rivail and co-workers.[9] In the LSCF method, the bonds between the QM and MM fragments are represented by strictly localized bond orbitals, which are obtained by separate quantum mechanical calculations of small model compounds. These localized bond orbitals are assumed to be transferable for use in proteins and are kept constant throughout the SCF calculation. Several studies indicate that the LSCF method can yield good results in energy minimization of reaction pathways in proteins, and the assumption of transferability of bond orbitals appears to be valid.[9] An elegant feature in the LSCF method is that it does not require the addition of link atoms into the system, although the parameters for the localized bond orbitals have to be determined from model studies for each new system in the LSCF treatment.

Finally, a double-iteration scheme has been used by Morokuma and co-workers and by Bersuker et al.[10] Here, the geometry of the QM fragment, which is saturated with link atoms as discussed above, is first optimized. Then, the MM fragment is optimized using molecular mechanics in the presence of the fixed QM fragment that was obtained in the previous step. The process continues until the energy of the entire system is converged. In the implementation by Morokuma and co-workers, Coulombic interactions between the QM and MM region are ignored without allowing for any polarization of the QM region.[10a] On the other hand, Bersuker et al. introduced an additional, intermediate QM fragment encompassing the first QM fragment to smooth out the transition from the QM to MM region.[10b]

In this article, we describe a generalized hybrid orbital (GHO) method for the treatment of QM/MM covalent bonds. In this method, we make use of hybrid atomic orbitals, analogous to Rivail's LSCF orbitals, as basis functions on the boundary atoms

† jiali@tams.chem.buffalo.edu.
‡ mjfield@ibs.ibs.fr.

S1089-5639(98)00989-X CCC: $15.00 © 1998 American Chemical Society
Published on Web 05/20/1998

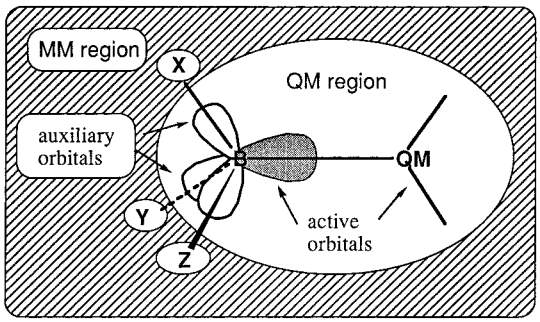

Figure 1. Schematic representation of the division of a QM-MM bond and partition of the hybrid orbitals on the MM boundary atom B.

of the MM fragment. These hybrid orbitals are divided into auxiliary and active orbitals, the latter of which are optimized along with all other atomic orbitals of the QM fragment in the SCF calculations. Consequently, the chemical bond connecting the QM and MM fragments is explicitly treated without introducing spurious "link atoms". Furthermore, in contrast to the LSCF approach, the GHO method does not need to be reparametrized every time when a new system is studied. In what follows, the theoretical background of the generalized hybrid orbital method will be presented along with computational details. The method is tested on some hydrocarbon model systems.

II. Method

We consider a molecular system partitioned into two fragments: one treated quantum mechanically and the other represented classically by an MM force field. The covalent bond connecting the QM and MM fragments is assumed to be a σ bond, with two boundary atoms denoted by Q and B. In particular, for applications to large organic molecules and biopolymers, the MM boundary atom B will most likely be chosen as an sp^3 carbon, which is also connected to three other MM atoms (groups), X, Y, and Z (Figure 1).

In this article, we limit our discussion to semiempirical methods, although extension to ab initio calculations can easily be made and will be presented in a later publication. The boundary atom has the standard valence s and p basis functions as all the other non-hydrogen atoms in the QM fragment. These four sp atomic orbitals are transformed into a set of orthogonal hybrid orbitals, $\{\eta_B, \eta_X, \eta_Y, \eta_Z\}$, which can be defined by the local geometry of the boundary atom.[11] These hybrid orbitals will be used along with the N atomic orbitals $\{\chi_\mu, \mu = 1, ..., N\}$ in the QM region to determine the QM energy. However, only one of the hybrid orbitals, η_B, which lies along the Q−B bond will be optimized. Therefore, η_B and $\{\chi_\mu\}$ are the active orbitals in SCF calculations, whereas $\{\eta_X, \eta_Y,$ and $\eta_Z\}$ act as a set of auxiliary orbitals, generating, along with the nucleus charge, an effective core potential (ECP) for the boundary atom. The valency of the QM fragment is fully satisfied by introducing the active hybrid orbital along with one "active electron" from the boundary atom. The ECP can be optimized by modifying the original semiempirical parameters for the boundary atom through examination of the bonding properties of the active hybrid orbital.

The use of mixed hybrid and atomic orbitals in combined QM/MM computations was initially described by Warshel and Levitt[2] and has been extensively explored by Rivail and coworkers.[9] Warshel and Levitt proposed the use of a single hybrid orbital for the MM boundary atom in SCF calculations, although the method was not developed further since the Warshel group has instead focused on the use of an empirical valence bond method.[1c,2] In Rivail's approach, three active orbitals on the QM "frontier" atom are included in the SCF optimization, while the density of the hybrid orbital in the direction to the MM frontier atom is kept frozen.[9] In the spirit of Rivail's approach, we have generalized the hybrid orbital method by recognizing the fact that the auxiliary orbitals may be used to mimic the effective core potential for the active electrons of the MM boundary atom. Rather than parametrizing the charge density of hybrid orbitals for each specific system, we have decided to optimize the semiempirical parameters for the boundary atom to reproduce bonding properties of full QM systems. As a result, the parameters in our generalized hybrid orbital (GHO) method are expected to be general and transferable in the same way as all the semiempirical parameters.

The Hartree−Fock wave function, Ψ, for the QM subsystem is written as a single Slater determinant of M doubly occupied molecular orbitals (MO) $\{\phi_i^H\}$, which are linear combinations of the N atomic orbitals $\{\chi_\mu\}$ plus one hybrid orbital η_B from the boundary atom:

$$\phi_i^H = \sum_\mu^N c_{\mu i}\chi_\mu + c_{Bi}\eta_B \tag{1}$$

where the superscript H is to emphasize the fact that the MO is constructed from a hybrid orbital basis set. In the neglect of diatomic differential overlap (NDDO) approximation that is made in the semiempirical MNDO, AM1, and PM3 models,[12−15] the MO coefficients must satisfy the orthonormal constraints:

$$\Lambda_{ij} = \sum_\mu^N c_{\mu i}c_{\mu j} + c_{Bi}c_{Bj} - \delta_{ij} \tag{2}$$

Note that the constraint conditions are independent of the overlap matrix. Consequently, the auxiliary hybrid orbitals $\{\eta_X, \eta_Y, \eta_Z\}$ are also orthogonal to the molecular orbitals $\{\phi_i^H\}$.

If the density matrix for the QM subsystem is $\mathbf{P}^H$, which has dimensions of $(N + 1) \times (N + 1)$, the total energy of the hybrid QM/MM system can be written as follows:

$$E = \sum_{\mu\nu}^{N+1} P_{\mu\nu}^H H_{\mu\nu}^{eff} + {}^1/_2 \sum_{\mu\nu}^{N+1} \sum_{\lambda\sigma}^{N+1} P_{\mu\nu}^H P_{\lambda\sigma}^H [(\mu\nu,\lambda\sigma) - {}^1/_2(\mu\sigma,\lambda\nu)] + E_{nuc}^{QM} + E_{nuc}^{QM/MM} + E_{aux} + E_{MM} \tag{3}$$

In eq 3, $H_{\mu\nu}^{eff}$ is an element of the "effective" one-electron matrix, E_{nuc}^{QM} and $E_{nuc}^{QM/MM}$ are the nuclear Coulombic energies for the QM region and between QM and MM atoms, E_{aux} is the interaction energy involving the auxiliary orbitals, and E_{MM} is the molecular mechanics energy for the MM region. The element of the "effective" one-electron matrix, $H_{\mu\nu}^{eff}$, is given in eq 4:

$$H_{\mu\nu}^{eff} = H_{\mu\nu}^H + I_{\mu\nu}^H + {}^1/_2 \sum_b^3 P_{bb}^H [(\mu\nu,bb) - {}^1/_2(\mu b,b\nu)] \tag{4}$$

where b specifies the boundary auxiliary orbitals, $H_{\mu\nu}^H$ is the "standard" one-electron matrix element for the QM subsystem, $I_{\mu\nu}^H$ is the QM/MM one-electron integral due to all classical partial charges in the MM region,[1] and P_{bb}^H is the charge density for the auxiliary hybrid orbitals (see below). The E_{aux} term in eq 3 includes electron−nucleus attraction and Coulomb-exchange repulsion interactions associated with the auxiliary

orbitals, and it is defined by eq 5:

$$E_{\text{aux}} = \sum_{b}^{3} P^{\text{H}}_{\text{bb}}(H^{\text{H}}_{\text{bb}} + I^{\text{H}}_{\text{bb}}) + {}^1/_2 \sum_{b,c}^{3} P^{\text{H}}_{\text{bb}} P^{\text{H}}_{\text{cc}}[(bb,cc) - {}^1/_2(bc,cb)] \quad (5)$$

where H^{H}_{bb} and I^{H}_{bb} are the electron−nucleus attraction integrals of the auxiliary orbitals for the QM and MM nucleus charges, respectively. Note that the second sum in eq 5 is constant for a given set of P^{H}_{bb} at a fixed geometry.

The electronic integrals in eqs 3−5 are expressed in terms of the mixed hybrid and atomic orbitals, which require appropriate integral transformations. Alternatively, the density matrices for the active and auxiliary orbitals can be combined by adding the auxiliary orbital density P^{H}_{bb} at the corresponding diagonal positions, resulting in a total density matrix $\mathbf{P}^{\text{H}}_{\mathbf{t}}$ in the generalized hybrid orbital basis set. Note that $\mathbf{P}^{\text{H}}_{\mathbf{t}}$ has dimensions of $(N + 4) \times (N + 4)$. If $\mathbf{P}^{\text{H}}_{\mathbf{t}}$ is transformed into a density matrix in the atomic orbital basis, $\mathbf{P}^{\text{AO}}_{\mathbf{t}}$, eq 3 can be simplified to the standard Hartree−Fock energy expression:

$$E = \sum_{\mu\nu}^{N+4} P^{\text{AO}}_{\mu\nu}(H^{\text{AO}}_{\mu\nu} + I^{\text{AO}}_{\mu\nu}) + {}^1/_2 \sum_{\mu\nu}^{N+4} \sum_{\lambda\sigma}^{N+4} P^{\text{AO}}_{\mu\nu} P^{\text{AO}}_{\lambda\sigma}[(\mu\nu,\lambda\sigma) - {}^1/_2(\mu\sigma,\lambda\nu)] + E^{\text{QM}}_{\text{nuc}} + E^{\text{QM/MM}}_{\text{nuc}} + E_{\text{MM}} \quad (6)$$

In eq 6, all electron integrals are expressed in terms of the atomic orbitals.

III. Computational Details

The performance of the GHO method is examined using the semiempirical AM1 model and the CHARMM-24 force field for the model compounds ethane, propane, and butane.[14,16] In this section, we summarize the parameters and rules developed for our GHO method.

It is clear from eqs 4 and 5 that the present GHO method requires a knowledge of the density matrix elements P^{H}_{bb} for the auxiliary hybrid orbitals, which are kept constant in the SCF calculation. The primary criterion for a proper value of P^{H}_{bb} is that it should reflect the bond polarity between the boundary atom B and the MM atoms. In the spirit of Mulliken population analysis,[17] P^{H}_{bb} is defined by eq 7:

$$P^{\text{H}}_{bb} = 1 - q_{\text{B}}/3 \quad (7)$$

where q_{B} is the atomic partial charge of the boundary atom if it were treated as an MM atom in the force field. Thus, the MM partial charge q_{B} on the boundary atom is equally distributed to the three auxiliary orbitals. It is interesting to note that, in molecular mechanical force fields such as CHARMM, the CH_n unit is typically treated as a neutral charge group.[16] The definition of eq 7 ensures such a charge neutrality by transfering the MM partial charge on the boundary atom to its auxiliary orbitals.

To define the QM and MM region, an important criterion is that charge separation between the QM and MM fragment should be minimal and should not extend farther beyond the boundary atom. Consequently, the small amount of charge transfer between the two fragments can be adequately represented by the charge polarization of the active hybrid orbital on the boundary atom. If this criterion cannot be satisfied, the size of the QM fragment should be increased.

Since the auxiliary orbitals are excluded from the orbital optimization, standard semiempirical parameters must be modified to reflect the characteristics of the effective core potential for the boundary atom. The semiempirical parameters β_{s} and β_{p} describe chemical bonding features, while the U_{ss} and U_{pp} values reflect the relative electronegativity of the atom. Consequently, these parameters may be adjusted to yield the desired geometry and partial charges associated with the boundary atoms. This has been accomplished by examining the model compound propane, for which one methyl group is treated as the QM fragment with the boundary atom at the C_2 position, and the remaining atoms are grouped into the MM region. The β_{s} and β_{p} parameters are adjusted such that the optimized $C_1(Q)-C_2(B)$ bond distance is in accord with the AM1 value when the entire molecule is treated quantum mechanically. In addition, the bond angles between the active and auxiliary orbitals on the boundary atom are determined by the hybridization of these orbitals, which imposes a second constraint on the optimization of the β parameters. Therefore, β_{s} and β_{p} are uniquely defined by these two geometrical constraints in our GHO method. We found that for a carbon boundary atom the standard AM1 parameters β_{s} and β_{p} need to be scaled by factors of 0.35 and 1.90, respectively. The remaining two parameters, U_{ss} and U_{pp}, are determined by matching the Mulliken population charge for the boundary atom C_2 and the MM partial charge (−0.18 e) used in the CHARMM force field. It turns out that only U_{pp} requires adjustment by a scaling factor of 0.977. These parameters are used in all calculations reported in this article (Table 1). We anticipate that the same procedure may be used to obtain boundary ECP parameters for heteroatoms.

TABLE 1: Modified Parameters for the Carbon Boundary Atom in the Combined Semiempirical GHO-AM1/CHARMM Potential; All Other QM and MM Parameters for Carbon Taken Directly from the Standard AM1 or CHARMM Parameter Sets

item	AM1 or CHARMM	GHO
β_{s}	−15.715 783	−5.500 524
β_{p}	−7.719 283	−14.666 638
U_{pp}	−39.614 239	−38.703 112
MM Bond Stretching Parameters (Å)		
$R_0(\text{CT2}-\text{C}_\text{B}2)$[a]	1.530	1.485
$R_0(\text{CT3}-\text{C}_\text{B}2)$[a]	1.528	1.478
$R_0(\text{C}_\text{B}-\text{HA})$[a]	1.111	1.091

[a] The subscript B specifies that the carbon atom is a boundary link atom. CT2 and HA are standard atom types in the CHARMM force field, whereas C_B2 corresponds to the CHARMM CT2 type and C_B indicates any carbon type.

Nonbonded van der Waals interactions between QM and MM atoms are treated classically using the standard force field convention,[16,18] whereas QM/MM electrostatic interactions are determined quantum mechanically by including the partial charges of *all MM atoms* in the QM/MM interaction Hamiltonian. This differs from the link-atom approach, in which MM atoms within one or two bonds of any QM atoms are typically excluded from the classical electric field, resulting in a highly distorted electric field at the QM/MM interface.[2,3,6,8] The transition from the QM region to the MM region in the our GHO method is much smoother since no electrostatic terms have been neglected. For bond stretch, angle bend, and torsional terms, those involving exclusively QM atoms are removed from the force field calculation because they are determined quantum mechanically. Thus, any bonded terms involving at least one MM atom, which are not included in the QM expression, are maintained classically.

It perhaps should be emphasized that the boundary atom in the GHO method is in essence both quantum mechanical and classical because the connection between the QM fragment and

the boundary atom is determined by the HF-SCF optimization, whereas its junction with the MM fragment is defined by the MM force field. The boundary atom serves as a channel, bridging the QM and MM regions.

Below, we outline an algorithm for the SCF calculation using the GHO method.

1. Determine the transformation matrix **T** for the interconversion between the atomic orbitals and the set of mixed atomic and hybrid orbitals, $\mathbf{C}^{H} = \mathbf{T}^{-1}\cdot\mathbf{C}^{AO}$. Calculate the $[(N+1) \times (N+1)]$-dimensional density matrix, $\mathbf{P}^{H}$, using only the active orbitals in the GHO basis set. In the present study, the hybrid orbitals on the boundary atom are kept to be sp^3 hybridized.

2. Expand $\mathbf{P}^{H}$ by adding the densities for the auxiliary orbitals to yield $\mathbf{P}_{t}^{H}$. Then, transform $\mathbf{P}_{t}^{H}$ into the atomic orbital basis, $\mathbf{P}_{t}^{AO} = (\mathbf{T}^{-1})^{\dagger}\cdot\mathbf{P}_{t}^{H}\cdot(\mathbf{T}^{-1})$. The subscript **t** indicates that the dimension of the matrixes are $(N+4) \times (N+4)$.

3. Construct the Fock matrix $\mathbf{F}_{t}^{AO}$ in the atomic orbital basis set with $\mathbf{P}_{t}^{AO}$.

4. Transform $\mathbf{F}_{t}^{AO}$, $\mathbf{F}_{t}^{H} = \mathbf{T}^{\dagger}\cdot\mathbf{F}_{t}^{AO}\cdot\mathbf{T}$.

5. Remove the columns and rows corresponding to the auxiliary hybrid orbitals to yield the Fock matrix, $\mathbf{F}^{H}$. Diagonalize $\mathbf{F}^{H}$ and compute the new density matrix $\mathbf{P}^{H}$.

6. Test convergence. If not satisfied, go to step 2.

Here we point out that the matrix transformations in steps 2 and 4 are particularly simple since the **T** matrix has nondiagonal elements only for the boundary atom, which takes a negligible amount of CPU time. All calculations are carried out using the program CHARMM, in which the GHO method has been implemented. Geometry optimizations are performed using the conjugated gradient and adopted basis Newton−Raphson methods.[18]

IV. Results and Discussion

Three model compounds, ethane, propane, and butane, have been chosen to assess the performance of the GHO method for the treatment of the division of covalent bonds in combined QM/MM calculations. In each case, the QM and MM division is made across a C−C bond with the QM region treated by the semiempirical AM1 model[14] and the MM region represented by the CHARMM force field.[18] The carbon atom directly linked to the QM fragment is chosen as the boundary atom. Thus, only three hydrogen atoms in ethane remain in the MM region. For propane, the GHO boundary atom can be placed at the C_2 or C_3 position. These two situations provide an assessment of the QM-B-MM and QM-QM-B bond angle representations in the GHO method. Three cases exist in butane, corresponding to the location of the boundary atom at the C_2, C_3, and C_4 positions, which provide a test of the torsional energy in various QM/MM representations.

The geometries determined using the combined GHO-AM1/CHARMM method are shown in Figures 2−4, along with those calculated at the HF/6-31G*, AM1, and MM/CHARMM levels. Of particular interest are comparisons of the QM/MM and AM1 results for the QM part of the molecule and of the QM/MM geometries for the MM region with that predicted using the CHARMM force field. For alkanes, the AM1 model underestimates the C−C bond lengths by ca. 0.02 Å and overestimates the C−H distances by about 0.03 Å in comparison with the HF/6-31G* results. On the other hand, the MM/CHARMM values are in better accord with the ab initio data for C−C distances. In the GHO-AM1 model, the QM-MM bond, C_Q-C_B, is fully represented quantum mechanically. Thus, the ability to reproduce the AM1 value for the C_Q-C_B bond distance is essential in the present GHO method. For the six C_Q-C_B bonds

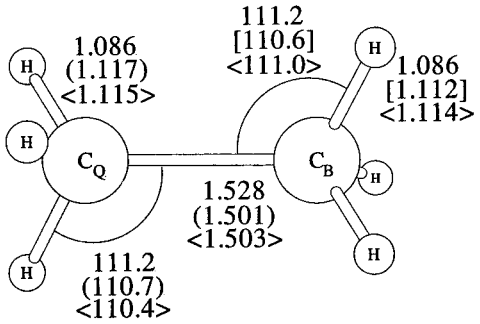

Figure 2. Computed geometries for ethane at the HF/6-31G*, AM1 (in parentheses), CHARMM [square parentheses], and GHO <angular parentheses> levels. Only the relevant AM1 or CHARMM values are listed for comparison with the GHO results. The fragment to the left-hand side of the boundary atom C_B is the QM fragment. This convention is used in all other figures.

present in the test cases (Figures 2−4), we obtained an average unsigned error of only 0.003 Å in comparison with the corresponding AM1 data. To illustrate the transferability of the β_s and β_p parameters for the boundary atom, Table 2 lists the C_Q-C_B bond lengths for a series of organic compounds. The agreement with the original AM1 values is also good. Importantly, the trend of bond length variations is clearly reflected in the GHO-AM1 calculations.

With the introduction of the QM fragment and the boundary atom, the geometry of the MM fragment may also be affected. We found that the only significant effect is the bond distance directly connected between the boundary atom and MM atoms. Specifically, without modification of the force field parameters, the C_B-C_{MM} distance is elongated from a typical 1.53 Å to about 1.56 Å in the GHO approach, and the C_B-H_{MM} distance changes from 1.11 to 1.13 Å. These changes are small enough not to warrant serious concerns, although they can easily be corrected, as has been done in the present study, by reducing the corresponding bond stretch parameters, R_0(C−C) and R_0(C−H), by values of about 0.05 and 0.02 Å, respectively (Table 1). With these corrections, the overall unsigned errors from the combined GHO-AM1/CHARMM method are ca. 0.002 Å for bond lengths and 0.3° for bond angles.

Another important criterion for the treatment of the QM-MM covalent bond is that the electronic structure of the full QM system should be retained in the QM fragment. Several properties, including the molecular electrostatic potential and charge density, may be used to illustrate this. Here, however, the Mulliken population analysis is adopted since it contains the essential features of charge polarization in a molecular system.[17] The aim is to produce a charge distribution for the QM fragment that is comparable to that of the entire molecule treated quantum mechanically and a charge density for the boundary atom analogous to the MM partial charge for that atom. It turns out that the parametrization of the ECP for the boundary atom is rather simple. The only parameter to be changed is the U_{pp} value in the original AM1 model, which is reduced by a factor of 0.977 (Table 1). This scaling factor is derived by producing a Mulliken charge of −0.18 e on the C_2 boundary carbon in $H_3C_{(Q)}-C_{(B)}H_2-CH_3$, a value used in the CHARMM force field for this atom. Since the charge densities for the three auxiliary orbitals are 1.06, 1.06, and 1.06 for the C−H, C−H, and C−C bonds from eq 7, the Mulliken charge on the boundary atom is entirely due to the "excess" bond density from the auxiliary orbitals. The implication of this charge distribution is that there is a balanced electronegativity between the QM fragment and the boundary atom with minimal

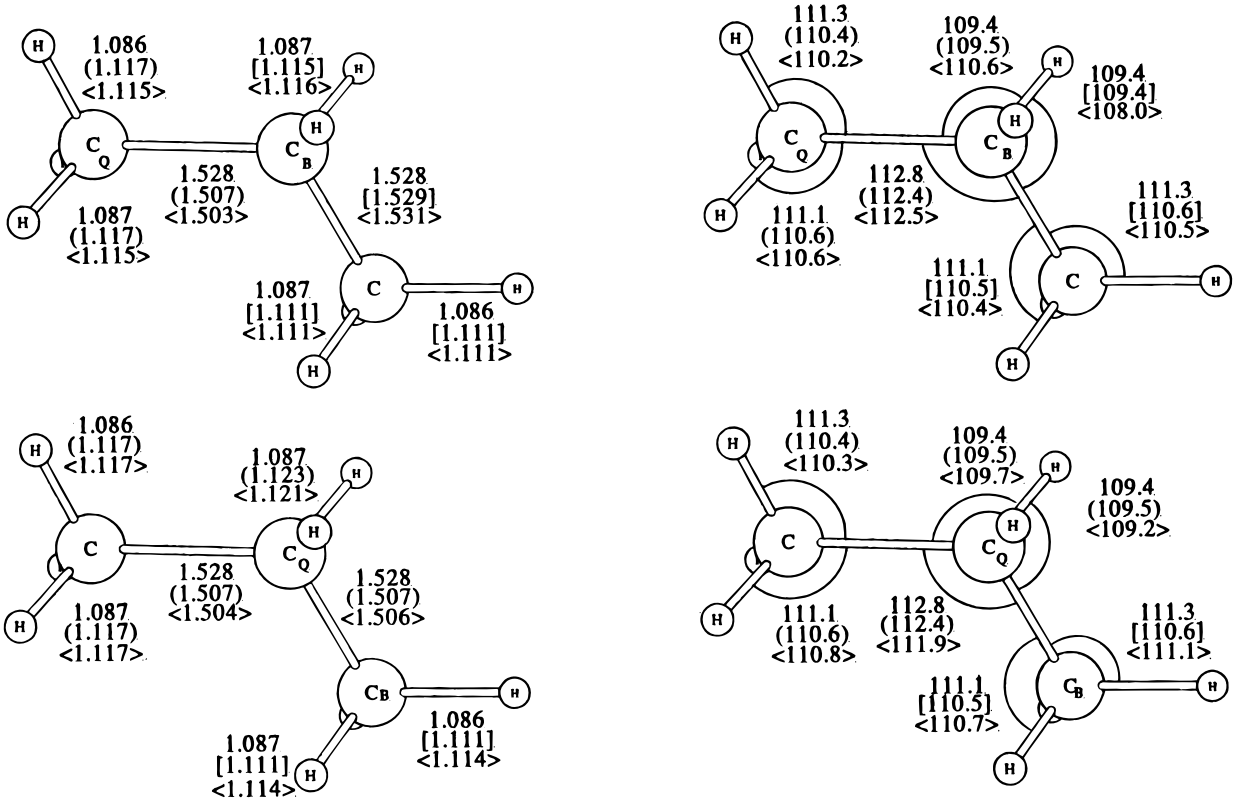

Figure 3. Computed geometries for propane for the boundary atom at (a, top) the C_2 position and (b, bottom) the C_3 position. See caption of Figure 2 for details of the notation.

charge transfer in these two regions, as it should be, for the hydrocarbon model systems. This is consistent with the results of full QM calculations. Charge polarization will occur, of course, for polar molecules, as demonstrated for acetic acid when the carboxyl group is treated as a QM fragment and the methyl carbon as a boundary atom (Table 3). In this case, the total charge of the methyl group is computed to be 0.068 e using the GHO method, which is slightly smaller than the AM1 value (0.133 e). As depicted in Figure 5, the Mulliken charges for the QM fragments in the hydrocarbon compounds are all in reasonable accord with the full QM AM1 results.

Bersuker et al. described a double self-consistent (DSC) procedure to allow charge transfer between the QM and MM fragments.[10b] In this approach, a central QM fragment is first selected, and its valency is saturated by the use of link atoms. The geometries of the QM and MM fragments are optimized separately. Then, an intermediate fragment adjacent to the "border" atom from the MM fragment is separated for a second QM-SCF calculation with link atoms to saturate its valency. The density matrix elements for the border atom are adjusted by the difference of the densities from the two calculations, which are subsequently used in the next iteration of the SCF calculation for the central QM fragment. In a test study of picket-fence porphyrin, it is found that each phenyl substituent loses 0.5−0.6 electrons to the iron porphyrin, with a total of 2−2.4 electrons transferred from the substituents.[10b] Although this finding is surprising in view of the amount of charge transfer, the method is complicated by problems of link atoms.[6,8] In addition, it is not clear how the MM charges are treated and adjusted as the QM fragment gains substantial electron density.

Finally, it is essential to be able to use the GHO method to predict conformational energies for a hybrid QM/MM system. For ethane and propane, the internal torsions only involve the rotation of a methyl group. The barrier height for ethane is predicted to be 2.94 kcal/mol using the GHO method, which may be compared with the CHARMM value of 2.91 kcal/mol. For propane, the methyl torsional barriers are 3.12 and 3.07 kcal/mol from the GHO and CHARMM calculations, respectively. The torsional energy profile around the central C_2-C_3 bond of butane, which is shown in Figure 6, is particularly interesting because of its relevance to large molecular systems. For the two cases in which the link atom is located at the C_2 (L2) and C_3 (L3) position in butane, the QM fragment does not contain the torsional terms around the C_2-C_3 bond. Thus, the torsional energy has the largest contribution from MM terms, and indeed, the agreement with the potential energy profile predicted with CHARMM is good, in both the shape of the profile and barrier heights. The only noticeable discrepancy is the predicted energy for the gauche conformer, which is 0.54−0.56 kcal/mol above the anti conformer in the GHO calculation. This may be compared with a value of 0.88 kcal/mol from CHARMM, 0.71 kcal/mol from AM1, and 0.60 kcal/mol at the MP3/6-311G** level.[20]

When the boundary atom is placed at the terminal C_4 (L4) position, the torsional potential energy surface is largely determined quantum mechanically by the GHO-AM1 method. As a result, the torsional energy profile (long dashed line in Figure 6) should be compared with the AM1 result. The agreement between the two computations is also reasonable. However, the AM1 model itself yields poor conformational energies for most organic molecules including butane in comparison with ab initio or experimental data. The barrier heights are proportionally too small from the AM1 calculations in comparison with high-level ab initio data. This is a documented problem in semiempirical models.[19] Thus, it is desirable to extend the present approach to ab initio calculations. Importantly, the present results demonstrate that the GHO method is capable of reproducing the energies for the individual

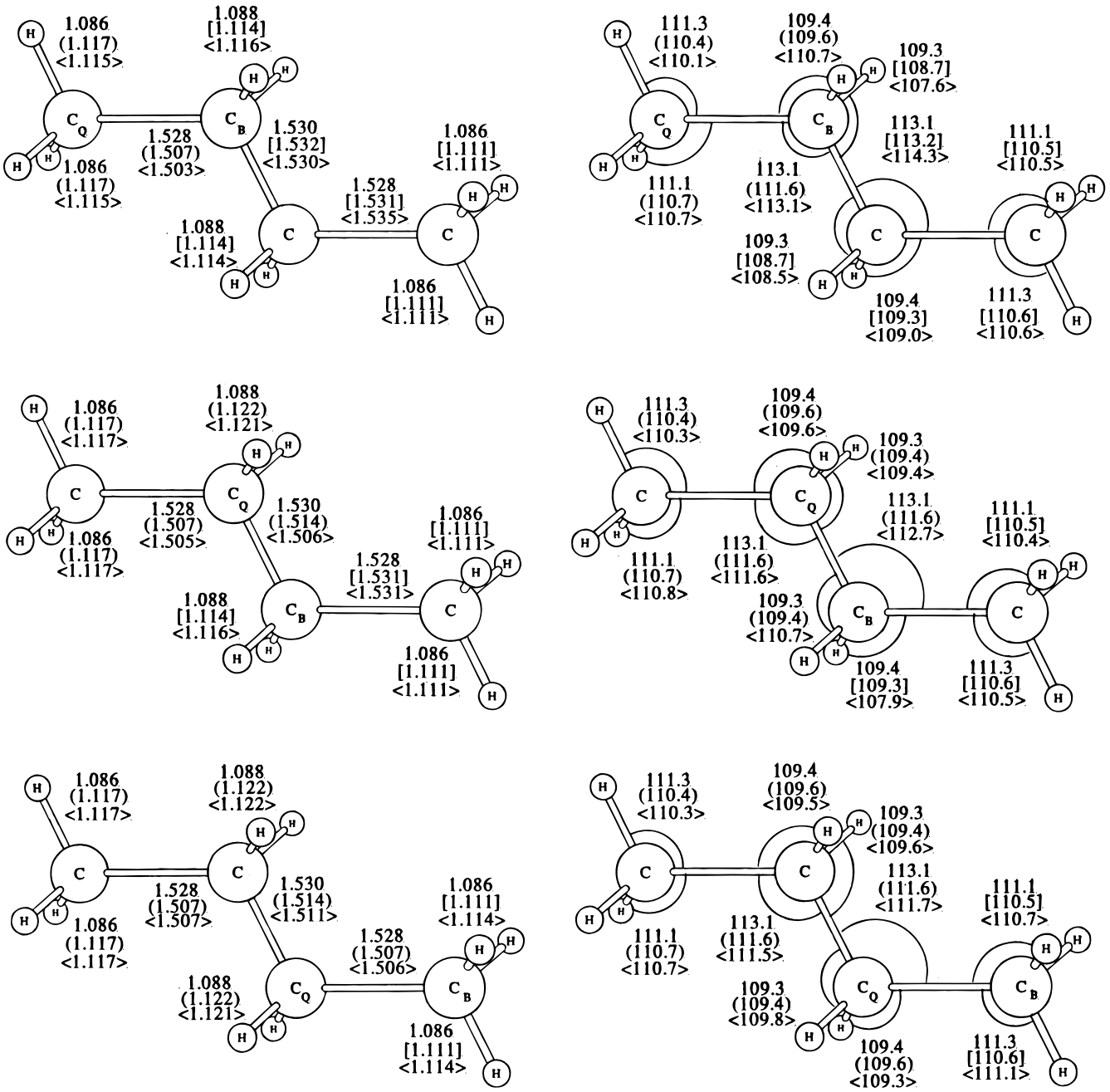

Figure 4. Computed geometries for butane corresponding to the boundary atom located at (a, top) the C_2 position, (b, middle) the C_3 position, and (c, bottom) the C_4 position. See caption of Figure 2 for details of the notation.

TABLE 2: Comparison of Optimized Bond Distances between the QM and Boundary Carbon Atoms for a Series of Organic Molecules Using the AM1 and Combined GHO-AM1/CHARMM Potentials (Å)

molecule[a]	AM1	GHO-AM1/CHARMM
$H_3C_B-CH_2NH_2$	1.522	1.517
$H_3C_B-CH_2OH$	1.512	1.511
$H_3C_B-COCH_3$	1.506	1.504
$H_3C_B-CH_2COOH$	1.506	1.504
$H_3C_B-CH_2SH$	1.500	1.502
H_3C_B-COOH	1.486	1.492
$H_3C_B-C_6H_5$	1.481	1.488

[a] The subscript B indicates that the methyl carbon is the boundary link atom. The functional groups on the right-hand side of the specified bond are the QM fragments.

QM and MM fragments as well as the QM-MM covalent bond in comparison with the corresponding QM and MM data.

TABLE 3: Atomic Charges Determined from AM1 and GHO-AM1 Calculations along with Those Used in the CHARMM Force Field for Acetic Acid (e)

atom	CHARMM	AM1	GHO-AM1/CHARMM
O(=C)	−0.55	−0.361	−0.329
O(−C)	−0.61	−0.321	−0.299
H(O)	0.44	0.243	0.249
C(=O)	0.75	0.306	0.311
H_3C(boundary)	−0.03	0.133	0.068

V. Conclusions

The interest in the development of the present GHO method lies in the fact that the connection between the QM and MM fragments can be adequately treated in a unified manner in combined QM/MM calculations. In this method, the "frontier" or boundary atom of the MM fragment is represented by an effective core potential, with one hybrid orbital participating in

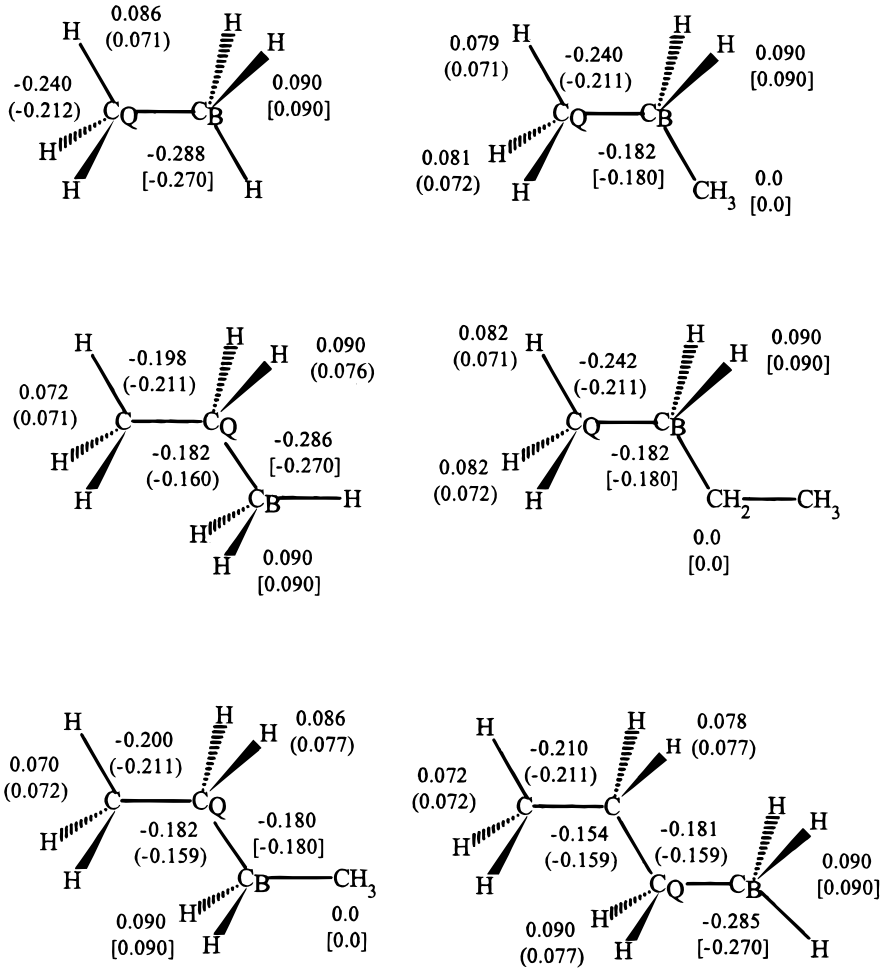

Figure 5. Atomic charges for ethane, propane, and butane from the AM1 (in parentheses) and hybrid GHO-AM1/CHARMM calculations. Empirical charges used in the CHARMM force field are listed in square parentheses for the MM fragment.

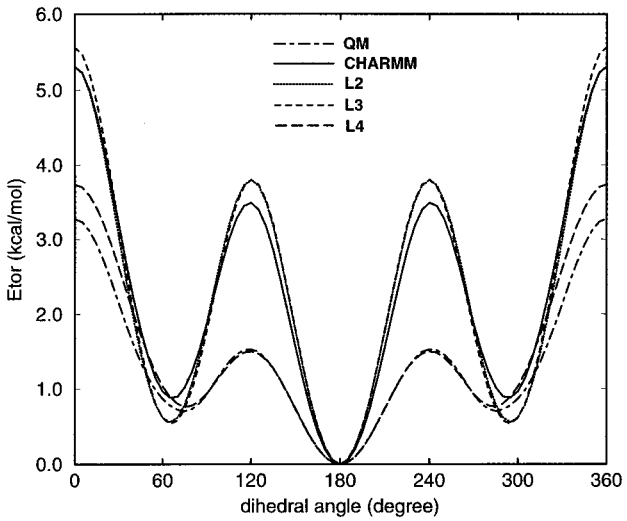

Figure 6. Torsional barriers about the C2−C3 bond in butane. Complete energy minimization was performed at each constrained value of the dihedral angle. The labels, L2, L3, and L4, indicate that the boundary atom is placed at the C2, C3, and C4 carbons, which correspond to increasing size of the QM fragment from a methyl to an ethyl and to a propyl group.

explicit SCF calculations in the QM region. *The GHO method provides a well-defined potential energy surface for a hybrid QM/MM system*, avoiding problems of bond collapsing in geometry optimizations and arbitrary energy corrections using the link-atom approach.[8] Since the use of the auxiliary orbitals on the boundary atom is recognized as a pseudopotential, the atomic parameters in the GHO approach are general and transferable. This is in contrast to the LSCF method, which requires specific parametrization for each new system. The test cases illustrated in this study indicate that the GHO method can yield reasonable structural, energetic, and electronic results. We anticipate that this method will help significantly the application of combined QM/MM methods to biochemical systems.

Acknowledgment. This work has been supported in part by a joint grant from the NSF and CNRS. Gratitude is also expressed to the National Institutes of Health (J.G.) for partial support and to the Institut de Biologie Structurale, the Commissariat à l'Energie Atomique, and the Centre National pour la Recherche Scientifique (M.J.F., P.A.).

References and Notes

(1) (a) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298. (b) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 7, pp 119−185. (c) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and in Solutions*; Wiley: New York, 1991.

(2) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(3) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

(4) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(5) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

(6) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189.

(7) (a) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826. (b) Karplus, M.; Evanseck, J. D.; Joseph, D.; Bash, P. A.; Field, M. *Faraday Discuss.* **1992**, *93*, 239. (c) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans. 2* **1991**, *225*, 1819, 2025. (d) Hartsough, D. S.; Merz, K. M. *J. Phys. Chem.* **1995**, *99*, 384. (e) Chatfield, D. C.; Brooks, B. R. *J. Am. Chem. Soc.* **1995**, *117*, 5561. (f) Mulholland, A. J.; Richards, W. G. *Protein* **1997**, *27*, 9.

(8) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.

(9) (a) Thery, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269. (b) Monard, G.; Loos, M.; Thery, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153. (c) Gorb, L. G.; Rivail, J.-L.; Thery, V.; Rinaldi, D. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1996**, *30*, 1525. (d) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.

(10) (a) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170. (b) Bersuker, I. B.; Leong, M. K.; Boggs, J. E.; Pearlman, R. S. *Int. J. Quantum Chem.* **1997**, *63*, 1051.

(11) Bernardi, F.; Olivucci, M.; Robb, M. A. *J. Am. Chem. Soc.* **1992**, *114*, 1606.

(12) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

(13) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(14) Dewar, M. J. S.; Zeobisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(15) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209, 221.

(16) MacKerell, A. D.; Bashford, D.; Bellot, M.; Dunbrack, R. L.; 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.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Wiorkievicz-Kuczera, J.; Karplus, M. *FASEB J.* **1992**, *6*, A143.

(17) Mulliken, R. S. *J. Chem. Soc.* **1955**, *23*, 1833.

(18) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(19) Stewart, J. J. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1990; Vol. 1, pp 45−81.

(20) Raghavachari, K. *J. Chem. Phys.* **1984**, *81*, 1383.