

*Regular article*

# Generalized hybrid orbital for the treatment of boundary atoms in combined quantum mechanical and molecular mechanical calculations using the semiempirical parameterized model 3 method

Mireia Garcia-Viloca, Jiali Gao

Department of Chemistry and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN55455, USA

Received: 11 March 2003 / Accepted: 3 April 2003 / Published online: 5 December 2003  
© Springer-Verlag 2003

**Abstract.** The application of combined quantum mechanical (QM) and molecular mechanical methods to large molecular systems requires an adequate treatment of the boundary between the two approaches. In this article, we extend the generalized hybrid orbital (GHO) method to the semiempirical parameterized model 3 (PM3) Hamiltonian combined with the CHARMM force field. The GHO method makes use of four hybrid orbitals, one of which is included in the QM region in self-consistent field optimization and three are treated as auxiliary orbitals that do not participate in the QM optimization, but they provide an effective electric field for interactions. An important feature of the GHO method is that the semiempirical parameters for the boundary atom are transferable, and these parameters have been developed for a carbon boundary atom consistent with the PM3 model. The combined GHO-PM3/CHARMM model has been tested on molecular geometry and proton affinity for a series of organic compounds.

**Keywords:** Generalized hybrid orbital – Combined quantum mechanical and molecular mechanical method

## Introduction

Recently, we described a generalized hybrid orbital (GHO) method for the treatment of covalent bonds across a quantum mechanical (QM) region and a molecular mechanical (MM) region in combined QM/MM calculations for large molecular systems [1, 2]. The GHO method was initially developed for the semiem-

pirical Austin model 1 (AM1) [3] method [1, 2], and it was a generalization of the strictly localized self-consistent-field (LSCF) method that was used by Rivail and coworkers [4, 5, 6, 7, 8]. In the GHO approach, the boundary atom, typically an  $sp^3$  hybridized carbon atom that separates the QM and MM regions, is represented by a set of four hybrid orbitals. These hybrid orbitals are obtained by transforming the valence  $s$  and  $p$  orbitals on the basis of the local geometry of the boundary atom and its four substituents. The GHO method has been shown to yield good results in computed geometries, partial charges and proton affinities for model compounds [1, 2], and has been successfully applied to a number of enzyme systems [9, 10, 11, 12, 13]. In this article, we report the semiempirical parameters of the GHO atom for the parameterized model 3 (PM3) method [14].

A number of methods have been proposed to terminate the valency of the molecular fragment that is treated quantum mechanically in combined QM/MM calculations [1, 4, 5, 6, 7, 8, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. An early approach that made use of a hydrogen atom to saturate the QM valency was the hydrogen-link atom method, and this is still widely used owing to its simplicity, and it often gives reasonable results if used carefully [15, 16, 18, 19, 20, 21, 22, 23]. However, a major drawback of the method is that it introduces unnecessary extra degrees of freedom into the system and the potential energy of the system is no longer uniquely defined in the presence of the extra atoms [26]. Furthermore, some of the MM partial charges close to the QM region must be deleted to avoid convergence problems. Methods that employ hybrid orbitals to cap the valence at the QM/MM boundary were mainly developed by Rivail and coworkers [4, 5, 6, 7, 8, 24, 25]. The GHO method is a generalization of Rivail's work by Gao and coworkers [1, 2]. Other techniques that deal with the QM/MM boundary include those by Maseras and Morokuma [27], Bersuker et al. [28], Bakowies and Thiel [26], Antes and Thiel [29],

Contribution to the Jacopo Tomasi Honorary Issue

Correspondence to: J. Gao  
e-mail: gao@chem.umn.edu

Zhang et al. [20], Das et al. [23], Murphy et al. [24], Philipp and Friesner [25], and DiLabio et al. [30].

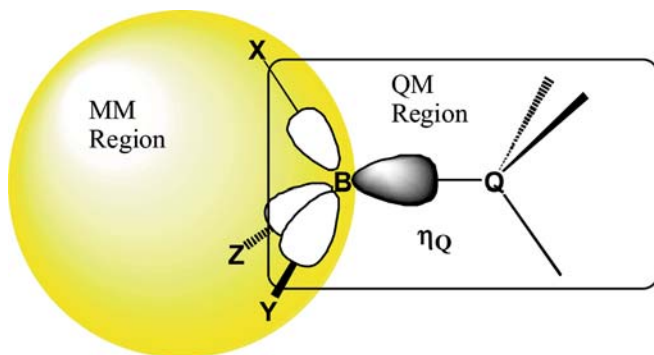
In the LSCF approach that has been popularized by Rivail and coworkers [4, 5, 24], the transformed hybrid orbitals are fixed and three of them are used as atomic basis orbitals in SCF calculations for the QM fragment. The remaining hybrid orbital, which points towards the MM fragment along the covalent bond between the two regions, is strictly frozen with a charge density determined from a separated calculation of a model compound that mimics the full QM/MM system [4, 5]. The GHO method differs from LSCF calculations in that only one hybrid orbital, which represents the bond between the “boundary” atom and the MM subsystem, is optimized in SCF calculations, whereas the density of the rest of the hybrid orbitals is kept fixed for a given geometry [1, 2]. This fixed density which is determined by the MM partial charge of the atom, along with the charge of the nucleus, generates an effective core potential for the boundary atom (Fig. 1). In contrast to the LSCF approach, the parameters needed in the definition of the GHO model are transferable in exactly the same way as all other semiempirical parameters. Another important feature is that the hybridization of the four hybrid orbitals is allowed to vary as the structure of the molecular system changes during the molecular dynamics simulation.

The outline of this paper is as follows. The GHO model is briefly described in Sect. 2. The results of our calculations using the GHO model are presented in Sect. 3, and Sect. 4 concludes with a summary of the present work.

## Method

### The GHO method

Given a hybrid molecular system composed of two subsets of atoms that are treated, respectively, by QM and MM methods, we define the boundary atom, B, as an  $sp^3$  carbon atom that is placed at the boundary between the two regions (Fig. 1). Of course, there is no reason to restrict the boundary atom to be a carbon atom with a specific hybridization, but for most biological applications, this is perhaps the best choice. The boundary atom is considered to be both QM and MM because it is



**Fig. 1.** Schematic representation of the division of a quantum mechanical (QM)-molecular mechanical (MM) bond and partition of the hybrid orbitals on the boundary atom B. The figure shows that the B atom belongs to both QM and MM regions

explicitly represented in the QM subsystem and its connection to the MM region is treated by the standard force field [1]. Specifically, the boundary atom B consists of four  $s-p$  atomic valence orbitals, which are transformed into a set of orthogonal hybrid orbitals  $\{\eta_Q, \eta_X, \eta_Y, \eta_Z\}$ , where X, Y and Z are MM atoms and Q is a QM atom, which form covalent bonds with B. The hybrid orbital  $\eta_Q$  forms a  $\sigma$  bond with the atom Q in the QM treatment. The rest of three hybrid orbitals are not optimized in the SCF calculation of the QM fragment although they participate in the interactions. For simplicity, we describe the case with only one boundary atom between the QM and the MM region. We further define that there are  $N$  atomic orbital (AO) basis functions in the QM fragment with the addition of the active hybrid orbital,  $\eta_B$ , resulting in a total of  $N+1$  basis functions that are used to construct the molecular orbitals (MOs) of the QM subsystem. This basis set comprises mixed atomic and hybrid orbitals to distinguish it from the standard  $s-p$  atomic orbital basis. The Hartree-Fock (HF) wave function,  $\Psi$ , for the QM fragment is written as a single Slater determinant of  $M$  doubly occupied MOs  $\{\phi_u^H\}$ , which are linear combinations of the hybrid basis orbitals:

$$\phi_i^H = \sum_{\mu}^N c_{\mu i} \chi_{\mu} + c_{Q_i} \eta_Q, \quad (1)$$

where the superscript H is to indicate that the MO is constructed from the hybrid orbitals. The orbital coefficients in Eq. (1) are obtained by diagonalizing the truncated  $(N+1) \times (N+1)$  Fock matrix in the mixed atomic and hybrid basis set, which is obtained by dropping the columns and rows corresponding to the auxiliary orbitals in the full  $(N+4)$  dimension [1].

The energy of the QM/MM system is calculated in terms of the AO basis functions that include all the QM and boundary orbitals [1, 2]:

$$E = \sum_{\mu\nu}^{N+4} P_{\mu\nu}^{AO} H_{\mu\nu}^{\text{eff}} + \frac{1}{2} \sum_{\mu\nu}^{N+4} \sum_{\lambda\sigma}^{N+4} P_{\mu\nu}^{AO} P_{\lambda\sigma}^{AO} \left[ (\mu\nu, \lambda\sigma) - \frac{1}{2} (\mu\sigma, \lambda\nu) \right] + E_{\text{QM}}^{\text{nuc}} + E_{\text{QM/MM}}^{\text{vdW}} + E_{\text{MM}}, \quad (2)$$

where the first two summation terms are the QM/MM electronic energy, the third term is the nuclear repulsion energy for the QM nuclei, the fourth term is the van der Waals interaction energy between QM and MM atoms, and the last term is the potential energy of the MM region.

Expressions for the analytical gradients of Eq. (2) have also been derived [2].

### Computational details

The goal of this study is to develop a set of transferable semiempirical parameters for the GHO atom that can be used in combined PM3/CHARMM calculations of protein systems. The structures of the model compounds used in the parameterization of the GHO boundary atom were optimized both using the semiempirical PM3 and the combined GHO-PM3/MM(CHARMM) methods. The latter model has been implemented into the CHARMM program [31], version c28b3. For comparison, the HF/6-31G(d) and molecular mechanics (CHARMM) results are given [1, 2]. The optimization criteria used for the GHO-PM3 model is the same as that used in Ref. [1], i.e., a root-mean-square value of the gradient lower than  $0.001 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ .

In computing the total energy of the QM/MM system with GHO atoms, Eq. (2) requires the knowledge of the density matrix elements,  $P_{bb}^H$ , for the auxiliary hybrid orbitals, which are constant in the SCF calculation, and are transformed into the AO representation in Eq. (2). With a valence basis function used in a semiempirical method [3, 14], there are four valence electrons for carbon, each of which is assigned to a hybrid orbital, and the effective nucleus charge is 4 [1]. For the auxiliary hybrid orbitals, in

the spirit of Mulliken population analysis  $P_{bb}^H$  takes into account the effect of the MM partial charge on the boundary atom, and it is defined by Eq. (3):

$$P_{bb}^H = 1 - \frac{q_B}{3}, \quad (3)$$

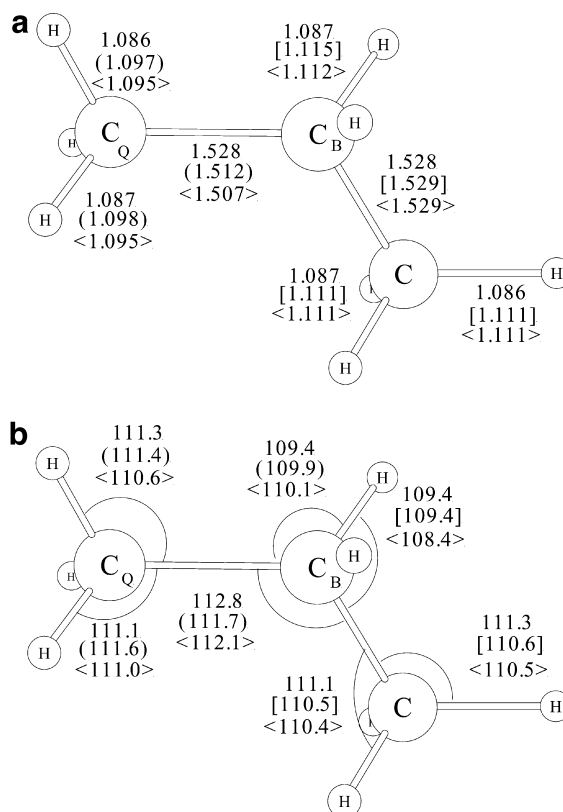
where  $q_B$  is the atomic partial charge of the boundary atom in the force field. Thus, the MM partial charge on the boundary atom is equally distributed to the three auxiliary orbitals. It is noted that when the GHO atom is the carbon atom in a  $\text{CH}_n$  group, which is typically treated as a neutral group in the CHARMM force field [32], the definition of  $P_{bb}^H$  in Eq. (3) ensures such a charge neutrality.

## Results

### Optimization of the PM3 parameters for GHO atoms

The optimization of suitable semiempirical parameters for the GHO carbon atom in combined PM3/CHARMM calculations was achieved by comparing results obtained using the full PM3 calculation and the combined GHO-PM3/CHARMM representation for propane. Three main criteria are used in the optimization [1]. First, the boundary atom must be parameterized to mimic the electronegativity of the carbon atom of the original QM model, i.e., the semiempirical PM3 Hamiltonian. This will ensure the transferability of the GHO boundary model in an environment different from the model compound that is used in the parameter optimization. Hydrocarbon compounds provide an ideal system for this purpose because they are nonpolar, and a balanced charge distribution from the GHO-PM3/CHARMM potential guarantees the electron-withdrawing power of the GHO atom similar to that in the PM3 model. In semiempirical models such as AM1 and PM3, the parameters  $U_{ss}$  and  $U_{pp}$  are closely related to this property [3, 14]. The second criterion is to ensure that the chemical bonding characters between the boundary atom B and the QM atom Q be adequately treated in comparison with the full PM3 calculation. The bonding properties are mainly geometrical parameters, including the bond distances and bond angles involving the boundary atom. The semiempirical resonance integral parameters  $\beta_s$  and  $\beta_p$  are most relevant to these quantities [3, 14]. Finally, chemical properties including acidity and basicity of simple organic molecules should be reasonably reproduced by the combined GHO-QM/MM model in comparison with the results from the full QM representation [14, 32].

The QM/MM partition for propane is shown in Fig. 2, in which one of methyl groups (indicated by  $\text{H}_3\text{C}_Q$ ) is treated as the QM fragment with the boundary atom placed at the second carbon  $\text{C}_2$  position (specified by  $\text{C}_B$ ), and the remaining atoms belong to the MM region. The quantum carbon atom  $\text{C}_Q$  is covalently bonded to the boundary atom  $\text{C}_B$  through a  $\sigma$  bond. The MM charge on the boundary atom  $\text{C}_B$  in the CHARMM force field was set to a value of  $-0.18e$  [32], which is equally partitioned into the fractional density of the auxiliary orbitals. Thus,  $P_{bb}^H = 1.06$  for each auxiliary orbital.



**Fig. 2a,b.** Computed geometries for propane using the HF/6-31G(d), semiempirical parameterized model 3 (PM3) (in parentheses), CHARMM (in square brackets), and generalized hybrid orbital (GHO)-PM3 (in angled brackets) methods. Only the relevant PM3 or CHARMM values are listed for comparison with the GHO-PM3 results. The fragment to the left-hand side of the boundary atom  $\text{C}_B$  is the QM fragment. This convention is used in Fig. 3

In hydrocarbons, there is a small bond polarization between a C–H bond, giving rise to the “standard” partial atomic charge of  $+0.09e$  on an aliphatic hydrogen atom in the CHARMM force field [32]. The carbon atom, then, carries a fractional charge to balance the charge neutrality of the alkyl group. For example, a methyl group for an aliphatic sidechain in the CHARMM force field has partial charges of  $+0.09e$  on hydrogen atoms and  $-0.27e$  on the carbon atom. This convention was made to be consistent with results from QM calculations, indicating that there is little charge transfer between alkyl groups in hydrocarbon compounds [32]. Since the electronegativity of the carbon atom is defined by the semiempirical PM3 model for the QM fragment, a methyl group, in propane, the GHO boundary atom for carbon should possess the same electron-withdrawing power [14]. Therefore, there shall be no (or little) charge transfer between the QM fragment and the boundary atom in hydrocarbon model compounds, and the most convenient way of testing this is to compare the Mulliken population charges [33] from the QM/MM calculation against the full PM3 results. For propane, the population charge density on the boundary atom does not come from the polarization of

the QM methyl group, but it results from the “excess” of bond density in the auxiliary orbitals. Of course, this “excess” auxiliary density is balanced by the MM partial charges from the two methylene hydrogen atoms. To achieve this charge balance, the PM3 parameters  $U_{ss}$  and  $U_{pp}$  for carbon were adjusted for the GHO boundary atom by matching the Mulliken population charge for the boundary atom  $C_B$  and the MM partial charge ( $-0.18$  au), which is equivalent to having a total charge of zero for the QM methyl group. Satisfactory results were obtained when  $U_{pp}$  was scaled by a factor of 0.973, and  $U_{ss}$  remains the same as in PM3 [14]. The computed partial charges on the atoms of the  $C_QH_3-C_B$  fragment from Mulliken population analysis are compared in Table 1, when only this fragment is treated quantum mechanically using the GHO-PM3 method, and when the entire propane is represented by the PM3 model.

The  $\beta_s$  and  $\beta_p$  parameters, which mainly determine chemical bonding features, were adjusted such that the optimized  $C_Q-C_B$  bond distance and the angles  $H-C_Q-C_B$  and  $C_Q-C_B-H$  were in accord with the PM3 value when the entire molecule is treated quantum mechanically. We found that for a carbon boundary atom the standard PM3 parameters  $\beta_s$  and  $\beta_p$  need to be scaled by factors of 0.2 and 1.5, respectively [14]. The computed geometrical variables obtained using the GHO method with the optimized parameters and using the PM3 model are listed in Fig. 2.

In addition, as explained in Ref. [1], minor modifications of the force field parameters associated with the boundary atom can result in better agreement in the computed bond distance with the MM results. Without these modifications, these bond distances are typically overestimated by 0.02–0.03 Å with respect to the original CHARMM value, and these small changes are generally insignificant compared to dynamic fluctuations of these variables. We found that the MM parameters developed for the AM1 boundary atom to correct this effect are also suitable for the PM3 boundary atom. Specifically, the bond stretch parameters,  $R_0(C-C)$  and  $R_0(C-H)$ , are reduced by values of 0.05 and 0.02 Å, respectively. The set of optimized parameters for the carbon boundary atom in a hybrid potential using the PM3 semiempirical quantum level and the CHARMM force field is shown in Table 2. These are the parameters

**Table 1.** Mulliken atomic charges (atomic units) determined at the semiempirical parameterized model 3 (PM3) and generalized hybrid orbital (GHO)-PM3 levels for the  $C_QH_3-C_B$  fragment of propane

Atoms	PM3	GHO-PM3/CHARMM
$C_Q$	-0.12	-0.12
$H_1$	0.04	0.04
$H_2$	0.04	0.04
$H_3$	0.04	0.04
$C_B$	-0.18 <sup>a</sup>	-0.18

<sup>a</sup>The target value corresponds to the molecular mechanical charge assigned to the boundary atom (see the text)

**Table 2.** Modified parameters for the carbon boundary atom in the combined GHO-PM3/CHARMM potential. All other quantum-mechanical and molecular mechanical parameters for carbon were taken directly from the standard PM3 or CHARMM parameters sets. The subscript  $B$  specifies that the carbon atom is a boundary atom.  $C$ ,  $CT2$ ,  $NH1$ ,  $HA$  and  $HB$  are standard atom types in the CHARMM force field, whereas  $C_{B2}$  and  $C_{B1}$  correspond to boundary atoms equivalent to the CHARMM CT2 and CT1 type, and  $C_B$  indicates a boundary atom equivalent to any CHARMM aliphatic carbon type

Item	PM3 or CHARMM	GHO-PM3
$\beta_s$	-11.9100150	-2.3820030
$\beta_p$	-9.8027550	-14.7041325
$U_{pp}$	-36.2669180	-35.2877112
Molecular mechanical bond stretching parameters (Å)		
$R_0(CT2-C_{B2})$	1.530	1.485
$R_0(CT3-C_{B2})$	1.528	1.478
$R_0(C-C_{B1})$	1.490	1.470
$R_0(NH1-C_{B1})$	1.430	1.400
$R_0(C_B-HA)$	1.111	1.091
$R_0(C_{B1}-HB)$	1.080	1.050

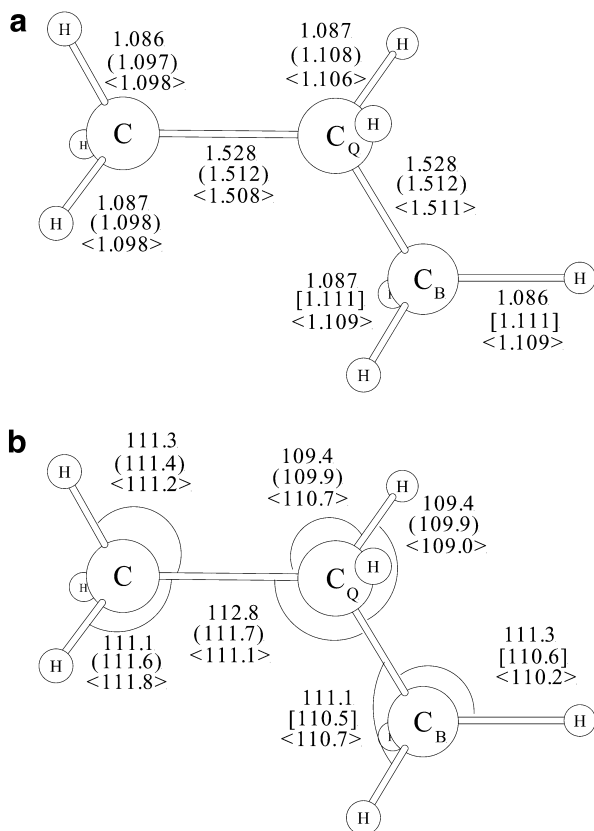
used in this paper for all the calculation at the GHO-PM3 level.

### Validation

A series of test calculations were carried out to validate the GHO-PM3 model.

### Geometry

For propane, the GHO atom can also be placed at the  $C3$  position, which gives a good assessment of the QM-QM-B type of bond angle. This QM/MM partition extends the investigation that was used to parameterize the GHO-PM3 model when B is located at the  $C2$  position, corresponding to the QM-B-MM situation. The geometries determined using the combined GHO-PM3/CHARMM method are shown in Figs. 2 and 3, along with those calculated at the HF/6-31G(d) [34], PM3 [14], and MM/CHARMM levels. Of particular interest is the comparison of geometries obtained from combined QM/MM calculation with geometries optimized by the full PM3 or the full CHARMM force field in the corresponding regions. In the GHO model, the QM/MM bond,  $C_Q-C_B$ , is fully represented quantum mechanically. Figures 2 and 3 show that the GHO-PM3 values for this bond differ only by 0.005 Å from the ones obtained by treating the whole molecule at the PM3 level. For propane, the overall unsigned errors from the combined GHO-PM3/CHARMM method are about 0.002 Å for the bond length and 0.5° for bond angles. In general, the PM3 level overestimates the C-H bond length and underestimates the C-C bond length in comparison with the HF/6-31G(d) results. Since the GHO-PM3 method was parameterized to reproduce the QM model—the PM3 results—the intrinsic PM3 error is the main cause of



**Fig. 3a,b.** Computed geometries for propane with the boundary atom at the C3 position. See also the caption of Fig. 2

the difference between the ab initio HF results and the QM(GHO-PM3)/MM results.

The two normal mode vibrational frequencies corresponding to the two C–C stretching modes in propane are 1,427 and 1,425  $\text{cm}^{-1}$  from the PM3 model, and 1,411 and 1,411  $\text{cm}^{-1}$  from the CHARMM force field. For comparison, the two frequencies are 1,491 and 1,418  $\text{cm}^{-1}$  in the GHO-PM3/CHARMM model in which the boundary atom is the  $C_2$  atom. The first frequency corresponds to the QM-B bond type, which is entirely determined by the GHO-PM3 QM calculation, and it is overestimated by about 60  $\text{cm}^{-1}$  compared with the PM3 result. This is consistent with the geometrical results in Fig. 2, which shows that the  $C_Q$ – $C_B$  bond distance is shorter by 0.02 Å than the PM3 value. When the CHARMM bond parameters were modified to improve the agreement with CHARMM for the B-MM bond type, the second C–C stretching frequency increased to 1,446  $\text{cm}^{-1}$  without affecting the QM-B stretch. These results suggest that the GHO-PM3/CHARMM(mod) model yields results consistent with pure QM and MM data on vibrational frequencies, although quantitative agreement may still be improved.

To illustrate the transferability of the  $\beta_s$  and  $\beta_p$  parameters for the boundary atom, the  $C_Q$ – $C_B$  bond lengths for a series of organic compounds are listed in Table 3. The trend is adequately reproduced in most

**Table 3.** Comparison of optimized bond distances between the quantum mechanical and boundary atoms for a series of organic molecules using the PM3 and combined GHO-PM3/CHARMM potentials (angstroms). The subscript *B* indicates that the methyl carbon is the boundary atom. The functional group on the right-hand side of the specified bond is the quantum mechanical fragment in each compound

Molecule	PM3	GHO-PM3/CHARMM
$\text{H}_3\text{C}_B$ – $\text{CH}_2\text{OH}$	1.520	1.519
$\text{H}_3\text{C}_B$ – $\text{CH}_2\text{NH}_2$	1.516	1.515
$\text{H}_3\text{C}_B$ – $\text{CH}_2\text{COOH}$	1.513	1.513
$\text{H}_3\text{C}_B$ – $\text{CH}_2\text{COCH}_3$	1.513	1.511
$\text{H}_3\text{C}_B$ – $\text{CH}_2\text{SH}$	1.508	1.508
$\text{H}_3\text{C}_B$ – $\text{COOH}$	1.497	1.512
$\text{H}_3\text{C}_B$ – $\text{C}_6\text{H}_5$	1.486	1.498

**Table 4.** Mulliken partial charges determined from PM3 and GHO-PM3 calculations along with those used in the CHARMM forcefield for acetic acid (atomic units)

Atom	CHARMM	PM3	GHO-PM3/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
$\text{H}_3\text{C}$ (boundary)	–0.03	0.133	0.068

cases and the agreement with the values obtained at the PM3 level is also good.

### Electronic structure

In a polar compound there will be some amount of charge transfer through the boundary of the QM and MM region, which will result in the polarization of the active hybrid orbital on the boundary atom. As a consequence, the Mulliken charge on this atom will not be equal to the assigned MM partial charge. This capability of the GHO model for acetic acid is demonstrated in Table 4, when the carboxyl group is treated as the QM fragment and the methyl carbon as the boundary atom. It is clear from Table 4 that the charge distribution obtained at the GHO-PM3 level for the QM fragment is comparable to that obtained when the entire molecule is treated quantum mechanically.

### Proton affinity

The calculation of proton affinities of organic bases was used as a more stringent test of the accuracy of the QM(GHO)/MM method using the PM3 semiempirical method as the quantum level and the CHARMM force field to represent the MM region. We report the computed proton affinities for some selected compounds, along with the experimental data, in Table 5. The proton affinity of a base, *X*, is defined as the negative value of the enthalpy of reaction at 25 °C for the protonation reaction:

**Table 5.** Computed and experimental proton affinities for model compounds. The subscript *B* indicates the boundary atom for the GHO calculations and atoms on the right of  $C_B$  are treated quantum mechanically. All energies are in kilocalories per mole

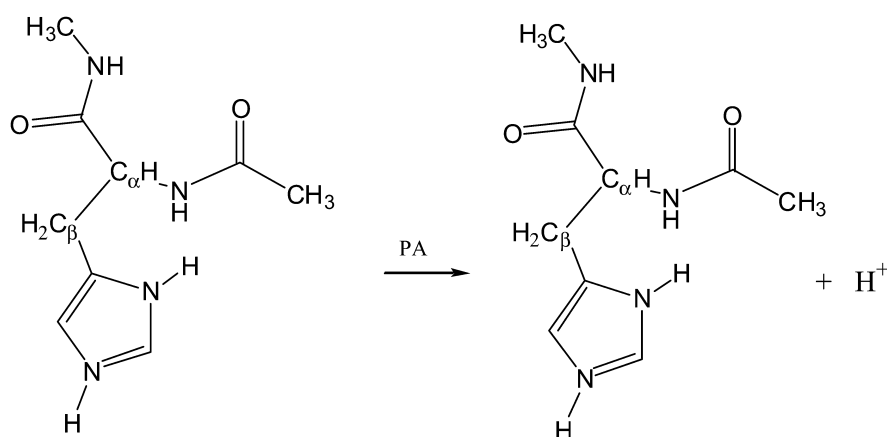
Molecule	Exp. [2]	PM3	GHO-PM3/CHARMM
$H_3C_B-COO^-$	348.5	348.1	344.7
$H_3C-C_BH_2-CH_2-COO^-$	346.6	347.1	346.8
$H_3C_B-CH(CH_3)-COO^-$	346.3	346.2	345.9
$H_3C_B-CH_2O^-$	376.1	379.7	378.5
$H_3C_B-CH(CH_3)-O^-$	374.1	377.8	376.9
$H_3C_B-NH^-$	403.2	392.6	398.5
$H_3C_B-CH_2NH^-$	399.4	390.3	390.8
$H_3C_B-CH_2NH_2$	217.0	210.3	206.9

**Table 6.** Computed proton affinities for a blocked histidine residue in the gas phase (kilocalories per mole). *B* indicates the location of the GHO boundary atom. *MM* denotes that standard CHARMM force field is used and *MM(opt)* indicates that the bond parameters associated with the boundary atom have been modified as in Table 2

Model	GHO-PM3/MM	GHO-PM3/MM(opt)	PM3
$B=C_\alpha$	223.0	223.8	224.6
$B=C_\beta$	218.9	218.9	224.6

**Table 7.** Computed bond distances associated with the boundary atom for a blocked histidine residue in the gas phase (angstroms)

Model	GHO-PM3/MM		GHO-PM3/MM(opt)		CHARMM	
	HisH <sup>+</sup>	His	HisH <sup>+</sup>	His	HisH <sup>+</sup>	His
$B=C_\alpha$						
$C_B1-C$	1.534	1.518	1.536	1.518	1.516	1.516
$C_B1-NH1$	1.488	1.465	1.482	1.457	1.462	1.447
$C_B1-HB$	1.106	1.079	1.106	1.079	1.082	1.082
$B=C_\beta$						
$C_B2-CT1$	1.566	1.558	1.564	1.556	1.563	1.556
$C_B2-HA$	1.113	1.112	1.131	1.130	1.112	1.113
$C_B2-HA$	1.113	1.111	1.131	1.128	1.112	1.109



**Scheme 1.**

$$\text{Proton affinity}(X) = \Delta H_f^0(H^+) + \Delta H_f^0(X) - \Delta H_f^0(HX^+), \quad (4)$$

where  $\Delta H_f^0(J)$  are standard state heats of formation for the species in parentheses. We calculated all the quantities in Eq. (4) using the PM3 and the GHO-PM3 models, except for the heat of formation of the proton,  $\Delta H_f^0(H^+)$ , which is poorly estimated by semiempirical methods. Instead, we used the experimental value of  $365.7 \text{ kcal mol}^{-1}$ , which was also used in the original development of the AM1 and PM3 models [3, 14]. The errors of both calculations when compared to the experimental data are of similar magnitude, suggesting that the introduction of a GHO boundary atom in combined PM3/MM calculations has not diminished the quality of the original QM model. The difference between them is even smaller when the GHO atom is placed farther away from the proton acceptor atom. It is also apparent that the boundary atom should be placed at least one carbon unit away from the functional group involved in a chemical process. In general, the results in Table 5 indicate that the GHO-PM3 model does not introduce unrealistic perturbations into the electronic structure of the system.

### Histidine residue

To extend the scope of the test cases, we examined the proton affinity for a histidine residue that is blocked by amide bonds both at the C and N terminal end. In this test case, we carried out several calculations. The proton affinity for this system was computed using the PM3 model for the entire system, and the GHO-PM3/CHARMM force field with the boundary atom placed at the  $C_\alpha$  and  $C_\beta$  carbon positions, respectively. In addition, we have presented the results of QM/MM calculations that include or exclude the small modifications of the MM parameters to show the effects of these small structural variations. The computational results for histidine are summarized in Tables 6 and 7. Consistent with the results listed in Table 5, when the GHO boundary atom is placed at the  $C_\alpha$  position, one meth-

ylene group away from the conjugated reaction center, the computed proton affinity is in good agreement with the PM3 model. The energy difference is 1.6 kcal mol<sup>-1</sup>, or 0.7% of the total PA value, and the agreement is further improved to a difference of only 0.8 kcal mol<sup>-1</sup> if the parameters of the MM bonds connected to the boundary atom are adjusted. The latter makes the B-MM bond types shorter by 0.02 Å, to be in close accord with the results from the CHARMM force field.

## Conclusions

In this paper, we reported the results of a set of optimized semiempirical parameters for a GHO carbon atom to be used in combined PM3/CHARMM calculations. The model was further tested on geometrical, electronic and chemical properties for a set of organic compounds. The results computed using the combined GHO-PM3/CHARMM model were compared with those obtained when the whole system is treated as a QM or a MM system. The results obtained for the test calculations demonstrate that the GHO-PM3 model is reasonable to represent the boundary between the QM and the MM parts of a hybrid system.

*Acknowledgement* We thank the National Institutes of Health for support of this research.

## References

- Gao J, Amara P, Alhambra C, Field MJ (1998) *J Phys Chem A* 102:4714
- Amara P, Field MJ., Alhambra C, Gao J (2000) *Theor Chem Acc* 104:336
- Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107:3902
- Thery V, Rinaldi D, Rivail JL, Maigret B, Ferenczy GGJ (1994) *Comput Chem* 15:269
- Assfeld X, Rivail J.-L (1996) *Chem Phys Lett* 263:100
- Monard G, Loos M, Thery V, Baka K, Rivail J.-L (1996) *Int J Quantum Chem* 58:153
- Assfeld X, Ferre N, Rivail J.-L (1998) *ACS Symp Ser* 712:234
- Ferre N, Assfeld X, Rivail J.-L (2002) *J Comput Chem* 23:610
- Alhambra C, Corchado J, Sanchez ML, Gao J, Truhlar DG (2000) *J Am Chem Soc* 122:8197
- Garcia-Viloca M, Truhlar DG, Gao J (2003) *J Mol Biol* 327:549
- Alhambra C, Corchado J, Sanchez ML, Garcia-Viloca M, Gao J, Truhlar DG (2001) *J Phys Chem B* 105:11326
- Alhambra C, Sanchez ML, Corchado JC, Gao J, Truhlar DG (2002) *Chem Phys Lett* 355:388
- Devi-Kesavan LS, Gao J (2003) *J Am Chem Soc* 125:1532
- Stewart JJP (1989) *J Comput Chem* 10:209
- Singh UC, Kollman PA (1986) *J Comput Chem* 7:718
- Field MJ, Bash PA, Karplus M (1990) *J Comput Chem* 11:700
- Warshel A, Levitt M (1976) *J Mol Biol* 103:227
- Antes I, Thiel W (1999) *J Phys Chem* 103:9290
- Reuter N, Dejaegere A, Maigret B, Karplus M (2000) *J Phys Chem A* 104:1720
- Zhang Y, Lee T-S, Yang W (1999) *J Chem Phys* 110:46
- Swart M (2003) *Int J Quantum Chem* 91:177
- Amara P, Field MJ (2003) *Theor Chem Acc* 109:43
- Das D, Eurenus KP, Billings EM, Sherwood P, Chatfield DC, Hodosek M, Brooks BR (2002) *J Chem Phys* 117:10534
- Murphy RB, Philipp DM, Friesner RA (2000) *Chem Phys Lett* 321:113
- Philipp DM, Friesner RA (1999) *J Comput Chem* 20:1468
- Bakowies D, Thiel W (1996) *J Phys Chem* 100:10580
- Maseras F, Morokuma K (1995) *J Comput Chem* 16:1170
- Bersuker IB, Leong MK, Boggs JE, Pearlman RS (1998) *ACS Symp Ser* 712:66
- Antes I, Thiel W (1998) *ACS Symp Ser* 712:50
- DiLabio GA, Hurley MM, Christiansen PA (2002) *J Chem Phys* 116:9578
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) *J Comput Chem* 4:187
- MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) *J Phys Chem B* 102:3586
- Mulliken RS (1964) *J Chem Phys* 61:20
- Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) *Ab initio molecular orbital theory*. Wiley, New York