Applications and assessment of QM:QM electronic embedding using generalized asymmetric Mulliken atomic charges

Priya V. Parandekar, Hrant P. Hratchian, and Krishnan Raghavachari[a)]
Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA

Hybrid QM:QM (quantum mechanics:quantum mechanics) and QM:MM (quantum mechanics: molecular mechanics) methods are widely used to calculate the electronic structure of large systems where a full quantum mechanical treatment at a desired high level of theory is computationally prohibitive. The ONIOM (our own *N*-layer integrated molecular orbital molecular mechanics) approximation is one of the more popular hybrid methods, where the total molecular system is divided into multiple layers, each treated at a different level of theory. In a previous publication, we developed a novel QM:QM electronic embedding scheme within the ONIOM framework, where the model system is embedded in the external Mulliken point charges of the surrounding low-level region to account for the polarization of the model system wave function. Therein, we derived and implemented a rigorous expression for the embedding energy as well as analytic gradients that depend on the derivatives of the external Mulliken point charges. In this work, we demonstrate the applicability of our QM:QM method with point charge embedding and assess its accuracy. We study two challenging systems—zinc metalloenzymes and silicon oxide cages—and demonstrate that electronic embedding shows significant improvement over mechanical embedding. We also develop a modified technique for the energy and analytic gradients using a generalized asymmetric Mulliken embedding method involving an unequal splitting of the Mulliken overlap populations to offer improvement in situations where the Mulliken charges may be deficient. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2976570]

I. INTRODUCTION

The electronic structure calculation of large molecular systems is computationally prohibitive due to the rapid scaling of *ab initio* molecular orbital techniques with size.[1] Hybrid energy methods,[2] in which different parts of the system are treated with different model chemistries, are frequently employed to treat such large molecules. Typically, a desired high level of theory is used to describe the region of interest where the chemical phenomenon is occurring, while the rest of the system is treated with a low level of theory to save computational costs. Hybrid methods are broadly classified into summation (or connection)[3] and extrapolation schemes.[4] The ONIOM (our own *N*-layer integrated molecular orbital molecular mechanics) hybrid method of Morokuma and coworkers is an extrapolation scheme,[4,5] which involves separate subcalculations, each on a well defined molecular system. In this method, if the regional boundary cuts through covalent bonds, hydrogen link atoms are used to cap the unsaturated dangling bonds. For the case where two regions are used—region I being the region of interest (also called as the model system) and region II, its surroundings (the rest of the molecule)—the general form for the ONIOM energy can be expressed in terms of real and model system energies,

$$E_{\text{ONIOM}} = E_{\text{RL}}(\text{I} + \text{II}) - E_{\text{ML}}(\text{I}) + E_{\text{MH}}(\text{I}), \quad (1)$$

where the subscripts RL, ML, and MH denote the real

system low-level, model system low-level, and model system high-level subcalculations. The ONIOM method can combine any two quantum mechanics (QM) levels as well as the standard QM and molecular mechanics (MM). When the interaction between the two regions is treated entirely by the low level of theory to derive the ONIOM extrapolated energy expression in Eq. (1), such models are known as *mechanical embedding* (ME).[6] In ONIOM-ME calculations, there is no direct coupling between the wave function of region I and the surrounding region II in the model system subcalculations. Electronic effects such as polarization and charge transfer between the regions are included only at the low level of theory, when we perform the real system subcalculation at the low level.

An improvement over the ONIOM-ME calculations is to include the region II charge distribution in the model system evaluations, which is referred to as *electronic embedding* (EE).[6] The ONIOM-EE approach allows the region II charge distribution to polarize the model system wave function and includes the electrostatic interaction between the regions by integrating over the QM charge density. Certain QM:MM-EE models embed the QM wave function of the model system in the distribution of fixed atomic charges employed in the electrostatic contribution to the MM potential, although polarizable force field methods have also been incorporated within a QM:MM-EE scheme.[7] For example, Gao has developed a molecular orbital–derived empirical potential for liquid (MODEL) simulations,[8] which is derived on the basis of

---

a)Electronic mail: kraghava@indiana.edu.

quantum chemical representation of individual molecules in a liquid, where the interactions of each solvent molecule with the surrounding molecules are determined by the QM:MM approach. Scaled Mulliken population charges, obtained by scaling the Mulliken charges by a universal constant associated with a particular quantum chemical model, were used in the MODEL potential to describe interactions in bimolecular complexes. In later work, Xie *et al.* have proposed an electronic structure based polarization method, called the X-POL potential[9] to treat many body polarization and charge delocalization effects in polypeptides. The bonded interactions are described by QM, and the nonbonded interactions are modeled by an iterative QM:MM method, in which the MM partial charges are derived from the molecular wavefunctions of the individual fragments. While such approaches have proven to be useful, when electronic coupling between the regions becomes important, MM itself may not be adequate as the low-level method even in the QM:MM-EE approach, or the needed MM parameters may not be available. In such cases, QM:QM approaches are better, and hence we have developed a QM:QM EE method within the ONIOM framework using Mulliken point charges in a recent publication, referred to as Paper 1.[10] Therein, we presented the development and implementation of an analytic gradient method where the energy derivative expressions give rise to a set of equations similar to the coupled-perturbed Hartree–Fock (CPHF) equations that appear in post-self-consistent-field (SCF) gradient theories.[11] As in the case of those forces, application of the *z*-vector method of Handy and Schaefer[12] reduced the CPHF work to a single set of SCF response equations.

In this work, we demonstrate the applicability and perform an initial assessment of the QM:QM Mulliken point charge embedding method developed in Paper 1. We test and analyze some challenging model reactions involving zinc metalloenzymes and silicon oxide clusters using ONIOM QM:QM embedding (electronic as well as mechanical). We also present a new method with modified Mulliken charges, in which the Mulliken overlap populations are divided unequally between the two constituent atoms and incorporate these asymmetric Mulliken charges in our QM:QM point charge embedding scheme (energy and gradients).

II. METHODS

In Paper 1, we developed the gradient theory for an ONIOM QM:QM point charge embedding model using Mulliken atomic charges.[10] Noting a key deficiency of Mulliken population analysis—electron density shared by two atomic centers is equally apportioned regardless of electronegativity differences between the two centers—in this work we introduce an asymmetric Mulliken scheme. Using this proposed modified population analysis model, we derive and implement the related ONIOM QM:QM point charge embedding energy derivatives. As in the case of QM:QM embedding using conventional Mulliken atomic charges, these forces are implemented by framing the resulting gradient expression in a post-SCF form. This allows use of the Handy–Schaefer

z-vector treatment[12] and yields an efficient gradient code. In this section, we now modify the ONIOM QM:QM point charge embedding gradient theory introduced previously[10] for asymmetric Mulliken charge QM:QM embedding.

A. Generalized asymmetric Mulliken embedding (GAME): Energy and gradients

The Mulliken charge on an atom is defined as[13]

$$q_A = z_A - \sum_{\alpha \in A} \sum_{\beta} P_{\alpha\beta} S_{\alpha\beta}, \tag{2}$$

where $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{z}$ are the density matrix, overlap matrix, and nuclear charges. As noted earlier, electron overlap populations are split equally between the two constituent atoms in deriving atomic charges in this model. Because of this approximation, the calculated Mulliken atomic charges may not be a good representation of the partial atomic charges to be used in EE; hence modifications to conventional Mulliken analysis have been proposed.[14] This is particularly true in cases where the two atoms sharing electron density have different electronegativities. For example, if we consider a molecule containing a Si–O bond, it may not be a reasonable approximation to divide the overlap population equally between Si and O. It may be more appropriate to assign a greater proportion of the overlap population to oxygen due to its higher electronegativity. To incorporate more realistic embedding charges, we propose an asymmetric Mulliken population model where overlap populations are unequally divided. We use the general notation D_{AB} to denote the fraction of the overlap population between atomic centers A and B that have been assigned to atom A. Similarly, D_{BA} denotes the corresponding fraction assigned to atom B. In standard Mulliken population analysis, $D_{AB} = D_{BA} = 0.5$. If we assign the overlap population differently, $D_{AB} \neq D_{BA}$ though, by definition, $D_{AB} + D_{BA} = 1$. Thus, as an alternative to Eq. (2), we consider the charge of atomic center A as given by

$$q_A = z_A - \sum_{\alpha \in A} \sum_{\beta} 2D_{\alpha\beta} P_{\alpha\beta} S_{\alpha\beta}, \tag{3}$$

where the matrix $\mathbf{D}$ gives the overlap population fractions (weight factors) between atomic centers of basis functions α and β. For convenience, we have used an atomic basis function representation for $\mathbf{D}$, whereby $D_{\alpha\beta} = D_{AB}$ for $\alpha \in A$ and $\beta \in B$. In conventional Mulliken atomic charge analysis, since all elements of $\mathbf{D}$ are 0.5, Eq. (3) reduces to Eq. (2).

ONIOM QM:QM point charge embedding energies using asymmetric Mulliken charges can be computed by using the charges from the population analysis as defined in Eq. (3). Analytic forces using the asymmetric charge model can also be written in terms of a straightforward modification of our previous theory, provided that $\mathbf{D}$ *is constant with respect to nuclear coordinate displacements*. This requirement is reasonable and is employed throughout this work.

The ONIOM energy for a two-layer system is given in Eq. (1), where E_{RL} is the energy of the real system determined at the low level of theory and E_{ML} and E_{MH} are the model system energies determined at the low and high levels of theory, respectively. Following our previous analysis,[10] we

TABLE I. Modified Mulliken charge analysis for orthosilicic acid (SiO_4H_4) with 75% and 100% of the electron overlap populations assigned to oxygen in a O–Si bond. Also shown are the regular Mulliken analysis (50% of the electron populations assigned to each O and Si), NBO, ESP derived charges, and Bader charges. The calculations are at the B3LYP/6-31G level.

| Atom | Mulliken | Modified Mulliken | | NBO | ESP | AIM |
		O-75 Si-25	O-100 Si-0			
Si	1.47	1.97	2.46	2.29	1.93	3.03
O	−0.78	−0.91	−1.03	−1.1	−1.02	−1.31
H	0.41	0.41	0.41	0.52	0.53	0.55

now derive an expression for the ONIOM energy with EE using these modified Mulliken charges given in Eq. (3),

$$E_{ONIOM} = E_{RL} - \tilde{E}_{ML} + \tilde{E}_{MH} - \sum_A q_A \phi_{ML;A}$$
$$+ \sum_A q_A \phi_{MH;A}, \tag{4}$$

$\tilde{E}_{ML}$ and $\tilde{E}_{MH}$ in Eq. (4) are the energies of the model system, self-consistently obtained in the field of the external point charges. The embedding contributions, which are the interaction terms of the external point charges with the model system nuclei and electrons, are written out separately as the last two terms of Eq. (4). The electrostatic potentials (ESPs) ϕ in Eq. (4) at the position of atom A exerted by the model system (atoms M), computed from the model system wave function optimized in the presence of the external point charges, are given by

$$\phi_{ML;A} = \sum_M \frac{Z_M}{R_{MA}} - \sum_{\mu_L \nu_L} \langle \mu_L | r_{IA}^{-1} | \nu_L \rangle P_{\mu_L \nu_L}, \tag{5}$$

$$\phi_{MH;A} = \sum_M \frac{Z_M}{R_{MA}} - \sum_{\mu_H \nu_H} \langle \mu_H | r_{IA}^{-1} | \nu_H \rangle P_{\mu_H \nu_H}. \tag{6}$$

The expression for the QM:QM energy gradient using the modified Mulliken charges with EE becomes

$$E_{ONIOM}^x = E_{RL}^x + \sum_A \phi_A^\Delta \sum_{\alpha \in A} \sum_\beta 2D_{\alpha\beta}[P_{\alpha\beta}^x S_{\alpha\beta}$$
$$+ P_{\alpha\beta} S_{\alpha\beta}^x] - \tilde{E}_{ML}^x + \tilde{E}_{MH}^x - \sum_A q_A \phi_{ML;A}^x$$
$$+ \sum_A q_A \phi_{MH;A}^x, \tag{7}$$

where

$$\phi_A^\Delta = \phi_{ML;A} - \phi_{MH;A}. \tag{8}$$

As before, terms in the ONIOM QM:QM energy derivative involving embedding charge derivatives can be grouped together with conventional RL gradient terms to yield an effective RL gradient,

$$\tilde{E}_{RL}^x = E_{RL}^x + \sum_A \phi_A^\Delta \sum_{\alpha \in A} \sum_\beta 2D_{\alpha\beta}[P_{\alpha\beta}^x S_{\alpha\beta} + P_{\alpha\beta} S_{\alpha\beta}^x]$$
$$= E_{RL}^x + \tilde{E}_{emb}^x. \tag{9}$$

We recast the real system energy gradient expression in the generalized post-SCF gradient form[15] as

$$\tilde{E}_{RL}^x = \sum_{\alpha\beta\delta\gamma} {}^{emb}_{\alpha\beta\delta\gamma}(\alpha\beta|\delta\gamma)^x + \sum_{\alpha\beta} P_{\alpha\beta}^{emb} H_{\alpha\beta}^x + \sum_{\alpha\beta} W_{\alpha\beta}^{emb} S_{\alpha\beta}^x$$
$$+ V_{nuc}^x. \tag{10}$$

In Eq. (10), derivatives of the internuclear potential and the two-electron integrals are denoted by their typical notation. $\mathbf{P}^{emb}$ and $\mathbf{W}^{emb}$ are the effective embedding density and energy-weighted density matrices in terms of the basis functions and are given by sums of the corresponding Hartree–Fock matrices and the embedding corrections $\mathbf{P}^\Delta$ and $\mathbf{W}^\Delta$. $\mathbf{\Gamma}^{emb}$ is the embedding effective two-particle density matrix and is written as a sum of products of the Hartree–Fock density matrix and embedding corrections. The general post-SCF gradient form of Eq. (10) can still be employed, using new definitions for the Lagrangian and energy-weighted density matrix correction. For QM:QM point charge embedding using the asymmetric Mulliken scheme we now have the Lagrangian as

$$L_{ai} = \sum_A \sum_{\alpha \in A} \sum_\beta 2\phi_A^\Delta S_{\alpha\beta} D_{\alpha\beta}(C_{\alpha i} C_{\beta a} + C_{\alpha a} C_{\beta i}). \tag{11}$$

The energy-weighted density matrix correction is given by

$$W_{\alpha\beta}^\Delta = (\phi_\alpha^\Delta D_{\alpha\beta} + \phi_\beta^\Delta D_{\beta\alpha}) P_{\alpha\beta}^{HF} + \sum_{pq} C_{\alpha p} C_{\beta q} \bar{W}_{pq}^\Delta, \tag{12}$$

$$\bar{W}_{ij}^\Delta = -\sum_{\alpha\beta} (\phi_\alpha^\Delta D_{\alpha\beta} + \phi_\beta^\Delta D_{\beta\alpha}) S_{\alpha\beta} C_{\alpha i} C_{\beta j}$$
$$- \sum_{ak} P_{ak}^\Delta (aj\ ki), \tag{13}$$

$$\bar{W}_{ab}^\Delta = 0, \tag{14}$$

$$\bar{W}_{ai}^\Delta = -P_{ai\ i}^\Delta. \tag{15}$$

Again, ϕ_α^Δ is equal to ϕ_A^Δ where the basis function α is centered on nucleus A. Note that when all elements of $\mathbf{D}$ are equal to 0.5, the equations for asymmetric Mulliken embedding developed here become identical to those for Mulliken point charge embedding in our previous work.[10]

TABLE II. Modified Mulliken charge analysis for the siloxide anion (H_3SiO^-) with 75% and 100% of the electron overlap populations assigned to oxygen in a O–Si bond. Also shown are the regular Mulliken analysis (50% of the electron populations assigned to each O and Si), NBO, ESP derived charges, and Bader charges. The calculations are at the B3LYP/6-31G level.

| Atom | Mulliken | Modified Mulliken | | NBO | ESP | AIM |
		O-75 Si-25	O-100 Si-0			
Si	0.61	0.84	1.06	1.03	0.97	2.37
O	−0.84	−1.06	−1.29	−1.14	−0.96	−1.43
H	−0.26	−0.26	−0.26	−0.3	−0.34	−0.65

The unequal splitting of the Mulliken overlap populations is illustrated in the modified Mulliken charge analysis of orthosilicic acid, SiO_4H_4, and the siloxide anion, H_3SiO^-, as shown in Tables I and II. We have obtained different charges by using values of 0.5 (default Mulliken analysis), 0.75, and 1.0, for D_{AB} with A as the oxygen atom and B as the silicon atom. Weight factors of 1.0 for oxygen and 0.0 for silicon imply that the electron overlap population is assigned to the oxygen atom completely in a Si–O atom pair. Here we assign more than 50% of the overlap population to oxygen to account for the greater electronegativity of oxygen than silicon in a Si–O bond. To demonstrate the modified Mulliken point charges as a function of the weight factors between one specific pair of atoms (i.e., Si–O, in this case), the O–H and Si–H overlap populations are assumed to be divided equally. The charge analyses are done at the B3LYP/6-31G optimized geometries. The charge on the Si atom in orthosilicic acid varies between 1.47 and 2.46 as we change the weight factor on O from 0.5 to 1.0 in dividing the Si–O overlap population. We include other popular charge analysis schemes such as natural bond orbital (NBO),[16] ESP derived charges according to the Merz–Kollman–Singh scheme,[17] and atoms in molecules (AIMs or Bader)[18] charges for comparison. Note that the Bader charges result in a much more ionic Si–O bond as compared to other charge analyses. In Table II, when the weight factor on O is 1.0, and on Si is 0.0, the modified Mulliken point charges on Si and O are more ionic than other charge analysis schemes except Bader charges, as expected. Similarly, in Table I, the charge on the Si atom is more ionic than other charge analysis schemes except Bader charges for the 100-0 splitting. The charge on the oxygen atom is not as ionic as compared to NBO and ESP because the overlap population between O and H is split equally, regardless of the fact that the regular Mulliken point charges give a less ionic O–H bond. We note that the general use of the asymmetric Mulliken charges requires the weight factors between all pairs of atoms. However, in practice, we plan to use a default value of 0.5 (as in normal Mulliken analysis) except between pairs of atoms with significantly different electronegativities (as in Si–O above). While it is possible to assign the weight factors based on an electronegativity scale, considering that the Mulliken charges are method and basis set dependent, the weight factors can also be decided in an empirical manner. For instance, we can modify the Mulliken charges using this method by a proper adjustment of the weight factors to concur with another desired set of charges such as ESP derived charges, which are popularly used in

MM. This would be important in QM:QM point charge embedding schemes using modified Mulliken charges made to simulate ESP charges, since analytic gradients of ESP charges are not available. It is also possible to derive weight factors that yield satisfactory performance for a representative molecule that can then be used for applications involving larger but similar systems. In the following section, we will illustrate the application of the asymmetric Mulliken point charge embedding scheme.

III. ASSESSMENT OF THE ONIOM-QM:QM MULLIKEN POINT CHARGE EMBEDDING SCHEME

In this section, we illustrate some test cases to evaluate the performance of the QM:QM Mulliken point charge embedding method developed in Paper 1. In addition, we also apply the asymmetric Mulliken point charge embedding method discussed above. We examine two situations: one where the QM:QM regional boundary cuts across covalent bonds and hydrogen link atoms are used to cap the model system at the boundary, and another where no covalent bonds are cut by the region boundary. We demonstrate a case where mechanical embedding is not adequate to describe the reaction and completely breaks down, and EE is necessary for proper physical behavior of the system.

The first example is the proton transfer between the two water molecules in a water dimer, catalyzed by zinc (II) ion in a $[ZnIm_3(H_2O)_2]^{2+}$ complex, where "Im" denotes the imidazole ligand. This is a simple model of metal catalyzed reactions common in metalloenzymes, recently studied by Hong et al.[19] using constrained and frozen density functional theory (DFT) methods, and is a challenging system to be treated by layered ab initio calculations. In reality, the complex should be surrounded by solvent molecules to mimic the enzymatic reaction; however, we consider the gas phase complex for simplification of the model. We compute the potential energy surface (PES) for the proton transfer reaction between the two water molecules for the molecular system as shown in Fig. 1(a). In our ONIOM-ME and ONIOM-EE calculations, we include the Zn (II) ion and the two waters in the high-level region, and the imidazole ligands that serve as spectators in the reaction, in the low-level region. In the ONIOM-EE calculations, all Mulliken atomic charges from the low-level region are included in the model system Hamiltonian without scaling. We compare the PES obtained from ONIOM-ME and ONIOM-EE methods to that obtained by treating the entire system at the target
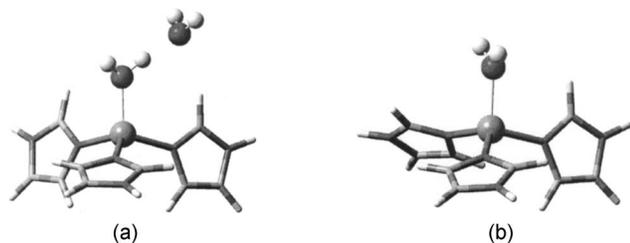
FIG. 1. Models used for the ONIOM QM:QM EE method—the high-level region is denoted as ball and stick representation and low-level region as tube rendering. (a) Proton transfer in a $[ZnIm_3(H_2O)_2]^{2+}$ complex. (b) Removal of water from a $[ZnIm_3(H_2O)]^{2+}$ complex.

high level of theory to assess the performance of the hybrid methods. The model chemistries considered are MP2/6-31G*:HF/6-31G, B3LYP/6-31+G**:HF/6-31G, and B3LYP/6-311+G**:BLYP/3-21G. The PESs are obtained by scanning the proton coordinate while keeping the geometry of the complex fixed at the high-level optimized geometry. The PESs are obtained at the high-level target, ONIOM-EE, ONIOM-ME, and the low level of theory for comparison, and are shown in Fig. 2. Clearly, the ONIOM-EE PES is in excellent agreement with the target high-level PES for all three model chemistries considered and is most accurate for the MP2:HF combination. The mean absolute deviations of the potential energy from the high-level target for the MP2/6-31G*:HF/6-31G combination are 2.8 kcal/mol with ME and 0.1 kcal/mol with EE, for the B3LYP/6-31+G**:HF/6-31G method are 3.4 kcal/mol with ME and 1.9 kcal/mol with EE, and for the B3LYP/6-311+G**:BLYP/3-21G combination are 5.6 kcal/mol with ME and 2.0 kcal/mol with EE. The ONIOM-EE method thus shows a better performance than the ONIOM-ME method. This shows that although the surrounding imidazole ligands do not participate directly in the proton transfer reaction, the electronic effects exerted by the ligands on the model system are very important. The Mulliken point charge approximation for the ONIOM QM:QM EE is very good in this case.

Next, we consider the PES for the removal of water from a $[ZnIm_3H_2O]^{2+}$ complex shown in Fig. 1(b). In this reaction the bond between metal and ligand is breaking, and therefore it is a more stringent test than the case when the ligands serve as spectators to the reaction with the metal-ligand bonds intact.[19] For the ONIOM calculations, we included the Zn (II) ion and H_2O in the high-level region and the three imidazole ligands in the low-level region. The PESs shown in Fig. 3 are computed at the ONIOM B3LYP/6-31+G**:HF/6-31G level for mechanical embedding and point charge embedding and compared to the high-level B3LYP/6-31+G** target and low level HF/6-31G. It is clear that EE agrees well with the high-level target, with the energy increasing monotonically with the Zn–O distance, approaching an asymptote, but mechanical embedding has a catastrophic failure at large Zn–O distances. An analysis of the component energies shows that this happens because, during the model system subcalculations, the bare system $Zn(H_2O)^{2+}$, in the absence of stabilizing ligands, dissociates incorrectly into partially charged species (a known failure of
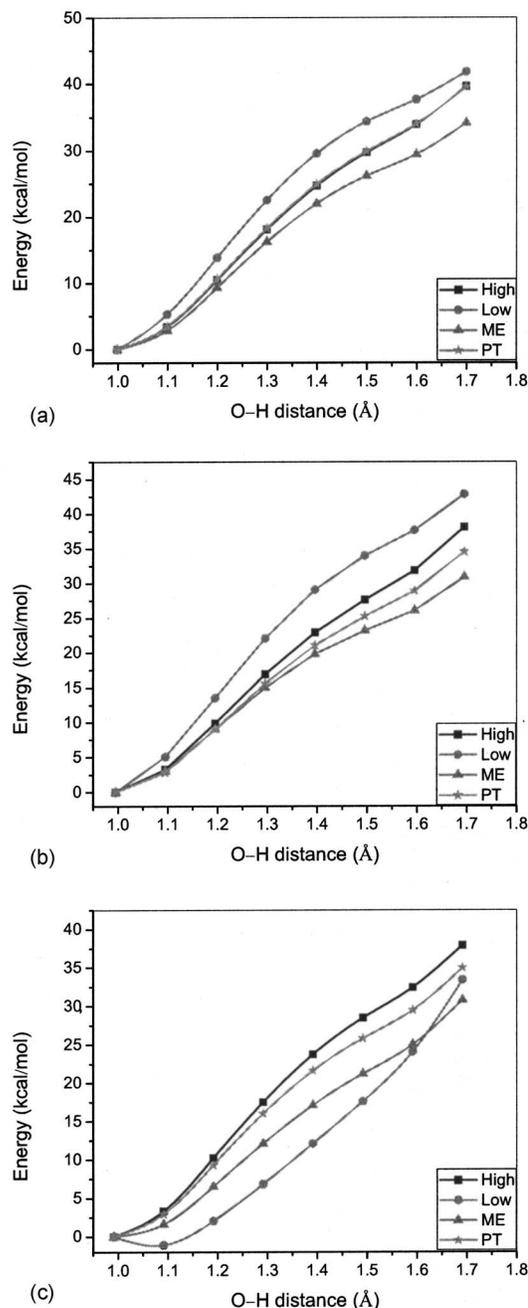
FIG. 2. PESs for the proton transfer between two water molecules in a $[ZnIm_3(H_2O)_2]^{2+}$ complex at different model chemistries; high-level target, low-level, ONIOM mechanical embedding (ME) and ONIOM point charge embedding (PT). (a) MP2/6-31G*:HF/6-31G, (b) B3LYP/6-31+G**:HF/6-31G, and (c) B3LYP/6-311+G**:BLYP/3-21G. In (a), the PES calculated by ONIOM point charge embedding almost coincides with the high-level target PES.

DFT[20] compounded by the fact that dissociation into Zn^{2+} and H_2O is significantly less favorable than dissociation to Zn^+ and H_2O^+). Only after the electronic effects of the ligands are included is the correct physical behavior of the PES observed with increasing Zn–O distance.

In the $[ZnIm_3H_2O]^{2+}$ complex, there is a strong polarization of the surrounding imidazole ligands due to the Zn(II) ion, so that some of its positive charge is transferred to the ligands. To take this into account we have used the ONIOM-
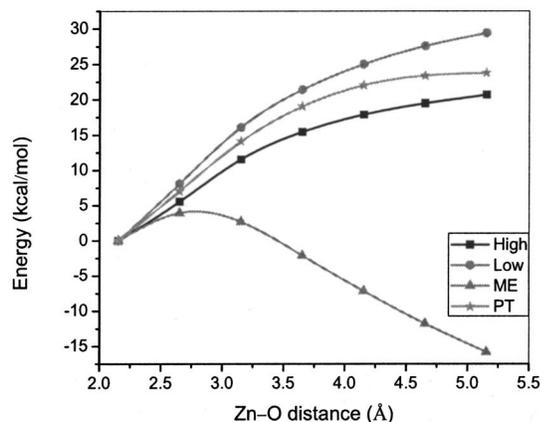
FIG. 3. PESs for the removal of water molecule from a $[ZnIm_3H_2O]^{2+}$ complex at the B3LYP/6-31+G**:HF/6-31G model chemistry. Note the catastrophic failure of the ONIOM-ME method with increasing Zn–O distance.

GAME method (ONIOM with generalized asymmetric Mulliken point charge embedding) applied to the $[ZnIm_3H_2O]^{2+}$ complex. As before, for the ONIOM calculations, Zn(II) ion and H_2O are in the high-level region and three Imidazole ligands in the low-level region. The PESs for the removal of water are calculated for conventional Mulliken point charge embedding and the case where we split the overlap populations as 70% Zn and 30% on N and 30% Zn and 70% on N. The PESs obtained by both point charge embedding variations and the regular Mulliken point charge embedding are indicated in Fig. 4. It is observed that in the case of the overlap populations being partitioned as 30% Zn and 70% N, the energy with point charge embedding is close to Mulliken point charge embedding; whereas in the case of 70% Zn and 30% N, which physically indicates the greater electronegativity of Zn^{2+} ion as compared to N, the energy is closer to the high-level target value than conventional Mulliken point charge embedding. This illustrates that a well-chosen un-
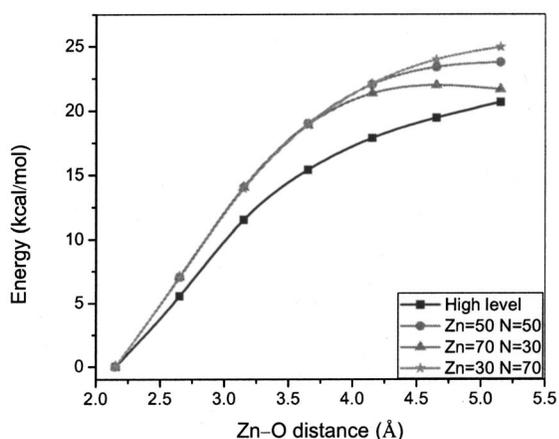
FIG. 4. PESs for the removal of H_2O from the $[ZnIm_3H_2O]^{2+}$ complex in the gas phase. The squares indicate the B3LYP/6-31+G** surface that is the high-level target, the circles indicate regular Mulliken point charge embedding at the B3LYP/6-31+G**:HF/6-31G level, the triangles show point charge embedding with 70% of the electron overlap population on Zn and 30% on N of the imidazole ligands, and the stars indicate the case with 30% on Zn and 70% on N.
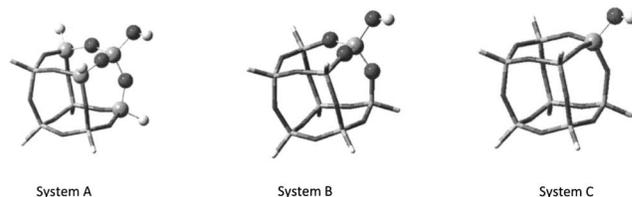
FIG. 5. The three ways [(A)–(C)] used to divide the entire hydroxylated spherosiloxane cluster ($Si_8O_{12}H_7OH$) into two layers: The high-level region is indicated as ball and stick representation and the low-level region as tube rendering.

equal splitting of the Mulliken overlap population may provide a better description of the atomic charges and yield better physical quantities.

The above two examples illustrate how the ONIOM Mulliken charge embedding energies offer significant improvement over mechanical embedding for a metalloenzyme system. In the next example, we perform full geometry optimizations to demonstrate the performance of the energy gradients.

The final example is the calculation of the deprotonation energy of a hydroxylated spherosiloxane cluster $Si_8O_{12}H_7OH \rightarrow Si_8O_{12}H_7O^-$. This cluster model is used to represent an isolated hydroxyl group on a silica surface. In this paper, ONIOM mechanical and point charge embedding calculations have been carried out using MP2/6-31G(d):HF/6-31G(d), B3LYP/6-31G(d):HF/6-31G(d), B3LYP/6-31G(d,p):BLYP/6-31G, and MP2/6-31G(d):BLYP/6-31G model chemistry combinations. These four combinations illustrate the typical combinations that are likely to be used for practical applications, viz., post-SCF:HF, DFT:HF, DFT:DFT, and post-SCF:DFT. The model systems for the reactants have been defined in three different ways as $(H_3SiO)_3SiOH$ (system A), $Si(OH)_4$ (system B), and H_3SiOH (system C), as shown in Fig. 5. This illustrates the point that selecting an appropriate model system involves a choice in such bonded cases since multiple ways of cutting the covalent bonds are possible. (This is somewhat different from the previous example involving the metalloenzyme where including the imidazole molecules in the low-level region seemed to be a logical choice.) In all three cases, severed covalent bonds in model system subcalculations have been capped by H link atoms. In models A and C, Si–O bonds are cut at the boundary and replaced by Si–H bonds during the model system subevaluations. The scale parameter[4] for the H link atoms is chosen to be 0.9 for models A and C, which corresponds to the ratio of Si–H and Si–O bond lengths in such silicon oxide clusters. In model B, O–Si bonds are cut at the boundary and replaced by O–H bonds. The scale parameter for the H link atoms is chosen to be 0.59 for model B, which is the ratio of O–H and O–Si bond lengths. Except for the atoms replaced by link atoms in the model systems, all atomic charges from the low-level region have been included in the embedding potentials without scaling. All geometries are optimized with the QM:QM Mulliken embedding gradient code. The calculated proton affinities are summarized in Table III. The deviations of the proton affinity calculated by the ONIOM-ME and ONIOM-EE methods

TABLE III. Proton affinity (kcal/mol) of the spherosiloxane cube anion ($Si_8O_{12}H_7O^-$) calculated by ONIOM-ME and ONIOM-EE. Also included for comparison are the target high-level and low-level methods. The mean absolute deviations of the ONIOM-ME and ONIOM-EE methods are 5.3 and 3.2 kcal/mol, respectively.

ONIOM combination	Model	High	Low	ME	EE	Error ME	Error EE
MP2/6-31G(d): HF/6-31G(d)	A	340.2	353.4	343.1	339.5	2.9	−0.7
	B	340.2	353.4	347.2	349.1	7.1	8.9
	C	340.2	353.4	346.0	343.1	5.8	2.9
B3LYP/6-31G(d): HF/6-31G(d)	A	344.0	353.4	345.3	340.9	1.3	−3.1
	B	344.0	353.4	349.5	351.9	5.5	7.9
	C	344.0	353.4	348.0	345.2	4.0	1.2
B3LYP/6-31G(d,p): BLYP/6-31G	A	348.1	335.2	341.7	346.9	−6.4	−1.2
	B	348.1	335.2	350.1	347.8	2.1	−0.3
	C	348.1	335.2	337.5	342.3	−10.6	−5.8
MP2/6-31G(d): BLYP/6-31G	A	340.2	335.2	335.5	341.4	−4.7	1.2
	B	340.2	335.2	344.9	341.6	4.7	1.4
	C	340.2	335.2	332.0	336.1	−8.2	−4.1
Mean absolute deviation			10.1			5.3	3.2

from the high-level target values are given in the last two columns of Table III. The mean absolute deviations for ONIOM-ME and ONIOM-EE methods from the target values are 5.3 and 3.2 kcal/mol, respectively, which indicates that ONIOM-EE has a better overall improvement over ONIOM-ME by 2 kcal/mol. More importantly, the performance for model A (perhaps the most appropriate model since it involves up to second nearest neighbors in the model system) for the four combinations improves significantly from a mean absolute deviation of 3.8 kcal/mol for ONIOM-ME to 1.6 kcal/mol for ONIOM-EE.

Also worth noting is the oscillatory nature of the proton affinity computed by the ONIOM ME and EE methods, as we proceed from model A to B to C. Such behavior was previously observed by Roggero *et al.*[21] in calculating the properties such as binding energies of ammonia on the isolated hydroxyl group of silica surface using ONIOM. This is clearly due to the alternating Si–O–Si bonds, which gives rise to a system of alternating positive and negative charges. In addition, the external charge in the surrounding region is either positive or negative depending on how we choose the model system in the ONIOM calculations. Hence the effect of EE when compared to ME is similar in models A and C, and exactly opposite in model B.

There are three cases in Table III where the error due to ONIOM-EE is worse than ONIOM-ME. For the MP2/6-31G(d):HF/6-31G(d) and B3LYP/6-31G(d):HF/6-31G(d) combinations and model B of the spherosiloxane cluster, the error due to ONIOM-EE increases as compared to ONIOM-ME. However, EE with model B yields better results for the other two ONIOM-combinations, viz., B3LYP/6-31G(d,p):BLYP/6-31G and MP2/6-31G(d):BLYP/6-31G. Clearly, future work is needed to understand the origin of such differences. For the B3LYP/6-31G(d):HF/6-31G(d) combination, and model A, there seems to be an overcorrection in the proton affinity calculated by ONIOM-EE, and the

value is worse than ONIOM-ME. Currently, we are evaluating the performance of our embedding models for more systems and different ONIOM combinations.

Lastly, to account for the greater electronegativity of oxygen than silicon, we divide the electron overlap populations between O and Si unequally as 70% on O and 30% on Si and 60% on O and 40% on Si. Using these charges with the ONIOM-GAME method, we have calculated the deprotonation energy for model A of the hydroxylated spherosiloxane cluster. The results are summarized in Table IV, which indicates the errors obtained from the high-level target for mechanical embedding, regular Mulliken point charge embedding (50% on O and 50% on Si), and the other two asymmetric Mulliken charge analyses. The mean absolute deviations of the proton affinity for model A are 2.7 kcal/mol for the 70-30 splitting and 2.1 kcal/mol for the 60-40 splitting (compared to 1.6 kcal/mol for regular Mulliken). It is unexpected that the asymmetric Mulliken embedding with a more logical splitting yields poorer results than regular Mulliken embedding. However, this is mainly due to the fact that one of the ONIOM combinations [B3LYP/6-31G(d):HF/6-31G(d)] performs significantly worse for asymmetric Mulliken embedding. This appears to be due to the different

TABLE IV. Errors in the proton affinity of the spherosiloxane cube anion (model A) calculated by ONIOM-ME and ONIOM-EE with regular Mulliken point charges (50-50) and asymmetric Mulliken charges obtained by dividing the electron overlap populations between O and Si unequally as 60% on O, 40% on Si and 70% on O, 30% on Si.

ONIOM combination	ME	50-50	60-40	70-30
MP2/6-31G(d): HF/6-31G(d)	2.9	−0.7	−1.5	−2.4
B3LYP/6-31G(d): HF/6-31G(d)	1.3	−3.1	−4.2	−5.6
B3LYP/6-31G(d,p): BLYP/6-31G	−6.4	−1.2	−0.5	0.1
MP2/6-31G(d): BLYP/6-31G	−4.7	1.2	2.0	2.8
Mean absolute deviation	3.8	1.6	2.1	2.7

TABLE V. Geometrical parameters for the hydroxylated spherosiloxane cube ($Si_8O_{12}H_7OH$) of C_s symmetry, (model A) for various model chemistry combinations including ONIOM-ME and ONIOM-EE.

Method	Si–O Bond distances (Å)					MAD (Å or °)	Maximum deviation (Å or °)
B3LYP/6-31G(d)	1.634	1.642	1.643	1.629	1.639		
B3LYP/6-31G(d,p)	1.632	1.642	1.643	1.629	1.639		
MP2/6-31G(d)	1.639	1.647	1.647	1.631	1.642		
MP2/6-31G(d): HF/6-31G(d), ME	1.639	1.646	1.648	1.634	1.645	0.001	0.003
MP2/6-31G(d): HF/6-31G(d), EE	1.643	1.649	1.654	1.635	1.650	0.005	0.008
B3LYP/6-31G(d): HF/6-31G(d), ME	1.634	1.642	1.642	1.629	1.639	0.000	0.001
B3LYP/6-31G(d): HF/6-31G(d), EE	1.639	1.644	1.648	1.631	1.644	0.004	0.005
B3LYP/6-31G(d,p): BLYP/6-31G, ME	1.630	1.644	1.638	1.632	1.635	0.003	0.005
B3LYP/6-31G(d,p): BLYP/6-31G, EE	1.626	1.642	1.631	1.633	1.630	0.006	0.010
MP2/6-31G(d): BLYP/6-31G, ME	1.638	1.649	1.645	1.636	1.642	0.002	0.005
MP2/6-31G(d): BLYP/6-31G, EE	1.632	1.648	1.638	1.637	1.637	0.006	0.009
	Bond angles (degrees)						
	Si–O–Si	Si–O–Si	Si–O–H	O–Si–O	O–Si–O		
B3LYP/6-31G(d)	145.8	152.8	116.5	106.6	110.6		
B3LYP/6-31G(d,p)	145.9	152.8	117.0	106.7	110.7		
MP2/6-31G(d)	144.2	155.4	116.0	106.6	110.5		
MP2/6-31G(d): HF/6-31G(d), ME	145.8	150.0	115.9	106.3	110.6	1.5	5.4
MP2/6-31G(d): HF/6-31G(d), EE	144.5	148.5	115.6	105.7	110.1	1.8	7.0
B3LYP/6-31G(d): HF/6-31G(d), ME	146.3	150.5	116.5	106.6	110.7	0.6	2.3
B3LYP/6-31G(d): HF/6-31G(d), EE	144.5	148.6	115.7	106.0	109.9	1.6	4.3
B3LYP/6-31G(d,p): BLYP/6-31G, ME	149.5	153.5	117.6	106.4	110.8	1.0	3.6
B3LYP/6-31G(d,p): BLYP/6-31G, EE	149.7	150.4	116.5	106.0	109.7	1.7	3.7
MP2/6-31G(d): BLYP/6-31G, ME	148.6	153.1	116.5	106.1	110.7	1.6	4.4
MP2/6-31G(d): BLYP/6-31G, EE	148.7	152.0	116.1	105.8	110.0	1.9	4.5

charge distributions obtained by Hartree–Fock and DFT methods offsetting the advantages of an asymmetric Mulliken splitting. In addition, it could partly be due to any deficiency associated with the charge imbalance created by cutting the system at a Si–O bond to cap it with a H atom. Clearly, the selection of ONIOM combinations plays an important role in the choice of appropriate asymmetric scale factors. Overall, however, we note that all EE variants are better than mechanical embedding, which has a mean absolute error of 3.8 kcal/mol.

The geometrical parameters of the neutral hydroxylated spherosiloxane cluster and its anion for model A, obtained by the different methods, are summarized in Tables V and VI. The mean absolute deviations and maximum deviations are also indicated. For the two compounds, the bond lengths obtained by both ONIOM-ME and ONIOM-EE calculations have deviations less than 0.01 Å, although the Si–O distances calculated using ONIOM-EE are typically too large by 0.005 Å. The bond angles obtained by the ONIOM-ME and ONIOM-EE calculations have a mean absolute deviation of less than 2° with respect to the target high-level values. The geometrical parameters for ONIOM-ME and ONIOM-EE are thus in close agreement with the high-level target values.

IV. CONCLUSIONS

In a previous publication,[10] we developed a QM:QM EE scheme within the ONIOM framework, where the model system is embedded in the Mulliken point charges of its sur-rounding region during the model system subcalculations. We rigorously implemented the energies as well as energy gradients in an efficient manner. In this paper, we demonstrate the applicability of the ONIOM-EE method and show that the method is robust and, in several cases, offers improvement over the existing ONIOM mechanical embedding scheme. We also introduce an asymmetric Mulliken population analysis scheme for incorporation into the ONIOM QM:QM point charge embedding energy and gradient, for cases where Mulliken charges may not be a good representation of the partial atomic charges. Two challenging cases are studied—zinc metalloenzymes and silicon oxide cages. In the proton transfer reaction between the water dimer of the $[ZnIm_3(H_2O)_2]^{2+}$ complex, the PESs obtained by ONIOM-EE are in excellent agreement with the target high-level PES for all three model chemistries considered. When the bond to the metal center is breaking, as in the removal of water from the $[ZnIm_3H_2O]^{2+}$ complex, the effects of the external point charges are very important in describing the proper physical behavior of the system at large Zn–O distances, when using the ONIOM approximation. Mechanical embedding undergoes a catastrophic failure in this case. When we split the overlap population between Zn(II) and the imidazole ligands unequally as 70% on Zn(II) and 30% on N to get modified Mulliken charges, we see a slight improvement in the point charge embedding energy. In the case of the silicon oxide cages, there is a modest overall improvement in the computed proton affinities for EE relative to ME. However, caution is required in such applications since we

TABLE VI. Geometrical parameters for the deprotonated hydroxylated spherosiloxane anion $(Si_8O_{12}H_7O^-)$ of C_{3v} symmetry, (model A) for various model chemistry combinations including ONIOM-ME and ONIOM-EE.

Method	Si–O bond distances (Å)			MAD	Maximum deviation
B3LYP/6-31G(d)	1.538	1.699	1.599		
B3LYP/6-31G(d,p)	1.537	1.699	1.599		
MP2/6-31G(d)	1.545	1.704	1.603		
MP2/6-31G(d): HF/6-31G(d), ME	1.545	1.703	1.605	0.001	0.002
MP2/6-31G(d): HF/6-31G(d), EE	1.552	1.706	1.611	0.006	0.008
B3LYP/6-31G(d): HF/6-31G(d), ME	1.537	1.698	1.599	0.001	0.001
B3LYP/6-31G(d): HF/6-31G(d), EE	1.545	1.700	1.608	0.005	0.008
B3LYP/6-31G(d,p): BLYP/6-31G, ME	1.536	1.702	1.594	0.003	0.005
B3LYP/6-31G(d,p): BLYP/6-31G, EE	1.531	1.702	1.589	0.006	0.010
MP2/6-31G(d): BLYP/6-31G, ME	1.543	1.706	1.599	0.002	0.004
MP2/6-31G(d): BLYP/6-31G, EE	1.539	1.707	1.593	0.006	0.01

	Bond angles (degrees)			
	Si–O–Si	O–Si–O		
B3LYP/6-31G(d)	156.3	117.0		
B3LYP/6-31G(d,p)	156.4	117.0		
MP2/6-31G(d)	156.4	117.1		
MP2/6-31G(d): HF/6-31G(d), ME	155.7	117.0	0.4	0.7
MP2/6-31G(d): HF/6-31G(d), EE	154.1	116.6	1.5	2.3
B3LYP/6-31G(d): HF/6-31G(d), ME	156.0	117.0	0.2	0.3
B3LYP/6-31G(d): HF/6-31G(d), EE	153.6	116.2	1.7	2.7
B3LYP/6-31G(d,p): BLYP/6-31G, ME	159.5	117.2	1.7	3.1
B3LYP/6-31G(d,p): BLYP/6-31G, EE	159.0	116.9	1.4	2.7
MP2/6-31(d): BLYP/6-31G, ME	159.1	117.3	1.5	2.7
MP2/6-31G(d): BLYP/6-31G, EE	159.5	117.0	1.6	3.1

have shown that the ONIOM method combinations and the way of cutting the system at the QM:QM boundary will play a significant role in the accuracy of the method.

Within the ONIOM approximation, if the polarization effect of the model system from the surrounding region is not important, or if the polarization effect is similar for both model low and model high level of theories, mechanical embedding may be a reasonable approximation; hence the effect of the external point charges must be assessed to determine their importance in specific problems. We are performing more assessments on other systems to understand the advantages and deficiencies of our Mulliken embedding technique. Although we have shown results on relatively small benchmark systems, the larger the size of the system, the larger the expected effect of EE. Finally, in the cases where the Mulliken point charge approximation is not sufficient, the developments of better ONIOM electronic embedding models, such as full density embedding and fitted density embedding are needed, and work is currently in progress in developing such more sophisticated techniques for electronic embedding.[22]

[1] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).

[2] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990); F. Maseras and K. Morokuma, *ibid.* **16**, 1170 (1995); U. C. Singh and P. A. Kollman, *ibid.* **7**, 718 (1986); A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[3] V. Gogonea, L. M. Westerhoff, and K. M. Merz, J. Chem. Phys. **113**, 5604 (2000).

[4] S. Dapprich, I. Komaromi, K. S. Byun, K. Morokuma, and M. J. Frisch, J. Mol. Struct.: THEOCHEM **462**, 1 (1999).

[5] N. Rega, S. S. Iyengar, G. A. Voth, H. B. Schlegel, T. Vreven, and M. J. Frisch, J. Phys. Chem. B **108**, 4210 (2004); B. W. Hopkins and G. S. Tschumper, J. Comput. Chem. **24**, 1563 (2003); T. Vreven and K. Morokuma, Theor. Chem. Acc. **109**, 125 (2003); T. Vreven, B. Mennucci, C. O. da Silva, K. Morokuma, and J. Tomasi, J. Chem. Phys. **115**, 62 (2001); T. Vreven and K. Morokuma, J. Comput. Chem. **21**, 1419 (2000); P. B. Karadakov and K. Morokuma, Chem. Phys. Lett. **317**, 589 (2000); M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, J. Phys. Chem. **100**, 19357 (1996); S. Humbel, S. Sieber, and K. Morokuma, J. Chem. Phys. **105**, 1959 (1996).

[6] D. Bakowies and W. Thiel, J. Phys. Chem. **100**, 10580 (1996).

[7] J. L. Gao, J. Comput. Chem. **18**, 1061 (1997); Y. L. Lin and J. L. Gao, J. Chem. Theory Comput. **3**, 1484 (2007); M. J. Field, Mol. Phys. **91**, 835 (1997); C. J. R. Illingworth, S. R. Gooding, P. J. Winn, G. A. Jones, G. G. Ferenczy, and C. A. Reynolds, J. Phys. Chem. A **110**, 6487 (2006); S. R. Gooding, P. J. Winn, R. I. Maurer, G. G. Ferenczy, J. R. Miller, J. E. Harris, D. V. Griffiths, and C. A. Reynolds, J. Comput. Chem. **21**, 478 (2000).

[8] J. Gao, J. Phys. Chem. B **101**, 657 (1997).

[9] W. Xie and J. Gao, J. Chem. Theory Comput. **3**, 1890 (2007); W. Xie, L. Song, D. G. Truhlar, and J. Gao, J. Chem. Phys. **128**, 234108 (2008).

[10] H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, J. Chem. Phys. **128**, 034107 (2008).

[11] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem., Quantum Chem. Symp. **13**, 225 (1979).

[12] N. C. Handy and H. F. Schaefer III, J. Chem. Phys. **81**, 5031 (1984).

[13] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).

[14] H. Sato and S. Sakaki, Chem. Phys. Lett. **434**, 165 (2007).

[15] J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, J. Phys. Chem. **96**, 135 (1992).

[16] J. P. Foster and F. Weinhold, J. Am. Chem. Soc. **102**, 7211 (1980); A. E. Reed and F. Weinhold, J. Chem. Phys. **78**, 4066 (1983).

[17] B. H. Besler, K. M. Merz, and P. A. Kollman, J. Comput. Chem. **11**, 431 (1990); U. C. Singh and P. A. Kollman, *ibid.* **5**, 129 (1984).

[18] R. Bader, *Atoms in Molecules: A Quantum Theory* (Oxford University Press, Oxford, 1994).

[19] G. Y. Hong, M. Strajbl, T. A. Wesolowski, and A. Warshel, J. Comput. Chem. **21**, 1554 (2000).

[20] A. Ruzsinszky and J. P. Perdew, J. Chem. Phys. **125**, 194112 (2006).

[21] I. Roggero, B. Civalleri, and P. Ugliengo, Chem. Phys. Lett. **341**, 625 (2001).

[22] H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven (in preparation).