

Combining *ab initio* and density functional theories with semiempirical methods

Qiang Cui

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138 and Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin 53706

Hong Guo

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Martin Karplus

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138 and Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France

(Received 1 April 2002; accepted 25 June 2002)

For large reactive systems, the calculation of energies can be simplified by treating the active part with a high-level quantum mechanical (QM) (*ab initio* or density functional) approach and the environment with a less sophisticated semiempirical (SE) approach, as an improvement over the widely used hybrid quantum mechanical/molecular mechanical (QM/MM) methods. An example is the interaction between an active region of an enzyme and its immediate environment. One such method is the original “Our-own-N-layer Integrated molecular Orbital+Molecular Mechanics (ONIOM)” approach. In this paper, the interaction between the QM and SE region is described explicitly by two different schemes. In the *iterative* QM/SE schemes (QM/SE-I), the electrostatic interaction and polarization effects are introduced explicitly for both the QM and SE atoms by a self-consistent procedure based on either polarizable point charges or the electron density. In the *noniterative* QM/SE scheme, based on the ONIOM model (QM/SE-O), the exchange (Pauli repulsion) and charge transfer effects are taken into account at the SE level, in addition to the explicit electrostatic interaction and polarization between the two regions. Test calculations are made on a number of model systems (including small polar or charged molecules interacting with water and proton transfer reactions in the presence of polar molecules or an extended hydrogen-bond network). The quantitative accuracy of the results depend on several parameters, such as the charge-scaling/normalization factors for the SE charge and the QM/SE van der Waals parameters, which can be chosen to optimize the result. For the QM/SE-O approach, the results are more sensitive to the quality of the SE level (e.g., self-consistent-charge density-functional-tight-binding vs AM1) than the explicit interaction between QM and SE atoms.

© 2002 American Institute of Physics. [DOI: 10.1063/1.1501134]

I. INTRODUCTION

Combined quantum mechanical and molecular mechanical (QM/MM) methods^{1,2} are powerful tools for the study of reactions in large systems. When carefully calibrated, they have been shown to provide useful insights into the mechanism of chemical reactions in solution, on surfaces and in enzymes.^{1–5} Algorithms for computing molecular properties such as infrared (IR) spectra⁶ and nuclear magnetic resonance (NMR) chemical shifts⁷ in the QM/MM framework have recently extended the range of QM/MM applications.

Despite the overall success of the QM/MM approach, there are several issues associated with the common implementations. One aspect concerns the appropriate treatment of the QM and MM interface, when it is inside a molecule. A number of schemes have been proposed, which include the standard link-atom scheme, double link-atoms, frontier orbitals and effective boundary potentials.⁸ A comparison of several schemes showed that they gave satisfactory results when compared to full QM calculations if “over polarization” of

the QM boundary atoms is avoided.⁹ Another concern about QM/MM methods is that the environment of the QM region is usually treated with a non-polarizable force field (such as CHARMM), in which atoms are represented by *fixed* point-charges. Although this appears to be qualitatively adequate for many properties, such as the determination of ground-state structures, it is not quantitatively accurate, in general, especially in cases where many-body effects are important. An example of the latter is the cooperativity induced by hydrogen-bond networks found in the active site of many enzymes.¹⁰ Although these nonadditive many-body effects can be captured by extending the size of the QM region, the cost of high-level QM methods often makes this choice impractical. One possibility for improving the description of the interaction between the QM region and its immediate environment is to include polarization in the molecular mechanics method used to describe the latter. This can be done most straightforwardly by the introduction of atomic polarizabilities on the MM atoms.^{1(c),11,12} Another possibility is to in-

clude polarization on the MM atoms by use of the principle of chemical potential equalization.¹³ The concept has been used in the development of polarizable force fields,^{14,15} and has been introduced into the QM/MM framework.¹⁶ An alternative approach, termed the effective fragment potential (EFP) method, has been employed in the context of solution chemistry.¹⁷ There the solvent molecules are treated as *rigid* fragments, which are represented in terms of distributed multipoles and polarizability tensors calculated with *ab initio* theory. An approximate formula for including the Pauli repulsion¹⁸ and charge transfer¹⁹ between the solvent and solute has also been developed. Applications to reactions in small clusters²⁰ and to enzymes (e.g., mechanism of chorismate mutase,²¹ residue pKa calculations²² and the absorption spectrum of a vanadate complex at the RNase A active site²³) where either the solvent molecules or the active site residues are described by EFP have given promising results. A limitation is that the geometry of the fragments has to be rigid, which makes it difficult to apply in molecular dynamics studies of reactions in proteins and nucleic acids.

Another approach for improving the description of the immediate environment of the reactive region (e.g., the active site of an enzyme) is to employ multiple partitions, in which the core is described with a high-level *ab initio* or density functional theory (DFT) method, its immediate environment (first solvation layer) is treated with a lower-level quantum mechanical approach, such as a semiempirical (SE) method, and atoms further away are described at the MM level. This philosophy has been pursued by Morokuma and co-workers,²⁴ who pioneered a set of additive schemes, generally termed "ONIOM" (Our own N-layer Integrated molecular Orbital+molecular Mechanics). In the original ONIOM implementation,²⁴ the reactive region does not explicitly interact with the environment at the high QM level, and all the environmental effects such as polarization, Pauli repulsion, and charge transfer were described at a low quantum mechanical level (e.g., Hartree-Fock). Therefore, the ONIOM scheme tends to give reliable results only when the low QM level captures the environmental effects in a satisfactory manner.²⁵ In more recent developments, nonadditive schemes were explored in which the wave function (or density) of the core region at the high QM level is polarized explicitly by the environment (which is described at a low quantum mechanical level). In the method proposed by Carter *et al.*,²⁶ the core region is described by a high-level *ab initio* approach, such as configuration interaction (CI), and the environment is described with DFT; the two interact self-consistently through an embedding potential derived from the charge densities of the two region. In the approach of Merz and co-workers,²⁷ the core region is described by DFT and the environment is treated with a semiempirical approach (AM1); a consistent solution to the coupled electronic structure problem is found through a chemical potential scheme originally proposed in the context of a divide and conquer DFT treatment for large systems.²⁸

Finally, we note that methods exist to treat complex systems in a modular fashion, similar in spirit to the multilayer approaches, but with the same level of QM method. For example, Gao developed an approach (Ref. 42) for simulat-

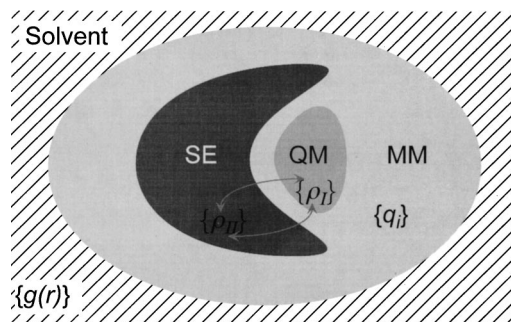


FIG. 1. A schematic representation of a multilayer partition of a large system in solution: the reactive region requires a high level quantum mechanical (QM) description, its immediate environment can be described at a less sophisticated quantum mechanical level such as a semiempirical (SE) approach, and molecular mechanics is sufficient for the rest systems. In some cases, a continuum model can be used to describe effects due to the bulk solvent [characterized, for example, by the radial distribution function $g(r)$].

ing water, in which each water molecule is treated with a semiempirical method (AM1). In the SCF cycle for each water molecule ("solute"), the other water molecules ("solvent") are represented by partial charges derived from AM1 calculations; the SCF procedure is iterated until the "solute" water and the "solvent" water molecules are polarized consistently. Another approach is based on the block-localized wave function method,²⁹ which can treat the electrostatic, Pauli repulsion (exchange) and charge transfer between the molecular fragments in a valence-bond framework.³⁰ This approach has been used for an energy decomposition of intermolecular interactions³¹ and chemical reactions in solution.³²

In the present work, we describe a number of multilayer methods that combine *ab initio* or density functional theories (QM) with semiempirical (SE) methods; MM can be included as the outer layer. To simplify notations, we refer to *ab initio* or DFT methods as the QM level and distinguish that from SE, although the latter is also quantum mechanical in nature. The methods introduced here are different from the original ONIOM²⁴ approach in that interactions between different partitions are taken into account explicitly; the specific coupling schemes are also different from other methods cited above.^{26,27} In Sec. II, we describe the QM/SE theory and numerical algorithms. Section III presents several test cases that illustrate the features of the new combined method. Tests include the interaction between several small polar or charged molecules and water, and the effect of polar molecules and hydrogen-bond networks on a proton transfer reaction related to triosephosphate isomerase. The conclusions are outlined in Sec. IV. More systematic explorations of the QM/SE methods and the necessary parameterization are left for a separate publication.

II. THEORY

The system is partitioned into three layers, as shown schematically in Fig. 1, though more general (multiple) partitions are possible. The immediate neighborhood of the chemically important QM region is treated with SE, while

the outer region is described with an MM force field. The total Hamiltonian of the system can be written in a self-evident notion as

$$\hat{H} = \hat{H}^{\text{QM}} + \hat{H}^{\text{SE}} + \hat{H}^{\text{QM/SE}} + \hat{H}^{\text{QM/MM}} + \hat{H}^{\text{SE/MM}} + E^{\text{MM}}. \quad (1)$$

The expressions for the interactions between MM part and the QM (SE) part have the usual form,^{1(b)}

$$\begin{aligned} \hat{H}^{\text{QM(SE)/MM}} &= \sum_{J \in \text{QM(SE)}} \hat{V}_{J, \text{Nuc}}^{\text{QM(SE)/MM}} \\ &+ \sum_{i=1}^{N_{\text{el}}^{\text{QM(SE)}}} \hat{V}_{\text{el}}^{\text{QM(SE)/MM}} + E_{\text{van}}^{\text{QM(SE)/MM}} \\ &= \sum_{I \in \text{MM}} \sum_{J \in \text{QM(SE)}} \frac{Q_I Z_J}{|\vec{R}_I - \vec{R}_J|} \\ &- \sum_{I \in \text{MM}} \sum_{i=1}^{N_{\text{el}}^{\text{QM(SE)}}} \frac{Q_I}{|\vec{R}_I - \vec{r}_i|} + E_{\text{van}}^{\text{QM(SE)/MM}}, \quad (2) \end{aligned}$$

where the first two terms on the rhs of Eq. (2) represent the electrostatic interaction between the MM atoms and the QM (SE) nuclei and electrons, respectively; the last term is the van der Waals term between the two sets of atoms that correspond to the Pauli repulsion at short range and dispersion at large distance. The interaction between the QM and SE region is not as well defined and will be considered in the following section. Two classes of methods will be described. The first uses iterative approaches that solve the coupled QM

and SE problem, and includes only the electrostatic and polarization effects between the two regions. The second class follows the spirit of the ONIOM method and is noniterative; they also include the charge transfer and exchange repulsion between the QM and SE regions at the SE level.

A. The iterative QM/SE scheme (QM/SE-I): Electrostatic and polarization effects

In this section, we treat the QM/SE coupling problem following the framework that has been used in analyzing weak intermolecular interactions. The methods obtained require iterative solutions and therefore are referred to as QM/SE-I (I for “iterative”). For simplicity, we assume that the QM and SE regions are not covalently bonded; more complicated cases will be considered elsewhere. We start by formally considering both the QM and SE regions at the Hartree–Fock (HF) level, and introduce necessary approximations to the SE region later; it is straightforward to extend the formulation to cases where the QM is DFT (see below). We label the QM part as region *A*, and the SE part as region *B*. Accordingly, the electronic structure problem of the composite system *A + B* is described by the standard generalized eigenvalue equations,

$$\mathbf{F}^{A+B} \mathbf{C}_m^{A+B} = \mathbf{S}^{A+B} \mathbf{C}_m^{A+B} \boldsymbol{\varepsilon}_m^{A+B} \quad (m = 1, N), \quad (3)$$

where *N* is the number of atomic basis functions; **F**, **C**, and **S** are the Fock, eigenvector and overlap matrix, respectively, and $\boldsymbol{\varepsilon}$ is a diagonal matrix with eigenvalues on its diagonal.

The Fock matrix in (3) can be written explicitly in terms of *intra*-region and *inter*-region contributions

$$\begin{aligned} F_{\mu\nu} &= T_{\mu\nu} + \left\langle \chi_\mu \left| \sum_{C \in A} \frac{-Z_C}{|\vec{R}_C - \vec{r}|} \right| \chi_\nu \right\rangle + \left\langle \chi_\mu \left| \sum_{D \in B} \frac{-Z_D}{|\vec{R}_D - \vec{r}|} \right| \chi_\nu \right\rangle + \langle \chi_\mu | \hat{V}_{\text{el}}^{A,B/MM} | \chi_\nu \rangle + \sum_{\lambda\sigma \in A} P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle \\ &+ \sum_{\lambda'\sigma' \in B} P_{\lambda'\sigma'} \langle \mu\nu | \lambda'\sigma' \rangle + \sum_{\lambda''\sigma'' \in B} P_{\lambda''\sigma''} \langle \mu\nu | \lambda''\sigma'' \rangle + \sum_{\lambda'''\sigma''' \in A} P_{\lambda'''\sigma'''} \langle \mu\nu | \lambda'''\sigma''' \rangle, \quad (4) \end{aligned}$$

where $\chi_{\mu(\nu)}$ is the atomic basis function ($\mu, \nu \in A+B$), and $P_{\lambda\sigma}$ is the one-particle density matrix element in the atomic orbital (AO) basis. Note that the electronic interaction with the MM atoms, if present, is included in Eq. (4); in accord with Eq. (2), it is directly separable and so is omitted for simplicity.

Now we formally introduce the NDDO (Neglect-Diatomic-Differential-Overlap) approximation³³ used in AM1 and PM3 for region *B*, which is treated with the semiempirical approach. Furthermore, we modify the terms corresponding to the *A–B* interaction. Specifically, we introduce the NDDO approximation not only for AO integrals involving only the SE region but also for the *inter-region* AO integrals,

$$\begin{aligned} \langle \mu^K \nu^{\bar{K}} | \lambda^{K+\bar{K}} \sigma^{K+\bar{K}} \rangle &\approx 0 \\ \left\langle \chi_{\mu^K} \left| \frac{-Z}{|\vec{R} - \vec{r}|} \right| \chi_{\nu^{\bar{K}}} \right\rangle, \quad T_{\mu^K \nu^{\bar{K}}} &\approx 0 \quad \{K, \bar{K}\} = \{A, B\}, K \cap \bar{K} = \emptyset. \quad (5) \end{aligned}$$

To further simplify the problem, we make the approximation that *the basis functions from the two regions do not mix* in the molecular orbitals, i.e., the off-diagonal block of the density matrix in Eq. (4) was neglected,

$$P_{\lambda''\sigma''} \equiv P_{\lambda'''\sigma'''} \equiv P_{\lambda\sigma} = 0. \quad (6)$$

This is a necessary simplification because usually quite different forms of basis functions are used in the QM (Gaussian) and SE (Slater) region. This would make the inter-region two electron integrals, especially the exchange type integrals, rather difficult to evaluate and violates the simplicity offered by the SE approach. Moreover, explicit mixing of the QM

and SE basis sets is likely to induce artifacts due to the large difference in the two types of methods; e.g., artificial charge transfer between the two regions. The inter-region integral approximation [Eqs. (5) and (6)] is not expected to cause large errors in describing the interactions between the two regions when the separation is larger than a hydrogen-bond distance (i.e., longer than 2.5 Å). This is based on the test of similar approximations in the frozen density functional method of Wesolowski *et al.*³⁴ (see below). In Ref. 34, they used the same basis sets for different regions and therefore can actually check the validity of the approximation; in the original ONIOM,²⁴ no explicit interaction between different layers is calculated.

With approximations Eqs. (5) and (6), the Fock matrix element in Eq. (4) reduces to

$$\begin{aligned}
 F_{\mu\nu}^{K,\text{NDDO}} &= \left[T_{\mu\nu} + \left\langle \chi_\mu \left| \sum_{C \in K} \frac{-Z_C}{|\vec{R}_C - \vec{r}|} \right| \chi_\nu \right\rangle \right. \\
 &+ \left. \sum_{\lambda\sigma \in K} P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle + \langle \chi_\mu | \hat{V}_E^{K/\text{MM}} | \chi_\nu \rangle \right] \\
 &+ \left[\left\langle \chi_\mu \left| \sum_{D \in \bar{K}} \frac{-Z_D}{|\vec{R}_D - \vec{r}|} \right| \chi_\nu \right\rangle \right. \\
 &+ \left. \sum_{\lambda'\sigma' \in \bar{K}} P_{\lambda'\sigma'} \langle \mu\nu | \lambda'\sigma' \rangle \right], \quad \mu, \nu \in K \\
 &= F'_{\mu\nu}{}^K + F''_{\mu\nu}{}^{K/\bar{K}}. \quad (7)
 \end{aligned}$$

Note that only the atomic basisfunctions centered on atoms in region K ($K=A, B$) are present. The first three terms in Eq. (7) represent the intraregion interaction, and the last two terms represent the inter-region electron–nuclei and electron–electron Coulombic interactions, respectively. In other words, only the Coulombic interaction between the QM and SE region is taken into account explicitly. The exchange-repulsion term and the two-center one-electron resonance terms that are present in the standard SE matrix elements are not included. They are modeled by van der Waals terms between the QM and SE regions in analogy to what is done in the QM/MM interaction contributions [Eq. (2), also see Eq. (9) below]. As a result, the eigenvalue problem Eq. (3) becomes partially decoupled, in the sense that the corresponding matrices are “region-diagonal,” although the matrix elements of each region depend on the solution of the other through the last two terms in Eq. (7). We have

$$\begin{aligned}
 &\begin{pmatrix} \mathbf{F}^{A,\text{NDDO}} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^{B,\text{NDDO}} \end{pmatrix} \begin{pmatrix} \mathbf{C}^A & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^B \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{S}^A & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^B \approx \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{C}^A & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^B \end{pmatrix} \begin{pmatrix} \boldsymbol{\varepsilon}^A & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varepsilon}^B \end{pmatrix}. \quad (8)
 \end{aligned}$$

The eigenvalue problem Eq. (8) has to be solved self-consistently (Fig. 1). One starts by solving the eigenvalue problem for block A (or B), which yields the (updated) density matrix that enters the Fock matrix element for block B (A) through the inter-region Coulombic terms in Eq. (7). The

eigenvalue problem for block B (A) is then solved which in turn updates the Fock matrix of block A (B). The procedure is carried out until self-consistency is reached.

With the converged Fock matrix from Eq. (8), the total energy of the system can be written

$$\begin{aligned}
 E_{\text{QM/SE-I}}^{\text{Tot}} &= \frac{1}{2} \sum_{\mu, \nu \in A} P_{\mu\nu} [h_{\mu\nu}^{\text{QM}} + h_{\mu\nu}^{\text{QM/SE}} \\
 &+ V_{\text{el}, \mu\nu}^{\text{QM/MM}} + F'_{\mu\nu}{}^{\text{QM}} + F''_{\mu\nu}{}^{\text{QM/SR}}] \\
 &+ \frac{1}{2} \sum_{\lambda\sigma \in B} P_{\lambda\sigma} [h_{\lambda\sigma}^{\text{SE}} + h_{\lambda\sigma}^{\text{SE/QM}} + V_{\text{el}, \lambda\sigma}^{\text{SE/MM}} \\
 &+ F'_{\lambda\sigma}{}^{\text{SE}} + F''_{\lambda\sigma}{}^{\text{SE/QM}}] + E_{\text{van}}^{\text{QM(SE)/MM}} \\
 &+ E_{\text{van}}^{\text{QM/SE}} + E_{\text{Nuc}}^{\text{QM(SE)/MM}} + E_{\text{Nuc}}^{\text{QM/SE}} + E^{\text{MM}}, \quad (9)
 \end{aligned}$$

where $h_{\mu\nu}^{\text{QM(SE)}}$ are the one-electron integrals for the QM (SE) region (kinetic plus nuclear–electron terms) and $h_{\mu\nu}^{\text{QM/SE}}$ ($h_{\mu\nu}^{\text{SE/QM}}$) are the integrals for QM (SE) electron and SE (QM) nuclei interactions [fifth term on the rhs of Eq. (7)]; the notation for the Fock matrix elements is indicated in Eq. (7). As mentioned above, the Pauli repulsion and charge transfer effects between QM and SE atoms are modeled with the effective van der Waals terms, $E_{\text{van}}^{\text{QM/SE}}$. In Eq. (9), the Coloumbic interaction between the QM and SE electrons are formally partitioned equally at the QM and SE levels, in the form of

$$\frac{1}{2} \sum_{\mu\nu \in A} P_{\mu\nu} \left[\sum_{\lambda\sigma \in B} P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle \right]$$

and

$$\frac{1}{2} \sum_{\lambda\sigma \in B} P_{\lambda\sigma} \left[\sum_{\mu\nu \in A} P_{\mu\nu} \langle \lambda\sigma | \mu\nu \rangle \right],$$

respectively. In the practical implementation, however, the polarization of the QM and SE wave functions was treated in an asymmetric manner because the QM region is more important (see Sec. II A 1). An alternative expression was used in which the QM/SE electron–electron Coloumbic interaction was evaluated at the QM level only. The required expression can be derived from Eq. (9) through a straightforward rearrangement,

$$\begin{aligned}
 E_{\text{QM/SE-I}}^{\text{Tot}} &= \frac{1}{2} \sum_{\mu, \nu \in A} P_{\mu\nu} [h_{\mu\nu}^{\text{QM}} + V_{\text{el}, \mu\nu}^{\text{QM/MM}} + F'_{\mu\nu}{}^{\text{QM}}] \\
 &+ \sum_{\mu, \nu \in A} P_{\mu\nu} F''_{\mu\nu}{}^{\text{QM/SE}} \\
 &+ \frac{1}{2} \sum_{\lambda\sigma \in B} P_{\lambda\sigma} [h_{\lambda\sigma}^{\text{SE}} + V_{\text{el}, \lambda\sigma}^{\text{SE/MM}} + F'_{\lambda\sigma}{}^{\text{SE}}] \\
 &+ \sum_{\lambda\sigma \in B} P_{\lambda\sigma} h_{\lambda\sigma}^{\text{SE/QM}} + E_{\text{van}}^{\text{QM(SE)/MM}} \\
 &+ E_{\text{van}}^{\text{QM/SE}} + E_{\text{Nuc}}^{\text{QM(SE)/MM}} + E_{\text{Nuc}}^{\text{QM/SE}} + E^{\text{MM}}. \quad (10)
 \end{aligned}$$

The derivations described above assumed a Hartree–Fock QM level, but the formulation can be readily extended to perturbation theories such as MP2, which use the converged MO coefficients from Eq. (9) or (10). An extension to other variational methods, such as CASSCF, can also be made, but we do not describe it here. The corresponding DFT treatment is straightforward, given the parallelism between Hartree–Fock and the Kohn–Sham formalisms.³⁵ We consider the Kohn–Sham matrix element for the $A+B$ problem at the DFT level,

$$\begin{aligned}
 F_{\mu\nu} = & T_{\mu\nu} + \left\langle \chi_{\mu} \left| \sum_{C \in A} \frac{-Z_C}{|\vec{R}_C - \vec{r}|} \right| \chi_{\nu} \right\rangle \\
 & + \left\langle \chi_{\mu} \left| \sum_{D \in B} \frac{-Z_D}{|\vec{R}_D - \vec{r}|} \right| \chi_{\nu} \right\rangle + \langle \chi_{\mu} | \hat{V}_{\text{el}}^{A,B/MM} | \chi_{\nu} \rangle \\
 & + \sum_{\lambda\sigma \in A} P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle + \sum_{\lambda'\sigma' \in B} P_{\lambda'\sigma'} \langle \mu\nu | \lambda'\sigma' \rangle \\
 & + \sum_{\lambda'' \in A, \sigma'' \in B} P_{\lambda''\sigma''} \langle \mu\nu | \lambda''\sigma'' \rangle \\
 & + \sum_{\lambda''' \in B, \sigma''' \in A} P_{\lambda'''\sigma'''} \langle \mu\nu | \lambda'''\sigma''' \rangle + V_{\mu\nu}^{\text{xc}}(\rho_1 + \rho_2),
 \end{aligned} \tag{11}$$

where the exchange integrals have been replaced by the exchange–correlation potential $V^{\text{xc}}(\rho)$. Invoking the NDDO approximation for the two regions [Eqs. (5) and (6)], the approximate Kohn–Sham matrix element becomes

$$\begin{aligned}
 F_{\mu\nu}^{K,\text{NDDO}} = & T_{\mu\nu} + \left\langle \chi_{\mu} \left| \sum_{C \in K} \frac{-Z_C}{|\vec{R}_C - \vec{r}|} \right| \chi_{\nu} \right\rangle + \langle \chi_{\mu} | \hat{V}_{\text{el}}^{K/MM} | \chi_{\nu} \rangle \\
 & + \sum_{\lambda\sigma \in K} P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle \\
 & + \left\langle \chi_{\mu} \left| \sum_{D \in \bar{K}} \frac{-Z_D}{|\vec{R}_D - \vec{r}|} \right| \chi_{\nu} \right\rangle \\
 & + \sum_{\lambda'\sigma' \in \bar{K}} P_{\lambda'\sigma'} \langle \mu\nu | \lambda'\sigma' \rangle + V_{\mu\nu}^{\text{xc}}(\rho_1 + \rho_2).
 \end{aligned} \tag{12}$$

This expression is very similar to the one derived by Wesolowski *et al.*,³⁴ except that the nonadditivity of the kinetic functional is not considered here because one region is treated with SE, and the associated effect (which is largely a short-range repulsion) is modeled by a van der Waals term between the DFT and SE region. When the SE method is also based on DFT (such as the SCC–DFTB approach³⁶), the exchange–correlation potential can, in principle, be evaluated on a set of grids at both levels, given an appropriate representation of $\rho_1(r)$ and $\rho_2(r)$. When the SE level is a Hartree–Fock based approach, such as AM1, although the exchange–correlation potential can be evaluated at the DFT level including a representation of the SE density (see Sec. II A 1), we choose to evaluate only $V_{\mu\nu}^{\text{xc}}(\rho_1)$ in the current implementation, based on the consideration that the quality of electron density is different in DFT and SE approaches.

This additional approximation is partially compensated by the van der Waals interaction between QM and SE atoms. In the DFT/AM1 combination, the AM1 Fock matrix takes the form of Eq. (7). With the expression of the Fock matrix elements in Eq. (12) for DFT and Eq. (7) for SE, the total energy of DFT/SE/MM formulation is again given by Eq. (10).

The remaining “technical” problem is the numerical calculation of the inter-region electron–electron interaction terms in Eqs. (7) and (12). This is not straightforward given that most QM methods use Gaussian-type basis functions while SE uses Slater orbitals. In the following sections, we describe two possible solutions.

1. The auxiliary basis method (QM/SE-I-A)—beyond the point charge model

To compute the QM/SE Coulombic integrals in Eqs. (7) and (12), one can fit the Slater orbitals in the semiempirical region with a STO–NG scheme, and then calculate the two electron integrals in the traditional manner.³⁷ However, we propose to use an alternative approach, which makes use of the fact that the required QM/SE Coulomb integrals can be written in the form

$$\begin{aligned}
 I_{\mu\nu} = & \int \int \chi_{\mu}(\vec{r}_1) \chi_{\nu}(\vec{r}_2) \frac{1}{r_{12}} \rho^{\bar{K}}(\vec{r}_2) d\vec{r}_1 d\vec{r}_2, \\
 & \mu, \nu \in K,
 \end{aligned} \tag{13}$$

where $\rho^{\bar{K}}(\vec{r})$ represents the electron density for the atoms in region \bar{K} . To compute $I_{\mu\nu}$, one can use a set of auxiliary basis functions of the same type as the atomic basis functions in region K to expand $\rho^{\bar{K}}(\vec{r})$ and then compute the resulting three-center two-electron ($3c,2e$) integrals. This density-fitting scheme was chosen in the current work based on several considerations. First, as shown in a number of previous studies,³⁸ the density fitting technique is numerically very efficient and reduces the formal scaling of the Coulomb integral calculations (from N^4 to N^3). Second, as shown by St-Amant *et al.*,³⁹ linear scaling can be achieved with the divide-and-conquer fitting scheme, which is an appealing feature for QM/SE calculations with a large SE region. Finally, once an appropriate representation of the density (and its gradient, if required) is available, the exchange–correlation potential can be evaluated as well.

As discussed by a number of authors,³⁸ the density fitting procedure requires solving the following set of linear equations:

$$\mathbf{c} = \mathbf{S}^{-1}(\mathbf{t} + \lambda \mathbf{n}), \tag{14}$$

where the \mathbf{c} is the fitting coefficient vector,

$$\rho^{\bar{K}}(\vec{r}) = \sum_i c_i \chi'_{i\bar{K}}(\vec{r}). \tag{15a}$$

The other quantities are defined as the following:

$$S_{ij} = \int \int \frac{\chi'_{i\bar{k}}(\vec{r}) \chi'_{j\bar{k}}(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}', \quad (15b)$$

$$t_i = \int \int \frac{\chi'_{i\bar{k}}(\vec{r}) \rho^{\bar{k}}(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' \\ = \sum_{\mu\nu \in \bar{k}} P_{\mu\nu}^{\text{STO}} \int \int \frac{\chi'_{i\bar{k}}(\vec{r}) \chi_{\mu}^{\text{STO}}(\vec{r}') \chi_{\nu}^{\text{STO}}(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}', \quad (15c)$$

$$n_i = \int \chi'_{i\bar{k}}(\vec{r}) d\vec{r}, \quad (15d)$$

$$\lambda = N_{\bar{k}} - \frac{\mathbf{nS}^{-1}\mathbf{t}}{\mathbf{nS}^{-1}\mathbf{n}}. \quad (15e)$$

The $\chi'_{i\bar{k}}(\vec{r})$ in the above equations is the auxiliary basis set centered on the atoms in region \bar{K} , which contains $N_{\bar{K}}$ electrons. Although one may follow the protocol to compute the inter-region Coulombic integrals in both the QM and SE Fock matrices, we only applied it to treat the QM Fock matrix in the current implementation and simply used partial charges on the QM atoms (see below) to polarize the SE wave function. This is based on the consideration that the polarization effect of the SE region is less important than that for the QM region, and point charge polarization is sufficient. The fitting basis was taken to be a set of uncontracted Gaussian orbitals developed by Salahub *et al.*,⁴⁰ however, d functions are not included to avoid over fitting. The density matrix for the semiempirical region is first transformed into the STO-3G basis, $P_{\mu\nu}^{\text{STO}}$, such that the integrals in Eq. (15c) can be evaluated with the Gaussian orbital $\chi_{\mu}^{\text{STO}}(\vec{r}')$. The fitted density [Eq. (15a)] is then used to calculate $I_{\mu\nu}$ [Eq. (13)].

2. The FLYQ method (QM/SE-I-F)—beyond the fixed point charge model

A simpler solution to avoid calculating the inter-region two electron integrals is to consider the interaction with a multipole expansion of the electron density in region \bar{K} when considering the Fock matrix of region K ; this was adopted in the effective fragment potential approach.¹⁷ Only one-electron inter-region integrals are required, so that the method is faster than the auxiliary basis approach. In the current work, we truncate the multipole expansion at lowest order, and uses partial charges to represent the electron density distribution; Mulliken charges were chosen here for simplicity, although charges from other schemes, such as natural orbital analysis⁴¹ can also be used. Since Mulliken charges from semiempirical methods are usually too small in magnitude, we introduce a uniform scaling factor, f_Q , which has to be determined empirically. Although scaling can also be done for the QM Mulliken charges, this was not done because their magnitude seems to be reasonable in the models tested here. Because the partial charges fluctuate as the geometry of the system varies, the method can be considered as an *ab initio* version of the FLUQ approach,¹⁴ with the fluctuating

charges derived on the fly. We, therefore, term the approximation the “FLYQ” approach (on the “FLY” as derived charges). The current approach does not require any fitting parameters other than f_Q (and the QM/SE van der Waals parameters). A similar scheme at the semiempirical level has been reported recently by Gao,⁴² who proposed to represent the electron density with either a set of partial charges or multipoles.

With the FLYQ approximation, the description of mutual polarization between the QM and the SE region is achieved by *iteratively* polarizing the wave function of each part with the other described by *flexible* point charges. This is expected to give rise to improvements over the standard QM/MM approach in which the MM is described with *fixed* charges. The auxiliary basis approach (Sec. II A 1) introduces further improvements by computing the QM/SE interaction beyond the monopole level. The quantitative difference between these models has to be established with test calculations, which are presented in Sec. III.

B. ONIOM based QM/SE method (QM/SE-O): Charge transfer and exchange effects

In the preceding sections, only the electrostatic and polarization effects between the QM and SE region are treated explicitly. The exchange effect (Pauli repulsion) is replaced with van der Waals terms, and charge transfer effect is not included explicitly (although one may consider it as part of the effective van der Waals term). Therefore, the *iterative* method (QM/SE-I) is applicable to systems with weak interaction between the QM and SE regions. In cases where there are QM and SE atoms interact strongly (or even are covalently bonded), the charge transfer and exchange repulsion between the two regions can become important and should be accounted for. Even for relatively weak interacting systems, it is likely that a specific set of van der Waals parameters has to be fitted for each QM and SE combination, which is somewhat cumbersome for general applications. Given these considerations, we implemented the following approach inspired by the ONIOM method of Morokuma *et al.*,²⁴ which will be referred to as QM/SE-O methods (O as in “ONIOM”).

We start by considering a two-layer system partitioned into $A+B$; including a third MM layer is straightforward. The ONIOM energy for the system, for which regions A and B are described with method I and II, respectively, is approximated by²⁴

$$E_{\text{ONIOM}}^{A+B} = E_{\text{I}}^A - E_{\text{II}}^A + E_{\text{II}}^{A+B} \\ = \langle \Psi_{\text{I}}^A | \hat{H}^A | \Psi_{\text{I}}^A \rangle - \langle \Psi_{\text{II}}^A | \hat{H}^A | \Psi_{\text{II}}^A \rangle \\ + \langle \Psi_{\text{II}}^{A+B} | \hat{H}^{A+B} | \Psi_{\text{II}}^{A+B} \rangle, \quad (16)$$

where the superscript of the wave function indicates the region, and the subscript denotes the level of calculation. As noted by the original authors, the QM region atoms do not interact explicitly with the environment at the QM level, and the environmental effect is fully described at the low (SE) level.²⁴ This simplification has caused substantial concern,^{26,27} and recent developments attempt to improve the

description by explicitly polarizing the QM wave function via an embedding potential²⁶ or allowing electrons to flow between the QM and SE region according to a chemical potential.²⁷ Although the later scheme is of interest, the fact that high level *ab initio* or DFT are so different from SE in terms of the orbital eigenvalues and qualities led us not to pursue the idea. We continue to follow the philosophy that no explicit basis mixture is allowed between the two regions, but allow the QM wave function to be explicitly polarized by the SE atoms. This leads to the following modification of the ONIOM approach. The total energy of the QM/SE system is first approximated by

$$\begin{aligned} E_1^{A+B} &= \langle \Psi_1^{A+B} | \hat{H}^A + \hat{H}^{A/B} + \hat{H}^B | \Psi_1^{A+B} \rangle \\ &\sim \langle \Psi_1^A \Psi_1^B | \hat{H}^A + \hat{H}^{A/B} + \hat{H}^B | \Psi_1^A \Psi_1^B \rangle \\ &= \langle \Psi_1^A | [\hat{H}^A + \langle \Psi_1^B | \hat{H}^{A/B} | \Psi_1^B \rangle] | \Psi_1^A \rangle \\ &\quad + \langle \Psi_1^B | \hat{H}^B | \Psi_1^B \rangle \\ &= \tilde{E}_1^{A+B}, \end{aligned} \quad (17)$$

where the formally “exact” wave function for the total system given in the first expression is assumed to be decoupled into $\Psi_1^A \cdot \Psi_1^B$ in the second step. A corresponding expression (\tilde{E}_1^{A+B}) can be written for the energy of $A+B$ at the SE level, E_{II}^{A+B} . We formally add E_{II}^{A+B} to and then subtract \tilde{E}_{II}^{A+B} from Eq. (17) to obtain

$$\begin{aligned} \tilde{E}_1^{A+B} &= \langle \Psi_1^A | [\hat{H}^A + \langle \Psi_1^B | \hat{H}^{A/B} | \Psi_1^B \rangle] | \Psi_1^A \rangle \\ &\quad + \langle \Psi_1^B | \hat{H}^B | \Psi_1^B \rangle + E_{II}^{A+B} \\ &\quad - \langle \Psi_{II}^A | [\hat{H}^A + \langle \Psi_{II}^B | \hat{H}^{A/B} | \Psi_{II}^B \rangle] | \Psi_{II}^A \rangle \\ &\quad - \langle \Psi_{II}^B | \hat{H}^B | \Psi_{II}^B \rangle. \end{aligned} \quad (18)$$

Assuming that region B can be well described at the SE (II) level (i.e., $\langle \Psi_1^B | \hat{H}^B | \Psi_1^B \rangle - \langle \Psi_{II}^B | \hat{H}^B | \Psi_{II}^B \rangle$) is nearly constant, we arrive at the expression for the modified ONIOM (QM/SE-O) energy,

$$\begin{aligned} \tilde{E}_1^{A+B} &\approx \langle \Psi_1^A | [\hat{H}^A + \langle \Psi_{II}^B | \hat{H}^{A/B} | \Psi_{II}^B \rangle] | \Psi_1^A \rangle \\ &\quad + E_{II}^{A+B} - \langle \Psi_{II}^A | [\hat{H}^A + \langle \Psi_{II}^B | \hat{H}^{A/B} | \Psi_{II}^B \rangle] | \Psi_{II}^A \rangle \\ &= E_{QM/SE-O}^{A+B}. \end{aligned} \quad (19)$$

Comparing Eq. (19) with Eq. (13), it is clear that the difference is that here the wave function of the QM region is polarized also at the high level (I). We note that to avoid double counting, the energy for the QM region at the SE level, also polarized by the environment, has to be subtracted. Therefore, Eq. (19) includes the difference in the QM and SE calculations in terms of the response of the wave function in the QM region to the environmental (SE) electrostatic polarization, while the original ONIOM represents the electrostatic and polarization effects only at the SE level. Thus, when the SE level is not satisfactory for describing these effects, the current QM/SE-O approach is expected to give better results than the original ONIOM formulation. The current scheme and the original ONIOM both describe ex-

change repulsion and charge transfer effects at the SE level; therefore, no specific van der Waals parameters have to be fitted for different QM and SE combinations.

The term that represents the electrostatic interaction and polarization between QM and SE atoms, $\hat{V}^{pol} = \langle \Psi_{II}^B | \hat{H}^{A/B} | \Psi_{II}^B \rangle$, is similar in nature to the embedding potential in other developments.²⁶ It can be evaluated at different levels, such as the scheme described for QM/SE-I in Sec. II A. Since it has to be evaluated at both the QM and SE levels, we took the simplest choice for the current implementation and used scaled Mulliken charge on the SE atoms from a calculation at the SE level for the $A+B$ region. With such a choice, the “embedding potential” is calculated only once and no iteration has to be performed. One short coming is that the Mulliken charges thus obtained do not sum exactly to the charge of region B . This could cause problems when there is substantial charge transfer between the QM and SE regions; no such cases were found in the current tests. More sophisticated schemes to construct the embedding term will be studied in the future.

Finally, it is straightforward to derive the expression for the QM/SE-O total energy for a three layer system described at the QM (I), SE (II), and MM level, respectively. The resulting expression is

$$\begin{aligned} E_{QM/SE-O}^{Tot} &\approx \langle \Psi_I^{QM} | [\hat{H}^{QM} + \langle \Psi_{II}^B | \hat{H}^{QM/SE} | \Psi_{II}^B \rangle \\ &\quad + \hat{H}^{QM/MM}] | \Psi_I^{QM} \rangle - \langle \Psi_{II}^{QM} | [\hat{H}^{QM} \\ &\quad + \langle \Psi_{II}^B | \hat{H}^{QM/SE} | \Psi_{II}^B \rangle + \hat{H}^{QM/MM}] | \Psi_{II}^{QM} \rangle \\ &\quad + \langle \Psi_{II}^{QM+SE} | \hat{H}^{QM+SE} \\ &\quad + \hat{H}^{(QM+SE)/MM} | \Psi_{II}^{QM+SE} \rangle + E^{MM}. \end{aligned} \quad (20)$$

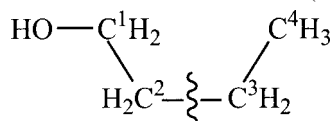
C. Implementations

The algorithms described above were implemented in the program CHARMM, which is interfaced with GAMESS and DeFT;⁴³ the latter programs allow one to use the *ab initio* methods (Hartree–Fock, MP2, CASSCF, CI, etc.) and density functional theories, respectively. The DeFT code was chosen because it is highly paralleled, and employs the charge-density fitting algorithm required for the auxiliary basis scheme in QM/SE-I. The available SE methods in CHARMM include MNDO, AM1, and SCC–DFTB;³⁶ the latter method was recently introduced into CHARMM,⁴⁴ and has been shown to give superior results for proton transfer reactions and metal ions than AM1 and PM3.^{36,44}

There are a number of parameters that can be optimized in the different QM/SE methods. In QM/SE-I, the Pauli repulsion and charge transfer effects are not included explicitly, and therefore requires a set of van der Waals parameters; it is likely that different sets would be optimal for different QM/SE combinations as in different QM/MM calculations.⁴⁵ Further, partial charges are used to polarize the SE atoms in QM/SE-I-A and to polarize both QM and SE atoms in QM/SE-I-F. One has the freedom to choose the optimal set of partial charges such as natural orbital charges, ESP charges, or scaled Mulliken charges; the last were chosen here for simplicity. In QM/SE-O, no van der Waals parameters are

needed because the Pauli repulsion and charge transfer effects are taken into account at the SE level. In the evaluation of the embedding interaction terms, however, specific parameters such as a scaling factor or charge normalization factor (see Sec. II A) for the Mulliken charge would be needed.

When the QM and SE regions are covalently bonded to MM regions, the interface can be treated with the standard link atom approach.^{1(b)} If the QM and SE regions are bonded together, a double link atom approach⁴⁶ is the most straightforward scheme to treat the boundary; i.e., link atoms are added to both QM and SE boundary atoms. Care is necessary to avoid over polarization of the boundary QM or SE atoms, as in the QM/MM method analyzed previously.⁹ For example, the partial charge or the density fitting coefficients on the boundary SE or QM atoms can be excluded in the QM/SE interaction, similar to the exclusion convention used in CHARMM for QM/MM.⁹ It is expected that the boundary issue is more delicate for QM/SE than QM/MM, and the best scheme has to be determined through test calculations. We emphasize that the QM/SE-I approach is mostly suited for weak intermolecular interactions between QM and SE regions, and the QM/SE-O approach is more appropriate for cases where the partition is done across a covalent bond. Consider the simple case of butanol, in which a partition is done between C² and C³ atoms (scheme I). Note



Scheme I

that in the QM/MM calculations, the bonded terms between the two regions such as the C²-C³ stretch and C⁴-C³-C² bending are included, which helps to maintain the overall structural stability. Since no such terms are included in QM/SE-I, the butanol molecule would be described by separated ethane and ethanol molecules, and thus would not be constrained to have the correct geometry. In the QM/SE-O scheme, however, there is a contribution corresponding to the entire butanol at the SE level [see Eq. (19)], and therefore the structure is expected to be qualitatively correct. In the current work, we restrict ourselves to systems where the QM and SE regions are not covalently bonded, and leave the more complicated cases to future studies.

III. TEST EXAMPLES

In this section, we illustrate the algorithms described in Sec. II by a number of applications. We determine how the different QM/SE methods compare to full QM calculations, and how sensitive the results are to the choice of the involved parameters. Also we examine whether the QM/SE methods give significant improvements over QM/MM calculations. Finally, the QM/SE methods proposed here are compared to determine which of them are most robust. The first set of models involves the interaction between small polar or charged molecules with water. The second set deals with the effect of polar molecules on the proton abstraction from the C α position of formaldehyde by a base (HCOO⁻); this is inspired by the first reaction step catalyzed by triosephos-

TABLE I. Water-Cl⁻ ion complexes calculated with different methods.^a

Method ^b	Cl ⁻		
	ΔE	$R_{\text{Cl-H}} (\text{\AA})$	$R_{\text{Cl-H}'} (\text{\AA})$
Full QM	-15.3	2.128	3.283
Full SE	-13.8	1.987	3.022
QM/MM	-13.2	2.782	2.782
SE/MM	-13.6	2.765	2.765
QM/SE-I-F ^c	-6.1	2.803	3.062
	(-7.9)	(2.688)	(3.160)
	[-10.9]	[2.561]	[3.239]
QM/SE-I-A	-13.5	2.116	3.215
QM/SE-O-0	-13.8	1.987	3.022
QM/SE-O-M ^c	-12.5	2.107	2.945
	(-12.4)	(2.132)	(2.920)
	[-12.3]	[2.174]	[2.874]

^aThe QM is BPW86 with 6-31+G** basis set, SE is AM1, MM is the TIP3P model for water. In the QM/MM(SE) calculations, the Cl ion was treated as QM and the water was described with MM (SE).

^bThe QM/SE-I indicates the iterative QM/SE algorithms described in Sec. II A; the QM/SE-O indicates the ONIOM based QM/SE methods described in Sec. II B. The suffix "-F" indicates that Mulliken charges from the SE region is used to polarize the QM wave function, "-A" means that auxiliary basis functions were used to fit the SE electron density which interacts with the QM wave function (see text). The QM/SE-O-0 method is the same ONIOM approach originally proposed by Morokuma and co-workers, where no (thus "0") explicit interaction was evaluated between QM and SE atoms (Ref. 19).

^cThe numbers without and with parentheses were obtained with a scaling factor (see text) for the Mulliken charge of 1.0 and 1.2, respectively; those with brackets were obtained with a scaling factor of 1.5.

phate isomerase,⁴⁷ which has been studied extensively by theoretical methods.⁴⁸ Finally, the effect of a hydrogen-bond network on the same proton transfer reaction is studied. This is based on the proposal of Guo *et al.*¹⁰ that cooperativity effects due to extensive hydrogen-bond network may play a role in enzyme catalysis. The relatively small sizes of those models allow us to perform full QM calculations, to which the QM/MM and QM/SE results can be compared. A DFT method with medium sized basis sets was used as QM for comparison, although more elaborate methods are likely to be explored in actual applications. The BP86 functional⁴⁹ was employed because it is the only nonlocal functional available in the DeFT program that we used.

A. Interaction between water and polar molecules

The first set of tests concerns the interaction between water and several polar or charge water molecules. These systems are analogous to studies in the validation of QM/MM methods with SE^{1(b)} and DFT⁵⁰ QM methods. The major purpose is to test the robustness of the proposed QM/SE methods, especially the schemes for calculating QM/SE interactions and the difference between the iterative and noniterative methods.

The first system is the complex between a Cl⁻ ion and a water molecule. This was chosen because it was noticed in previous studies⁵⁰ that although full QM calculations gave an asymmetric optimal structure (i.e., the two Cl⁻···H distances are different, see Table I), QM/MM calculations in which the Cl⁻ was treated with QM and the water was treated with MM gave a symmetric structure (see Table I). The result was

rationalized by the fact that the MM water (TIP3P) is not polarizable, and the two hydrogen atoms have the same strength of interaction which led to the symmetric structure. When the water is polarized, the asymmetric structure is favored because the electrostatic interaction between Cl^- and one hydrogen atom can be maximized.

The binding energy and the optimal $\text{Cl}^- \cdots \text{H}$ distances at different levels are shown in Table I; the Cl^- was treated by BP86/6-31+G(*d,p*),⁵¹ and the water was treated by either AM1 or TIP3P. Although the van der Waals parameters can be optimized, no such attempt was made here for the purpose of exploring the behavior of the methods. It is encouraging that the asymmetric structure was correctly reproduced at all the QM/SE levels. When the polarization was modeled by the simple Mulliken charge in the iterative scheme (QM/SE-I-F), the result depends sensitively on the scaling factor (f_Q). Without scaling ($f_Q=1.0$), the Mulliken charges on the water atoms are very small; i.e., the charge on the hydrogen is about 0.2, compared to the value of +0.417 in the TIP3P model and the value of +0.35 from a Mulliken analysis at the full QM level. With the small charge, the interaction energy at the QM/SE-I-F level is substantially underestimated (-6.1 kcal/mol vs -15.3 kcal/mol from full QM calculations), and the structure is only slightly different from being symmetric (the $\text{Cl}^- \cdots \text{H}$ distances are 2.803 and 3.062 Å, respectively, compared to the full QM values of 2.128 and 3.283 Å, respectively). As expected, the agreement with the full QM results becomes much better when the AM1 Mulliken charges are scaled. For example, with $f_Q=1.5$ (e.g., a value of 2.13 was found to be appropriate in Ref. 42), the interaction energy increases to -10.9 kcal/mol, and the short $\text{Cl}^- \cdots \text{H}$ distance drops to 2.561 Å; better agreements can be obtained if f_Q is optimized (e.g., to obtain exact agreement with the QM binding energy, $f_Q=1.95$ is required, which gives a $\text{Cl}^- \cdots \text{H}$ distance of 2.356 Å). In contrast to the FLYQ approach, the auxiliary basis set approach (QM/SE-I-A) works rather well without the need of adjusting parameters. The interaction energy was calculated to be -13.5 kcal/mol, and the $\text{Cl}^- \cdots \text{H}$ distances are 2.116 and 3.215 Å, all in good agreement with the full QM results. This indicates that the density-fitting scheme is a robust approach to describe the QM/SE interactions at hydrogen-bonding distances (but see below).

With the original ONIOM model,²⁴ the QM/SE-O-0 (“0” is used here to emphasize that no explicit interaction between the QM and SE atoms was included at the QM level) gave exactly the same results as full SE (AM1) for the particular case. This is true because both QM (BP86) and SE (AM1) gave zero gradient for a free Cl^- atom. With the embedding interaction included, the results become different from full AM1 values. This is due to the fact that the current scheme includes a term reflecting the difference in the response of the QM (BP86) and SE (AM1) wave functions to external electrostatic interactions and polarization (see Sec. II B). An interesting observation is that the QM/SE-O-F results are much less sensitive to the Mulliken scaling factor than QM/SE-I-F because both the QM (BP86) and SE (AM1) are polarized by the same (scaled) Mulliken charges. The results, however, are not significantly different from the

original ONIOM values; in fact the binding energies are somewhat further away from the full QM values. The observation that the binding energy decreases and the shorter $\text{Cl}^- \cdots \text{H}$ distance increases as the scaling factor increases indicates that the BP86 wave function (density) interact less strongly than AM1 wave function with the external charge. This is consistent with the result that BP86/MM calculation predicts a slightly weaker binding energy compared to AM1/MM (Table I) with the same water model (Table I).

To further test the robustness of the density-fitting scheme, we studied a number of complexes between water and molecules representing the fragments of amino acids. These include two neutral molecules, imidazole and N-methyl formamide (NMF), which model the side chain of histidine and the main-chain peptide unit, respectively, and two charged species, NH_4^+ and $\text{C}(\text{NH}_2)_3^+$, which model the side-chain of lysine and arginine, respectively. The interaction energy between these models and a water molecule was calculated at the full BP86/6-31G(*d,p*), BP86/CHARMM and BP86/AM1 level with the auxiliary basis function approach (QM/SE-A). Morokuma energy decomposition was carried out with HF at the full QM level to reveal different components; a full HF calculation was found to give very similar interaction energies as BP86 with the same basis set. In the hybrid calculations, the water molecule was treated with BP86, and the fragment was treated with SE. In all the models, the SE part is the hydrogen-bond donor and the QM part is the acceptor; this is motivated by the fact in most enzyme systems, protein residues serve as hydrogen-bond donors to polarize (or stabilize) charged substrates. In Fig. 2, the electrostatic and polarization components of the interaction from the full HF calculations according to the Morokuma energy decomposition scheme are shown as a function of the distance between the hydrogen-bond donor and acceptor. The corresponding components from BP86/MM and BP86/AM1 calculations, which are defined as the QM/MM (AM1) interaction without the van der Waals contribution, are also shown; the van der Waals term is not included because the dispersion energy is not described at the Hartree–Fock level. Clearly, there is striking agreement between the full QM and QM/SE calculations at all the distances considered here. The good agreement between QM/SE and full QM calculations is very encouraging, indicating that the SE density is reasonable for calculating hydrogen-bonding interactions. This is in accord with the previous observation that good hydrogen-bonding interaction energies can be obtained when the hydrogen-bond donor, such as water, is modeled by a charge density distribution fitted by a small set of basis functions.⁵² On the other hand, we realize that the density distribution in AM1 is far from optimal because the density itself was not directly involved in the fitting of AM1 parameters; in fact, many integrals were parametrized according to experimental values rather than calculated. Therefore, we expect that SE methods in which the density is well described, such as the SCC–DFTB approach, will work even better than AM1 in QM/SE-A calculations. Integration of DFT and SCC–DFTB in an iterative fashion with the auxiliary basis approach (Sec. II A 1) is currently underway. The QM/MM results are very different at short distances. The observation

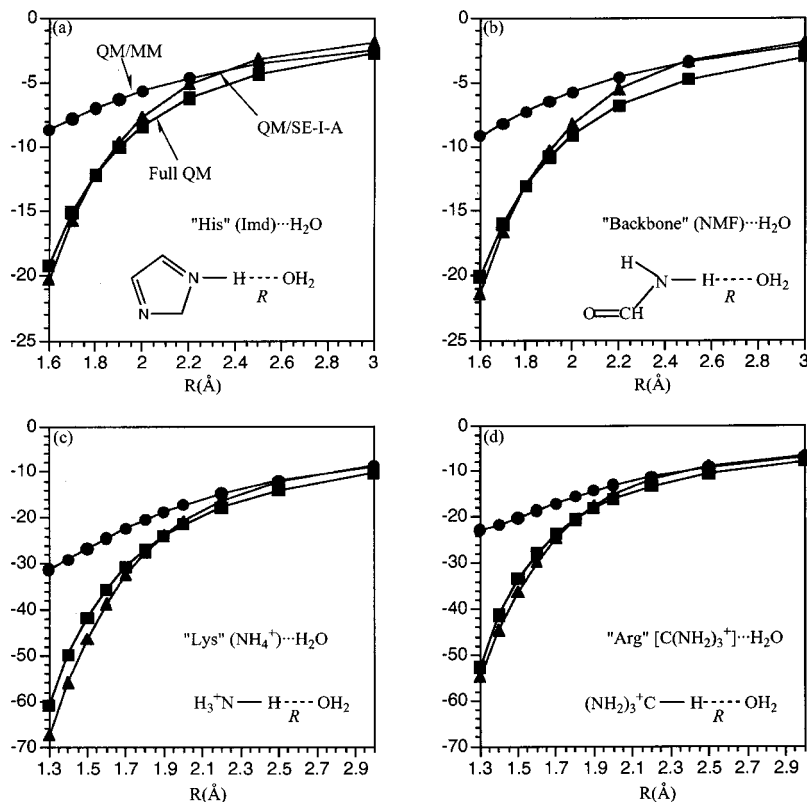


FIG. 2. The electrostatic and polarization energy (in kcal/mol) between a water molecule and a hydrogen-bond donor that models a unit in an amino acid [imidazole, N-methyl formamide, NH_4^+ and $\text{C}(\text{NH}_2)_3^+$] as a function of their separation. Such a quantity from a full QM calculation was obtained from a Morokuma energy decomposition analysis at the HF/6-31+G(*d,p*) level, which gives interaction energies close to BP86/6-31+G(*d,p*) calculations. The latter was used in the QM/MM and QM/SE-I-A calculations, for which the electrostatic and polarization energy is defined as the interaction between QM and SE (MM) atoms without the van der Waals contribution.

does not imply directly that QM/MM results for the interaction energy are poor, because the van der Waals contribution in the QM/MM calculations is empirical and does not necessarily reflect the full Pauli repulsion; in other words, error cancellation may lead to good interactions from QM/MM calculations. Finally, we note that because the electrostatic and polarization contributions in QM/SE-A calculations are very different from the QM/MM values, different van der Waals parameters are required in the two sets of calculations, especially at short distances.

B. The effect of polar molecules on a proton transfer reaction

As the second set of tests, we study the effect of small polar molecules on a proton transfer reaction inspired by triosephosphate isomerase (TIM).^{47,48} As shown in Fig. 3, the proton transfer reaction takes place between formic acid and formaldehyde, which model the catalytic glutamic acid (Glu 165) and substrate (dihydroxyacetone phosphate) in TIM, respectively. Several polar molecules hydrogen bonded to the carbonyl group in formaldehyde were studied. They are one and two water molecules, one NMF, and one water molecule plus one imidazole; the imidazole is chosen to mimic His 95 in TIM, and the others are selected for an analysis of polarization effects. For simplicity, the distance between the donor and acceptor in the proton transfer is fixed to be 3.0 Å. Geometry optimization was carried out for the reactant, product system in the presence of the constraint at the HF/6-31G(*d,p*) level, and the transition state was determined by an adiabatic mapping procedure at the same level; critical distances in the reactant, transition state and product

are shown in Fig. 3. Subsequently, single point energies were obtained by full QM (BP86/6-31+G(*d,p*)), BP86/CHARMM, and BP86/AM1 calculations; in the hybrid calculations, the environment (the added polar molecules) was treated with CHARMM or AM1 and the reacting groups were described with BP86/6-31+G(*d,p*). No specific van der Waals parameters were optimized for the MM or AM1 atoms, because our primary purpose is to examine the qualitative behavior of different QM/SE methods relative to QM/MM.

At the full QM level (see Table II), the effect of the environment increases in the order of one water, one NMF, two waters and one water plus one imidazole, and the product is stabilized more than the transition state. For example, one water molecule stabilizes the transition state and product by 3.7 and 6.8 kcal/mol, respectively, and the values for NMF are 5.0 and 9.2 kcal/mol, respectively. One imidazole and one water has a larger stabilizing effect compared to two water molecules, by ~ 2 kcal/mol for the transition state and ~ 4 kcal/mol for the product. These are consistent with the fact that the carbonyl group in the product is more negatively charged than in the transition state; imidazole and NMF have larger stabilizing effects than water because their dipole moments are better aligned with the lone pair of the carbonyl oxygen.

At the QM/MM level, the same trend was observed, although the magnitude of the increase in the stabilizing effect going from water to NMF and imidazole was reduced. For example, instead of a 2 kcal/mol increase in the stabilizing effect on the transition state for an imidazole compared to water, a value of 0.6 kcal/mol was obtained from single point QM/MM calculations. Better agreements might be obtained

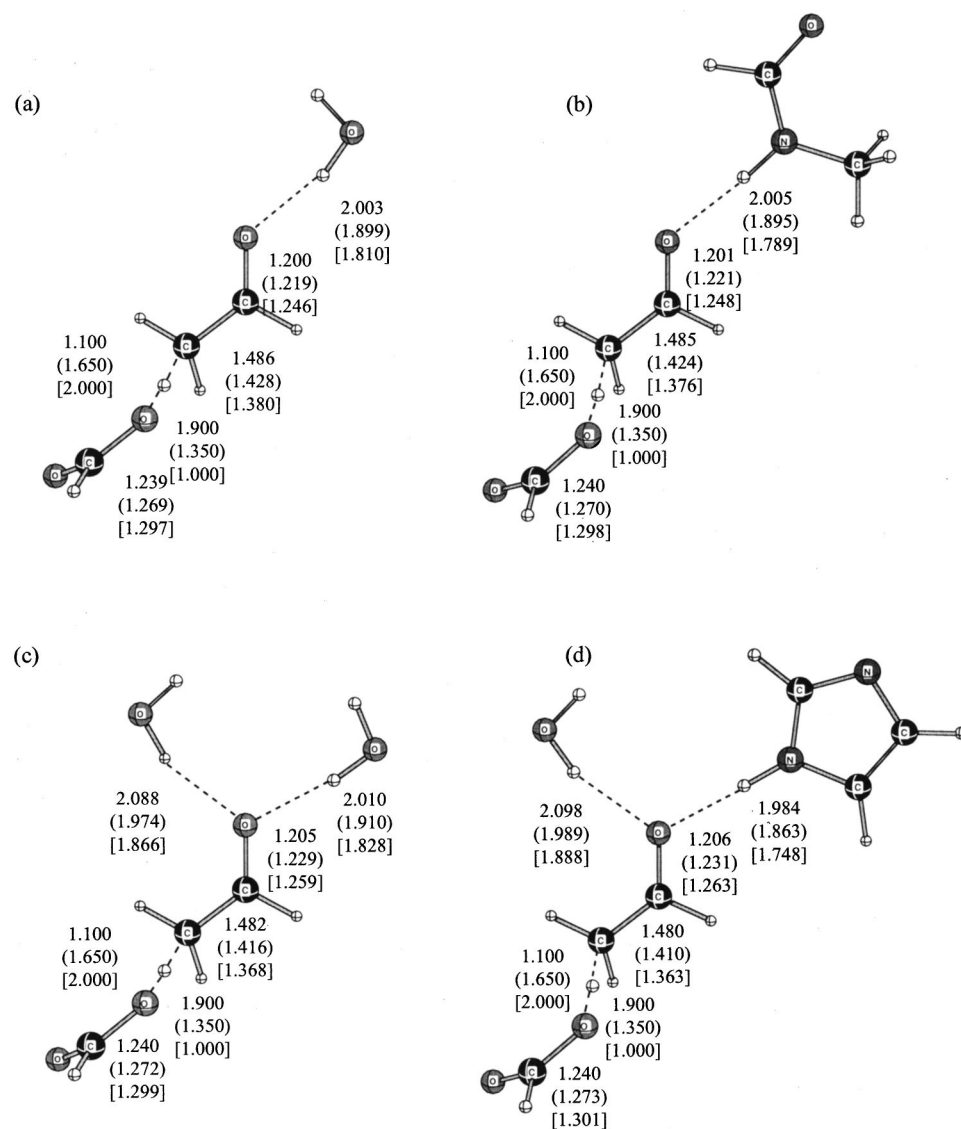


FIG. 3. Model systems used to study the effect of small polar molecules on a proton transfer reaction between formic acid and formaldehyde. The distance between the proton donor and acceptor was fixed to be 3.0 Å during the geometry optimizations, which were carried out at the HF/6-31+G(*d,p*) level. Critical distances in the reactant, transition state (in parentheses) and product (in brackets) are shown in Å. Selected Mulliken charges for the polar ligands are also shown for full QM [BP86/6-31+G(*d,p*)] and full SE single point calculation for the transition states to illustrate the difference between QM and SE Mulliken charges.

TABLE II. Comparison of full QM, QM/MM and different QM/SE algorithms for the effect of different molecules on the transition state and product energetics of a model proton transfer reaction.^a

Method ^b	H ₂ O		NMF		2×H ₂ O		Imd+H ₂ O		RMS error	
	TS	Prod	TS	Prod	TS	Prod	TS	Prod	TS	Prod
Full QM	-3.7	-6.8	-5.0	-9.2	-7.5	-12.5	-9.3	-16.9		
QM/MM	-2.8	-4.8	-3.3	-5.6	-6.1	-10.3	-6.7	-11.0	1.8	3.8
QM/SE-I-F	-2.3	-4.2	-5.0	-9.9	-5.0	-9.0	-8.6	-16.3	1.5	2.2
QM/SE-I-A ^b	-4.5	-9.0	-6.7	-14.2	-9.1	-17.7	-12.4	-25.0	2.0	5.5
	(-3.0)	(-5.6)	(-5.0)	(-10.2)	(-6.5)	(-11.5)	(-9.4)	(-17.5)	(0.6)	(1.0)
QM/SE-O-0 ^c	-1.6	-3.3	-3.1	-6.0	-4.6	-8.1	-6.7	-11.9	2.4	4.1
	(-2.9)	(-6.1)	(-4.6)	(-9.7)	(-5.7)	(-11.9)	(-7.7)	(-16.1)	(1.3)	(0.7)
QM/SE-O-M	-1.8	-3.1	-2.9	-5.4	-4.3	-7.1	-6.2	-10.7	2.6	4.9

^aThe QM is BPW86, SE is AM1, MM is CHARMM22 force field for proteins. The proton transfer fragments (formic acid and formaldehyde) were treated with QM, and the ligands were treated with either MM or SE in the QM/MM and QM/SE calculations, respectively. The values without ligands are 14.3 and 14.0 kcal/mol, for the TS and Prod, respectively. See footnote of Table I for notations for the different QM/SE methods, and Fig. 3 for the structures, which were optimized with the HF/6-31+G(*d,p*) method (see text). In the QM/SE-I-F and QM/SE-O-M calculations, a scaling factor of 1.2 was used for the SE Mulliken charges.

^bThe values in parentheses were obtained by considering the fact that the QM/SE van der Waals contribution based on the standard CHARMM van der Waals parameters is substantially underestimated, and an approximate correction was made based on the difference in the electrostatic and polarization components in QM/SE-I-A and QM/MM calculations for small molecules and water (Fig. 2).

^cThe values in parentheses were obtained with SCC-DFTB as the semiempirical level.

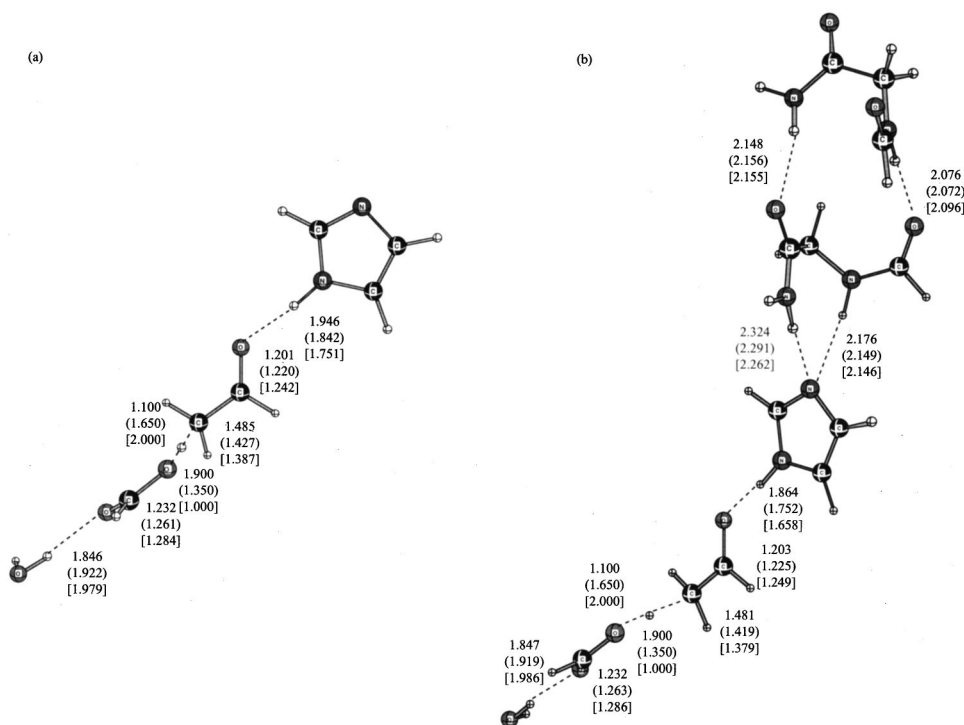


FIG. 4. Model system used to illustrate the performance of QM/SE methods in the presence of long-range non-additive polarization effects. See Fig. 3 for notations.

when the geometry is allowed to relax in the QM/MM calculations because it has been observed that shorter hydrogen bond distances (compared to full QM) are often required to give reliable QM/CHARMM interaction energies, due in part to the parametrization philosophy of the CHARMM force field to represent the situation in the condensed phase.⁵³

At QM/SE levels, the same qualitative trend was obtained although the quantitative results depend significantly on how the QM and SE levels are combined. At the QM/SE-I-M level, a scaling factor of 1.2 was found to give good results for the effect of NMF and imidazole (see Table II); the effect of water with the same scaling factor underestimates the magnitude of interaction substantially because AM1 gives a very small charge on the water molecules (see Fig. 3 for a comparison of QM and SE Mulliken charges of the ligands from full QM and SE calculations). At the QM/SE-I-A level, the effect of the environment was found to be overestimated, and the degree of overestimation is larger when the interaction is stronger. For example, the stabilizing effect of one water on the product is 6.8 and 9.0 kcal/mol at the full QM and QM/SE-I-A level, respectively, and the values for one water plus one imidazole are 16.9 and 25.0 kcal/mol, respectively. These results can be understood when one realizes that the van der Waals contribution in QM/SE-I-A is expected to be very different from QM/MM or full MM calculations, as discussed in the preceding section. When the interaction between the ligand and reacting group is strong, the hydrogen-bond distance becomes very short; e.g., the hydrogen bond between the imidazole and the carbonyl oxygen in the product is only 1.75 Å. According to the calculations in Sec. III A, the van der Waals contribution in QM/SE-I-A is expected to be underestimated by as much as 7 kcal/mol at such a distance (see Fig. 2). Introducing this as an approximate correction to the van der Waals QM/SE-I-A contribu-

tions, the agreement with full QM becomes much better (see Table II).

Using the ONIOM approach, the results are not encouraging with both the original scheme of Morokuma *et al.* (RMS error is 2.4 and 4.1 kcal/mol for the TS and product, respectively) and the one proposed here which includes explicit polarization of the QM part (RMS error is 2.6 and 4.9 kcal/mol for the TS and product, respectively); the effects are typically worse compared to QM/MM or QM/SE-I results. This observation reflects the fact that AM1 is rather poor at representing the interaction from the environment, and a better semiempirical approach has to be used. Indeed, very good results (RMS error is 1.3 and 0.7 kcal/mol for the TS and product, respectively) were obtained when SCC-DFTB was used in the ONIOM calculations even when the QM region is not explicitly polarized.

C. The effect of hydrogen-bond network on a proton transfer reaction

Although hydrogen-bonding interactions of medium strength are dominated by electrostatic interactions, nonadditive effects have been shown to be involved in strong hydrogen-bonding interactions.⁵⁴ As discussed by Guo and Salahub,¹⁰ and by Luo *et al.*,⁵⁵ the extensive hydrogen-bonding networks found in the active site of many enzymes can induce such nonadditive effects. Since the effects are substantially larger when the hydrogen-bond acceptor is charged compared to when it is neutral, it was proposed that *differential* cooperative effects can play a role in enzyme catalysis where the charge of the substrate changes during the reaction. Because the nonadditive effects are quantum mechanical in nature, we expect that a hybrid calculation in which the hydrogen-bond network is represented by SE,

TABLE III. Comparison of full QM, QM/MM and different QM/SE algorithms for the model proton transfer system plus a model di-peptide.^a

Method ^b	Model + His/H ₂ O		Di-peptide		Effect of hydrogen bond network	
	TS	Prod	TS	Prod	TS	Prod
Full QM	11.4	9.6	8.3	4.7	-3.1	-4.9
QM/MM	13.3	11.7	11.9	10.6	-1.4	-1.2
QM/SE-I-F	10.6	6.0	8.2	2.4	-2.4	-3.6
QM/SE-I-A ^b	11.3 (11.6)	5.4 (7.2)	8.3 (9.6)	1.6 (4.6)	-3.0 (-2.0)	-3.8 (-2.6)
QM/SE-O-0 ^c	12.5 (12.8)	9.7 (11.9)	10.9 (10.3)	6.9 (7.9)	-1.6 (-2.5)	-2.7 (-4.0)
QM/SE-O-M	12.4	9.5	11.1	7.8	-1.3	-1.7

^aSee footnote of Table I for the notations for the different QM/SE methods. See Fig. 4 for the geometries, which were optimized with HF/6-31+G(*d,p*). In the QM/SE-I-F and QM/SE-O-M calculations, a scaling factor of 1.2 was used for the SE Mulliken charges.

^bThe values in parentheses were obtained by considering the fact that the QM/SE van der Waals contribution based on the standard CHARMM van der Waals parameters is substantially underestimated, and an approximate correction was made based on the difference in the electrostatic and polarization components in QM/SE-I-A and QM/MM calculations for small molecules and water (Fig. 2).

^cThe values in parentheses were obtained with SCC-DFTB as the semiempirical level.

rather than by MM, will give a better description in this case. The proton transfer model shown in Fig. 4 was constructed to examine this behavior. This model differs from those discussed in Sec. III B in that the environment is replaced by an extended hydrogen-bond network consisting of one imidazole molecule and two diglycine peptides, which represent the backbone of a half-turn “helix;” a water molecule was added on the proton acceptor end to stabilize the charged reactant. Similar to Sec. III B, the geometries were optimized at the HF/6-31+G(*d,p*) level with the distance between the proton donor and acceptor fixed at 3.0 Å. Single point energies were then calculated at the BP86/6-31+G(*d,p*), BP86/CHARMM, and BP86/AM1 levels.

At the full QM level, the barrier for the proton transfer is substantially lower when the di-glycine is present to form an extended hydrogen-bond network; the values are 8.3 and 11.4 kcal/mol with and without the di-glycine, respectively. The distance between the imidazole and the carbonyl group in the reacting group is also much shorter in the presence of the di-glycine; in the transition state, the distance is 1.752 and 1.842 Å with and without the di-glycine, respectively. The product is also further stabilized by the di-glycine, despite the fact that the latter is 6.7 Å (shortest distance) from the reacting center (the C β in the formaldehyde); the product is 4.7 and 9.6 kcal/mol above the reactant with and without the di-glycine, respectively. In other words, the stabilizing effect of the di-glycine is 3.1 and 4.9 kcal/mol for the transition state and product, respectively.

The QM/MM single point energetics are discouraging; although a lower barrier (by 1.4 kcal/mol) was reproduced in the presence of the di-glycine, the larger stabilizing effect for the product was not found (see Table III) and the error in the endothermicity of the proton transfer was nearly 6 kcal/mol. As discussed in Sec. III B, the error might be reduced when geometry optimization is allowed. The current results do indicate, however, that QM/MM calculations are not fully satisfactory for such nonadditive effects.

At the QM/SE levels, the qualitative trends in the stabilizing effect are well reproduced. When the SE region is represented by the (scaled) Mulliken charges, as in QM/SE-I-M, a scaling factor of 1.2 was found to give satisfactory

results, as in the pair-wise type of models discussed in Sec. III B. When the SE density is represented by a set of auxiliary basis functions, as in QM/SE-I-A, the effects of the diglycine on the transition state and product are also described well. The absolute endothermicities of the reactions are underestimated. This is related to the fact that the van der Waals contribution has not been optimized for QM/SE and large errors can arise at short hydrogen bond distances between QM and SE atoms, as also discussed in the preceding section. With the ONIOM approaches, the results are also satisfactory, especially when the environment was treated with SCC-DFTB, rather than AM1.

IV. CONCLUDING DISCUSSIONS

To improve the description for the interaction between an active region in large systems, such as an enzyme, and its immediate environment over that obtained with the standard QM/MM methods,¹ we have proposed several algorithms for treating the active part with a high level QM (*ab initio* or DFT) approach and the environment with a less sophisticated semiempirical (SE) approach. In the iterative QM/SE schemes (QM/SE-I), the electrostatic interaction and polarization effects are introduced *explicitly* for both the QM and SE atoms by a self-consistent procedure. In the noniterative QM/SE scheme based on the ONIOM model (QM/SE-O), the exchange (Pauli repulsion), and charge transfer effects are taken into account at the SE level, in addition to the explicit electrostatic interaction and polarization between the two regions.

From the test calculations, it is evident that the performance of QM/SE calculations are encouraging, and that the quantitative accuracy of the results depend on several parameters; they are the charge-scaling/normalization factors (see Sec. II A) and the QM/SE van der Waals parameters. When the SE region is represented by the Mulliken charges that polarize the QM atoms, the iterative QM/SE-I-M is sensitive to the scaling factor because the SE Mulliken charges are in general too small (at least for the cases tested here, see Fig. 3). In the modified ONIOM approach, because the QM atoms are polarized at both the QM and SE levels and only a

differential effect is considered, the results are less sensitive to the scaling factor. The auxiliary basis function approach (QM/SE-I-A) was found to be satisfactory for the description of hydrogen-bonding donor groups in all the test cases examined, including both neutral and charged species. This is very encouraging considering the fact that density is not optimized in the standard SE approaches, such as AM1, and suggests that semiempirical approaches in which electron density is considered in the parametrization, such as the SCC-DFTB method,^{36,44} are well-suited for QM/SE-I-A type of calculations.

A practical issue related to the general application of QM/SE-I-M(A) approaches is that the van der Waals parameters have to be fitted to give reliable geometries and energetics for each level of implementation. This is more important for QM/SE-I-A, because the van der Waals contribution is closer to that in the full QM calculation, which is very different from the typical values in MM or QM/MM calculations. In other words, although using the standard MM van der Waals parameters in QM/MM calculations can give meaningful results, very poor values result in QM/SE calculations.

In the current studies, no dramatic improvements in the ONIOM results were found when the QM region is explicitly polarized. This is because the major part of the environmental effects are described at the SE level, and explicit polarization of the QM region only provides a correction to the SE treatment of the QM atoms and their response to the electrostatic interaction and polarization. Therefore, the quality of the SE description plays a more important role in the ONIOM results, as illustrated by the observation that BP86/SCC-DFTB, in general, gives significantly better results than BP86/AM1. Finally, we note that although self-consistency in the iterative schemes is usually achieved in less than six or seven cycles, the iterative procedure makes them more expensive (by a factor of N , where N is the number of iterations) than the ONIOM based approach, which does not require iteration in the current approximation for the embedding polarization.

¹(a) A. Warshel and M. Karplus, *J. Am. Chem. Soc.* **94**, 5612 (1972); (b) M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700 (1990); (c) J. Gao, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1995), Vol. 7, p. 119; (d) U. C. Sighn and P. A. Kollman, *J. Comput. Chem.* **7**, 718 (1996); (e) D. Bakowies and W. Thiel, *J. Phys. Chem.* **100**, 10580 (1996).

²F. Maseras and K. Morokuma, *J. Comput. Chem.* **16**, 1170 (1995); T. Matsubara, F. Maseras, N. Koga, and K. Morokuma, *J. Phys. Chem.* **100**, 2573 (1996).

³See, for example, P. A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe, and M. Karplus, *Biochemistry* **30**, 5826 (1991); P. A. Bash, M. J. Field, and M. Karplus, *J. Am. Chem. Soc.* **109**, 8092 (1987); J. Gao, *Acc. Chem. Res.* **29**, 298 (1996); M. A. Thompson and G. K. Schenter, *J. Phys. Chem.* **99**, 6374 (1995); R. D. Froese, S. Humbel, M. Svensson, and K. Morokuma, *J. Phys. Chem. A* **101**, 227 (1997); S. Humbel, S. Sieber, and K. Morokuma, *J. Chem. Phys.* **105**, 1959 (1996).

⁴V. Gogonea, D. Suarez, A. Van der Vaart, and K. M. Merz, Jr., *Curr. Opin. Struct. Biol.* **11**, 217 (2001).

⁵E. V. Stefanovich and T. N. Truong, *J. Phys. Chem. B* **102**, 3018 (1998).

⁶Q. Cui and M. Karplus, *J. Chem. Phys.* **112**, 1133 (2000).

⁷Q. Cui, M. Karplus, *J. Phys. Chem. B* **104**, 3721 (2000).

⁸Y. Zhang, T. Lee, and W. Yang, *J. Chem. Phys.* **110**, 46 (1999); J. Gao, P. Amara, C. Alhambra, and M. Field, *J. Phys. Chem. A* **102**, 4714 (1998); S.

Antonczak, G. Monard, M. F. Ruiz-Lopez, and J.-L. Rivail, *J. Am. Chem. Soc.* **120**, 8825 (1998).

⁹N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, *J. Phys. Chem. A* **104**, 1720 (2000).

¹⁰H. Guo and D. R. Salahub, *Angew. Chem. Int. Ed. Engl.* **37**, 2985 (1998).

¹¹J. Gao, *J. Comput. Chem.* **18**, 1062 (1997); J. Gao and K. Byun, *Theor. Chem. Acc.* **96**, 151 (1997).

¹²M. A. Thompson, *J. Phys. Chem.* **100**, 14492 (1996).

¹³D. York and W. Yang, *J. Chem. Phys.* **104**, 159 (1996).

¹⁴S. W. Rick *et al.*, *J. Chem. Phys.* **101**, 6141 (1994); **108**, 4739 (1998); H. A. Stern *et al.*, *ibid.* **115**, 2237 (2001).

¹⁵H. A. Stern, G. A. Kaminski, J. L. Bans, R. H. Zhou, B. J. Berne, and R. A. Friesner, *J. Phys. Chem. B* **103**, 4730 (1999).

¹⁶M. J. Field, *Mol. Phys.* **91**, 835 (1997).

¹⁷P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. R. Garmer, H. Basch, and D. Cohen, *J. Chem. Phys.* **105**, 1968 (1996).

¹⁸J. H. Jensen and M. S. Gordon, *J. Chem. Phys.* **108**, 4772 (1998); *Mol. Phys.* **89**, 1313 (1996).

¹⁹J. H. Jensen, *J. Chem. Phys.* **114**, 8775 (2001).

²⁰J. H. Jensen, P. N. Day, M. S. Gordon, H. Basch, D. Cohen, D. R. Garmer, M. Krauss, and W. J. Stevens, in *Modeling the Hydrogen Bond*, edited by D. A. Smith, ACS Symposium Series 569 (American Chemical Society, Washington, DC, 1994), Chap. 9.

²¹S. E. Worthington, A. E. Roitberg, and M. Krauss, *J. Phys. Chem. B* **105**, 7087 (2001).

²²R. M. Minikis, V. Kairys, and J. H. Jensen, *J. Phys. Chem. A* **105**, 3829 (2001).

²³M. Krauss and B. D. Wladkowski, *Int. J. Quantum Chem.* **69**, 11 (1998).

²⁴M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *J. Phys. Chem.* **100**, 19357 (1997).

²⁵T. Vreven and K. Morokuma, *J. Comput. Chem.* **21**, 1419 (2000).

²⁶N. Govind, Y. A. Wang, and E. A. Carter, *J. Chem. Phys.* **110**, 7677 (1999).

²⁷V. Gogonea, L. M. Westerhoff, and K. M. Merz, Jr., *J. Chem. Phys.* **113**, 5604 (2000).

²⁸W. Yang, *Phys. Rev. Lett.* **66**, 1438 (1991).

²⁹Y. Mo and S. D. Peyerimhoff, *J. Chem. Phys.* **109**, 1687 (1998).

³⁰Y. Mo and J. Gao, *J. Phys. Chem. A* **104**, 3012 (2000).

³¹Y. Mo, J. Gao, and S. D. Peyerimhoff, *J. Chem. Phys.* **112**, 5530 (2000); Y. Mo and J. Gao, *J. Phys. Chem.* **105**, 6530 (2001).

³²Y. Mo and J. Gao, *J. Comput. Chem.* **21**, 1458 (2000).

³³M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.* **99**, 4899 (1977).

³⁴T. A. Wesolowski, *J. Chem. Phys.* **106**, 8516 (1997).

³⁵W. Yang and R. Parr, *Density Function Theory for Atoms and Molecules* (Oxford University Press, New York, 1989).

³⁶M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* **58**, 7260 (1998); D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, *ibid.* **51**, 12947 (1995).

³⁷See, for example, A. Szabo and N. Ostlund, *Modern Electronic Structure Theory* (Dover, New York, 1998).

³⁸B. J. Dunlap, J. W. Connolly, and J. R. Sabin, *J. Chem. Phys.* **71**, 3396 (1979); A. St-Amant and D. R. Salahub, *Chem. Phys. Lett.* **169**, 387 (1990).

³⁹R. T. Gallant and A. St-Amant, *Chem. Phys. Lett.* **256**, 569 (1996); S. K. Goh, R. T. Gallant, and A. St-Amant, *Int. J. Quantum Chem.* **69**, 405 (1998).

⁴⁰N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, *Can. J. Chem.* **70**, 560 (1992).

⁴¹A. E. Reed, L. A. Curtiss, and F. A. Weinhold, *Chem. Rev.* **88**, 899 (1988).

⁴²J. Gao, *J. Phys. Chem. B* **101**, 657 (1997); *J. Chem. Phys.* **109**, 2346 (1998).

⁴³DeFT, A. St-Amant, University of Ottawa, Canada.

⁴⁴Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, *J. Phys. Chem. B* **105**, 569 (2001).

⁴⁵P. A. Bash, L. L. Ho, A. D. MacKerell, D. Levine, and P. Hallstrom, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3698 (1996).

⁴⁶D. Das, B. R. Brooks, and E. M. Billings (unpublished).

⁴⁷For a recent review, see J. R. Knowles, *Nature (London)* **350**, 121 (1991).

⁴⁸Q. Cui and M. Karplus, *J. Am. Chem. Soc.* **123**, 2284 (2001); **124**, 3093 (2002); *J. Phys. Chem. B* **106**, 1768 (2002).

- ⁴⁹A. D. Becke, Phys. Rev. A **38**, 3098 (1988); J. P. Perdew and Y. Wang, Phys. Rev. B **22**, 8800 (1986).
- ⁵⁰P. Lyne, M. Hodoscek, and M. Karplus, J. Phys. Chem. A **103**, 3462 (1998).
- ⁵¹R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971); W. J. Hehre, R. Ditchfield, and J. A. Pople, *ibid.* **56**, 2257 (1972); P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).
- ⁵²C. M. Smith and G. G. Hall, Theor. Chim. Acta **69**, 63 (1986); **69**, 71 (1986); K. Strasburger, Comput. Chem. (Oxford) **22**, 7 (1998); W. Sokalski and R. A. Poirier, Chem. Phys. Lett. **98**, 86 (1983).
- ⁵³A. Mackerell *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
- ⁵⁴H. Guo and M. Karplus, J. Phys. Chem. **96**, 7273 (1992); H. Guo, N. Gresh, B. P. Roques, and D. R. Salahub, J. Phys. Chem. B **104**, 9746 (2000).
- ⁵⁵L. Luo, K. L. Taylor, H. Xiang, Y. Wei, W. Zhang, and D. Dunaway-Mariano, Biochemistry **40**, 15684 (2001).