Redistributed Charge and Dipole Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations

Hai Lin and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455-0431

Received: November 24, 2004

Special care is needed in carrying out combined quantum mechanical and molecular mechanical (QM/MM) calculations if the QM/MM boundary passes through a covalent bond. The present paper discusses the importance of correctly handling the MM partial point charges at the QM/MM boundary, and in particular, it contributes in two aspects: (1) Two schemes, namely, the redistributed charge (RC) scheme and the redistributed charge and dipole (RCD) scheme, are introduced to handle link atoms in QM/MM calculations. In both schemes, the point charge at the MM boundary atom that is replaced by the link atom is redistributed to the midpoint of the bonds that connect the MM boundary atom and its neighboring MM atoms. These redistributed charges serve as classical mimics for the auxiliary orbitals associated with the MM host atom in the generalized hybrid orbital (GHO) method. In the RCD scheme, the dipoles of these bonds are preserved by further adjustment of the values of the redistributed charges. The treatments are justified as classical analogues of the QM description given by the GHO method. (2) The new methods are compared quantitatively to similar methods that were suggested by previous work, namely, a shifted-charge scheme and three eliminated-charge schemes. The comparisons were carried out for a series of molecules in terms of proton affinities and geometries. Point charges derived from various charge models were tested. The results demonstrate that it is critical to preserve charge and bond dipole and that it is important to use accurate MM point charges in QM/MM boundary treatments. The RCD scheme was further applied to study the H atom transfer reaction $CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH$. Various QM levels of theory were tested to demonstrate the generality of the methodology. It is encouraging to find that the QM/MM calculations obtained a reaction energy, barrier height, saddle-point geometry, and imaginary frequency at the saddle point in quite good agreement with full QM calculations at the same level. Furthermore, analysis based on energy decomposition revealed the quantitatively similar interaction energies between the QM and the MM subsystems for the reactant, for the saddle point, and for the product. These interaction energies almost cancel each other energetically, resulting in negligibly small net effects on the reaction energy and barrier height. However, the charge distribution of the QM atoms is greatly affected by the polarization effect of the MM point charges. The QM/MM charge distribution agrees much better with full QM results than does the unpolarized charge distribution of the capped primary subsystem.

I. Introduction

The combined quantum mechanical and molecular mechanical (QM/MM) method[1−80] is a powerful tool for studying many chemical and biochemical processes such as enzyme reactions. A QM/MM model treats a relatively localized region (e.g., where bond breaking/forming or electronic excitation occur) with QM methods and includes the influence of the surroundings at the MM level. In some cases, such as the treatment of solvation, the boundary between the QM and MM subsystems is between solute and solvent molecules, and no covalent bond is cut. In many other cases, however, passing the boundary through covalent bonds is desirable, and special care is required to treat the boundary.

Treatments of the boundary between QM and MM regions can be largely grouped into two classes. The first is called the link-atom approach, where a "link atom" or "cap atom" is used to saturate the dangling bond at the "frontier atom" of the QM fragment. This link atom is usually taken to be a hydrogen atom[3,11,12,14,16−18] or a parametrized atom, e.g., a one-free-valence atom in the "connection atom",[27] "pesudobond",[61] and "quantum-capping potential"[62] schemes, which involve a parametrized semiempirical Hamiltonian[27] or a parametrized effective core potential (ECP)[61,62] adjusted to mimic the properties of the original bond being cut. The link atom method is straightforward and is widely used. However, it introduces additional degrees of freedom (the coordinates of the link atom) that are not present in the original molecular system, and this makes the definition of the QM/MM energy more complicated. It also presents complications in optimizations of geometries. In addition, it is found, at least in the original versions of the link-atom method, that polarization of the bond between the QM frontier atom and the link atom is unphysical due to the nearby point charge on the MM "boundary atom" (an MM boundary atom is the atom whose bond to a frontier QM atom is cut). In early work,[2] the point charges on the MM boundary atoms and on some of the atoms directly bonded to it were deleted, and in second-generation link-atom methods, these point

* To whom correspondence should be addressed. E-mail: truhlar@umn.edu.

10.1021/jp0446332 CCC: $30.25
Published on Web 04/12/2005

charges are treated in a special manner[25,27,48,61,69] to avoid this unphysical polarization, for example, some (or all) of the point charges might be redistributed, scaled, or zeroed. Extensive discussions of these problems can be found in the literature.[25,48,65] For example, one recent study[80] concluded that "the QM/MM interface is not free of introducing artifacts" and "improvement in the effective operator describing the QM/MM link is an important subject of further research". Another study[13,14] suggested that users of such methods "are strongly advised to test, calibrate, and confirm for themselves the validity of the method combination and the model subsystem for the properties they want to calculate." This is undoubtedly good advice, and one goal of further research is to assess which methods have broad robustness such that they are suitable starting points for such further testing and calibrating on specific systems. Studies that improve the link-atom approach continue to appear. An example of a proposed improvement is the so-called "double-link-atom" approach,[30] where two link atoms are employed to cap the QM and MM fragments, respectively, to reduce electrostatic unbalance in the standard single-link-atom scheme.

The second class of QM/MM methods consists of methods that use localized orbitals instead of link atoms at the boundary of the QM and MM regions. An example is the so-called local self-consistent field (LSCF) algorithm,[6−10] where the bonds connecting the QM and MM fragments are represented by a set of strictly localized bond orbitals (SLBOs) that are determined by calculations on small model compounds and assumed to be transferable. The SLBOs are excluded from the SCF optimization of the large molecule to prevent their admixture with other QM basis functions. Recently, specific force-field parameters have been developed for the LSCF method.[10] Another approach in the spirit of the LSCF method is the generalized hybrid orbital (GHO) method.[38,40−45] In this approach, a set of four sp^3 hybrid orbitals is assigned to each MM boundary atom. The hybridization scheme is determined by the local geometry of the three MM atoms to which the boundary atom is bonded, and the parametrization is assumed to be transferable. The hybrid orbital that is directed toward the frontier QM atom is called the active orbital, and the other three hybrid orbitals are called auxiliary orbitals. All four hybrid orbitals are included in the QM calculations, but the active hybrid orbital participates in the SCF optimizations, while the auxiliary orbitals do not.

The methods using local orbitals are theoretically more fundamental than the methods using link atoms, since they provide a quantum mechanical description for the charge distribution around the QM/MM boundary. The delocalized representation of charges in these orbitals helps to prevent or reduce the overpolarization that, as mentioned above, is sometimes found in the link-atom methods. However, the local-orbital methods are much more complicated than the link-atom methods. Test calculations showed that reasonably good accuracy can be achieved by both approaches if they are used with special care,[10,30,43,48,65] and we envision that both the link-atom and local-orbital methods will continue to be applied in various forms.

It is natural to ask if there is a way to combine some of the merits of the link-atom and local-orbital approaches. Such a combination would be attractive if it retains the simplicity of the former and the theoretical justification of the latter. Some progress in this direction has already been made by utilizing delocalized Gaussian functions in the link-atom method to overcome the strong polarization near the QM/MM boundary region.[30,65] In this paper we explore even simpler ways to incorporate delocalization into the link-atom picture. In particular, we introduce a redistributed charge (RC) scheme and a redistributed charge and dipole (RCD) scheme, each of which can be viewed in one sense as a point charge analogue to the GHO method. Both schemes use redistributed charges as classical mimics for the auxiliary orbitals associated with the MM boundary atom in the GHO method. The methods may also be considered as an attempt to further refine the procedure introduced earlier by de Vries and co-workers.[69]

In practice, the treatment of the QM/MM boundary is not a critical issue if the boundary is sufficiently far away from the active center. However, in many applications one is limited to the use of a small QM subsystem due to expensive computational costs, and the QM/MM boundary can be quite close to the active center. Furthermore, since one does not wish to put a boundary in the interior of a conjugated or aromatic subsystem, or perhaps even in a nonconjugated ring, one does not always have the option to move the boundary just one or two atoms farther away from the site of bond breaking. It has been suggested that, when one cannot afford to treat a large subsystem by a high-level quantum method, one can use the three-layer ONIOM (MO:MO:MM) method,[12] where the second (middle) layer is treated by an appropriate lower-level QM theory (e.g., semiempirical molecular orbital theory), which is computationally less expensive. The second QM layer is designed to allow a consistent treatment of the polarization of the active center by the environment, but the actual performance depends on how large the second layer region is and whether the second layer includes the prominent polar or charged groups. If the groups carrying significant partial charges are far from the active center, one might need to use a large second QM layer, and the computation costs grow. It is therefore worthwhile to have a two-layer QM/MM method that treats the polarization effect on the active center due to the environment as well as possible. This motivates us to examine several schemes for manipulating MM point charges, as well as the RC and RCD schemes developed in this work. We will study the proton affinities and geometries for a series of molecules, and we will compare the QM/MM results to full QM calculations. Point charges derived from various charge models for the MM subsystem will be tested. We aim to answer two specific questions: (1) How critical is it to preserve the charge and bond dipole at the QM/MM boundary? (2) How much do the values of MM point charges affect the results? The second question is raised because, in most of the validation tests for QM/MM methods, one essentially works on model systems in the gas phase. The use of gas-phase models is understandable, since it is not practical to employ an extensive training/testing set in liquid solution. However, the point charges in many MM force fields such as CHARMM[81] and OPLS-AA[82−87] are designed for simulations in condensed phases, and strictly speaking, they are not suitable for validation tests in the gas phase.

The motivation to include the interacting MM environment is to provide an improved description for the system under investigation. The change of electronic structure during a reaction may involve only a small number of atoms, but the electronic structure of these atoms might be perturbed by the rest of the atoms. How significant can the perturbation be? Would it have a significant effect on reaction energy and barrier height? Using the RCD scheme, we will study an H atom transfer reaction. The goal of studying this reaction is 2-fold: First, it serves as a demonstration of the generality and reliability of the RCD scheme. For this purpose, we will test various QM

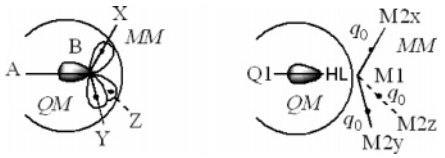

Figure 1. Schematic representations of the QM/MM boundary treatments in (a) the GHO scheme and (b) the RC scheme. The frontier atom is denoted as A in (a) and as Q1 in (b). The boundary atom is denoted as B in (a) and as M1 in (b). The MM atoms bounded to the boundary atoms are denoted as X, Y, and Z in (a) and as M2x, M2y, and M2z in (b). The link atom is denoted as HL. The redistributed charge is denoted as q_0.

levels of theory in the QM/MM computations. Second, it serves as an example, based on which we will examine the changes in primary-system charge distribution due to the interactions between the QM and MM subsystems for reactions that do not involve significant charge transfer. (The study of proton affinity, considered first, provides an examination on the reaction energy for reactions involving charge transfer.) Deeper insight will be obtained through energy decomposition as well as comparison of the charge distribution in the QM subsystem as obtained by capped-primary-system calculations, QM/MM calculations, and full QM calculations for the entire system.

The organization of the present paper is as follows: The RC and RCD algorithms are presented in section II, which begins with the simpler RC scheme and then proceeds to the RCD scheme. The computations and results are presented in section III, and the discussion will be given in section IV, followed by concluding remarks in section V.

II. Theory

II.A. General Description of the QM/MM Setup. As mentioned above, the redistributed charge scheme can be considered as a point-charge analogue of the GHO method. This is illustrated in Figure 1a. In the GHO scheme,[38,40,43] the valence electron distribution on the boundary atom B is represented by the active orbital η_Q pointing toward the QM frontier atom A and by three auxiliary orbitals (η_X, η_Y, and η_Z) pointing toward the MM atoms (X, Y, and Z) directly bonded to B. These four GHO orbitals are constructed by hybridization of the atomic *s* and *p* valence basis functions of B. The atom B (which is always a carbon atom) has a formal MM partial charge q_B consisting of an effective nuclear charge of 4 and electronic charges in valence orbitals that sum to $-4 + q_B$. To accomplish this charge distribution, the GHO method assigns an occupation of $n_{aux} = 1 - q_B/3$ to each auxiliary orbital, which therefore carries a charge of $q_{aux} = -n_{aux}$ (a typical value of q_B is -0.18, for which $q_{aux} = -1.06$); the remaining single electronic charge of B is added to the active electron pool treated by the SCF process and is primarily in the active orbital but also possibly delocalized over the quantum system.

The proposed RC scheme is illustrated in Figure 1b. We will find it convenient to label the atoms according to "tiers". Thus the B atom of GHO, i.e., the MM boundary atom, will be denoted as M1 in this paper, and the partial charge q_B becomes q_{M1}. Those MM atoms directly bonded to M1 will be called second-tier molecular mechanics atoms or M2 (e.g., M2*x*, M2*y*, and M2*z* in Figure 1b). The QM atom that is directly connected to M1 is still called the frontier atom, but now it is labeled Q1. One continues the numbering in this way: M3 atoms are the third-tier molecular mechanics atoms, i.e., those MM atoms bonded to M2 atoms. Similarly, one defines Q2 and Q3 atoms in the QM subsystem. In the RC treatment, a link atom HL (which denotes hydrogen link) is used to represent the active hybrid orbital η_Q, and the MM partial charge on M1 ($q_{M1} = q_B$) is delocalized evenly as *n* point charges q_0 with $q_0 = q_{M1}/n$, where *n* is the number of M1−M2 bonds, usually three. The delocalized point charges q_0 are located on the M1−M2 bonds, as discussed in more detail in the next subsection. This is illustrated in Figure 1b for the case of $n = 3$. These redistributed point charges (q_0) serve as mimics for the auxiliary hybrid orbitals. These are in approximately the same location as the q_{aux} charges of the GHO method; however, because there is no nuclear charge on M1, they are much smaller in magnitude, e.g., $q_0 = -0.06$ when $q_B = -0.18$ and $n = 3$. By construction, the HL atom does not carry any MM point charge, which is consistent with the requirement that adding the link atom to the QM subsystem should not change the charge for the QM subsystem, e.g., should not make a neutral QM subsystem partially charged. Apparently, further improvement can be achieved by associating the link atom with a parametrized ECP as in the pesudobond and related schemes,[27,61,62] but the goal of the present paper is to test the simpler treatment where no ECP is used. One reason to keep the method as simple as possible is to facilitate its incorporation into a wide variety of electronic structure codes and make it universally applicable to all electronic structure methods.

It should be noted that the redistributed point charges provide only classical mimics for calculating the Columbic interactions of the auxiliary orbitals, and quantum mechanical exchange interactions are not recovered. However, we hope that such a treatment will not cause unacceptably large errors.

II.B. Redistributed Charges and Dipoles. In considering the details of charge redistribution, the first question to ask is precisely where to locate the redistributed point charges. The most physical choice for a simple model is to place the redistributed charges at points along each M1−M2 bond, i.e., at the nominal centers of the bond charge distributions. One would expect that the precise position along each M1−M2 bond would depend on the nature of the bond; in practice, though we found that QM/MM calculations such as geometry optimizations are not very sensitive to the actual location, provided that the location is sufficiently far from M1. For simplicity, we choose the locations to be the midpoints of the M1−M2 bonds. Schematically, these places are indicated in Figure 1b by dots on the M1−M2 bonds. This is the only redistribution that occurs in the RC scheme.

The second question to ask is how large a perturbation is introduced to the MM subsystem by redistributing charge from M1 to the M1−M2 bond. Clearly, the RC method reduces the contribution of q_0 to each M1−M2 bond dipole q_0R, where *R* is the M1−M2 bond distance, by 50%. Therefore, we consider a second method called the RCD method, which is the same as the RC method except that the values of redistributed charges q_0 and of the charges on M2 atoms (labeled $k = 1, 2, ...$) are further modified such that these contributions to the M1−M2 bond dipoles are preserved

$$q_0^{\mathrm{RCD}} = 2q_0 \tag{1}$$

$$q_{\mathrm{M2},k}^{\mathrm{RCD}} = q_{\mathrm{M2},k} - q_0 \tag{2}$$

This is the only difference between the RC and RCD treatments.

On the other hand, because the RC method does not modify the point charges on M2, the M2−M3 bond dipoles are preserved in RC while this is not the case in RCD. Since the QM/MM method is designed to treat the most important region by QM and the less important region by MM, and since the

M1−M2 bonds are closer to the QM region than are the M2−M3 bonds, one might expect the RCD method to be more accurate for the most important region of the problem, but actual tests are required to validate either or both methods, and such tests are reported in section III.

II.C. Link Atoms. The position of the link atom is another important issue in QM/MM models, and it has been investigated extensively.[14,25,48,55] In accordance with the argument in section II.A that the active orbital is represented by a link-atom HL, a natural location for HL is on the Q1−M1 bond with the Q1−HL distance depending on the Q1−M1 distance. Hence, we adopt the scaled-bond-distance method proposed by Maseras, Morokuma, and co-workers.[11,14] In this approach, the link atom is placed along the Q1−M1 bond. The Q1−HL distance, R(Q1−HL), is related to the Q1−M1 distance, R(Q1−M1), by a scaling factor

$$R(\mathrm{Q1-HL}) = C_{\mathrm{HL}} R(\mathrm{Q1-M1}) \quad (3)$$

During a QM/MM geometry optimization or a molecular dynamics or reaction path calculation, the equilibrium Q1−HL and Q1−M1 distances are constrained to satisfy eq 3.

The scaling factor, C_{HL}, depends on the nature of the bonds being cut and constructed. It has been suggested[14] that it should be the ratio of standard bond lengths of the Q1−HL and Q1−M1 bonds, which is close to 0.71 for replacement of a C−C single bond by a C−H bond. We follow this scale-bond-distance treatment in the present work and set the scaling factor by

$$C_{\mathrm{HL}} = R_0(\mathrm{Q1-H})/R_0(\mathrm{Q1-M1}) \quad (4)$$

where R_0(Q1−H) and R_0(Q1−M1) are the MM bond distance parameters for the Q1−H and Q1−M1 stretches, respectively. Notice that eq 2 does not introduce any new parameters. We examined more complicated alternatives to eq 4, but we did not find that the complications made the calculations significantly more accurate.

II.D. QM/MM Energy. The QM/MM energy is defined by

$$E(\mathrm{QM/MM;ES}) = E(\mathrm{MM;ES}) - E(\mathrm{MM;CPS^*}) + E(\mathrm{QM;CPS^{**}}) \quad (5)$$

where ES and CPS denote the entire system and capped primary system, respectively, the CPS is the primary system capped by the link-atom HL, the asterisk (*) denotes that the CPS is embedded in the electrostatic field of the secondary subsystem (SS), and the double asterisks (**) denote such an embedding in an appropriately modified electrostatic field of the SS; the SS is defined as

$$\mathrm{SS} = \mathrm{ES} - \mathrm{PS} \quad (6)$$

where PS denotes the primary system. The PS is the QM subsystem, and the SS is the MM subsystem.

In eq 5, the MM energy for the ES, E(MM;ES), is a sum of the valence (stretch, bend, and torsion) energy E(val;ES), the van der Waals energy E(vdW;ES), and the Coulombic energy E(Coul;ES)

$$E(\mathrm{MM;ES}) = E(\mathrm{val;ES}) + E(\mathrm{vdW;ES}) + E(\mathrm{Coul;ES}) \quad (7)$$

The term E(MM;CPS*) is the MM energy for the CPS that is embedded in the background charge distribution of the SS. It includes both the MM energy for CPS itself, E(MM;CPS), and the Coulombic interaction energy between the CPS and the SS, E(Coul;CPS|SS)

$$E(\mathrm{MM;CPS^*}) = E(\mathrm{MM;CPS}) + E(\mathrm{Coul;CPS|SS}) \quad (8)$$

The term E(MM;CPS) also consists of three contributions, i.e., the valence energy E(val;CPS), the van der Waals energy E(vdW;CPS), and the Coulombic energy E(Coul;CPS)

$$E(\mathrm{MM;CPS}) = E(\mathrm{val;CPS}) + E(\mathrm{vdW;CPS}) + E(\mathrm{Coul;CPS}) \quad (9)$$

Special scaling factors are often used in the MM force field for calculations of Coulombic interactions for pairs of atoms that are connected by a valence potential. For example, Coulombic interactions between neighboring or geminal atoms are neglected, and Coulombic interactions between vicinal atoms may be neglected[81] or scaled by 0.5.[82−87] This feature is retained in the calculations for E(Coul;ES) and E(Coul;CPS|SS) in eqs 7 and 8.

The term E(QM;CPS**) is the QM energy for the CPS that is obtained with a background charge distribution of the SS that has been modified by an appropriate boundary treatment; in particular, the M1 charge has been redistributed in the RC and RCD schemes, and the M1 and M2 charges have been modified to restore the contribution of q_0 to the M1−M2 bond dipole if the RCD scheme is adopted. The only modification required for the electronic structure program to calculate E(QM;CPS**) is that it can carry out a calculation in the presence of a background point-charge distribution; many electronic structure codes already have this capability, and the required integral types are the ones already required for the nuclear attraction term.

One notices that both E(MM;ES) and E(MM;CPS*) include the MM energies for the PS, and thus those MM terms that involve only the PS atoms are completely canceled in computations employing eq 5. On the other hand, those terms involving only SS atoms survive, and they provide the MM descriptions of the SS system. Also surviving in E(MM;CPS*) are the MM terms for interactions between the PS atoms and the HL atom, which can be considered as corrections to the QM calculations for the CPS.[25] (Note that the PS−HL Coulombic terms vanish because $q_{\mathrm{HL}} = 0$.) The final group of MM energy terms that survive are interactions between CPS and SS; they are more complicated as discussed next.

First, we consider at the valence interactions$E$(val;CPS|SS). The surviving terms are the Q1−M1 stretch, the Q2−Q1−M1 and Q1−M1−M2 bends, and the Q3−Q2−Q1−M1, Q2−Q1−M1−M2, and Q1−M1−M2−M3 torsions. The second kind of MM interactions that we consider is the nonbonded van der Waals interactions. We retain the contributions in E(vdW;PS|SS) that describe the interactions between the PS and SS, but the HL atom is not seen by the SS. The final type of CPS|SS interaction is Coulombic. We note that

$$E(\mathrm{Coul;ES}) = E(\mathrm{Coul;SS}) + E(\mathrm{Coul;PS}) + E(\mathrm{Coul;PS|SS}) \quad (10)$$

By use of eqs 8−10 and the fact that the MM charge on the HL atom is zero, one finds that the E(Coul;PS) and E(Coul;PS|SS) terms cancel exactly. This is what we might have expected, because now the electrostatic interactions between PS and SS are handled at the QM/MM level, i.e., by the E(QM;CPS**) computations. Since the E(Coul;PS) and E(Coul;PS|SS) terms cancel exactly, the calculations can be simplified by setting all the MM charges on the PS atoms to zero.

The final expression for the QM/MM energy is therefore given as follows

$$E(\mathrm{QM/MM;ES}) = [E(\mathrm{val;ES}) - E(\mathrm{val;CPS})] + [E(\mathrm{vdW;ES}) - E(\mathrm{vdW;CPS})] + E(\mathrm{Coul;SS}) + E(\mathrm{QM;CPS}^{**}) \quad (11)$$

II.E. MM Parameters for the PS. The QM/MM schemes tested in this paper (see next section) are designed to be applicable with any MM method that employs atom-centered partial charges. Some QM/MM methods, such as the GHO method and the pseudobond method, require new parameters for boundary atoms, integral scaling factors in the QM calculations, or specially parametrized ECPs. Such parameters usually require reconsideration if one switches MM scheme (e.g., from CHARMM[81] to OPLS-AA[82−87]), QM scheme (e.g., from semiempirical molecular orbital methods to density functional theory or post-Hartree−Fock ab initio methods), or QM basis set. A central objective in the present work is to avoid introducing any new parameters. Thus, for example, no MM parameters are changed, no integrals are scaled, and the link atom is an ordinary hydrogen atom with a standard basis set.

The key issue discussed in this section is how to select MM parameters for the atoms in the PS. As discussed in section II.D, we do not need the partial charges of PS atoms, but we do need stretch parameters for the Q1 atoms, bend parameters for Q1 and Q2 atoms, torsion parameters for Q1, Q2, and Q3 atoms, and van der Waals parameters for all Q atoms. This presents a problem since reaction is allowed to occur in the PS, and therefore the atom types of the Q atoms are not uniquely defined. An example is the deprotonation of RCH_2COOH to form RCH_2COO^-, for which the R group is the SS, and the CH_2-COOH subunit is the PS. The COOH group becomes a COO^- group upon deprotonation; therefore, the atom types for the Q2 carbon atom and the Q3 oxygen atoms are different at different points along the reaction path. Which set of MM parameters should we use when carrying out molecular dynamics calculations or following the reaction path, those for the protonated form or those for the deprotonated form? Switching between these two sets of parameters during a dynamics calculation or along the reaction path is not convenient. Moreover, even if the switching between parameters could be done, one does not know at which point along the reaction path it should be done. There is no unambiguous answer, but for all tests in this paper, the decision that we make is to use the MM parameters for the protonated form, even for calculations on the deprotonated reagent. Although our treatment is not a perfect solution, it is very practical, and it appears to be reasonable as discussed next for the protonation of RCH_2COO^-.

First consider the valence interactions, in particular, those for the Q1, Q2, and Q3 atoms, since, as seen above, certain valence interactions involving these atoms do not cancel. In principle, this can be avoided if one uses a larger QM subsystem,[70] such that the atoms types for the Q1, Q2, and Q3 atoms do not change. However, a larger QM subsystem is not always feasible, e.g., in the RCH_2COOH case where R is a naphthyl group. Generally speaking, the Q1−M1 stretch is the most important interaction among those surviving valence interactions due to its large force constant; the Q2−Q1−M1 and Q1−M1−M2 bends are less significant, and the Q3−Q2−Q1−M1, Q2−Q1−M1−M2, and Q1−M1−M2−M3 torsions are the least critical. Fortunately, the Q1 atom type does not change in this case (and in most applications), thus the Q1−M1 stretch, the Q1−M1−M2 bend, and the Q1−M1−M2−M3 torsion are unambiguous. The parameters for the other bends and torsions often remain the same or change just slightly. The OPLS-AA[82−87] force field (in the TINKER[88] implementation that we used in this work) uses the same parameters for CH_3CH_2-COOH and $CH_3CH_2COO^-$ for the Q2−Q1−M1 bend and for the Q2−Q1−M1−M2 torsion. There are two kinds of Q3−Q2−Q1−M1 torsion in CH_3CH_2COOH: (a) the O−C−C−C torsion where the O bonds to the H atom and (b) the O=C−C−C torsion with a double bond between the O and C atoms. There is only one kind of Q3−Q2−Q1−M1 torsion in $CH_3CH_2COO^-$, the O−C−C−C torsion. The (a) torsion in CH_3-CH_2COOH also uses the same parameters as the O−C−C−C torsion in $CH_3CH_2COO^-$, and only the (b) torsion uses a different one. Because of the very small force constants (the torsional barrier height is less than 0.9 kcal/mol) for all Q3−Q2−Q1−M1 torsions, using a single set of valence parameters along the reaction path does not seem to produce unacceptably large uncertainty in comparison with the errors produced by other approximations that are introduced into the QM/MM framework.

Next, we examine the nonbonded interactions. For the van der Waals interactions, any PS atoms that change atom types are ambiguous, and in principle, this problem cannot be avoided even if a larger QM subsystem is adopted. Fortunately, in practice it is not a serious problem, since the van der Waals interactions are significant only at short distances, and the use of only one set of van der Waals parameters is often adequate.

Turning to the electrostatic interactions, this is not a problem at all. In our RC and RCD schemes, as well as all other electrostatic embedding schemes tested in this paper, the electrostatic contributions to E(MM;ES) and E(MM;CPS*) cancel exactly, and they do not need to be evaluated.

II.F. Mechanical Embedding. The treatments discussed so far are sometimes called electric embedding. Another commonly used QM/MM scheme is the so-called mechanical embedding (ME) scheme,[25] which is the same as the original integrated MO/MM (IMOMM) scheme.[11−13] In the ME scheme, the CPS calculations are performed in gas phase, i.e., without the background charge distribution for the SS. The QM/MM energy is defined by

$$E(\mathrm{QM/MM;ES}) = E(\mathrm{MM;ES}) - E(\mathrm{MM;CPS}) + E(\mathrm{QM;CPS}) \quad (12)$$

The electrostatic interactions between PS (or CPS) and SS are taken care of by the E(MM;ES) term, i.e., they are handled at the MM level and require MM charge parameters for the ES. Thus in contrast to the electrostatic embedding schemes, where one does not require the MM charge parameters for the PS atoms, the ME treatment relies on the availability of MM charge parameters for the PS atoms. This creates a problem for studying reactions, and the problem is especially serious for processes accompanied by charge transfer, such as a proton-transfer reaction or an electron-transfer reaction. Unlike the van der Waals interactions, the electrostatic interaction is long-range, and the use of inappropriate charge parameters can cause a serious error. For this reason, we do not employ the ME scheme in this article.

II.G. Implementation of RC and RCD Schemes. The RC and RCD schemes are implemented in the QMMM package,[89] which was developed on the basis of MULTILEVEL.[90] The QMMM package performs QM/MM calculations by interfacing the electronic structure program *Gaussian03*[91] with the force field program TINKER.[88] Briefly, QMMM invokes *Gaussian03* for doing QM calculations in order to get QM energy and energy derivatives (gradient and Hessian) when required. Similarly,

QMMM invokes TINKER to do MM calculations to get MM energy and energy derivatives. The QM and MM energies and energy derivatives are integrated by QMMM to produce the final output; in particular, the energy derivatives are obtained by the chain rule [taking account of eq 4] as described in ref 14. No modification to the electronic structure program *Gaussian03* or to the force-field program TINKER is needed. Thus, in principle, QMMM is automatically upgraded whenever *Gaussian03* and/or TINKER are upgraded.

II.H. Other QM/MM Schemes in QMMM. In addition to the RC and RCD methods, QMMM also contains some other schemes for charge manipulation in the link atom approaches. The first one is the straight electrostatic embedding (SEE) where no special treatment for the background MM charges is performed; in particular, there is no redistribution, scaling, or zeroing of MM partial charges.

Four additional methods are included in QMMM and in the present tests; these methods, as employed here, differ from the RC and RCD schemes only in the treatment of the electrostatic embedding in the E(QM;CPS**) term of eq 5. Three of these other methods are eliminated charge[2] schemes where selected MM point charges are eliminated. If only the M1 charge is zeroed, it is called the Z1 scheme. If both M1 and M2 charges are zeroed, it is called the Z2 scheme; and Z3 denotes the treatment where all M1, M2, and M3 charges are zeroed. It should be noted that the Z1, Z2, and Z3 schemes may not preserve the overall charge for the system under study, e.g., a neutral system may become partially negatively charged. It is also interesting to note that Z3 is the default scheme in *Gaussian03*.

The final scheme employed for comparison is the shifted charge ("Shift") scheme,[69] where the M1 charge is evenly shifted onto all M2 atoms, and a pair of point charges is added in the vicinity of M2 to preserve M1−M 2 bond dipole. The distance between this pair of point charges is set to 20% of the M1−M2 bond distance.

In all these schemes, q_{HL} is zero.

We note that since we made the schemes as simple as possible to promote clarity and portability, our implementations for these schemes might not be exactly the same as other groups' implementations of their schemes. For example, the parameters selected and also the treatments for locating the link atom position could be different. It is therefore not possible to make direct comparisons of our results with other groups' works based on the published literature, but the schemes are compared on a consistent basis here.

For the sake of clarifying the differences between the methods, it is useful at this point to consider the limit of the Zn schemes as $n \rightarrow \infty$; we call this Z∞. In particular, we point out that the Z∞ scheme is not the same as mechanical embedding for two reasons. First, in the presence of a solvent or other nonbonded environment (e.g., a protein or a supramolecular cage), the Z∞ method does not zero out all charges but only those connected by a sequence of bonds to Q1. Second, the mechanical embedding scheme differs from the Z∞ scheme in the middle terms of eqs 5 and 12. Thus, in the absence of nonbonded moieties, electrostatic interactions between QM and MM subsystems would cancel out in Z∞ but not in ME.

III. Computations and Results

All the QM/MM schemes of sections II.G and II.H can be applied with arbitrary quantum mechanical levels, including density functional theory (DFT) method and post-Hartree−Fock *ab initio* methods. However, most tests in the present paper are based on a single level of theory since that is sufficient to demonstrate the method. In particular, sections III.A and III.B are based on comparing the QM/MM results to full QM calculations, where the QM computations are carried out at the Hartree−Fock level of theory,[92] and the MIDI![93] basis set is employed. More advanced QM treatments by post-Hartree−Fock theory or by DFT are straightforward, and two examples are given in section IV.C. First, though, we want to focus on the critical issue of handling MM point charges at the QM/MM boundary, and it is sufficient to illustrate the nature of the problem by the HF/MIDI! calculations. Except for the point charges that will be discussed in detail below, the OPLS-AA force field was used for pure MM or combined QM/MM computations. Pure QM calculations were performed by use of the *Gaussian03* program, and QM/MM computations were carried out by use of the QMMM package, which combines QM calculations carried out by *Gaussian03* with MM calculations carried out by TINKER. The pure MM calculations were done by invoking TINKER through the interface of the QMMM package. Geometry optimizations were accomplished by use of the Berny optimizer[94] in *Gaussian03*. In all QM/MM calculations, we used $C_{HL} = 0.713$ as determined by eq 4 using the OPLS-AA force field parameters [R_0(C−H) = 1.09 Å and R_0(C−C) = 1.529 Å].

III.A. Proton Affinity and MM Point Charges. In this section, we tested the electrostatic embedding schemes explained in section II. We will study the proton affinities for a series of molecules, and we will also examine the optimized geometries, in particular, the Q1−M1 bond distances. The proton affinity is defined in this paper as the zero-point-exclusive energy difference between a chemical species (denoted $X^- + H^+$ or $X + H^+$) and its protonated form (denoted XH or XH^+). Table 1 lists the *protonated* form for the selected species, for which the boundary between QM and MM subsystems is indicated by a dash between the MM (left) and QM (right) fragments. We note that for some molecules, e.g., CF_3-CH_2OH, there are MM atoms with significant charges very close to the QM/MM boundary. Such cases are included in the present study to provide difficult tests for validation of the method. In general, one is advised to avoid such locations for the QM/MM boundary if a more suitable place is possible, but for testing, it is instructive to push the envelope.

In principle, the sum of the MM point charges for the MM subsystem and the QM charge for the QM subsystem should be equal to the charge for the entire system, to preserve the overall charge. Since the QM charge for the QM subsystem is zero for a neutral species, it follows that the summed MM charge for the MM subsystem should also be zero. For a charged species, the QM charge for the QM subsystem will be −1 or +1, and the MM subsystem is again required to be neutral. Except for CF_3-CH_2OH, the standard OPLS-AA charges that we use for all the MM subsystems in Table 1 satisfy these neutrality conditions. In the CF_3-CH_2OH case, the net OPLS-AA point charge on the CF_3 group is −0.08, i.e., the CF_3 group is slightly negatively charged. Consequently, one should be aware of that the use of OPLS-AA charges for CF_3-CH_2OH in QM/MM calculations slightly violates charge neutrality.

As already mentioned in introduction, we are motivated to examine how much the QM/MM energetics depend on the MM point charges. To this end, we test not only the original OPLS-AA point charges for the MM subsystem but also point charges derived from other charge model computations, namely, the CM2[95] and CM3[96,97] charges, as well as the charges obtained by fitting the electrostatic potential (ESP) using the Merz−

TABLE 1: Comparisons of the OPLS-AA Charges and the CM2, CM3, and ESP Charges for Selected Molecules[a]

molecule (MM-QM)		charges for the MM subsystem			
		OPLS-AA	CM2	CM3	ESP
CH_3-CH_2OH					
CH_3-CH_2SH	q_C	−0.1800	−0.1920	−0.2960	−0.0548
CH_3-$CH_2NH_3^+$	q_H	0.0600	0.0640	0.0990	0.0183
CH_3-CH_2COOH					
CF_3-CH_2OH	q_C	0.5323	0.5380	0.4490	0.4255
	q_F	−0.2067	−0.1793	−0.1497	−0.1416
CH_2OH-CH_2OH	q_C	0.1450	0.1260	−0.0040	0.2209
	q_H	−0.60600	−0.0710	−0.0780	−0.0131
CH_2OH-CH_2SH	q_O	−0.6830	−0.5900	−0.4640	−0.6290
	$q_{H(O)}$	0.4183	0.3220	0.3120	0.3818

[a] Only the protonated form is listed in the first column. The OPLS-AA charges were taken from the TINKER 4.1 implementation. The CM2, CM3, and ESP charges were derived from MM subsystem dimer calculations, e.g., by using C_2F_6 for the CF_3 group (the MM subsystem in CF_3CH_2OH). The QM level of theory employed was HF/MIDI!. The charge on the H atom bonded to the O atom is given as $q_{H(O)}$.

TABLE 2: Proton Affinities (kcal/mol)[a]

molecule (MM-QM)	QM	QM/ MM								
			CPS	SEE	RC	RCD	Shift	Z1	Z2	Z3
CH_3-CH_2O H	416. 8	OPLS-AA	421.0	432.7	427.2	431.5	430.4	399.3	422.5	422.5
		CM2	421.0	433.3	427.5	432.1	430.9	397.7	422.5	422.5
		CM3	421.0	438.2	429.9	436.5	435.0	383.8	422.5	422.5
		ESP	421.0	425.8	423.9	425.4	425.0	415.5	422.5	422.5
CH_3-CH_2SH	381. 5	OPLS-AA	383.5	389.4	386.7	389.5	389.2	363.1	383.8	383.8
		CM2	383.5	389.8	386.9	389.8	389.5	361.7	383.8	383.8
		CM3	383.5	392.4	388.4	392.5	392.3	349.5	383.8	383.8
		ESP	383.5	385.6	384.7	385.6	385.5	377.5	383.8	383.8
CH_3-$CH_2NH_3^+$	232. 8	OPLS-AA	229.9	236.7	233.3	236.5	235.8	209.4	230.0	230.0
		CM2	229.9	237.2	233.5	236.9	236.1	208.0	230.0	230.0
		CM3	229.9	240.6	235.2	240.2	239.2	196.0	230.0	230.0
		ESP	229.9	232.1	231.0	232.0	231.7	223.7	230.0	230.0
CH_3-CH_2COOH	375. 3	OPLS-AA	377.3	382.2	379.9	382.2	382.0	358.4	377.4	377.4
		CM2	377.3	382.5	380.0	382.5	382.3	357.1	377.4	377.4
		CM3	377.3	384.9	381.3	384.8	384.7	345.8	377.4	377.4
		ESP	377.3	378.9	378.1	378.9	378.8	371.6	377.4	377.4
CF_3-CH_2OH	396. 8	OPLS-AA	421.0	388.9	415.5	398.7	403.0	492.7	422.3	422.3
		CM2	421.0	377.3	404.9	387.1	391.7	483.9	422.3	422.3
		CM3	421.0	385.8	408.0	393.8	397.5	474.0	422.3	422.3
		ESP	421.0	388.0	408.8	395.4	398.9	471.3	422.3	422.3
CH_2OH-CH_2OH	413. 2	OPLS-AA	421.0	417.9	424.2	420.0	421.3	446.1	385.2	422.6
		CM2	421.0	416.9	422.4	418.7	419.8	441.4	393.9	422.6
		CM3	421.0	427.4	427.2	427.3	427.3	426.6	394.9	422.6
		ESP	421.0	411.9	422.0	415.3	417.5	455.2	388.4	422.7
CH_2OH-CH_2SH	376. 5	OPLS-AA	383.5	380.4	383.4	380.8	381.4	402.2	350.8	383.8
		CM2	383.5	379.3	381.9	379.6	380.1	398.2	358.5	383.8
		CM3	383.5	385.8	385.7	385.8	385.8	385.2	359.3	383.8
		ESP	383.5	377.5	382.4	378.3	379.3	410.9	353.7	383.8

[a] Only the protonated form is listed in the first column. See section II in text for notation. Except for the point charges that are given explicitly in Table 1 for the MM subsystem, the OPLS-AA force field is used for the MM calculations. The QM level of theory employed was HF/MIDI!. $C_{HL} = 0.713$.

Singh−Kollmann scheme.[98,99] To ensure a neutral MM subsystem, the CM2, CM3, and ESP charge model charges were derived from model calculations for the MM subsystem in dimers. For example, C_2F_6 was used to derive point charges for the CF_3 group, and the neutrality of the CF_3 group is assured by symmetry. Similarly, 1,2-ethanediol is used to derive charges for CH_2OH and so forth. The CM2, CM3, and ESP charges are compared with the OPLS-AA charges in Table 1.

In application of the ESP fitting scheme, it is well known[60,100−102] that the results are sometimes unphysical. To avoid unphysical charges, we used the option in *Gaussian03* by which one constrains the charge to give the correct dipole moment as well as to fit the ESP. (The only case where this was required is one of the calculations in section IV.G.)

The computed proton affinities are tabulated in Table 2. In addition to the electrostatic embedding schemes addressed in section II, i.e., the SEE, Z1, Z2, Z3, RC, RCD, and Shift schemes, we also carried out calculations for the CPS. A CPS calculation can be considered as a very special kind of QM/MM scheme, in which HL atoms substitute the whole MM subsystem. The ME scheme is however left out, because it is not appropriate for such a test, as discussed in section II.

The overall performance of a given QM/MM scheme is gauged by the mean unsigned errors (MUEs). An MUE is calculated for each charge model by averaging over the species in the testing set, and an averaged MUE (AMUE) is obtained by averaging the MUEs over all charge models. All MUEs and AMUEs are listed in Table 3. Since the MM point charges are not involved in CPS calculations, the proton affinities by use of the CPS scheme do not depend on the choices for charges, and the MUE is the same for all charge models. Table 3 actually gives two different rows for MUE. The first MUE row is for all four charge models. However, as we will see in the discussion (in particular, in section IV.B), we conclude that the CM2 and

TABLE 3: MUEs for Proton Affinities (kcal/mol) by QM/MM Calculations in Comparison to Full QM Calculations[a]

charge model	CPS	SEE	RC	RCD	Shift	Z1	Z2	Z3
OPLS-AA	7.1	7.0	8.5	6.6	8.6	30.8	15.0	8.6
CM2	7.1	8.6	6.7	7.6	8.4	28.9	12.3	8.6
CM3	7.1	11.5	9.5	10.5	12.7	29.0	12.0	9.0
ESP	7.1	4.0	6.1	3.2	3.7	23.9	14.0	8.1
averaged MUE[b]	7.1	7.7	7.7	7.0	8.4	28.2	13.3	8.6
averaged MUE[c]	7.1	5.5	7.3	4.9	6.2	27.4	14.5	8.4

[a] See section II in text for notation. See also footnote *a* in Table 2 for computation setup. MUE was obtained by averaging over the molecules listed in Table 2 for each charge model and a given QM/MM scheme. [b] Averaged over all charge models for a given QM/MM scheme. [c] Averaged over the OPLS-AA and ESP charge models for a given QM/MM scheme.

TABLE 4: QM/MM Optimized Q1−M1 Bond Distances (Å) by Use of the ESP Charges in Comparison to Full QM Calculations[a]

molecule (MM-QM)	QM	QM/MM						
		SEE	RC	RCD	Shift	Z1	Z2	Z3
Neutral Species								
CH_3-CH_2OH	1.527	1.498	1.523	1.522	1.524	1.526	1.525	1.525
CH_3-CH_2SH	1.539	1.498	1.523	1.521	1.524	1.525	1.524	1.524
CH_3-CH_2NH_2	1.543	1.504	1.530	1.528	1.531	1.533	1.532	1.532
CH_3-CH_2COOH	1.533	1.499	1.524	1.523	1.525	1.526	1.525	1.525
CF_3-CH_2OH	1.498	1.721	1.537	1.558	1.540	1.524	1.529	1.529
CH_2OH-CH_2OH	1.521	1.618	1.526	1.531	1.525	1.519	1.531	1.527
CH_2OH-CH_2SH	1.529	1.618	1.527	1.532	1.526	1.521	1.529	1.526
Charged Species								
CH_3-CH_2O^-	1.594	1.540	1.571	1.567	1.571	1.579	1.575	1.575
CH_3-CH_2S^-	1.550	1.504	1.531	1.528	1.531	1.533	1.533	1.533
CH_3-$CH_2NH_3^+$	1.528	1.496	1.520	1.519	1.521	1.523	1.521	1.521
CH_3-CH_2COO^-	1.533	1.505	1.531	1.529	1.532	1.533	1.533	1.533
CF_3-CH_2O^-	1.521	1.858	1.613	1.661	1.631	1.570	1.579	1.579
CH_2OH-CH_2O^-	1.561	1.710	1.587	1.605	1.591	1.565	1.591	1.579
CH_2OH-CH_2S^-	1.525	1.639	1.537	1.546	1.537	1.532	1.536	1.535
MUE		0.094	0.019	0.028	0.021	0.012	0.016	0.015

[a] See section II in text for notation. See also footnote *a* in Table 2 for computation setup. MUE was obtained by averaged over molecules for a given QM/MM scheme.

CM3 models are less realistic than the other two models for several of the cases included in the present tests, at least for the purpose of this article; therefore the MUEs including these models are not the best test of the QM/MM models per se because they also reflect the inappropriateness of the partial charges. For this reason, the last row of Table 3 gives MUEs in which the results obtained with CM2 and CM3 charges are excluded.

III.B. Geometries. Table 4 gives the QM/MM-optimized Q1−M1 bond distances in comparison with full QM calculations for all species involved in the tests for proton affinities.

III.C. Reaction: More QM Levels. To illustrate the power of the general formulation presented here, this section presents RCD calculations of the barrier height for the reaction (see also Figure 2)

$$CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH \quad \text{(R-1)}$$

when the following QM levels are employed: HF/MIDI!, MP2/6-31G(d),[103−105] MPW1K/6-31+G(d,p),[106−108] and CCSD/6-311G(d,p)//MPW1K/ 6-31+G(d,p).[109−111] The primary system is CH_3 + CH_3CH_2, giving rise to CH_3 + CH_3CH_3 as the CPS, and the secondary system is CH_2OH. In each case, we optimized the geometry at both the full QM and QM/MM levels. We also carried out normal-mode analysis for the saddle-point geometry,

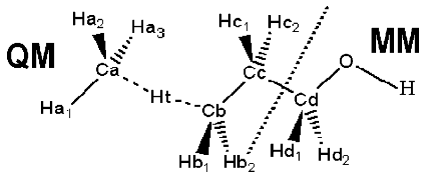

Figure 2. The QM/MM boundary setup for the H atom transfer reaction CH_3 + $CH_3CH_2CH_2OH$ → CH_4 + $CH_2CH_2CH_2OH$. The primary system is CH_3 + CH_3CH_2, giving rise to a capped primary system as CH_3 + CH_3CH_3, which is treated quantum mechanically, and the secondary system is CH_2OH, which is referred to as the MM subsystem. The transferring H atom between the Ca and Cb atoms is denoted Ht.

and in each case, we found only one imaginary frequency mode, which corresponds to the H atom transfer. Table 5 summarizes the key energetic and geometric data, as well as the imaginary frequencies.

III.D. Reaction: Effects due to MM Environment. In this section, we analyze the effects on the CPS of the MM environment for the reaction that was studied in section III.C. The analysis was performed for the HF/MIDI! calculations for simplicity. Computations at the other QM levels are expected to be qualitatively similar.

The energy difference between the QM energies for the CPS in the gas phase and the QM energies for the CPS in an interacting MM environment is defined by

$$E_{CPS/MM} = E(QM;CPS^{**}) - E(QM;CPS) \quad (13)$$

where E(QM;CPS**) is the QM energy for the CPS embedded in the background point charges and E(QM;CPS) is the QM energy for the CPS in the gas phase. In either case, the geometry is fully optimized at the corresponding level of theory, i.e., at the QM/MM level for E(QM;CPS**) and at the QM level for E(QM;CPS). Equation 13 provides a measure of the magnitude of the perturbation on the QM subsystem due to the MM subsystem. Generally speaking, the two geometries in eq 13 are different because of the interaction between the CPS and the SS in E(QM;CPS**). We further decompose $E_{CPS/MM}$ into two contributions: the energy due to the polarization of the background point charges (E_{pol}) and the energy due to the geometry distortion from the CPS in the gas phase (E_{steric}), which are defined as

$$E_{pol} = E(QM;CPS^{**}) - E(QM;CPS^{dis}) \quad (14)$$

$$E_{steric} = E(QM;CPS^{dis}) - E(QM;CPS) \quad (15)$$

$$E_{CPS/MM} = E_{pol} + E_{steric} \quad (16)$$

where E(QM;CPS^{dis}) is the gas-phase single-point CPS energy for the QM/MM optimized geometry, i.e., we took the CPS geometry that resulted from QM/MM optimization, and we removed the MM point charges. Although such an energy decomposition is approximate, it is informative and provides us deeper understanding of the QM/MM calculations. The energy decomposition is illustrated in Figure 3.

We also computed and listed in Table 6 the ESP-fitted point charges for the CPS atoms (or, in the case of full QM calculations for the ES, for the PS atoms). Figure 4 shows the ESP charges for selected atoms at the saddle-point geometry.

IV. Discussion

IV.A. Proton Affinities: Overall Performance. By examination of Table 3, one finds that the CPS calculations produce

TABLE 5: Results for Reaction $CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH$[a]

QM		ΔE	$V^{\ddagger}$	$R^{\ddagger}$(Ca−Ht)	$R^{\ddagger}$(Cb−Ht)	$\omega^{\ddagger}$
HF/MIDI!	full QM	−2.7	27.2	1.371	1.350	2491*i*
	QM/MM	−2.9	26.9	1.372	1.351	2487*i*
	CPS	−2.9	27.0	1.375	1.349	2486*i*
MP2/6-31G(d)	full QM	−2.4	19.7	1.347	1.319	2086*i*
	QM/MM	−2.9	19.7	1.349	1.318	2098*i*
	CPS	−2.9	19.8	1.352	1.316	2097*i*
MPW1K/6-31G+(d,p)	full QM	−2.4	15.1	1.357	1.315	1765*i*
	QM/MM	−2.9	14.9	1.359	1.314	1764*i*
	CPS	−2.9	14.9	1.363	1.312	1759*i*
CCSD/6-311G(d,p)//	full QM	−2.5	18.1	1.357	1.315	N/A
MPW1K/6-31G+(d,p)	QM/MM[b]	−2.6	18.0	1.359	1.314	N/A
	CPS	−2.8	18.0	1.363	1.312	N/A
MUE[c]	QM/MM	0.3	0.2	0.002	0.001	6*i*
	CPS	0.4	0.2	0.005	0.002	7*i*

[a] See section II in text and Figure 2 for notation. The QM/MM boundary setup is illustrated in Figure 2. The RCD scheme was used in QM/MM calculations, and $C_{HL} = 0.713$. The OPLS-AA force field is used, except for the partial charges for the SS, for which the ESP-fitted charges shown in Table 1 are adopted. The zero-point-energy exclusive reaction energy ΔE and barrier height $V^{\ddagger}$ are given in kcal/mol, the bond distances for the breaking and forming bonds at the saddle point are given in Å, and the imaginary frequency at the saddle point is given in cm^{-1}. [b] The CCSD/6-311G(d,p):OPLS-AA single-point energy calculations on the MPW1K/ 6-31+G(d,p):OPLS-AA optimized geometries. [c] Averaged over the first three QM levels, i.e., HF/MIDI!, MP2/6-31G(d), and MPW1K/ 6-31+G(d,p).

TABLE 6: ESP-Fitted Charges of the CPS or PS Atoms of the Reactant, Product, and Saddle Point for reaction $CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH$[a]

	reactant				saddle point				product			
	CPS	CPS^{dis}	CPS**	FullQM	CPS	CPS^{dis}	CPS**	FullQM	CPS	CPS^{dis}	CPS**	FullQM
Ca	−0.52	N/A	N/A	−0.52	−0.5 8	−0.58	−0.58	−0.59	−0.52	N/A	N/A	−0.52
Ha1	0.17	N/A	N/A	0.17	0.16	0.16	0.16	0.16	0.13	N/A	N/A	0.13
Ha2	0.17	N/A	N/A	0.17	0.15	0.15	0.15	0.16	0.13	N/A	N/A	0.13
Ha3	0.17	N/A	N/A	0.17	0.16	0.16	0.16	0.16	0.13	N/A	N/A	0.13
Ht	0.02	0.03	0.06	0.11	0.11	0.13	0.15	0.20	0.13	N/A	N/A	0.13
Cb	−0.05	−0.08	−0.20	−0.46	−0.05	−0.10	−0.21	−0.46	−0.32	−0.32	−0.36	−0.53
Hb1	0.02	0.03	0.05	0.11	0.03	0.04	0.06	0.11	0.12	0.12	0.13	0.17
Hb2	0.02	0.03	0.05	0.11	0.03	0.04	0.06	0.12	0.12	0.12	0.13	0.17
Cc	−0.06	−0.05	0.00	0.18	−0.12	−0.10	−0.02	0.21	0.02	0.04	0.09	0.28
Hc1	0.02	0.02	0.04	0.00	0.03	0.04	0.06	0.00	0.02	0.02	0.04	−0.01
Hc2	0.02	0.02	0.04	0.00	0.04	0.04	0.06	0.00	0.02	0.02	0.04	−0.01
HL	0.02	0.02	−0.07	N/A	0.04	0.03	−0.06	N/A	0.03	0.02	−0.07	N/A
Sum[b]	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.06

[a] See section II as well as Figures 2 and 3 for notation. See also footnote *a* in Table 5 for the computation setup. Only the HF/MIDI! calculations are analyzed. [b] Sum over the ESP charges for the CPS atoms in CPS, CPS^{dis}, and CPS** calculations, and sum over PS atoms in full QM calculations.

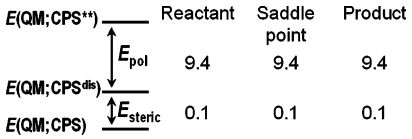

Figure 3. Schematic representation of the decomposition of $E_{CPS/MM}$, i.e., the QM energy difference between unperturbed CPS calculations and calculations for the CPS coupled to the MM environment, as defined by eq 13, into two contributions: the energy due to the polarization by the background point charges (E_{pol}) and the energy due to the geometry distortion from the CPS (E_{steric}). The energies are shown in kcal/mol for the reactant, the saddle point, and the product, respectively, for the H atom transfer reaction $CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH$.

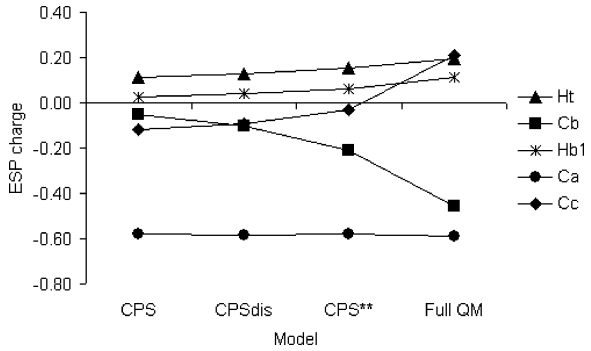

Figure 4. ESP charges for the CPS atoms in CPS, CPS^{dis}, and CPS** calculations and for the PS atoms in full QM calculations (E_{steric}) at the saddle-point geometry for the H atom transfer reaction $CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH$.

an MUE of 7 kcal/mol. There is no polarization of the QM subsystem in the CPS calculations. However, polarization effects must be included properly, as the outcome depends on how one handles the MM point charges and also on whether appropriate MM point charges are employed.

The eliminated-charge schemes Z1, Z2, and Z3 actually make the result even worse, as indicated by their MUEs, which are larger than about 24, 12, and 8 kcal/mol, respectively. One expects (for these solvent-free calculations, see discussion in section II.H) that the results obtained by eliminated-charge schemes will eventually converge to the calculations where all MM point charges are zeroed out. The poor performances by the Z1, Z2, and Z3 schemes show that it can be dangerous to arbitrarily eliminate MM point charges. The remaining four schemes (SEE, RC, RCD, and Shift) preserve overall charges, and except for the RC scheme, the other three schemes also preserve the M1−M2 bond dipoles. These four schemes yield generally smaller MUEs than the "best" eliminated-charge scheme (the Z3 scheme) when the OPLS-AA, CM2, or ESP charges are used, but larger MUEs are produced if the CM3 charges are employed. In most of the remaining discussions,

we focus on these four schemes as well as the CPS method, and we discuss the results in more detail.

It should be noted that our QM/MM boundary treatments are validated for the cutting of a single bond, in particular, a C−C bond. A recent study[112] by Ferré and Olivucci investigated the behavior of using a link atom for treating QM/MM boundaries that cut bonds with some double-bond character, such as an amide (C−N) bond. These authors found that cutting an amide bond can be dangerous.

IV.B. Proton Affinities: The CH_3 Group. We start by looking at the first four species in Table 2: CH_3-CH_2OH, CH_3-CH_2SH, $CH_3-CH_2NH_3^+$, and CH_3-CH_2COOH. These four species have the same MM subsystem, a CH_3 group. For CH_3-CH_2OH, CH_3-CH_2SH, and CH_3-CH_2COOH, the QM proton affinities are smaller than the CPS values, indicating a decrease of proton affinity if an H atom is replaced by a CH_3 group. However, none of the QM/MM schemes that we consider here predicts the correct trend for the substituent effects; instead, they all predict that the proton affinity increases. (In the $CH_3-CH_2NH_3^+$ case, the QM/MM schemes are correct in predicting that the proton affinity increases if an H atom is replaced by a CH_3 group.) Moreover, for all four species, when we compare four different methods (see Table 1) for assigning MM charges, we find that, the smaller charges on the CH_3 group, the closer the agreement with the full QM calculations. The error decreases in the order of CM3 (−0.30) > CM2 (−0.19) > OPLS-AA (−0.18) > ESP (−0.055), where the charge on the methyl carbon atom is given in parentheses. It is not realistic to ask for accuracy better than 2 kcal/mol from these QM/MM calculations, because of the intrinsic limitations in the QM/MM approach itself. For example, charge transfer between the QM and MM subsystems is not allowed, while it surely takes place in the real QM system. However, the correlation between the errors and the MM charges on the CH_3 group suggests that the charges on the CH_3 group (and possibly on CH_2 group, too) are probably overestimated in these charge models. In other words, the alkanes seem to be very unpolar in the gas phase; and the CM2 and CM3 methods seem to overestimate their polarity, at least for the purpose of electrostatic embedding calculations.

The OPLS-AA charges are designed for simulations in liquids instead of in the gas phase. In liquid, the alkanes can be more polar than in the gas phase. (We notice that a recent re-parametrization[113] of the OPLS-AA force field suggests reducing the original OPLS-AA charges for alkanes by 25% of their magnitude for improved simulations in liquids.) As we pointed out in the Introduction, these kinds of charges might not be suitable for gas-phase modeling, and more appropriate charges are desirable. Unfortunately, we do not know what the accurate charges are in the gas phase, although the ESP charges seem to be candidates. The ESP fitting procedure can be problematic for systems with buried atoms,[60,100,101] but it is sometimes stable for very small compounds. The ESP charges computed from gas-phase molecules at least have the advantage that they are not parametrized for the liquid phase. The very small gas-phase ESP charges on the CH_3 group do imply that the alkanes are very unpolar in gas phase.

In a recent work by Amara and Field,[65] the M1 and M2 point charges at the QM/MM boundary were represented by Gaussian functions. This kind of delocalization of MM point charges was shown to improve the QM/MM geometries and energetics in many situations. However, the use of OPLS-AA charges for the CH_3 and CH_2 groups for validation tests in the gas phase produced the same kinds of errors in that paper as those produced in this work. For example, as shown in the fifth column in Table 4 in ref 65, the proton affinities for CH_3-CH_2OH, $CH_3CH_2-CH_2OH$, and $CH_3CH_2CH_2-CH_2OH$ decrease in the order of 405.3, 404.2, and 403.9 kcal/mol, respectively, by full QM calculations at the HF/6-31G* level of theory. (Please note that the first column in that table lists the *deprotonated* form rather than the *protonated* form that is listed in the present study, and also note that energies in ref 65 are given in kJ/mol.) However, the trend was not reproduced by the corresponding QM/MM calculations even with the Gaussian-delocalization scheme: if only the M1 point charge is represented by Gaussian functions, the proton affinities for CH_3-CH_2OH, $CH_3CH_2-CH_2OH$, and $CH_3CH_2CH_2-CH_2OH$ are 400.9, 406.1, and 405.5 kcal/mol, respectively; and if both the M1 and M2 point charges are represented by Gaussian functions, the respective proton affinities are 407.2, 406.5, and 407.5 kcal/mol. In other words, the predictions for the substituent effects ($CH_3 \rightarrow CH_3CH_2 \rightarrow CH_3CH_2CH_2$) are incorrect. We are not arguing against using Gaussian functions to delocalize MM point charges; rather, we think that physically it is a good idea (although it does not satisfy the objective of the present paper which is to find a robust method that does not require any changes in standard electronic structure codes). What we want to point out here is that the error introduced by inaccurate MM point charges in validation tests can spoil gas-phase tests of QM/MM boundary treatments, a problem that seems to have been underappreciated so far, despite the extensive use of alkyl groups in many QM/MM method validations. To minimize the effect of overpolar alkyl group representations on our conclusions, we suggest that the reader's final conclusions about the validity of the QM/MM boundary treatment should be based on the MUEs for the ESP charge model in Table 3 or on the MUEs averaged over the results for the OPLS-AA and ESP charge models, as given in the last row of Table 3.

IV.C. Proton Affinities: The CF_3 and CH_2OH Groups. In sharp contrast to the very unpolar CH_3 group, the atoms in a CF_3 group carry very large charges, as can be seen from Table 1. Neglecting the CF_3 charges causes large errors in proton affinities, as illustrated by the CPS calculations, which have errors bigger than 24 kcal/mol for CF_3-CH_2OH in comparison with full QM studies. This large energetic effect provides a challenging test for QM/MM boundary treatments; it also allows us to draw conclusions that are not compromised by the 2 kcal/mol intrinsic uncertainty that we discussed in section IV.B.

As shown in Table 2, for these more polar examples (which are more typical of real practical applications), the SEE, RC, RCD, and Shift schemes are in generally better agreement than the CPS method with full QM results. The RCD and Shift schemes appear to be superior to the SEE and RC methods. Again, the best agreement is obtained with ESP charges, which we regard as the most reliable charges in the present study; by employment of the ESP charges, the RCD and Shift schemes give errors less than or close to 2 kcal/mol, respectively. The CM3 charges are very close to the ESP charges for CF_3, and it is thus not surprising to see similarly small errors when the CM3 charges are utilized. The CF_3-CH_2OH test case demonstrates the importance of preserving charge and dipole in QM/MM boundary treatments. It also confirms the criticalness of using accurate MM point charges.

The CH_2OH group is less polar than the CF_3 group, and this is reflected in the smaller errors (roughly 8 kcal/mol) in proton affinity by CPS calculations with respect to full QM computations. However, we found in Table 1 that various charge models predict quite different charges for CH_2OH; the CM3 model

predicts that the C atom has negligible charge, while the other three charge models do not agree. As judged from the results in Table 2, the ESP charges seem to be the most reliable. For example, employing the ESP charges, the RCD and Shift schemes produce results in agreement with full QM calculations within 2 and 4 kcal/mol, respectively.

IV.D. Optimized Geometries. Table 4 shows that the QM/MM optimized Q1−M1 distances agree better with full QM calculations for the neutral molecules than for charged species. The largest deviations occur for $CF_3CH_2O^-$, for which all QM/MM geometry optimizations show deviations close to or larger than 0.05 Å. This is not unexpected, since there are unusually large charges on a group (CF_3) located close to the QM/MM boundary, making this case very challenging for QM/MM boundary tests. In general, one avoids such a QM/MM boundary setup and gets smaller errors for the Q1−M1 bond distance by the QM/MM schemes in real practical applications.

The preceding sections provided the interesting result that the SEE scheme does not seem to be particularly poor for the energetics of the proton affinities, especially in the CH_2OH-CH_2OH and CH_2OH-CH_2SH cases. Therefore, it is important to emphasize that the SEE method does very poorly for optimized geometries. This is illustrated by the QM/MM optimized Q1−M1 bond distances in Table 4 in comparison with full QM calculations. Table 4 shows that the SEE scheme usually yields large errors for the Q1−M1 bond distances, as indicated by the MUE (0.094 Å). The MUEs for the other QM/MM methods are about 4−6 times smaller. The SEE scheme produces especially large errors in the Q1−M1 bond distances for CH_2OH-CH_2OH and CH_2OH-CH_2SH and their deprotonated forms, while the other schemes all give reasonable agreement with QM calculations. Therefore, the surprisingly good results in proton affinities for CH_2OH-CH_2OH and CH_2OH-CH_2SH by the SEE scheme seem to result largely from error cancellations.

IV.E. Reaction: High-Level QM Methods. Although some previous work using QM/MM methods with high-level QM methods such as coupled-cluster theory have appeared,[80,114] most attention has been devoted to DFT and low-level theories. To illustrate that the methods presented here are general, Table 5 presents higher-level QM calculations. This illustrates not only the generality of the method but also the generality of the QMMM computer program.

As can be seen from Table 5, the QM/MM calculations by the RCD scheme yielded energetic and geometric data very similar to the full QM computations, within 0.5 kcal/mol in energy and within 0.002 Å in bond distances. The QM/MM vibrational frequency for the imaginary frequency mode also agrees well with full QM analysis (within $12i$ cm^{-1}). These results are very encouraging in that they demonstrate that good accuracy can be achieved by the QM/MM calculations for these kinetically important quantities in comparison with full QM calculations at the same level.

The CPS calculations also produce reasonably good results, although less accurate (especially in the geometry) than the QM/MM calculations in comparison with full QM calculations. However, although CPS calculations often provide competitive accuracy to QM/MM for gas-phase molecules, they do not provide an acceptable treatment of charge polarization effects, as required for modeling in the condensed phase. This important point will be discussed further in the next two sections.

IV.F. Reaction: Reaction Energy and Barrier Height Affected by MM Environment. It may be surprising to see (in the previous section) that the CPS and QM/MM calculations give similar energetic results for the $CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH$ reaction, although the CPS interacts with the MM environment in the QM/MM calculations but not in gas-phase computations. The reason is that the interaction energies (with the MM environment) are very similar at all three critical geometries, the reactant, the saddle point, and the product, resulting a large cancellation when computing the relative energies.

As shown in Figure 3, the interactions with the MM environment can be decomposed into two contributions: the steric effect and the polarization effect. The steric effect is rather small (0.1 kcal/mol) for the present example, since the distortion of geometry for the CPS from the fully relaxed geometry in the gas phase is rather small. (The steric effect can be much more significant in a more complex MM environment, e.g., in a protein environment.) On the other hand, the polarization effect is dominant (9 kcal/mol) for this reaction, due to the nearby polar group, CH_2OH, of the MM subsystem. However, the energies due to geometry distortion and polarization are so similar that they almost cancel out, giving rise to negligibly small net contributions to the reaction energy and barrier height.

Recently, Thiel, Shaik, and co-workers reported a series of comprehensive studies[115,116] of C−H hydroxylation by the P450cam enzyme.[117] The CPS in their studies included compound I (an oxoiron porphyrin radical cation), the camphor (substrate), and the proximal sulfur ligand (Cys357). (See the original reference for details.) Both QM model calculations in the gas phase and full QM/MM calculations in the (solvated) protein environments were performed. The polarization effect for the CPS was found to be significant:[115] The gas-phase calculations showed that the sulfur ligand carried more than 60% unpaired spin density and that the porphyrin carried less than 40%. That is, compound I was predicted to be mainly a sulfur-centered radical. The QM/MM calculations predicted that compound I is a porphyrin-centered radical with ∼70% unpaired spin density on the porphyrin; the difference was mainly due to the polarization effect. However, these authors observed similar reaction barrier heights for the H atom abstraction in gas phase and in QM/MM computations. The reported similar reaction barriers could be rationalized by the cancellation effect mentioned above, i.e., the energy contribution due to this significant polarization effect cancels to a large extent for the reactant and saddle point. Although the cancellation is more complete in the reaction in the present study, it is very informative to see how this cancellation of the energetic effect can occur even in a prototypical simple system with much less complex MM environments, and this shows how the cancellation may be a somewhat general effect. Often (but of course not always), the electrostatic interactions between the CPS and the MM environment will have a more pronounced energetic effect for reactions accompanied by significant charge transfer than for reactions without much charge transfer. The study of proton affinity in previous sections provides examples of reactions with significant charge movement, and we did find large effects. The H atom transfer reaction (R1) involves less charge movement, and we obtained less significant effects.

Although the MM environment does not significantly change the reaction barrier height, one should not conclude that the MM environment does not contribute to the kinetics and/or dynamics. For example, there are well-recognized effects of the MM environment on P450cam kinetics, such as lowering the entropic cost, controlling the access of water to the active site, high regio- and stero-selectivity of the reaction, and facilitation[116,118] of product release (i.e., the dissociation of the

hydroxylated camphor, by modulating the stability of spin states through the polarization effect and by stabilizing the dissociated product through favorable interaction with the residues in the pocket.

IV.G. Reaction: Effect of the MM Environment on the Primary System Charge Distribution. Although the MM environment does not have a large net effect on the relative energies of the H atom transfer reaction (R1), it does have effects on the electronic structure of the CPS through polarization. This can be illustrated by examining quantities that depend on the electronic structure, e.g., atom-centered partial charges.

The ESP-fitted charges tabulated in Table 6 and illustrated in Figure 4 clearly show a trend of stepwise change from the unperturbed CPS (denoted as CPS), to the CPS with distorted geometry (CPS^{dis}), then to the CPS embedded in the background point charge distribution (CPS**), and finally to the ES, as modeled by full QM calculations. For example, the charge on the Cb atom at the saddle point geometry is only −0.05 e in the CPS calculations, and it increases to −0.10 e in CPS^{dis}, to −0.21 e in CPS**, and finally to −0.46 e in full QM calculations. The charge on the Cc atom even changes sign when moving from CPS to full QM calculations, and the CPS** result lies in between. It is interesting to note that the Cb−Cc bond seems to be very unpolar according to the CPS calculations, with a small bond dipole pointing from the Cb to the Cc atom. This disagrees qualitatively with the full QM results. The CPS** calculations predict that the Cb−Cc bond is more polar, with a larger and *inverse* bond dipole pointing from the Cc to the Cb atom, in qualitative agreement with full QM calculations. Although such an analysis is very approximate and the ESP charges might not be very accurate, it appears that the CPS** result is generally closer to the full QM results, suggesting that QM/MM calculations provide a more realistic description for the QM subsystem than the isolated gas-phase QM model calculations. The change of the point charges also implies that the alkyl group could be more polar in water or other solvents of large dielectric constants than in the gas phase, which is consistent with our previous discussion on the choice of point charges for the proton affinity calculations.

The inclusion of charge polarization effects on the PS charge distribution may partly cancel in calculating reaction energies, but it will not cancel in calculating the interaction energy of the PS with solvent or with a protein active site. Thus, it is preferable, when modeling large systems, to use a method like RCD, where these effects are included, than a method like Z3, where key polarization effects are eliminated.

V. Concluding Remarks

In this work we developed two new schemes, namely, the RC and RCD schemes, for handling the charges on boundary atoms by a classical simplification of the GHO method. Redistributed point charges are applied to mimic the auxiliary hybrid orbitals in GHO theory, and link atoms are used to represent the active hybrid orbital. The values of the redistributed charges and the values of charges on the second-tier molecular mechanics (M2) atoms are further adjusted to preserve the M1−M2 bond dipoles in the RCD method. Both the RC and RCD schemes combine the merits of the link-atom and frozen-orbital methods, and they offer the following advantages: First, the ways that they handle the MM point charges near the QM/MM boundary are justified as a classical analogue to the QM description, and second, the simplicity of the methods allows direct incorporation into most electronic structure programs in a general way. These schemes are completely general in that they may be used with any quantum mechanical level, with any molecular mechanics method that is based on atom-centered point charges for the electrostatics, and with any electronic structure program that allows both positive and negative point charges. There are no new parameters, no pseudopotentials, and no integral scaling. The protocol requires validation, and that is one of the most important purposes of this paper.

In all QM/MM methods, there are choices about which cross-boundary valence and electrostatic terms to include or exclude and choices about which set of MM parameters to use. This is especially a concern in doing dynamics calculations when following a reaction path or in which any molecules react during the course of the simulation. In some cases, in previously published work, it was not possible to be sure precisely which terms and/or parameters were selected in a given application and why such choices were made. In the presentation here, we have made a special effort to make all such choice explicit and systematic.

The RC and RCD schemes, together with five other QM/MM schemes, were applied to study the proton affinities for a set of selected species. Although one could present a variety of tests (and in fact we have carried out many more calculations than are presented here), we believe that a systematic study of proton affinities is sufficient to illustrate our major conclusions about the treatment of the QM/MM boundary in the absence of solvent. In this regard, it is worth emphasizing that proton affinities with the protonation site close to the boundary provide one of the most severe tests one can imagine in that the initial and final states (e.g., $RO^- + H^+$ and ROH) have very different charge distributions. Our test set includes difficult cases where large MM point charges are close to the QM/MM boundary. Comparisons of the QM/MM results with full QM computations revealed that it is important to preserve the charge and dipole when handling the QM/MM boundary and that it is necessary to employ accurate MM point charges.

Validating QM/MM methods by comparison to high-level calculations or experiment is essential, since the use of the unvalidated method is unacceptable. Although the motivation for developing QM/MM methods is to apply them to large systems (e.g., reactions in the condensed phase, including liquids, enzymes, nanoparticles, and solid-state materials), most of the validation studies, including those in the present article, have been based on small gas-phase model systems, where a "model system" is a small- or medium-sized molecule. We believe that it is important, in interpreting such validation tests, to keep two important issues in mind. First, the molecular mechanics parameters, especially partial charges, are often designed for treating condensed-phase systems where partial charges are systematically larger due to polarization effects in the presence of dielectric screening; thus electrostatic effects of the MM subsystem may be overemphasized in the gas phase. Second, while it is often essential to employ QM/MM calculations in modeling large systems, because it is unaffordable to apply high-level QM methods (as required, for example, for quantitative prediction of reactive barrier heights) to the whole system, the goal of the QM/MM treatment is not usually to predict MM substituent effects on the QM subsystem but rather to stitch the whole system together without artifacts. Any part of the QM subsystem that has a significant energetic (as opposed to structural) effect on the QM subsystem should probably be incorporated in the QM subsystem by moving the boundary farther from the reactive site (this is sometimes accomplished by treating a buffer region by a lower-level QM method).[12,15] Thus the main goal of validation tests should usually be to

ensure that no unacceptably large energetic or structural artifacts are introduced rather than to achieve high quantitative accuracy for MM substituent effects (such substituent effects, as well as strong first-solvation-shell effects, can be treated quantitatively at high-level-QM/low-level-QM boundaries[15,114,119−125]). In this regard, tests such as those in the present paper are examples of testing the methods for problems more difficult than those that are normally asked to handle, as a way of ascertaining where the performance envelope lies. Among the schemes that we tested in this way, the RCD scheme and the shifted charge scheme[69] provided the best QM/MM boundary treatments for the test set.

The RCD scheme was further applied to investigate the H atom transfer reaction $CH_3 + CH_3CH_2CH_2OH \rightarrow CH_4 + CH_2CH_2CH_2OH$. Various QM levels of theory were tested to demonstrate the generality of the methodology. Previous workers[80,114] have already illustrated the power of combining MM methods with coupled-cluster theory, and the present article presents an example how the same QM/MM formalism can be used with Hartree−Fock theory, Møller−Plesset perturbation theory, DFT, and coupled-cluster theory. It is encouraging to find that the QM/MM calculations obtained a reaction energy, barrier height, saddle-point geometry, and imaginary frequency at the saddle point in quite good agreement with full QM calculations at the same level. Furthermore, analysis based on energy decomposition revealed quantitatively similar interaction energies between the QM subsystem and the MM environment for the reactant, for the saddle point, and for the product. These interaction energies approximately cancel each other, giving rise to negligibly small net effects on the reaction energy and barrier height. Finally, the examination on the ESP-fitted charges for the atoms of the primary subsystem illustrated the polarization effect due to the MM background point charges. QM/MM calculations give a charge distribution that agrees much better with full QM results than do calculations without the MM point charges. This suggests that the QM/MM calculations provide a more realistic description for the QM subsystem.

A next higher level of theory would be to include polarization in the MM force field. As emphasized elsewhere, inclusion of polarization in the MM subsystem introduces new difficulties not considered here.[25,80] While polarizable MM is a promising technique for the future, most QM/MM calculations employ standard nonpolarizable MM methods, which have demonstrated their usefulness in many applications. The present article presents a well-defined and very portable (no nonstandard code requirements, no new parameters) QM/MM method that may be employed with any QM method and with any MM method that treats the solute electrostatics in terms of partial charges. Validation of the new RCD method on test cases where reaction occurs very close to the QM/MM boundary show that structural and energetic predictions are robust (no unphysical artifacts). In particular, even without solvent present, the RCD method shows, on average, quantitative improvement for the proton affinity over the third-nearest-neighbor charge elimination (Z3) scheme and even larger quantitative improvement over the neighbor and near-neighbor charge elimination (Z1 and Z2) schemes. One expects the RCD scheme to be even more satisfactory than the charge elimination schemes in the presence of solvent or a protein active site, because no charges are eliminated, first-tier bond dipole contributions are preserved, and the primary system polarization by the MM subsystem is not eliminated (either fully or from the closest MM tiers). This should provide a more realistic description of the QM/MM boundary region with nonbonded surroundings. More detailed comparisons with the shift method would also be interesting because, like the RCD method, the shift scheme does not eliminate any charges, it preserves the M1−M2 bond dipole contributions, and it performs relatively well in our tests.

Acknowledgment. We thank Chris Cramer, Jiali Gao, Casey Kelly, Keiji Morokuma, Jingzhi Pu, Jason Thompson, and Yan Zhao for helpful discussions. This work was supported in part by the U.S. Department of Energy, Office of Basic Energy Sciences, and the National Science Foundation. Hai Lin thanks the Minnesota Supercomputing Institute for a Research Scholarship.

References and Notes

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(2) Singh, U. C.; Kollmann, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(3) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(4) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(5) Gao, J. *Rev. Comput. Chem.* **1996**, *7*, 119.
(6) Ferenczy, G. G.; Rivail, J.-L.; Surjan, P. R.; Naray-Szabo, G. *J. Comput. Chem.* **1992**, *13*, 830.
(7) Thery, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.
(8) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.
(9) Monard, G.; Loos, M.; Thery, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153.
(10) Ferre, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610.
(11) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
(12) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
(13) Froese, R. D. J.; Musaev, D. G.; Morokuma, K. *J. Am. Chem. Soc.* **1998**, *120*, 1581.
(14) Dapprich, S.; Komiromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *THEOCHEM* **1999**, *461−462*, 1.
(15) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.
(16) Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, *115*, 62.
(17) Morokuma, K. *Philos. Trans. R. Soc. London, Ser. A* **2002**, *360*, 1149.
(18) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355*, 257.
(19) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Comput. Chem.* **1995**, *16*, 113.
(20) Monard, G.; Merz, K. M., Jr. *Acc. Chem. Res.* **1999**, *32*, 904.
(21) Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr. *J. Chem. Phys.* **2000**, *113*, 5604.
(22) Gogonea, V. *Internet Electron. J. Mol. Design* **2002**, *1*, 173.
(23) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374.
(24) Thompson, M. A. *J. Phys. Chem.* **1995**, *99*, 4794.
(25) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
(26) Bakowies, D.; Thiel, W. *J. Comput. Chem.* **1996**, *17*, 87.
(27) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
(28) Lennartz, C.; Schaefer, A.; Terstegen, F.; Thiel, W. *J. Phys. Chem. B* **2002**, *106*, 1758.
(29) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189.
(30) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.
(31) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1997**, *18*, 1496.
(32) Titmuss, S. J.; Cummins, P. L.; Rendell, A. P.; Bliznyuk, A. A.; Gready, J. E. *J. Comput. Chem.* **2002**, *23*, 1314.
(33) Bersuker, I. B.; Leong, M. K.; Boggs, J. E.; Pearlman, R. S. *Int. J. Quantum. Chem.* **1997**, *63*, 1051.
(34) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1997**, *101*, 6299.
(35) Woo, T. K.; Cavallo, L.; Ziegler, T. *Theor. Chem. Acc.* **1998**, *100*, 307.
(36) Woo, T. K.; Blöchl, P. E.; Ziegler, T. *THEOCHEM* **2000**, *506*, 313.
(37) Gao, J.; Thompson, M. A.; Ed. *Combined Quantum Mechanical and Molecular Mechanical Methods: ACS Symp. Ser. 712*; American Chemical Society: Washington, DC, 1998.
(38) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
(39) Byun, K.; Gao, J. *J. Mol. Graphics Model.* **2000**, *18*, 50.
(40) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336.
(41) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

(42) Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M. L.; Villa, J. *Acc. Chem. Res.* **2002**, *35*, 341.
(43) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.
(44) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 5454.
(45) Devi-Kesavan, L. S.; Garcia-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2003**, *109*, 133.
(46) Aaqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.
(47) Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462.
(48) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
(49) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
(50) Riccardi, D.; Li, G. H.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.
(51) Hillier, I. H. *THEOCHEM* **1999**, *463*, 45.
(52) Hall, R. J.; Hindle, S. A.; Burton, N. A.; Hillier, I. H. *J. Comput. Chem.* **2000**, *21*, 1433.
(53) Nicoll, R. M.; Hindle, S. A.; MacKenzie, G.; Hillier, I. H.; Burton, N. A. *Theor. Chem. Acc.* **2001**, *106*, 105.
(54) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656.
(55) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452.
(56) Röthlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. *J. Biol. Inorg. Chem.* **2000**, *5*, 236.
(57) Laio, A.; VandeVondele, J.; Röthlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941.
(58) Colombo, M. C.; Guidoni, L.; Laio, A.; Magistrato, A.; Maurer, P.; Piana, S.; Rohrig, U.; Spiegel, K.; Sulpizi, M.; VandeVondele, J.; Zumstein, M.; Röthlisberger, U. *Chimia* **2002**, *56*, 13.
(59) Sulpizi, M.; Laio, A.; VandeVondele, J.; Cattaneo, A.; Röthlisberger, U.; Carloni, P. *Proteins* **2003**, *52*, 212.
(60) Laio, A.; Gervasio, F. L.; VandeVondele, J.; Sulpizi, M.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 7963.
(61) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
(62) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.
(63) Yang, W.; Drueckhammer, D. G. *J. Phys. Chem. B* **2003**, *107*, 5986.
(64) Field, M. J.; Albe, M.; Bret, C.; Martin, F. P.-D.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088.
(65) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.
(66) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.
(67) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442.
(68) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *Chem. Phys. Lett.* **2000**, *321*, 113.
(69) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. *J. Phys. Chem. B* **1999**, *103*, 6133.
(70) Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; NIC-Directors: Princeton, 2000; Vol. 3, p 285.
(71) Turner, A. J.; Moliner, V.; Williams, I. H. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323.
(72) Moliner, V.; Williams, I. H. *Chem. Commun.* **2000**, 1843.
(73) Hu, H.; Elstner, M.; Hermans, J. *Proteins* **2003**, *50*, 451.
(74) Pitarch, J.; Pascual-Ahuir, J. L.; Silla, E.; Tunon, I.; Ruiz-Lopez, M. F. *J. Comput. Chem.* **1999**, *20*, 1401.
(75) Swart, M. *Int. J. Quantum Chem.* **2003**, *91*, 177.
(76) Loferer, M. J.; Loeffler, H. H.; Liedl, K. R. *J. Comput. Chem.* **2003**, *24*, 1240.
(77) Mordasini, T.; Curioni, A.; Andreoni, W. *J. Biol. Chem.* **2003**, *278*, 4381.
(78) Worthington, S. E.; Krauss, M. *J. Phys. Chem. B* **2001**, *105*, 7096.
(79) Poteau, R.; Ortega, I.; Alary, F.; Solis, A. R.; Barthelat, J. C.; Daudey, J. P. *J. Phys. Chem. A* **2001**, *105*, 198.
(80) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Phys. Chem. A* **2003**, *107*, 2578.
(81) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(82) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(83) McDonald, N. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 8049.
(84) Jorgensen, W. L.; McDonald, N. A. *THEOCHEM* **1998**, *424*, 145.
(85) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827.
(86) Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. *J. Comput. Chem.* **2001**, *22*, 1340.
(87) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
(88) Ponder, J. W. TINKER, Version 4.1; Washington University: St. Louis, MO, 2003.
(89) Lin, H.; Truhlar, D. G. QMMM, Version 1.0; University of Minneapolis: Minneapolis, 2004. The software is available at http://comp.chem.umn.edu/qmmm.
(90) Rodgers, J. M.; Lynch, B. J.; Fast, P. L.; Chuang, Y.-Y.; Pu, J.; Zhao, Y.; Truhlar, D. G. MULTILEVEL, Version 3.1/G03; University of Minnesota: Minneapolis, 2003.
(91) Frisch, M. J., et al. *Gaussian03*; Gaussian, Inc.: Pittsburgh, PA, 2003.
(92) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.
(93) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1996**, *93*, 281.
(94) Schlegel, H. B. *J. Comput. Chem.* **1982**, *3*, 214.
(95) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.
(96) Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 10707.
(97) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*.
(98) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
(99) Besler, B. H., Jr.; K. M. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
(100) Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. *J. Comput. Chem.* **1996**, *17*, 367.
(101) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(102) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*.
(103) Møller, C. M. S.; Plesset, M. S. *Phy. Rev.* **1934**, *46*, 618.
(104) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.
(105) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.
(106) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.
(107) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
(108) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.
(109) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35.
(110) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.
(111) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.
(112) Ferré, N.; Olivucci, M. *THEOCHEM* **2003**, *632*, 71.
(113) Kahn, K.; Bruice, T. C. *J. Comput. Chem.* **2002**, *23*, 977.
(114) Re, S.; Morokuma, K. *J. Phys. Chem. A* **2001**, *105*, 7185.
(115) Schöneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 8142.
(116) Schöneboom, J. C.; Cohen, S.; Lin, H.; Shaik, S.; Thiel, W. *J. Am. Chem. Soc.* **2004**, *126*, 4017.
(117) *Cytochrome P450: Structure, Mechanisms and Biochemistry*, 2nd ed.; Ortiz de Montellano, P. R., Ed.; Plenum Press: New York, 1995; Vol. 2.
(118) Lin, H.; Schöneboom, J. C.; Cohen, S.; Shaik, S.; Thiel, W. *J. Phys. Chem. B* **2004**, *108*, 10083.
(119) Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654.
(120) Coitiño, E. L.; Truhlar, D. G.; Morokuma, K. *Chem. Phys. Lett.* **1996**, *259*, 159.
(121) Noland, M.; Coitino, E. L.; Truhlar, D. G. *J. Phys. Chem. A* **1997**, *101*, 1193.
(122) Coitiño, E. L.; Truhlar, D. G. *J. Phys. Chem. A* **1997**, *101*, 4641.
(123) Froese, R. D. J.; Morokuma, K. *J. Phys. Chem. A* **1999**, *103*, 4580.
(124) Torrent, M.; Vreven, T.; Musaev, D. G.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B. *J. Am. Chem. Soc.* **2002**, *124*, 192.
(125) Rickard, G. A.; Karadakov, P. B.; Webb, G. A.; Morokuma, K. *J. Phys. Chem. A* **2003**, *107*, 292.