JCTC Journal of Chemical Theory and Computation

Including Charge Penetration Effects in Molecular Modeling

Bo Wang and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, Minnesota 55455-0431, United States

Received July 11, 2010

Abstract: Electrostatic effects are often the dominant component of intermolecular interactions, but they are often modeled without accounting for charge penetration effects due to the finite extent of electronic orbitals. Here, we propose a new scheme to include charge penetration effects in electrostatic modeling, and we parametrize it and illustrate it by employing the electronically embedded combined quantum mechanical and molecular mechanical (QM/MM) method. It can also be extended to other molecular modeling approximations that include electrostatic effects. The method, which is based on introduction of a single parameter for each element, is simple in concept and implementation, modest in cost, and easily incorporated into existing codes. In the new scheme, the MM atomic charge density of an atom in a molecule is represented by a screened charge rather than by a point charge. The screened charge includes a point charge for the nucleus, core electrons, and inner valence electrons, and a smeared charge for the outer valence electron density, which is distributed in a Slater-type orbital representing the outer part of the atomic charge distribution such that the resulting pairwise interactions are still analytic central potentials. We optimize the exponential parameters of the Slater-type orbitals for 10 elements, in particular H, C, N, O, F, Si, P, S, Cl, and Br, to minimize the mean unsigned error (MUE) of the QM/MM electrostatic and induction energies with respect to the Hartree−Fock electrostatic energies and the sum of induction and induction-exchange energies calculated by symmetry-adapted perturbation theory (SAPT). The resulting optimized exponential parameters are very physical, which allows one to assign parameters to all nonmetal elements (except rare gases) with atomic number less than or equal to 35. For a test set of complexes, the improved description of MM charge densities reduces the error of electrostatic interactions between QM and MM regions in the QM/MM method from 8.1 to 2.8 kcal/mol and reduces the error of induction interactions from 1.9 to 1.4 kcal/mol.

1. Introduction

The combined quantum mechanical and molecular mechanical (QM/MM) method is a useful tool to model large and complex systems.[1−5] It has been widely used in modeling complex molecules, condensed-phase chemistry, materials, and homogeneous, enzymatic, and heterogeneous catalysis. The electronically embedded QM/MM approach, in which interactions between QM electrons and MM partial charges are added as one-electron integrals into the QM Hamiltonian, allows the polarization of the QM electron density by the MM environment; therefore, it is a more accurate embedding scheme than mechanical embedding, in which such polarization is neglected.[6] The electrostatic interactions between the QM and MM regions are usually written as:

$$H_{\mathrm{QM/MM}}^{\mathrm{el}} = -\sum_{i,A}\frac{q_A}{r_{iA}} + \sum_{\alpha,A}\frac{Z_\alpha q_A}{R_{\alpha A}} \qquad (1)$$

where q_A are the MM point charges; the indices i and α run over all QM electrons and nuclei, respectively; and r_{iA} and $R_{\alpha A}$ are the distances between the QM electrons and the MM

* Corresponding author e-mail: truhlar@umn.edu.

10.1021/ct1003862
Published on Web 10/11/2010

point charges and the distances between the QM nuclei and the MM point charges, respectively.

The use of MM partial atomic charges as point charges at the positions of the nuclei, as in eq 1, is a very popular way to parametrize the electrostatics in molecular modeling, and it is sometimes called the distributed monopole approximation. However, an MM point charge need not be a good model for the electron density of the MM subsystem. Four possible improvements can be considered, in particular, the addition of higher-order multipole contributions at each nuclear center,[7−9] the use of off-nuclei point charges,[10,11] the inclusion of penetration effects,[12−14] and the treatment of additional quantum mechanical effects associated with the distributed charge distribution.[15−27] In a general way, the first three approaches all account for the same effect, namely, that the actual electron density has more structure than a collection of point charges. The distributed multipole method accounts for the asymmetry of the charge distribution of an atom in a molecule, the use of off-nuclei charge centers accounts for both asymmetry and finite orbital extent, and the penetration modeling accounts for finite orbital extent. One can include both higher multipole moments and penetration effects,[20,28−30] but in the present study, we concentrate solely on the penetration effects. This has the advantage that we retain the radial, pairwise functional form for the contribution of each atom to the molecular electrostatic potential. The fourth approach is beyond the present scope, but it can include, for example, exchange repulsion or the effects of embedding atoms on the matrix elements of the QM atoms.

The essence of penetration effects, which cannot be described by MM point charges or distributed multipoles, is that, when two atoms are close enough, their charge densities can overlap, and the shielding of the nuclear charge of each atom by its own electron density decreases. Various approaches have been suggested to include this effect in both MM calculations and QM/MM calculations.[13,14,16,20,28−41] In the MM studies, the MM charge densities were represented by Gaussian multipoles,[32] point charges with damping functions,[14,20,30,41] a set of s-type Gaussian functions,[29,35,36] spherical atomic charge densities using Hartree−Fock-limit wave functions,[37] and single Slater-type contracted Gaussian multipole charge densities.[40] The coefficients of Gaussian-type functions or damping functions were chosen to fit the electron density and electrostatic potential,[20,30,32,35−37,39,40] the electrostatic energy,[14,41] or liquid-state properties,[33] or they were based on the intermolecular overlap.[29] Similar strategies have also been used in QM/MM calculations in which charges with damping functions[14,16] and a set of s-type Gaussian functions[38] have been employed to include the charge penetration effects. Gaussian-type charge densities were used to mimic the real charge distributions of MM atoms in the link atom and double link atom methods.[13,34]

In the present work, we describe a simple scheme to include penetration effects in QM/MM calculations. Because the inclusion of distributed multipoles complicates the implementation in QM/MM calculations, and the MM point charges fitted to electrostatic-potential (ESP) can effectively include some contributions due to higher-order multipoles,[3] we do not add multipole refinements to the MM charges in this study. Moreover, it was pointed out by Cisneros et al. that damping of atom-centered point charges is more important than using distributed multipoles.[14] The basic idea of the method advanced here is using a charge screened by a Slater-type orbital[42] (STO) to represent the outermost portion of the charge density of an MM atom. The parameter used to specify the spatial extent of the screened charge is optimized to give the best agreement with Hartree−Fock electrostatic energies and with induction energies computed by symmetry-adapted perturbation theory[43] (SAPT). In section 2, we will derive an expression for the screened MM charge. In section 3, we will present the method to optimize the exponential parameters of the STOs for different elements, and we will present the test suite and implementation details. Section 4 gives the optimized parameters and an analysis of the resulting accuracy. Finally, section 5 summarizes the main conclusions.

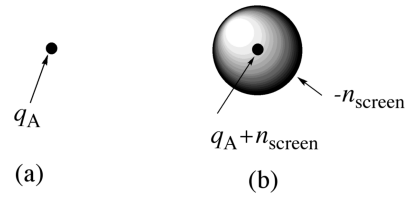

Figure 1. Comparison between (a) a point charge model and (b) a screened charge model of an MM atom A. The total smeared charge in model (b) is $-n_{screen}$, representing n_{screen} electrons.

2. Theory

All variables and equations are in atomic units. We define a normalized STO by

$$\varphi = Ar^{n-1}\exp(-\zeta r) \tag{2}$$

where A is the normalization constant (normalizing φ^2 to integrate to unity), r is the distance of the electron from the nucleus, and n is the highest principal quantum number of the element. In particular, $n = 1$ for H and He, $n = 2$ for Li through Ne, $n = 3$ for Na through Ar, and $n = 4$ for K through Kr. The exponential parameter ζ is a parameter that depends on the atomic number.

We make the assumption that the charge density of an atom in the MM subsystem can be represented by two components: (i) a smeared charge Q distributed like electrons in the orbital φ and (ii) the rest of the charge, which is located at the nucleus. Since the net charge on an atom A in the MM subsystem is a parameter q_A, the charge at the nucleus is $q_A - Q$. Since the smeared charge represents the outer electrons, Q is negative and it is convenient to define $n_{screen} = -Q$. Then the charge density $\rho_A(r)$ of the smeared charge on atom A is given by

$$\rho_A(r) = -n_{screen}[Ar^{n-1}\exp(-\zeta r)]^2 \tag{3}$$

Figure 1 compares the point charge model and the screened charge model.

Table 1. Clementi−Raimondi Exponential Parameters for the Outermost Orbitals[a]

atom	H	C	N	O	F
parameter	1.00	1.57	1.92	2.23	2.55
atom	Si	P	S	Cl	Br
parameter	1.43	1.63	1.83	2.04	2.26

[a] Reference 44.

The exponential parameter ζ for the STO determines the spatial extent of the smeared charge distribution of the atom. It will be optimized for several elements (H, C, N, O, F, Si, P, S, Cl, and Br) in the present study. Our initial guess of the ζ values for these elements are the exponential parameters of their outermost orbitals (that is, 1s for H and He, 2s for Li through Be, 2p for B through Ne, 3s for Na through Mg, 3p for Al through Ar, 4s for K through Zn, and 4p for Ga through Kr) as optimized by Clementi and Raimondi,[44] and they are shown in Table 1 for the elements considered in this study. Another possible choice of initial guess would be the ζ values optimized by Cusachs et al. to reproduce the overlap integrals.[45]

It can be shown that the electrostatic potential $U(r)$ at a distance r from the nucleus of an atom can be written as[46]

$$U(r) = \frac{Z}{r} - 4\pi\left[\frac{1}{r}\int_0^r \rho(r')r'^2 dr' + \int_r^\infty \rho(r')r'dr'\right] \quad (4)$$

where $\rho(r)$ is the atomic electron density and Z is the nuclear charge. In our screened charge model, eq 4 becomes

$$U(r) = \frac{q_A + n_{\text{screen}}}{r} + 4\pi\left[\frac{1}{r}\int_0^r \rho_A(r')r'^2 dr' + \int_r^\infty \rho_A(r')r'dr'\right] \quad (5)$$

where ρ_A is given by eq 3. [Notice that $\rho(r)$, being an electron density, is positive, whereas $\rho_A(r)$, being a smeared charge density, is negative, just as Q is negative and n_{screen} is positive.] Integrating eq 5 yields

$$U(r) = \frac{q_A + n_{\text{screen}} f(\zeta r)\exp(-2\zeta r)}{r} \quad (6)$$

where the polynomial factor f is

$$\begin{aligned} f(\zeta r) &= 1 + \zeta r && n = 1 \\ &= 1 + \frac{3}{2}\zeta r + (\zeta r)^2 + \frac{1}{3}(\zeta r)^3 && n = 2 \\ &= 1 + \frac{5}{3}\zeta r + \frac{4}{3}(\zeta r)^2 + \frac{2}{3}(\zeta r)^3 + \frac{2}{9}(\zeta r)^4 + \frac{2}{45}(\zeta r)^5 && n = 3 \\ &= 1 + \frac{7}{4}\zeta r + \frac{3}{2}(\zeta r)^2 + \frac{5}{6}(\zeta r)^3 + \frac{1}{3}(\zeta r)^4 + \frac{1}{10}(\zeta r)^5 + \frac{1}{45}(\zeta r)^6 + \frac{1}{315}(\zeta r)^7 && n = 4 \end{aligned} \quad (7)$$

The electrostatic potential of a point charge is

$$U_A(r) = \frac{q_A}{r} \quad (8)$$

and the electrostatic potential of a screened charge can be written as

$$U_A(r) = \frac{q_A^*}{r} \quad (9)$$

where

$$q_A^* = q_A + n_{\text{screen}} f(\zeta r)\exp(-2\zeta r) \quad (10)$$

We substitute q_A by q_A^* as the MM charge in eq 1, and the first term of eq 1 enters into the QM Hamiltonian in our QM/MM calculations.

Our formula is similar to eq 3 in ref 14, but there are some differences. Their eq 3 can be rewritten as

$$q_A^* = q_A - (N_{\text{val}} - q_A)\exp(\Omega_{ij}(r)) \quad (11)$$

where N_{val} is the number of valence electrons, and $\Omega_{ij}(r)$ is a factor depending on the distance between atoms i and j, which satisfies

$$\Omega_{ij}(r) \rightarrow \begin{cases} 0 & \text{as } r \rightarrow 0 \\ -\infty & \text{as } r \rightarrow \infty \end{cases} \quad (12)$$

When r approaches infinity, both eqs 10 and 11 become $q_A^* = q_A$; however, when r approaches 0, q_A^* in eq 10 becomes more positive, but q_A^* in eq 11 becomes more negative. When the MM atom becomes close to a QM nucleus, the screening effect of the electrons of the MM atoms decreases, and the QM region experiences a more positive electrostatic potential; therefore, it is more physical to use eq 10. It seems likely that the first negative sign in eq 11 arose from an algebra error.

We explored two possible ways to parametrize eq 10. In one way, we assume that all N_{val} valence electrons are described by the STO, and we define n_{screen} as

$$n_{\text{screen}} = N_{\text{val}} - q_A \quad (13)$$

In the other way, since eq 10 will be parametrized to reproduce Hartree−Fock electrostatic and induction energies in the van der Waals region, we recognize that the parameter ζ will be suitable for describing the outermost fringe of an electron density, which may be appropriate for only one electron or a proper fraction of an electron. Moreover, the charge located at the MM nucleus equals $q_A + n_{\text{screen}}$, and it can be quite large if we use eq 13 (e.g., 6.0 for an oxygen). Since no exchange repulsion is added in the QM/MM self-consistent field (SCF) optimizations, this large charge at the MM nucleus can cause overpolarization of the QM region, especially for basis sets with a large number of diffuse functions. Therefore, we reduce the number of electrons that we treat as smeared to

$$n_{\text{screen}} = \min\{N_{\text{val}} - q_A, 1\} \quad (14)$$

In our tests, we found that the second choice is able to represent the penetration effects while avoiding overpolarization, so we only give detailed results for this choice. Looking ahead, we note that for nine of the 10 elements parametrized in this paper, eq 14 will yield $n_{\text{screen}} = 1$. The exception is hydrogen. Moreover, we found that the screening of metals (Na and Al in the test suite) does not systematically improve the results, which is possibly because the penetration effects are small for the positively charged cations. Therefore,

only the nonmetals are screened in the screened charge schemes and the metals are always treated as point charges, even when they appear in the same molecule in which nonmetal charges are screened.

3. Methods

3.1. Theoretical Framework. To calculate the accurate electrostatic energy and induction energy, we use symmetry-adapted perturbation theory[43] (SAPT) to partition the total interaction energy E_{int} into a sum of several physically distinct contributions. We will use three of these contributions: $E_{elst}^{(10)}$ represents the Hartree−Fock electrostatic interaction of the unperturbed monomers' charge distributions; $E_{ind}^{(20)}$ is the Hartree−Fock induction energy and equals the sum of $E_{ind}^{(20)}$(A←B) and $E_{ind}^{(20)}$(B←A), where $E_{ind}^{(20)}$(A←B) and $E_{ind}^{(20)}$(B←A) are the Hartree−Fock induction energy of monomer A with the static field of the monomer B, and vice versa; and $E_{ind\text{-}exch}^{(20)}$ is the induction exchange energy. The sum of $E_{ind}^{(20)}$ and $E_{ind\text{-}exch}^{(20)}$ is called the damped-induction energy and is denoted by $E_{damp\text{-}ind}^{(20)}$ in this study.

The first question is how to compare these quantities to the components of a QM/MM calculation in which the QM method is a self-consistent-field (SCF) calculation. The SCF calculation can be either the Hartree−Fock approximation or a density functional calculation; we will use Hartree−Fock in the present paper. We consider a dimer comprised of two monomers, A and B. We can place either monomer A or monomer B in the QM region, and the other one is placed in the MM region. $E_{elst}^{QM/MM}$ is the interaction energy between the screened MM charges and the QM unpolarized electron density from a separate monomer calculation. $E_{ind}^{QM/MM}$ is derived in two steps. First, we place monomer A in the QM region and monomer B in the MM region, and we compare the interaction between the screened MM charges and the QM unpolarized electron density to the interaction between the screened MM charges and the QM electron density that is relaxed in the presence of the screened MM charges. The difference is $E_{ind}^{QM/MM}$(A←B). Second, we switch the QM and MM regions, and calculate $E_{ind}^{QM/MM}$(B←A) in the same way as $E_{ind}^{QM/MM}$(A←B). The induction energy $E_{ind}^{QM/MM}$ is the sum of these two terms.

The comparison of $E_{elst}^{QM/MM}$ and $E_{ind}^{QM/MM}$ from QM/MM calculations to $E_{elst}^{(10)}$, $E_{ind}^{(20)}$, and $E_{ind\text{-}exch}^{(20)}$ from SAPT calculations needs special attention. The $E_{elst}^{QM/MM}$ of the QM/MM calculation can be directly compared with $E_{elst}^{(10)}$ of the SAPT calculation. However, $E_{ind}^{QM/MM}$ in QM/MM calculations does not have the same meaning as $E_{ind}^{(20)}$ in SAPT. In the QM/MM calculations, the induction energy $E_{ind}^{QM/MM}$ includes all orders of the perturbation of the induction, while only the second-order perturbation energy is calculated in $E_{ind}^{(20)}$ in SAPT. Moreover, in SAPT calculations, the induction exchange energy $E_{ind\text{-}exch}^{(20)}$ is always evaluated and added to the induction energy $E_{ind}^{(20)}$ to give a reasonable estimation of the total interaction energy. In QM/MM calculations, it is hard to add this induction exchange energy in an empirical way, as may more readily be done for the static exchange energy, because the induction exchange energy is not even approximately pairwise additive. Therefore, we decide to compare the QM/MM induction energy $E_{ind}^{QM/MM}$ with $E_{damp\text{-}ind}^{(20)}$, that is, with the sum of the induction $E_{ind}^{(20)}$ and induction-exchange $E_{ind\text{-}exch}^{(20)}$ energies of the SAPT calculations. The same strategy has sometimes been adopted in the development of polarizable MM force fields.[36,47]

3.2. Basis Sets and MM Charges. Two Gaussian-type basis sets were used for the QM/MM and SAPT calculations, namely, the aug-cc-pVTZ basis set of Dunning and co-workers[48,49] and the def2-TZVP basis set from TURBOMOLE.[50] Although in applications one might use a smaller number of diffuse functions than are present in the aug-cc-pVTZ basis set, it is important to use a large diffuse space during parametrization to be sure that the parametrized model is stable against overpolarization catastrophes.

The MM partial atomic charges are Hartree−Fock ChElPG charges[51] for the separated monomers using the same basis set as the QM method for each of the QM/MM calculations. For example, when the aug-cc-pVTZ basis set is used as the QM method in the QM/MM calculations, the MM charges are also derived using the aug-cc-pVTZ basis set.

3.3. Optimization Methods and Software. We optimize the ζ values for the 10 elements (H, C, N, O, F, Si, P, S, Cl, and Br) in our test suite in order to fit the QM/MM electrostatic energies to SAPT results. The error function is based on the difference between the QM/MM and SAPT Hartree−Fock electrostatic and damped-induction energies

$$\mathrm{MUE} = \sum_{B=1}^{2}\sum_{M=1}^{\text{molecules}}\sum_{G=1}^{3}\Big(\sum_{i=1}^{2}\big|E_{elst}^{QM/MM}(\mathrm{QM/MM};B,M,\mathrm{G},i) - E_{elst}^{(10)}(\mathrm{SAPT};B,M,G)\big| + \big|E_{ind}^{QM/MM}(\mathrm{QM/MM};B,M,G) - E_{damp\text{-}ind}^{(20)}(\mathrm{SAPT};B,M,G)\big|\Big) \quad (15)$$

where B labels the basis sets, M labels the molecules in the database (see below), G labels the geometries for each molecule in the database (see below), and i denotes which monomer is treated as QM. Equation 15 is minimized in the parametrization.

The SAPT calculations are performed with SAPT2008 program package[52] interfaced to the version E.01 of *Gaussian 03*[53] integral and self-consistent-field package. All QM/MM calculations are carried out using our own QMMM program, which is based on the version D.01 of *Gaussian 03*[53] and TINKER[54] programs. We use a modified version of 1.3.5.[55] The optimized structures of some of the dimers in the test suite (see below) are acquired using a locally modified version of *Gaussian 03* (MN-GFM[56]).

3.4. Test Suite. We included 40 dimers in our database, as shown in Figures 2−5. Because some molecules are too large for SAPT analysis using aug-cc-pVTZ, we do calculations with the aug-cc-pVTZ basis set for only 29 out of 40 molecules in the test suite; these 29 consist of those in Figures 3 and 5 plus the ones that are labeled by asterisks (*) in Figure 2. Thus the first two sums in eq 15 encompass 69 cases, not 80.

We considered three geometries for each of the dimers: the equilibrium geometry, a compressed geometry, and an extended geometry. The equilibrium geometry of all mol-

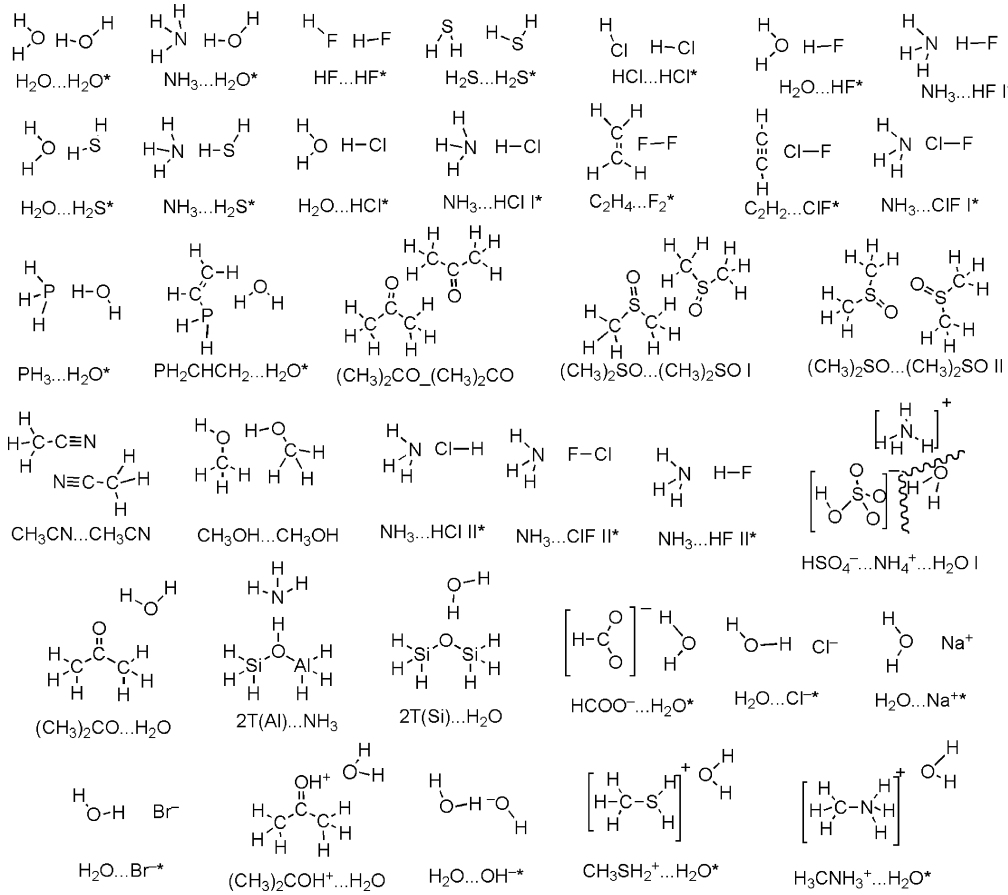

Figure 2. Thirty-six of 40 dimers in the test suite. We use * to label the molecules that are tested using both aug-cc-pVTZ and def2-TZVP basis sets.

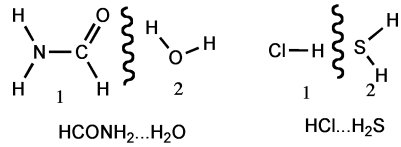

Figure 3. Geometry of the 37th and 38th dimers.

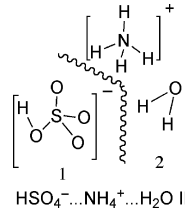

Figure 4. Geometry of the 39th dimer.

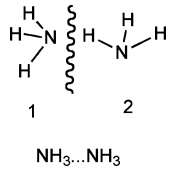

Figure 5. Geometry of the 40th dimer.

ecules is from several sources: (i) the HB6/04, CT7/04, and DI6/04 databases[57] are the first choice, when the dimer is present in one of them; (ii) $H_2O\cdots OH^-$ is the QCISD/MG3S-optimized structure;[58] and (iii) $NH_3\cdots$HCl II, $NH_3\cdots$ClF II, and $NH_3\cdots$HF II are based on the geometry of $NH_3\cdots$ClF I in the CT7/04 database;[57] For the HCl, ClF, and HF monomers in the $NH_3\cdots$HCl II, $NH_3\cdots$ClF II, and $NH_3\cdots$HF II dimers, the Cl, F, and H are placed at the same position as the Cl in $NH_3\cdots$ClF I, and the Cl−H, F−Cl, and H−F are placed along the same direction as Cl−F in $NH_3\cdots$ClF I with the bond lengths equal 1.282, 1.701, and 0.924 Å, respectively. (iv) Other dimers are M06-2X[59]-optimized structures; the basis set we used for these optimizations is MG3S[60] for H through Cl, and 6-311+G(3d2f)[61] for Br.

In the compressed geometry, we move the monomers closer, without changing their internal structures, along a line connecting their centers of mass until the distance between the centers of mass of the two monomers is 10% *shorter* than that in the equilibrium geometry. In the extended geometry, we move the monomers farther apart along a line connecting their centers of mass until the distance between the centers of mass of the two monomers is 10% *longer* than that in the equilibrium geometry.

3.5. Implementation in Gaussian 03. As is shown in eq 10, the screened MM charge contains a Slater-type function. To facilitate the implementation of the Slater-type function into the Hamiltonian of the QM/MM calculations, we expand a single Slater-type function in terms of three Gaussian-type functions, as follows:[62,63]

$$\exp(-\lambda r) = \sum_{i=1}^{3} C_i \exp(-\alpha_i \lambda^2 r^2) \qquad (16)$$

Table 2. Contraction Coefficients and Exponential Parameters[a]

	C_i	α_i
1	0.107 150	0.109 818
2	0.343 808	0.405 771
3	0.355 483	2.227 660

[a] References 62 and 63.

The contraction coefficients C_i and exponential parameters α_i in eq 16 are listed in Table 2. The electrostatic potentials generated by the screened MM charges are incorporated in version D.01 of *Gaussian 03*[53] as pseudopotentials using the keyword "pseudo". In the Gaussian implementation, the pseudopotentials do not interact with the QM nuclei. Therefore, in the QMMM program revised for this work, we calculate the interactions between the pseudopotentials and the QM nuclei, and we add them to the energy of the system.

4. Results and Discussion

4.1. SAPT Results. We computed averages over the SAPT results in order to give an indication of the typical values of the various terms. The averaged SAPT results of the electrostatic, induction, and damped-induction (induction plus induction-exchange) energies with two basis sets and three geometries are shown in Table 3. For the aug-cc-pVTZ basis set, they are averaged over 29 dimers. For the def2-TZVP basis set, they are averaged over 40 dimers. To compare the results for different basis sets, we also averaged the def2-TZVP results over only 29 dimers (the same molecules as those tested by the aug-cc-pVTZ basis set), with the results shown in Table 4. The results in Table 4 can be compared with the aug-cc-pVTZ results in Table 3 to see the energy differences when using different basis sets.

Table 3 shows that the average SAPT electrostatic, induction, and damped-induction energies over all geometries and basis sets are −19, −13, and −4 kcal/mol, respectively. Comparing the results from Tables 3 and 4, we found that for the electrostatic energy, the induction energy, and the damped-induction energy, the differences between two basis sets are less than 7% of the averaged values. The aug-cc-pVTZ basis set contains a larger number of diffuse functions than the def2-TZVP basis set. While these functions may be important for the dispersion interactions, the electrostatic and induction energies between the monomers can be described quite well without them. Using basis sets with smaller numbers of diffuse functions not only decreases the cost of the calculations but also it avoids the overpolarization in QM/MM calculations when large charges are placed near the boundary.

Comparing three different geometries, we found that both the electrostatic and induction energies become more negative when the two monomers are placed closer, while the exchange energy becomes more positive. The absolute value of the damped-induction energy is smaller than the induction energy, and it changes more slowly with geometry.

In the following discussion, we will use the electrostatic energy and damped-induction energy from SAPT calculations as benchmarks to evaluate how well different charge schemes treat charge penetration effects in the QM/MM calculations.

4.2. QM/MM Results with Point Charge Scheme and Screened Charge Scheme. We define the number of electrons in the STO using eq 14. The ζ value for each element is optimized to minimize the MUE in eq 15. The optimized values are shown in Table 5. Note that the metal elements (Na and Al) are always treated as point charges (i.e., $\zeta = \infty$). Contrary to our initial guess that the optimized ζ values would be close to the Clementi−Raimondi exponential parameters for the outermost orbitals, they are instead similar (except for H) to half of the Strand and Bonham[64] exponential parameters for the outermost component ($^a\lambda_1$ for H−Ar and $^b\lambda_3$ for K−Kr in their paper) of the density (the factor of 1/2 results from the fact that the screened charge used in this study has an exponential parameter of 2ζ because the orbital is squared, not ζ); these values are also shown in Table 5. The mean signed error (MSE) and mean unsigned error (MUE) of the electrostatic and damped-induction energies using MM point charges, MM screened charges with optimized parameters, and MM screened charges with modified Strand−Bonham (MSB) parameters (half of the Strand−Bonham parameters for all elements except H and the optimized parameter for H) are listed in Tables 6 and 7 respectively.

Table 6 shows that the MUE of the electrostatic energy using the screened charge scheme with optimized parameters is 2.8 kcal/mol, compared with 8.1 kcal/mol for the traditional point charge scheme. The screened charge scheme with the MSB parameters gives an MUE of 3.1 kcal/mol, which is quite close to the optimized result. Table 7 shows that the MUE of the QM/MM induction energy also decreases from 1.9 to 1.4 kcal/mol when using the screened charge schemes, although the improvement is not as significant as the electrostatic energy (a factor of 1.4 for the induction energy vs a factor of 2.9 for the electrostatic energy). The fact that our optimized STO exponential parameters are very close to the values we derived from the Strand−Bonham fits to atomic densities shows that the model parameters are very physical. The results clearly show that the inclusion of penetration effects improves the description of the MM electrostatic potential and its effect on the QM system. As the penetration effects decrease exponentially as a function of the distance between atoms, the improvement is most significant for the compressed geometry, in which the penetration effects are large. For the extended geometry, in which the monomers are placed farther from each other, the difference between the point charge scheme and the screened charge scheme is relatively small.

We also tested the screened charge scheme using eq 13, rather than eq 14, to define the number of screened electrons in the Slater-type orbital, and optimized the parameters to reproduce the SAPT electrostatic energies. The electrostatic energies can be well-reproduced. However, we found that the induction energy is greatly overestimated, especially when the aug-cc-pVTZ basis set is used. The error is hard to control; therefore, we abandoned eq 13.

One issue not so far considered is that in comprehensively parametrized force fields, one can, to some extent, make up

Table 3. Averaged SAPT Electrostatic, Induction, and Damped-Induction Energies (kcal/mol) Using Three Geometries and Two Basis Sets

	equilibrium/ acTZ[a]	equilibrium/ def2-TZVP	compressed/ acTZ	compressed/ def2-TZVP	extended/ acTZ	extended/ def2-TZVP	all
electrostatic	−14.50	−18.78	−25.58	−30.89	−9.20	−12.80	−18.98
induction	−9.61	−9.58	−24.86	−25.14	−4.14	−4.08	−12.91
damped-induction	−3.18	−3.49	−7.87	−8.58	−1.63	−1.77	−4.45

[a] acTZ represents aug-cc-pVTZ.

Table 4. Averaged SAPT Electrostatic, Induction, and Damped-Induction Energies (kcal/mol) Using Three Geometries and the def2-TZVP Basis Set over Only 29 Molecules

	equilibrium/ def2-TZVP	compressed/ def2-TZVP	extended/ def2-TZVP
electrostatic	−15.17	−26.32	−9.80
induction	−9.55	−24.82	−4.08
damped-induction	−3.07	−7.73	−1.53

Table 5. ζ Values Used in the Slater-type Orbital

atom	H	C	N	O	F
optimized parameters	1.32	0.92	0.92	1.20	1.16
Strand and Bonham[a]	1.00	0.87	1.01	1.12	1.24
atom	Si	P	S	Cl	Br
optimized parameters	0.73	0.68	0.90	0.98	0.91
Strand and Bonham[a]	0.74	0.81	0.88	0.95	1.01

[a] Half of the values in ref 64.

for errors in the electrostatics by parametrization of other MM parameters or by parametrization of the MM charges. However, neither approach is satisfactory. The first approach introduces systematic errors in the QM subsystem of combined QM/MM calculations because the electrostatic terms enter the QM Hamiltonian, but the other MM terms do not affect the quantum mechanical electronic Hamiltonian. The first approach is also unsatisfactory for MM calculations because electrostatic interactions have a different functional form than the usual Lennard-Jones or Buckingham forms used for other MM nonbonded terms and therefore cannot be completely mimicked by them. The second approach introduces systematic errors because the electrostatic potential due to an unscreened Coulomb interaction with a modified charge has a different dependence on geometry than a screened Coulomb interaction with the correct charge.

4.3. Case Studies. We use the $HCONH_2\cdots H_2O$, $HSO_4^-\cdots NH_4^+\cdots H_2O$ II, and $HCl\cdots H_2S$ dimers as three examples to illustrate the use of the new scheme. The equilibrium geometries of these three dimers are illustrated in Figures 3 and 4. In the calculations, either monomer 1 or 2 can be treated as the subsystem in the QM region. Tables 8−10 show the SAPT electrostatic and damped-induction energies and the QM/MM electrostatic and induction energies with the three charge schemes we considered, namely, unscreened point charges, screened charges with optimized STO parameters, and screened charges with MSB parameters for the STOs.

Table 8 shows the results for the $HCONH_2\cdots H_2O$ dimer. Including penetration effects in the screened charge scheme yields a much closer match to the SAPT results. The absolute error in electrostatic energy, averaged over three geometries and two basis sets, decreases from 5.4 to 1.3 kcal/mol. The point charge scheme always underestimates the magnitude of the electrostatic and induction energies. Screening the point charge with the optimized parameters makes the electrostatic energy more negative by 9.3 kcal/mol for the compressed geometry, 2.8 kcal/mol for the equilibrium geometry, and 0.7 kcal/mol for the extended geometry, when averaged over two QM/MM calculations with different QM regions and two basis sets.

Although the SAPT damped-induction energies are similar for the aug-cc-pVTZ and def2-TZVP basis sets, the QM/MM induction energies are not. Taking the compressed geometry as an example, the SAPT damped-induction energies are −4.6 and −4.5 kcal/mol for the aug-cc-pVTZ and def2-TZVP basis sets, while the QM/MM induction energies using the screened charge scheme with the optimized parameters are −6.3 and −3.7 kcal/mol for aug-cc-pVTZ and def2-TZVP basis sets. The induction energies using the aug-cc-pVTZ basis set are more negative than those using the def2-TZVP basis set in the QM/MM calculations. As the aug-cc-pVTZ basis set contains more diffuse functions than the def2-TZVP basis set, the QM region is more prone to be polarized by MM point charges when the aug-cc-pVTZ basis set is used, which leads to the increase of the induction energy. This effect is more significant for the screened charge scheme than the point charge scheme, because the charge at the MM nucleus is larger in the screened charge scheme than that in the point charge scheme, which causes greater polarization. Despite this sensitivity, the average error in the induction energies is reduced by a about a factor of 2 when screening is included.

We also discuss a dimer composed of HSO_4^- and $NH_4^+\cdots H_2O$, with the results shown in Table 9. Only the def2-TZVP basis set is used. This is a very challenging test for the charge scheme because both of the monomers are charged. From the results, we can see that the screened charge scheme improves both the electrostatic and induction energies significantly.

The results for the $HCl\cdots H_2S$ dimer listed in Table 10 are not as good as those for the two examples that we have discussed. The improvement is not very significant when H_2S is in the MM region. The performance of the screened charge scheme for individual elements will be discussed in section 4.4.B.

In the examples singled out for illustration in this section, we found that the screened charge scheme with modified Strand−Bonham (MSB) parameters sometimes gives a better result than that with optimized parameters. This is because the optimized parameters are optimized for all the molecules in the test suite, rather than optimized for the specific dimers.

Table 6. MSE and MUE of Electrostatic Energies (kcal/mol) Using the QM/MM Method[a]

	equilibrium/ acTZ		equilibrium/ def2-TZVP		compressed/ acTZ		compressed/ def2-TZVP		extended/ acTZ		extended/ def2-TZVP		all	
	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
Pt charge[b]	6.43	6.43	6.43	6.43	14.80	14.80	15.10	15.10	3.02	3.02	2.98	2.98	8.13	8.13
Opt[c]	1.58	1.97	1.32	2.17	3.64	5.09	2.61	5.31	1.06	1.11	1.03	1.16	1.84	2.81
MSB[d]	1.59	2.22	1.19	2.29	3.55	5.39	2.15	5.90	1.07	1.25	1.00	1.24	1.71	3.06

[a] Exact values are SAPT electrostatic energies. [b] Point charge scheme. [c] Screened charge scheme with optimized parameters. [d] Screened charge scheme with modified Strand-Bonham (MSB) parameters.

Table 7. MSE and MUE of the Induction Energies (kcal/mol) Using the QM/MM Method[a]

	equilibrium/ acTZ		equilibrium/ def2-TZVP		compressed/ acTZ		compressed/ def2-TZVP		extended/ acTZ		extended/ def2-TZVP		all	
	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
Pt charge[b]	0.67	1.09	1.12	1.18	3.46	3.98	4.34	4.46	0.12	0.45	0.35	0.38	1.72	1.94
Opt[c]	−0.74	1.00	0.69	0.82	−0.62	2.67	2.68	3.13	−0.34	0.45	0.25	0.30	0.46	1.40
MSB[d]	−0.79	1.02	0.68	0.80	−0.71	2.65	2.60	3.12	−0.36	0.46	0.25	0.29	0.42	1.39

[a] Exact values are SAPT damped-induction energies. [b] Point charge scheme. [c] Screened charge scheme with optimized parameters. [d] Screened charge scheme with modified Strand−Bonham (MSB) parameters.

Table 8. Electrostatic and Induction Energies (kcal/mol) of $HCONH_2\cdots H_2O$ Dimer in QM/MM Calculations Compared with SAPT Results, and MUE (kcal/mol) of QM/MM Calculations over Three Geometries and Two Basis Sets

		equilibrium/ acTZ	equilibrium/ def2-TZVP	compressed/ acTZ	compressed/ def2-TZVP	extended/ acTZ	extended/ def2-TZVP	MUE
				Electrostatic				
SAPT		−11.3	−11.6	−21.9	−22.3	−6.6	−6.9	
QM/MM	Pt charge	−7.4/−7.3[a]	−8.0/−7.5	−10.5/−11.3	−11.4/−11.5	−5.1/−5.0	−5.5/−5.1	5.4
QM/MM	Opt	−10.5/−9.8	−10.8/−10.0	−22.0/−18.6	−22.5/−18.8	−5.9/−5.7	−6.2/−5.8	1.3
QM/MM	MSB	−10.8/−10.5	−11.2/−10.7	−23.4/−19.5	−23.8/−19.8	−6.0/−6.0	−6.3/−6.1	1.1
				Induction				
SAPT (damped)		−1.7	−1.6	−4.6	−4.5	−0.7	−0.7	
QM/MM	Pt charge	−1.2	−1.1	−2.8	−2.5	−0.6	−0.5	0.9
QM/MM	Opt	−2.0	−1.3	−6.3	−3.7	−0.7	−0.5	0.5
QM/MM	MSB	−2.1	−1.3	−6.8	−3.9	−0.8	−0.5	0.6

[a] *x*/*y* denotes that the electrostatic energy is *x* when $HCONH_2$ is the QM region and is *y* when H_2O is the QM region.

Table 9. Electrostatic and Induction Energies (kcal/mol) of $HSO_4^-\cdots NH_4^+\cdots H_2O$ II in QM/MM Calculations Compared with SAPT Results, and MUE (kcal/mol) of QM/MM Calculations over Three Geometries

		equilibrium/def2-TZVP	compressed/def2-TZVP	extended/def2-TZVP	MUE
			Electrostatic		
SAPT		−144.1	−176.4	−124.0	
QM/MM	Pt charge	−126.4/−130.2[a]	−135.0/−147.6	−116.4/−117.5	19.3
QM/MM	Opt	−149.4/−138.7	−190.9/−161.5	−124.7/−121.6	7.2
QM/MM	MSB	−146.0/−139.4	−184.6/−162.2	−123.2/−122.3	5.2
			Induction		
SAPT (damped)		−15.7	−30.6	−9.8	
QM/MM	Pt charge	−12.2	−17.4	−8.8	5.9
QM/MM	Opt	−15.7	−26.7	−9.9	1.4
QM/MM	MSB	−14.9	−25.4	−9.5	2.1

[a] *x*/*y* denotes that the electrostatic energy is *x* when HSO_4^- is the QM region and is *y* when NH_4^+ (H_2O) is the QM region.

4.4. Discussion of ζ Parameters. *A. ζ Value Effects on Electrostatic and Induction Energies.* In order to understand how the ζ parameters of the STOs affect the electrostatic and induction energies, we here examine the effect of varying ζ on the $NH_3\cdots NH_3$ dimer with the equilibrium geometry and with the aug-cc-pVTZ basis set. The geometry is shown in Figure 5. The electrostatic and induction energies are most sensitive to the nearest MM atom, so we choose to vary the ζ parameter of the nearest MM atom. In Figure 6, we show the change of electrostatic energy with respect to the H ζ value with monomer 1 in the QM region and with the ζ value for N kept at its optimized value of 0.92. In Figure 7, we show the change of electrostatic energy with respect to the N ζ value with monomer 2 in the QM region and with the ζ value for H kept at its optimized value of 1.32. To make the comparison, we draw two lines representing the SAPT result and the QM/MM result without screening H and N, respectively.

We can see that, when the ζ values become larger and larger, the electrostatic energy gets closer to the result from the point charge scheme, as is easy to understand both physically and mathematically. The figures show how optimum ζ values that best reproduce the electrostatic energy

Table 10. Electrostatic and Induction Energies (kcal/mol) of HCl···H_2S Dimer in QM/MM Calculations Compared with SAPT Results, and MUE (kcal/mol) of QM/MM Calculations over Three Geometries and Two Basis Sets

		equilibrium/ acTZ	equilibrium/ def2-TZVP	compressed/ acTZ	compressed/ def2-TZVP	extended/ acTZ	extended/ def2-TZVP	MUE
				Electrostatic				
SAPT		−5.2	−5.4	−10.9	−11.1	−2.8	−3.0	
QM/MM	Pt charge	−1.0/−2.1[a]	−1.1/−2.4	−1.7/−2.7	−1.8/−3.3	−0.6/−1.5	−0.7/−1.7	4.7
QM/MM	Opt	−2.4/−4.5	−2.5/−4.7	−4.0/−10.5	−4.0/−11.1	−1.2/−2.2	−1.2/−2.3	2.2
QM/MM	MSB	−2.5/−4.6	−2.5/−4.9	−3.9/−11.0	−3.9/−11.6	−1.2/−2.2	−1.3/−2.4	2.2
				Induction				
SAPT (damped)		−1.1	−1.0	−3.0	−3.0	−0.5	−0.4	
QM/MM	Pt charge	−0.4	−0.3	−0.8	−0.6	−0.2	−0.1	1.1
QM/MM	Opt	−1.5	−0.5	−5.1	−1.9	−0.4	−0.2	0.7
QM/MM	MSB	−1.5	−0.5	−5.4	−2.0	−0.5	−0.2	0.8

[a] *x*/*y* denotes that the electrostatic energy is *x* when HCl is the QM region and is *y* when H_2S is the QM region.

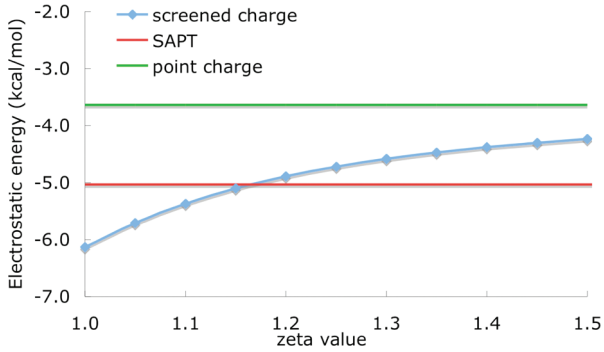

Figure 6. QM/MM and SAPT electrostatic energies (kcal/mol) with respect to the H ζ value.

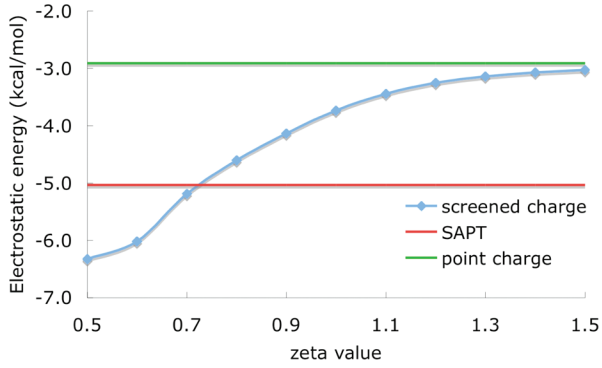

Figure 7. QM/MM and SAPT electrostatic energies (kcal/mol) with respect to the N ζ value.

are located in a reasonable range. The same trend is found for other dimers.

B. Consideration of Individual Elements. In order to understand how the screened charge scheme works for individual elements, we divided the 40 dimers into several groups. Two division schemes have been used. In the first scheme, the dimers are divided on the basis of the elements included in the MM regions. For example, for the group corresponding to element O, we selected all the QM/MM electrostatic calculations in which one or more oxygen atom is included in the MM regions. Because an MM region always contains several different elements, each QM/MM calculation is included in several element groups. Then we calculated the MUE of electrostatic energies for different elements in their groups. The drawback of the first division scheme is that the screening effects of an MM element are not properly reflected in the QM/MM electrostatic energy when the MM element is far from the QM region. Therefore, in the second scheme, only the MM atoms that are close to the QM region are considered. We calculated the distances between all QM and MM atom pairs and identified the shortest QM−MM distance. We define a close MM atom as one that has a distance from any QM atom that is less than the sum of 0.30 Å and the shortest QM−MM distance. Then we divided the 40 dimers on the basis of the elements among the close MM atoms. We carried out the division only for the equilibrium geometry, and we assigned the same group membership for the compressed and extended geometries as in the equilibrium geometry. Because the QM/MM electrostatic energy is most sensitive to the screening of the closest MM atoms, the second division scheme is more appropriate to test the contribution of individual elements. Note that if we change 0.30 Å in the criterion to define a MM atom close to infinity, the second division scheme becomes the first division scheme. We show the number of calculations included in each group and the MUE of the QM/MM electrostatic energy for both the two division schemes in Table 11. We define an improvement ratio (IR) as

$$\mathrm{IR} = \frac{\mathrm{MUE(point\ charge\ scheme)}}{\mathrm{MUE(screened\ charge\ scheme)}} \qquad (17)$$

IRs for individual elements using the screened charge scheme with the optimized parameters are also shown in Table 11, and it is gratifying that they are all greater than or equal to 1.7. The results of the two division schemes show similar trends (except for C, for which the first division scheme does not reflect the screening effect of the C atom).

For the point charge scheme, Cl and N have the largest MUEs. All of them are greater than 10 kcal/mol for both division schemes. The large error of the Si using the first division scheme is mainly due to other elements, as Si is not a close MM atom in any dimers.

For the screened charge scheme with optimized parameters, N is the only element with an MUE of more than 3 kcal/mol by either division, whereas when screening is not employed all elements have an MUE greater than or equal to 3.6 kcal/mol for both divisions. We found that the large error of N is mainly due to the inclusion of NH_3···HF, NH_3···HCl, and 2T(Al)···NH_3 dimers, in which the monomers are very close to each other.

Next, we compare the improvement ratios for individual elements. In the second division scheme, the elements that

Table 11. MUE (kcal/mol) of QM/MM Electrostatic Energies and Improvement Ratios (IR) for Individual Elements

	H	C	N	O	F	Si	P	S	Cl	Br
					First Division Scheme					
number of calculations[a]	372	78	87	171	54	6	12	54	60	6
Pt charge	7.9	6.9	10.2	7.5	9.6	11.0	4.9	7.3	13.7	3.6
Opt	2.9	2.4	5.1	2.6	1.6	1.7	2.9	2.5	2.1	1.4
MSB	3.1	2.7	5.5	2.7	2.0	2.2	3.1	2.6	2.6	1.6
IR (Opt)[b]	2.7	2.9	2.0	2.9	6.1	6.5	1.7	2.9	6.6	2.6
					Second Division Scheme					
number of calculations[a]	213	12	66	96	18	0	12	18	30	6
Pt charge	6.9	5.2	11.1	6.1	5.9		4.9	4.0	18.6	3.6
Opt	1.8	0.7	5.7	2.8	1.2		2.9	2.3	2.7	1.4
MSB	1.8	1.2	6.6	3.0	2.1		3.1	2.2	3.1	1.6
IR (Opt)	3.8	7.7	1.9	2.1	4.9		1.7	1.7	6.9	2.6

[a] This number includes all calculations using different geometries and different basis sets. [b] Improvement ratios (IRs) for individual elements using the screened charge scheme with optimized parameters.

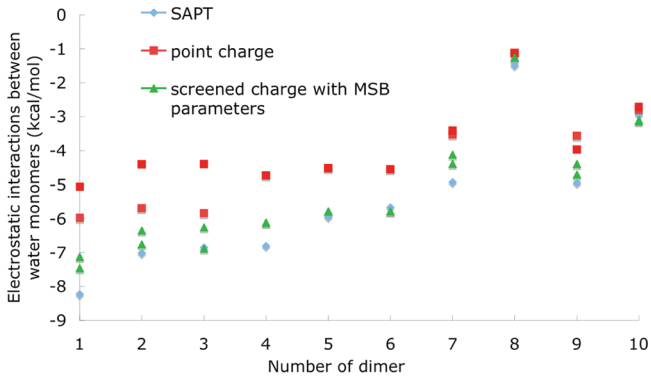

Figure 8. QM/MM and SAPT electrostatic energies (kcal/mol) for 10 water dimers.

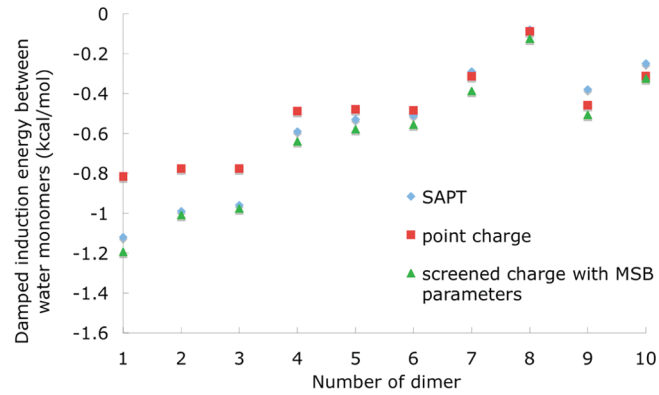

Figure 9. QM/MM induction energies and SAPT damped-induction energies (kcal/mol) for 10 water dimers.

have the largest improvement ratios are C (7.7), Cl (6.9), and F (4.9), followed by H (3.8), Br (2.6), and O (2.1); N (1.9), P (1.7), S (1.7) have the smallest improvement, but even in these cases, the improvement is significant. These findings show that for all elements, electrons distributed in an STO can mimic the charge distribution of the MM atoms well enough to improve the results as compared to a point charge model.

4.5. An Application to 10 Water Dimers with a Variety of Geometries. To test the broader usefulness of the new method, we applied it to water dimers with 10 geometries tested by Reinhardt et al.[65] They have carried out SAPT calculations on these dimers using the aug-cc-pVTZ basis set. We carried out QM/MM calculations with the aug-cc-pVTZ basis set on their geometries using the point charge scheme and the screened charge scheme with MSB parameters, and we compared the results with their benchmark results of electrostatic and induction energies. The point charges of the water monomer are Hartree−Fock ChElPG charges[51] calculated with the aug-cc-pVTZ basis set. Figures 8 and 9 compare the electrostatic and induction energies for these 10 geometries. Because either monomer in the dimer can be treated as the QM region, we have two QM/MM electrostatic energies at each geometry. We found that using the screened charge scheme decreases the MUE of the electrostatic energy from 1.42 to 0.42 kcal/mol and the MUE of the induction energy from 0.11 to 0.06 kcal/mol. Note that the averaged electrostatic and damped-induction energies of the 10 water dimers are −5.48 and −0.57 kcal/mol, respectively. Therefore, the percent error drops from 26% to 8% and from 19% to 11% for the QM/MM electrostatic and induction energies, respectively. These findings show that including penetration effects is important for accurate modeling. We also carried out calculations using the screened charge scheme with optimized parameters, in which the ζ value for O is 1.20 rather than 1.12 as in the MSB parameters. The MUE of electrostatic energy is 0.57 kcal/mol, which is larger than that using MSB parameters, but still much better than that for point charges, and the MUE of the induction energy is 0.05 kcal/mol. We found that if the two monomers in the dimer are not very close (two monomers are very close in the case of $H_2O\cdots OH^-$ in the test suite for parametrization), the ζ value of 1.12 for O gives better results than the optimized value from Table 5.

4.6. Limitations of the Scheme. We note five limitations.

First, only a single STO with a fixed ζ value is used for each element in our scheme. However, a single STO may not describe the electron distribution well in all regions. When the QM electron goes closer to the MM atom, it should feel a charge distribution with a larger ζ value. This can be improved by double-ζ STOs, in which a more diffuse STO is used to describe the ultimate outer layer and a less diffuse one is used to describe the penultimate layer. Moreover, the ζ value for an element changes when the element has different partial charges and bonding environments, as pointed out before,[66] so the ζ value we optimize here is only an average value for an element in different environments. If one is willing to treat ζ as an MM parameter and optimize

it separately for different hybridization states (e.g., sp^2 O would have a different value than sp^3 O) or different functional groups or charge states, one could do even better.

Second, to avoid overpolarization, we use at most one electron in the STO. A small number of electrons in the STO may underestimate the charge penetration effect, especially for the compressed geometry. One may include more electrons in the STO if the overpolarization of the QM region can be avoided (e.g., when def2-TZVP basis set is used for the QM calculations); this would, for example, improve the results in Table 10 when H_2S is the MM region and S is the close MM atom.

Third, when multiply charged cations (such as Zn^{2+} and Mg^{2+}) are placed in the MM region near the QM−MM boundary, we found that the QM region may be greatly overpolarized by the MM region if a large number of diffuse functions are used for the QM region, such as that in the aug-cc-pVTZ basis set. This is due to the lack of exchange and orthogonality interactions between MM and QM regions in the QM/MM SCF optimizations. Although the problem is much less severe with a less diffuse basis set, our method cannot completely solve this problem. Ab initio model potentials[67,68] (AIMP), smeared charges,[13,34] and damped charges[38] can be applied to alleviate the overpolarization. One can also avoid the problem by not placing a QM−MM boundary next to a multiply charged MM cation.

Fourth, we have not tested the method for rare gases.

Fifth, we have so far developed the method only for the case where a QM−MM boundary does not pass through a covalent bond.

If desired, further studies can be carried out to remove limitations 1, 2, 4, and 5 mentioned above. Despite limitations 1−3, the method already provides a much more realistic treatment of the electrostatic and induction energies in the general case.

5. Conclusions

In this paper, we proposed a general screening scheme to include charge penetration effects in the treatment of electrostatic interactions in molecular modeling, and we parametrized it and applied it by using electronically embedded QM/MM calculations. Our scheme utilizes a Slater-type orbital to mimic the outer portion of the electron distribution around the MM atom. By including the charge penetration effects, we can greatly improve the description of electrostatic interactions in the QM/MM method. The parameters for the STOs of several common elements (H, C, N, O, F, Si, P, S, Cl, and Br) are optimized to reproduce the SAPT electrostatic and damped-induction energies. For the metal elements, we suggest keeping the point charge scheme, as no systematic improvement has been found by treating them as screened.

We found that the optimal exponential parameters are very close to the values describing the outermost layer of the electron density of atoms in Strand and Bonham's fits to atomic electron densities. This is extremely encouraging in that it shows that the method is very physical. The finding that screening does not offer systematic improvement for metal atoms is also physical, since metal atoms in molecules usually have a partial positive charge and hence less diffuse electron density. These findings, combined with the availability of Strand and Bonham's fits for the first 36 elements, mean that parameters for the nonmetals (except rare gases) with $Z \leqslant 36$ that are not optimized here (B, Ge, As, and Se) can be obtained from their fits to electron densities.

Since the point charge model has been used in the overwhelming majority of MM parametrizations, but the present work shows that the point charge model leads to systematic errors in electrostatics and induction energies, we conclude that conventional MM parametrizations can only succeed by systematic cancellation of errors. Improving the electrostatics by including charge penetration effects can in principle lead to a new generation of more physical MM parameter sets.[35,69] By evaluating the electrostatic energies more accurately, we can derive more accurate and physical empirical parameters for the exchange repulsion and dispersion. The present work shows how this can be done in a practical way. In particular, the bare Coulomb interaction is replaced by a central potential centered at the nuclei with only one additional parameter per element.

Although the formulation of the electrostatic interactions in the current work applies only to QM/MM calculations, the formula for MM−MM electrostatic energy calculations can also be derived on the basis of the screened charge model. One obtains

$$E = \int \mathrm{d}r_1 \int \mathrm{d}r_2 \frac{\rho_A(r_1)\,\rho_B(r_2)}{|r_1 - r_2|} \tag{18}$$

for the interaction energy between delocalized charge distributions at sites A and B. Evaluating the integrals of eq 18 in MM/MM electrostatic energy calculations would increase the computational cost; so for low-cost calculations, one should develop efficient schemes to evaluate or approximate the integrals; for example, these integrals can be replaced by point charge interactions for sites separated by more than a certain distance.

Because the optimized values of the parameters in our treatment of screened electrostatics are physical, they can be used without reoptimization in new force fields or in other methods that incorporate electrostatic effects, such as the electrostatically embedded many-body method,[70,71] the electrostatically embedded multiconfiguration molecular mechanics method,[72,73] or fragment methods for large molecular systems.[74]

The present study only considered QM−MM boundaries that pass between nonbonded fragments. In future work, we will consider the use of this kind of scheme when QM−MM boundaries pass through covalent bonds.

Acknowledgment. We thank Hannah Leverentz for providing the geometry of several dimers. This work was supported in part by the National Science Foundation under grant no. CHE09-56776 and the Air Force Office of Scientific Research under grant no. FA9550-08-1-0183.

References

(1) *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J.; Thompson, M. A., Eds.; ACS

Symposium Series 712; American Chemical Society: Washington, DC, 1998.

(2) Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*, Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; p 285.

(3) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.

(4) Senn, H. M.; Thiel, W. *Angew. Chemie Int. Ed.* **2009**, *48*, 1198.

(5) Bernstein, N.; Kermode, J. R.; Csányi, G. *Rep. Prog. Phys.* **2009**, *72*, 026501.

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

(7) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(8) Sokalski, W. A.; Poirier, R. A. *Chem. Phys. Lett.* **1983**, *98*, 86.

(9) Leverentz, H.; Gao, J.; Truhlar, D. G. *Theor. Chem. Acc.*, in press.

(10) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(11) Hancock, G. C.; Truhlar, D. G.; Dykstra, C. E. *J. Chem. Phys.* **1988**, *88*, 1786.

(12) Stone, A. J. In *The Theory of Intermolecular Forces*; Oxford University Press Inc.: New York, 1996; p 94.

(13) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.

(14) Cisneros, G. A.; Tholander, S. N.-I.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J. P. *Int. J. Quantum Chem.* **2008**, *108*, 1905.

(15) Bredow, T.; Geudtner, G.; Jug, K. *J. Chem. Phys.* **1996**, *105*, 6395.

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

(17) Yudanov, I. V.; Nasluzov, V. A.; Neyman, K. M.; Rösch, N. *Int. J. Quantum Chem.* **1997**, *65*, 975.

(18) López, N.; Illas, F. *J. Phys. Chem. B* **1998**, *102*, 1430.

(19) Pacchioni, G.; Ferrari, A. M. *Catal. Today* **1999**, *50*, 533.

(20) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300.

(21) Soave, R.; Pacchioni, G. *Chem. Phys. Lett.* **2000**, *320*, 345.

(22) Gomes, J. R. B.; Illas, F.; Hernández, N. C.; Márquez, A.; Sanz, J. F. *Phys. Rev. B* **2002**, *65*, 125414.

(23) Gomes, J. R. B.; Illas, F.; Hernández, N. C.; Sanz, J. F.; Wander, A.; Harrison, N. M. *J. Chem. Phys.* **2002**, *116*, 1684.

(24) Gomes, J. R. B.; Lodziana, Z.; Illas, F. *J. Phys. Chem. B* **2003**, *107*, 6411.

(25) Cinquini, F.; Valentin, C. D.; Finazzi, E.; Giordano, L.; Pacchioni, G. *Theor. Chem. Acc.* **2007**, *117*, 827.

(26) Valero, R.; Gomes, J. R. B.; Truhlar, D. G.; Illas, F. *J. Chem. Phys.* **2008**, *129*, 124710.

(27) Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2010**, *6*, 2469.

(28) Köster, A. M.; Kölle, C.; Jug, K. *J. Chem. Phys.* **1993**, *99*, 1224.

(29) Kairys, V.; Jensen, J. H. *Chem. Phys. Lett.* **1999**, *315*, 140.

(30) Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353.

(31) Hall, G. G.; Smith, C. M. *Int. J. Quantum Chem.* **1992**, *42*, 1237.

(32) Wheatley, R. J.; Mitchell, J. B. O. *J. Comput. Chem.* **1994**, *15*, 1187.

(33) Guillot, B.; Guissani, Y. *J. Chem. Phys.* **2001**, *114*, 6720.

(34) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.

(35) Piquemal, J. P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.

(36) Torheyden, M.; Jansen, G. *Mol. Phys.* **2006**, *104*, 2101.

(37) Spackman, M. A. *Chem. Phys. Lett.* **2006**, *418*, 158.

(38) Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Phys. Chem. B* **2006**, *110*, 13682.

(39) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.

(40) Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190.

(41) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 014309.

(42) Slater, J. C. *Phys. Rev.* **1930**, *36*, 57.

(43) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.

(44) Clementi, E.; Raimondi, D. L. *J. Chem. Phys.* **1963**, *38*, 2686.

(45) Cusachs, L. C.; Trus, B. L.; Carroll, D. G.; Mcglynn, S. P. *Int. J. Quantum Chem. Symp.* **1967**, *1*, 423.

(46) Massey, H. S. W.; Burhop, E. H. S. In *The International Series of Monographs on Physics—Electronic and Ionic Impact Phenomena*; Marshall, W., Wilkinson, D. H., Eds.; University Press: Oxford, 1969; p 376.

(47) Archambault, F.; Chipot, C.; Soteras, I.; Luque, F. J.; Schulten, K.; Dehez, F. *J. Chem. Theory Comput.* **2009**, *5*, 3022.

(48) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(49) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(50) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

(51) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(52) Bukowski, R.; Cencek, W.; Jankowski, P.; Jeziorski, B.; Jeziorska, M.; Kucharski, S. A.; Lotrich, V. F.; Misquitta, A. J.; Moszynski, R.; Patkowski, K.; Podeszwa, R.; Rybak, S.; Szalewicz, K.; Williams, H. L.; Wheatley, R. J.; Wormer, P. E. S.; Zuchowski, P. S.; *SAPT2008: "An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies" version*, 2008.

(53) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.;

Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, version D.01 and E.01*, Gaussian, Inc: Wallingford, CT, 2004.

(54) Ponder, J. W. *TINKER, version 4.2*; Washington University: St. Louis, MO, 2004.

(55) Lin, H.; Zhang, Y.; Truhlar, D. G. *QMMM, version 1.3.5*, University of Minnesota: Minneapolis, 2007.

(56) Zhao, Y.; Truhlar, D. G. *MN-GFM: Minnesota Gaussian Functional Module, version 4.1*; University of Minnesota: Minneapolis, 2009.

(57) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.

(58) Dahlke, E. E.; Orthmeyer, M. A.; Truhlar, D. G. *J. Phys. Chem. B* **2008**, *112*, 2372.

(59) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(60) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.

(61) Curtiss, L. A.; McGrath, M. P.; Blaudeau, J. P.; Davis, N. E.; Binning, R. C.; Radom, L. *J. Chem. Phys.* **1995**, *103*, 6104.

(62) Stewart, R. F. *J. Chem. Phys.* **1969**, *50*, 2485.

(63) Szabo, A.; Ostlund, N. S. In *Modern Quantum Chemistry—Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc.: Mineola, NY, 1989; p 157.

(64) Strand, T. G.; Bonham, R. A. *J. Chem. Phys.* **1964**, *40*, 1686.

(65) Reinhardt, P.; Piquemal, J. P. *Int. J. Quantum Chem.* **2009**, *109*, 3259.

(66) Gill, P. M. W. *J. Phys. Chem.* **1996**, *100*, 15421.

(67) Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1988**, *89*, 5739.

(68) Nygren, M. A.; Pettersson, L. G. M.; Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1994**, *100*, 2010.

(69) Rotenberg, B.; Salanne, M.; Simon, C.; Vuilleumier, R. *Phys. Rev. Lett.* **2010**, *104*, 138301.

(70) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.

(71) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1.

(72) Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 790.

(73) Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 2925.

(74) Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. *J. Phys. Chem. B* **2009**, *113*, 9646.

CT1003862