

Combined Quantum Mechanical and Molecular Mechanical Methods for Calculating Potential Energy Surfaces: Tuned and Balanced Redistributed-Charge Algorithm

Bo Wang and Donald G. Truhlar*

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

Received July 15, 2009

Abstract: The combined quantum mechanical and molecular mechanical (QM/MM) method is one of the most powerful approaches for including correlation and polarization effects in simulations of large and complex systems, and the present article is concerned with the systematics of treating a QM/MM boundary that passes through a covalent bond, especially a polar covalent bond. In this study, we develop a new algorithm to treat such boundaries; the new method is called the balanced redistributed charge (balanced RC or BRC) scheme with a tuned fluorine link atom. The MM point charge on the MM boundary atom is modified to conserve the total charge of the entire system, and the modified charge is redistributed to the midpoints of the bonds between an MM boundary atom and its neighboring MM atoms. A pseudopotential is added to the fluorine link atom to reproduce the partial charge of the uncapped portion of the QM subsystem. We select proton affinities as the property used to validate the new method because the energy change associated with the addition of an entire charge (proton) to the QM system is very sensitive to the treatment of electrostatics at the boundary; we apply the new method to calculate proton affinities of 25 molecules with 13 different kinds of bonds being cut. The average proton affinity in the test set is 373 kcal/mol, and the test set provides a more challenging test than those usually used for testing QM/MM methods. For this challenging test set, common unbalanced schemes give a mean unsigned error (MUE) of 15–21 kcal/mol for H link atoms or 16–24 kcal/mol for F link atoms, much larger than the 5 kcal/mol obtained by simply omitting the MM region with either kind of link atom. Balancing the charges reduces the error to 5–7 kcal/mol for H link atoms and 4–6 kcal/mol for F link atoms. Balancing the charges and also tuning an F link atom lowers the MUE to 1.3–4 kcal/mol, with the best result for the balanced RC scheme. We conclude that properly tuning the link atom and correctly treating the point charges near the QM/MM boundary significantly improves the accuracy of the calculated proton affinities.

1. Introduction

The application of quantum chemistry to large and complex systems is one of the most challenging areas of current computational chemistry and also one that is seeing the most

progress.¹ An important tool for such applications is the combined quantum mechanical/molecular mechanical (QM/MM) method for calculating potential energy surfaces and interatomic forces; the reader is directed to several reviews and overviews for background information.^{2–23}

A stubborn issue in QM/MM calculations is the treatment of the boundary between the QM and MM regions when it

* Corresponding author phone: (612) 624-7555; fax: (612) 624-9390; e-mail: truhlar@umn.edu.

passes through a bond, which is practically unavoidable in the treatment of many solids, polymers, and complex systems. In general the QM region is capped to saturate dangling valences caused by the cut. Three different kinds of methods have been proposed to deal with capping the QM boundary atom. The first one is the link atom approach (LA).^{24,25} The dangling bond of the QM region is capped with an additional atom (usually a hydrogen atom) and the QM calculations are performed on this capped system. The second method is localized orbitals.^{26–28} The dangling bond is saturated by orbitals rather than by an atom. Examples of this approach are the local self-consistent field (LSCF) method²⁶ and the generalized hybrid orbitals (GHO) scheme.^{27,28} The third kind of method involves a pseudobond or an effective core potential (ECP). In this approach, a parametrized atom, modified to mimic the behavior of the original MM boundary atoms or groups, is used to cap the QM system; examples of this approach are tuned capping atoms,^{29–31} adjusted connection atoms,³² a pseudobond,^{33–35} an effective group potential,³⁶ a quantum capped potential,^{37–39} and a variationally optimized effective atom-centered potential.⁴⁰ This third class of methods may be considered to be a second-generation link-atom method in which the link atom is optimized or tuned.

Though much progress has been made, there are still many problems in the treatment of QM–MM boundaries that pass through a bond. Most attention has been devoted to the cutting of C–C bonds, especially for modeling enzymatic binding and reactions, but some procedures are more general. The methods that have been developed exhibit a wide variety of differences in the precise way in which they have been implemented.

Pople has emphasized the importance of theoretical models, where a theoretical model is “an approximate but well-defined mathematical procedure for simulation. . . The approximate mathematical treatment must be precisely formulated. It should be general. . . Particular procedures for particular molecules. . . should be avoided.”⁴¹ If tests of the model against a broad data set are successful, the model is said to be validated. The goal of this article is to develop and validate a new method, in the spirit of a theoretical model chemistry, for the treatment of a boundary between bonded atoms in QM/MM simulations. It is precisely defined in a general way applicable to all systems and all kinds of single bonds, and it is tested against a data set of 25 systems in which 13 different kinds of bonds are cut, in particular (where the atom listed first is in the QM subsystem, and the one listed second is in the MM subsystem): C–C, N–C, O–C, S–C, C–N, O–N, C–O, Al–O, Si–O, C–Si, O–Si, C–S, and S–S.

2. Methods

Our group has developed redistributed charge (RC) and redistributed charge and dipole (RCD) methods to treat the charges near a QM/MM boundary that passes through a bond.⁴² These methods give good results even when large charges are present near the boundary. In the current work, we improve the RC and RCD methods by adding two new elements, a charge balancing step and a tuned link atom. In

particular, the redistributed charges are used to conserve the charge of the entire system, and a tuned fluorine atom is used to saturate the free valence of the QM region and to reproduce the partial charge of the uncapped portion of the capped QM subsystem. The improved method is used to treat polar bonds between the QM and MM subsystems with large partial atomic charges near the boundary. In order to describe the algorithm, we label the atoms according to “tiers”. The definition is the same as what is used in previous work;^{4,42} in particular, the MM boundary atoms are denoted as M1 atoms, and the MM atoms directly bonded to M1 atoms are denoted as M2 atoms. M3 atoms are the third-tier MM atoms. The QM boundary atoms are denoted as Q1 atoms. The QM atoms directly bonded to Q1 atoms are labeled Q2 atoms. Q3 atoms are those bonded to Q2 atoms and so forth for Q4, Q5, etc. The QM region is also called the primary subsystem (PS) in this study. The sum of all QM atoms and MM atoms before the cutting and capping is called the original entire system. The sum of the capped QM subsystem and the whole MM subsystem after the charge redistribution is called the QM/MM entire system.

In the QM/MM calculations, we use an additive QM/MM scheme to define the total energy of the system:²³

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

$$E^{\text{QM/MM}} = E_{\text{el}}^{\text{QM/MM}} + E_{\text{vdW}}^{\text{QM/MM}} + E_{\text{val}}^{\text{QM/MM}} \quad (2)$$

where E^{QM} is the quantum mechanical energy of the QM region, E^{MM} is the molecular mechanical energy of the MM region, and $E^{\text{QM/MM}}$ accounts for the interaction energy between the QM and the MM regions. $E^{\text{QM/MM}}$ is decomposed into three terms; $E_{\text{el}}^{\text{QM/MM}}$ represents electrostatic interactions, $E_{\text{vdW}}^{\text{QM/MM}}$ represents van der Waals interactions, and $E_{\text{val}}^{\text{QM/MM}}$ represents valence interactions. In this study, we will concentrate on the electrostatic coupling term $E_{\text{el}}^{\text{QM/MM}}$, which is the most technically involved term. The $E_{\text{vdW}}^{\text{QM/MM}}$ and $E_{\text{val}}^{\text{QM/MM}}$ terms will cancel out in the present work because we study fixed-geometry proton affinities to isolate the electrostatic terms, but these other QM/MM terms will be studied later when we consider QM/MM geometry optimization.

2.A. Treatment of Boundary Charge. It has been found that it is important to conserve the total charge of the QM/MM entire system in QM/MM calculations,⁴³ that is, the sum of the MM partial atomic charges of the MM region and the QM charge of the capped QM region should equal the total charge of the original entire system, as shown in eq 3:

$$q^{\text{MM}} + q^{\text{QM}} = q^{\text{total}} \quad (3)$$

However, when the original entire system is divided into QM and MM regions, the sum of MM charges of the MM region does not necessarily equal zero or an integer. If MM charges are not modified, the total charge of the QM/MM entire system is not conserved. Several workers have recognized that this causes inaccuracies and have suggested various methods to remedy this.^{33,43–45} Sherwood et al.⁴⁴ adjusted the charge on an M1 atom to conserve the total charge of the QM/MM entire system,

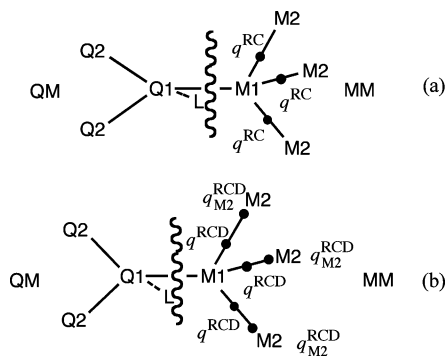


Figure 1. QM/MM boundary treatments in (a) the balanced RC scheme and (b) the balanced RCD scheme.

and they redistributed the adjusted charge on the M1 atom evenly to M2 atoms; point dipoles were added at the M2 atoms to compensate the changes in the M1–M2 bond dipoles due to the movement of the charges. Zhang et al.³³ zeroed the charges on all MM atoms that are in the same group as the M1 atom. Das et al.⁴⁵ used a double link atom approach combined with delocalized Gaussian MM charges. Walker et al.⁴³ added the charge difference to the nearest M2 atom or evenly to all the MM atoms except the M1 atom.

In the previous RC scheme,⁴² the charge on each M1 atom is redistributed to the midpoints of M1–M2 bonds. However, the total charge of the QM/MM entire system is not conserved when the sum of MM charges of the MM region is not zero or an integer. In the balanced RC scheme, introduced here, we first adjust the charge on the M1 atom so that

$$q_0 + \sum_i q_i + q^{\text{QM}} = q^{\text{total}} \quad (4)$$

where q_0 is the modified M1 charge, $\{q_i\}$ are the MM point charges of other MM atoms (except M1), q^{QM} is the charge of the QM region (that is, of the capped QM subsystem), and q^{total} is the charge of the original entire system. This charge balancing step conserves the charge of the QM/MM entire system.

Then the balanced RC scheme redistributes the charge q_0 evenly to the midpoints of all M1–M2 bonds, with each bond midpoint obtaining a charge $q^{\text{RC}} = (q_0/n)$, where n is the number of M1–M2 bonds. For the balanced redistributed charge and dipole (balanced RCD) method, we double the redistributed charges and adjust the charges $q_{\text{M2}}^{\text{RCD}}$ on M2 atoms to conserve the total charge of the QM/MM entire system, as shown in eqs 5 and 6:

$$q^{\text{RCD}} = 2q^{\text{RC}} = \frac{2q_0}{n} \quad (5)$$

$$q_{\text{M2}}^{\text{RCD}} = q_{\text{M2}} - q^{\text{RC}} \quad (6)$$

These two schemes are illustrated in Figure 1.

In this study, we compare balanced RC and balanced RCD to other methods that differ in how the redistributed charges are handled, e.g., to what location are they

redistributed. These methods include: balanced straight electrostatic embedding (SEE), balanced RC2, Amber-1,⁴³ balanced RC3, Amber-2,⁴³ and balanced Shift.⁴⁴ Amber-1 and Amber-2 are the options called *adjust_q* = 1 and *adjust_q* = 2 in *AMBER 10*. The distinction between these methods is in the position of the redistributed charges and whether the dipoles of the M1–M2 bonds are corrected. In balanced SEE, the charge on the M1 atom is set to q_0 , and it is not moved. In balanced RC2, we distribute q_0 evenly to all M2 atoms. In balanced RC3, we distribute q_0 evenly to all M2 and M3 atoms. In Amber-1, we move q_0 to the nearest M2 atom. In Amber-2, we distribute q_0 evenly to all MM atoms, except the M1 atom. (Note that Amber-2 is the default option in revision 10 of *AMBER*,⁴⁶ whereas Amber-1 can be selected in *AMBER 10* by specifying *adjust_q* = 1.) In balanced Shift, the redistributed charges are placed at M2 atoms, and dipoles are added around M2 atoms to compensate the movement of the charges. A summary of these charge schemes is shown in Table 1.

We call the methods in Table 1 balanced methods because they all conserve the total charge of the QM/MM entire system. Five unbalanced methods, in which the total charge of the QM/MM entire system is not necessarily conserved, are also tested, including SEE, Z1, Z2, Z3, and RC.⁴² SEE is straight electronic embedding that makes no change of the charges of MM boundary atoms, Z1 denotes that the charge of the M1 atom is zeroed (this can be chosen by specifying *adjust_q* = 0 in *AMBER 10*, and it is the default method in CHARMM⁴⁷), Z2 denotes that the charges of M1 and M2 atoms are zeroed (Z2 is the default scheme in both *Gaussian 03*⁴⁸ and *Gaussian 09*⁴⁹), and Z3 denotes that all the charges of all M1, M2, and M3 atoms are zeroed. RC denotes that the charge on the M1 atom is redistributed to the midpoints of M1–M2 bonds without the balancing step. Balanced methods and unbalanced methods are compared to test the importance of conserving the charge of the QM/MM entire system. To make a comparison, we also carry out calculations on the capped primary system (CPS), in which the whole MM region is substituted by the link atom.

2.B. Link Atom. Another issue in the boundary treatment is the choice of the link atom. A hydrogen atom can be used as the link atom when a C–C bond is cut. However, a Q1–H bond may be a poor model for the cut bond when the M1 atom is electronegative, such as in a Si–O or C–O bond. Therefore, we use a tuned capping atom as the link atom to mimic a cut polar bond and to reproduce the electronic structure of the QM subsystem. Redondo et al.²⁹ used a tuned hydrogen atom to replace a silicon atom. Koga et al.³⁰ added a shift operator on the hydrogen atom to reproduce the effect of the substitution. Zhang et al.³³ and Nasluzov et al.³¹ used tuned fluorine atoms and derived pseudopotentials for carbon and oxygen boundary atoms. Here, we provide a more general rule to tune a capping atom for boundary atoms. The capping atom is always a tuned F atom. We first replace the $1s^2$ core by a conventional pseudopotential U , and then a tuning pseudopotential $U_0'(r)$ is added to U . The con-

Table 1. Charge Schemes

	position of the redistributed charges	correction of bond dipole	ref
balanced SEE	M1 atom	no	present
balanced RC	midpoints of M1–M2 bonds	no	present
balanced RC2	M2 atoms	no	present
Amber -1	nearest M2 atom	no	Walker et al. ⁴³
balanced RC3	M2, M3 atoms	no	present
Amber -2	all MM atoms (except M1 atom)	no	Walker et al. ⁴³
balanced RCD	midpoints of M1–M2 atoms	yes	present
balanced Shift	M2 atoms	yes	Sherwood et al. ⁴⁴

Table 2. CRENBL Effective Core Potential^a

	n_{ij}	α_{ij}	C_{ij}
U_0	2	2.8835	12.685 306
	2	3.1077	−19.302 589
	1	5.6122	1.002 179
	0	2.8146	2.245 349
U_1	2	44.5166	−6.723 024
	2	12.9487	−0.929 649
	1	132.4967	−1.526 734

^a Reference 50.

ventional pseudopotential used here is the CRENBL effective core potential (ECP) for a fluorine atom developed by Pacios and Christiansen.⁵⁰

The form of this potential is

$$U = U_0(r) + \sum_{m=-1}^1 [U_1(r) - U_0(r)] |1m\rangle \langle 1m| \quad (7)$$

where

$$U_l(r) = r^{-2} \sum_j C_{lj} r^{n_{lj}} e^{-\alpha_{lj} r^2} \quad (8)$$

r is the distance of an electron from the capping nucleus, and $|lm\rangle$ is a spherical harmonic. The parameters for this pseudopotential are listed in Table 2. The form of $U_0'(r)$ is

$$U_0'(r) = C \exp[-(r/r_0)^2] \quad (9)$$

where C and r_0 are parameters. The basis set used for the tuned F atom is the same as for a conventional F atom. For example, if the QM subsystem is treated by the 6-31G* basis set, then the tuned F atom has the 6-31G* basis set of a conventional F atom. To find an appropriate pseudopotential, we set r_0 equal to 1 bohr and tune the parameter C of the pseudopotential.

The next key decision is how to choose C . In order to reproduce the electronic structure of the QM subsystem, we require that the total charge of the uncapped portion of the QM subsystem in a QM/MM calculation is equal to the total charge of the same subsystem in a QM calculation of the original entire system or, in practice, of a system that mimics the original entire system better than the capped QM subsystem does (see below for more details of this large system). Mulliken charges were used as the indicator. Because Mulliken charges become unphysical when large basis sets are used, we used small basis sets without diffuse functions for this tuning step, in particular, 6-31G* when M1 is from the second period (Li through F) and STO-3G

otherwise. Since the STO-3G basis set is defined for the entire periodic table, the tuning step is well-defined for the entire periodic table.

We can perform the tuning process on either the reactant or the product. For the validation suite, the reactant is a neutral molecule, and the product is a deprotonated anion. In this study, we used the protonated neutral reactant to tune the F atom. All the parameters C of the pseudopotentials are derived in the presence of MM background charges. Because the MM charges are redistributed differently in the various boundary charge schemes explained in section 2.A, the derived pseudopotentials are not the same for different charge schemes.

To enable the method to be applied to large systems for which it is difficult to perform a QM calculation on the original entire system, the tuning is performed on a model system created from the original entire system. This model system is called the tuning system or the entire system model (ESM). It consists of all QM atoms and all M1, M2, and M3 atoms, and all free valences on M3 atoms are capped by untuned H atoms. The scheme is illustrated in Figure 2.

The completely defined tuning process employed in the present study is as follows:

1. Choose a geometry and charge state for the tuning system and create the entire system model (ESM) by capping all M3 atoms with untuned hydrogens.
2. Do a full QM calculation on ESM and carry out Mulliken population analysis. For the basis set, use 6-31G*

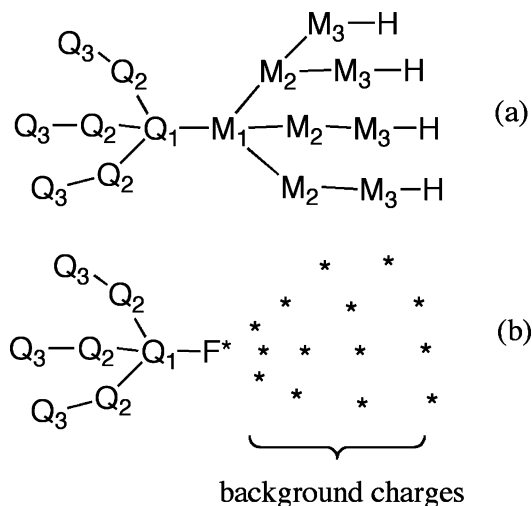


Figure 2. Determining the pseudopotential for the tuned F atom in the entire system model (ESM): (a) ESM and (b) CPS**, which is the capped QM subsystem with background charges to replace the rest of the ESM.

if M1 is from the second period (e.g., C, N, O) and use STO-3G otherwise (that is, if M1 belongs to the third or higher period, for example, Si). This yields the total charge on the primary system (PS); call this $q_{\text{PS}}^{\text{ESM,MPA}}$, where MPA denotes Mulliken population analysis. It also yields $q_{\text{SS}}^{\text{ESM,MPA}}$, where SS is the secondary subsystem, equal to all the atoms in ESM except the PS atoms. By construction, $q_{\text{PS}}^{\text{ESM,MPA}} + q_{\text{SS}}^{\text{ESM,MPA}} = q_{\text{ESM}}$ where q_{ESM} is the total charge of ESM.

3. Select an MM charge scheme. For the present calculations, the MM charge scheme is CM4M charges from a calculation on the ESM. The basis set used for the calculation of CM4M charges could in principle be the same as chosen for step 2, but in fact, we do not have CM4M charge schemes for STO-3G; therefore, the MM charges are always CM4M charges determined with M06-2X/6-31G* calculations on the ESM.

4. Define TSS as the truncated secondary system of the original entire system model, which includes all atoms in the secondary subsystem of the ESM except the M1 atom. For the chosen MM charge scheme of step 3, calculate $q_{\text{TSS}}^{\text{ESM,MM}}$, which equals the sum of the MM charges from step 3 (that is, the sum of CM4M charges) on all TSS atoms of ESM.

5. Cap the PS with an F* atom to create the capped primary system (CPS), where F* denotes a tuned F atom. Always set r_0 equal to 1 bohr in the pseudopotential. The other parameter (C) of the pseudopotential will be determined in step 7.

6. Select a charge modification scheme, for example, balanced RC or balanced RCD. For the balanced charge schemes, we set q_0 to make $q_{\text{CPS}} + q_0 + q_{\text{TSS}}^{\text{ESM,MM}}$ equal to q_{ESM} . In the usual case where $q_{\text{CPS}} = q_{\text{ESM}}$, then this yields $q_0 = -q_{\text{TSS}}^{\text{ESM,MM}}$.

7. Now, for a given MM charge scheme, and given the charge modification scheme, carry out a series of fixed-geometry CPS** calculations with various values of C . Note that CPS** here denotes the capped primary system in the modified charge environment of the secondary system of the ESM. Adjust C until $q_{\text{F}^*}^{\text{MPA}}$ equals $q_{\text{SS}}^{\text{ESM,MPA}}$, which was determined in step 2. Now the pseudopotential is known, so F* is properly tuned.

After the tuning step, the tuned F link atom can be used with the selected charge scheme on the QM/MM entire system to do QM/MM calculations on the proton affinities.

3. Details of Validation Calculations

We have implemented the proposed charge schemes and link atom treatments in the QMMM program,⁵¹ which is based on the Gaussian 03⁴⁸ and TINKER⁵² programs. Either density functional theory (DFT) or wave function theory (WFT) can be used for the QM calculations. In the study, the M06-2X density functional^{53,54} was used for all the QM calculations. Proton affinities were used as the criterion to evaluate the methods, as they are sensitive to the boundary treatment.⁴⁷ The 6-31G** basis set was used for aluminosilicate clusters, and the 6-31G* basis set was used for organic molecules.

The geometry of all the molecules was fixed at the protonated geometry, so in eqs 1 and 2, the QM/MM valence

term $E_{\text{val}}^{\text{QM/MM}}$, the QM/MM van der Waals term $E_{\text{vdW}}^{\text{QM/MM}}$ and the MM term E^{MM} are the same for the deprotonated and protonated forms, and they cancel out in the QM/MM calculations of proton affinities. The final expression for the proton affinity is

$$\begin{aligned} E(\text{proton affinity}) &= E(\text{deprotonated}) - E(\text{protonated}) \\ &= [E^{\text{QM}}(\text{deprotonated}) + E_{\text{el}}^{\text{QM/MM}}(\text{deprotonated})] - \\ &\quad [E^{\text{QM}}(\text{protonated}) + E_{\text{el}}^{\text{QM/MM}}(\text{protonated})] \end{aligned} \quad (10)$$

In this study, CM4M charges⁵⁵ were used for MM atoms. The protonated forms of the molecules in the test suite are illustrated in Figure 3. In each case, the QM region is on the left and the MM region is on the right. The selected molecules in the test suite contain different kinds of Q1–M1 bonds at the QM/MM boundary, in particular, C–C, N–C, O–C, S–C, C–N, O–N, C–O, Al–O, Si–O, C–Si, O–Si, C–S, and S–S. Both polar and nonpolar bonds are included in this test suite.

For the position of the link atom, the scaled bond distance method^{56,57} was used, that is, the link atom is placed along the Q1–M1 bond, and the ratio of the Q1–link atom distance to the Q1–M1 distance is set to be the ratio of the standard bond length of the Q1–link atom bond to the standard bond length of the Q1–M1 bond. The standard bond lengths used in this study are listed in Table 3. For the tuned F link atom, it is placed at the same position as that of an ordinary F atom.

4. Results

4.A. Example for Balancing the Charge and Tuning the Link Atom. We use molecule **w** as an example to demonstrate the proposed algorithm. The O–Si bond is cut. In the tuning process, we create the entire system model (ESM) from the original QM system. The original entire system and the ESM are shown in Figure 4. A full QM calculation is performed on the ESM to get the total Mulliken charge on the QM region $q_{\text{PS}}^{\text{ESM,MPA}}$ and on the MM region $q_{\text{SS}}^{\text{ESM,MPA}}$. As the M1 atom is silicon, the STO-3G basis set was used. CM4M charges were calculated for all the MM atoms in ESM; the M1 atom has a charge of $0.515e$, and the sum $q_{\text{TSS}}^{\text{ESM,MM}}$ of the MM charges in the truncated secondary system (TSS) is $-0.377e$. Therefore, in the balanced schemes, the redistributed charge $q_0 = -q_{\text{TSS}}^{\text{ESM,MM}} = 0.377e$. Then the QM system is capped with a tuned F atom, and the capped QM subsystem is embedded in the redistributed MM charges using a boundary charge scheme. The parameter C of the pseudopotential is adjusted to make the Mulliken charge $q_{\text{F}^*}^{\text{MPA}}$ of the tuned F atom equal to $q_{\text{SS}}^{\text{ESM,MPA}}$. For example, in the balanced RC scheme, the parameter of the pseudopotential is 0.80, as shown in Table 6.

The tuned F* link atom is used to cap the QM system in the QM/MM entire system. The same charge scheme is used for the tuning and the calculations of proton affinities. The results with the tuned F atom are compared to those with untuned H and F link atoms.

4.B. H and F Link Atoms. Tables 4 and 5 show the proton affinities of the 25 molecules by full QM calculations

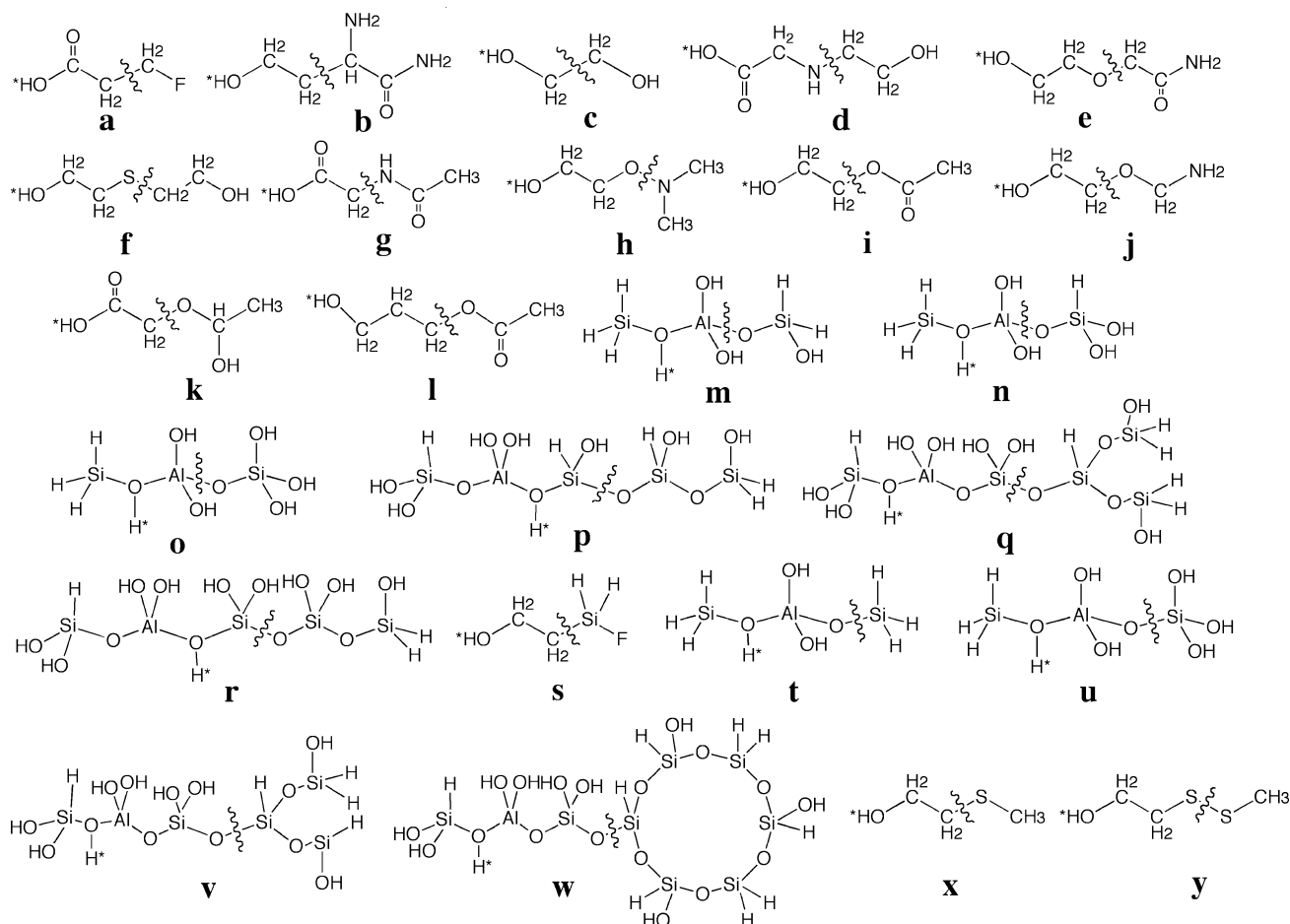


Figure 3. The 25 molecules used in the test suite. The QM subsystem is on the left, and the MM subsystem is on the right. The * represents the proton involved in the protonation process.

Table 3. Standard Bond Lengths (Å)

bond	distance	bond	distance	bond	distance	bond	distance
C-H	1.09	C-F	1.33	C-C	1.53	N-O	1.47
N-H	1.01	N-F	1.41	N-C	1.45	Al-O	1.72
O-H	0.95	O-F	1.41	O-C	1.42	Si-O	1.61
Al-H	1.55	Al-F	1.67	S-C	1.81	S-S	2.04
Si-H	1.45	Si-F	1.56	Si-C	1.86		
S-H	1.34	S-F	1.65				

and the signed error by QM/MM calculations using untuned H and F atoms as link atoms.

4.B.1. Balanced and Unbalanced Charge Schemes. The balanced methods (balanced SEE, balanced RC, Amber-1, balanced RC2, balanced RC3, Amber-2, balanced RCD, balanced Shift) give much smaller errors in proton affinities than the unbalanced ones (SEE, Z1, Z2, Z3, RC). Both the H link atom and the F link atom schemes have the same trends. The mean unsigned errors (MUEs) given by all the balanced methods are 4–7 kcal/mol, while the MUEs given by all the unbalanced methods are 15–24 kcal/mol. This is because in the unbalanced methods, a net partial charge is created near the QM region and the interactions between the QM and MM regions become unphysical.^{4,33,43} The CPS method, in which any polarization of the QM region by the MM region is excluded, does not change the total charge of the QM/MM entire system and gives a smaller MUE than the unbalanced methods. Therefore, the conservation of the

total charge of the QM/MM entire system is one of the key factors for the calculations of proton affinities.

4.B.2. Different Link Atoms and Charge Schemes. The comparison of the results using the H link atom (Table 4) and the F link atom (Table 5) shows that the proton affinities

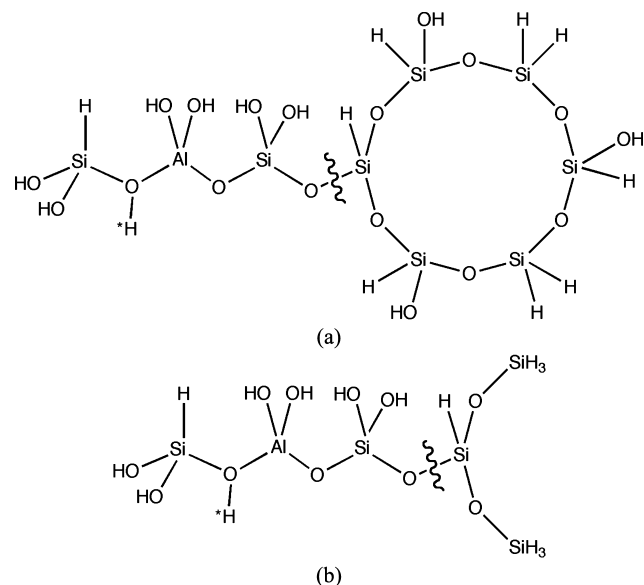


Figure 4. (a) The original entire system and (b) the entire system model (ESM) of the molecule w.

Table 4. Full-QM Proton Affinities (PA, in kcal/mol), QM/MM Signed Errors (in kcal/mol), and Mean Unsigned Errors (MUE) (in kcal/mol) Averaged over 25 cases Using the H Link Atoms

case	bond	PA	site	CPS	balanced SEE	balanced RC	balanced RC2	balanced Amber-1	balanced RC3	balanced Amber-2	balanced RCD	balanced Shift	SEE	Z1	Z2	Z3	RC
a	C-C	362.3	Q3	6.8	1.2	2.5	3.9	3.5	3.9	3.9	1.0	1.3	-2.4	13.4	6.8	6.8	-0.7
b	C-C	406.6	Q3	-0.6	3.9	5.1	6.3	5.5	6.9	7.3	3.9	4.2	4.8	13.1	-3.6	-40.7	5.8
c	C-C	402.6	Q2	9.0	12.0	13.0	14.0	13.7	14.5	14.5	11.9	12.1	12.0	19.7	-19.2	9.0	13.0
d	N-C	376.7	Q4	3.4	6.6	4.8	2.9	3.8	1.7	1.2	6.6	6.2	-14.7	-9.7	7.2	-15.1	-14.1
e	O-C	398.6	Q4	4.0	5.4	4.1	2.9	3.4	2.0	1.4	5.3	5.0	-7.8	-5.7	35.5	-26.7	-7.5
f	S-C	394.7	Q4	-1.1	2.2	0.6	-1.2	-0.5	-2.6	-3.2	2.3	1.9	-6.8	-15.0	2.0	-16.4	-7.6
g	C-N	355.2	Q3	13.0	17.0	12.5	7.4	11.3	2.5	-1.4	17.5	16.2	39.9	-26.9	47.0	-3.6	32.9
h	O-N	400.0	Q4	3.0	13.8	9.6	5.4	5.2	3.8	3.8	13.5	12.4	6.8	-19.1	-23.0	3.0	3.4
i	C-O	394.9	Q3	11.9	12.1	11.3	10.3	10.3	9.7	9.3	12.4	12.1	38.8	5.9	40.2	-5.9	34.4
j	C-O	401.0	Q3	5.5	13.9	11.1	7.7	7.7	6.1	5.1	14.4	13.5	37.5	-6.8	12.0	-30.5	31.3
k	C-O	366.7	Q3	5.8	10.8	7.5	4.2	4.2	2.5	1.6	10.8	10.0	27.0	-11.8	9.2	-35.4	21.3
l	C-O	398.3	Q4	7.1	7.6	7.1	6.5	6.5	6.2	5.9	7.7	7.6	28.1	3.5	30.6	-7.8	25.1
m	Al-O	340.7	Q2	5.1	6.0	3.8	1.1	1.1	-0.2	-0.3	6.5	5.7	33.6	-10.8	19.9	-18.2	28.5
n	Al-O	339.1	Q2	6.8	6.0	3.8	1.0	1.0	-0.4	-0.3	6.6	5.7	33.6	-11.1	26.1	-42.0	28.4
o	Al-O	348.0	Q2	-2.2	5.5	3.7	1.5	1.5	0.4	-0.2	5.9	5.1	35.1	-8.0	35.1	-61.5	29.9
p	Si-O	349.0	Q2	-4.9	4.0	0.7	-2.9	-2.9	-4.5	-6.3	4.3	2.8	27.2	-15.9	27.9	-35.5	20.1
q	Si-O	353.2	Q4	-4.0	0.7	-0.4	-1.8	-1.8	-2.5	-3.8	1.0	0.7	18.3	-11.0	22.8	-22.2	15.6
r	Si-O	348.1	Q2	-4.1	4.0	0.6	-3.2	-3.2	-5.0	-6.8	4.4	2.8	23.9	-17.3	30.7	-60.6	17.4
s	C-Si	397.0	Q3	8.6	-2.9	2.3	7.2	6.2	7.2	7.2	-3.0	-1.3	-20.4	33.4	8.6	8.6	-12.6
t	O-Si	342.7	Q3	5.8	3.9	4.9	5.8	6.0	5.8	5.8	4.0	4.3	-8.1	13.2	5.8	5.8	-5.7
u	O-Si	348.0	Q3	0.5	1.7	6.2	9.8	10.8	11.7	11.7	2.4	3.5	-9.1	38.7	-59.5	0.5	-3.1
v	O-Si	353.2	Q5	1.6	2.2	4.6	6.6	7.9	8.2	10.4	2.4	2.9	-8.1	29.2	-16.8	24.5	-4.7
w	O-Si	354.8	Q5	0.0	2.7	5.3	7.5	9.0	9.3	14.2	3.0	3.4	-5.0	32.5	-13.0	38.6	-1.6
x	C-S	395.8	Q3	10.5	16.1	14.4	12.6	12.6	12.1	12.1	16.2	15.5	25.5	5.4	-7.5	10.5	22.3
y	S-S	390.3	Q4	2.8	6.1	5.1	4.0	4.0	3.7	3.7	6.1	5.8	5.3	-2.8	-14.8	2.8	4.4
MUE				5.1	6.7	5.8	5.5	5.7	5.3	5.7	6.9	6.5	19.2	15.2	21.0	21.3	15.7

Table 5. Full-QM Proton Affinities (PA, in kcal/mol), QM/MM Signed Errors (in kcal/mol), and Mean Unsigned Errors (MUE, in kcal/mol) Averaged over 25 cases Using the F Link Atoms

case	bond	PA	site	CPS	balanced SEE	balanced RC	balanced RC2	balanced Amber-1	balanced RC3	balanced Amber-2	balanced RCD	balanced Shift	SEE	Z1	Z2	Z3	RC
a	C-C	362.3	Q3	-2.6	-7.8	-7.0	-5.6	-6.0	-5.6	-5.6	-8.4	-8.1	-11.1	3.8	-2.6	-2.6	-10.1
b	C-C	406.6	Q3	-10.4	-5.4	-4.6	-3.4	-4.2	-2.9	-2.5	-5.7	-5.4	-4.6	3.2	-12.4	-49.7	-3.9
c	C-C	402.6	Q2	-8.5	-5.3	-4.8	-3.7	-4.0	-3.3	-3.3	-5.8	-5.6	-5.3	1.8	-36.3	-8.5	-4.8
d	N-C	376.7	Q4	-11.5	-8.2	-9.6	-11.7	-10.8	-13.0	-13.5	-7.4	-7.7	-29.6	-24.7	-7.1	-30.3	-29.2
e	O-C	398.6	Q4	-10.4	-9.7	-10.7	-12.3	-11.7	-13.3	-14.0	-9.2	-9.4	-23.7	-21.5	22.6	-42.3	-23.5
f	S-C	394.7	Q4	-7.4	-3.7	-5.2	-7.3	-6.5	-8.8	-9.4	-3.1	-3.7	-12.9	-21.2	-4.0	-22.7	-13.6
g	C-N	355.2	Q3	4.0	5.4	3.4	-1.5	2.3	-6.2	-9.9	8.4	7.1	26.8	-34.7	37.4	-12.5	23.6
h	O-N	400.0	Q4	-11.0	2.2	-1.8	-7.7	-7.9	-9.8	-9.8	3.9	1.7	-5.5	-33.8	-37.9	-11.0	-8.7
i	C-O	394.9	Q3	3.1	3.3	2.9	1.9	1.9	1.3	1.0	3.8	3.6	27.6	-2.3	30.5	-14.3	25.1
j	C-O	401.0	Q3	-2.8	4.6	2.8	-0.6	-0.6	-2.1	-3.0	6.2	5.0	26.6	-14.5	3.5	-37.8	22.8
k	C-O	366.7	Q3	-0.8	3.5	1.4	-2.3	-2.3	-4.2	-5.1	5.2	4.0	19.2	-18.4	2.7	-41.9	15.7
l	C-O	398.3	Q4	1.5	1.8	1.5	0.9	0.9	0.6	0.3	2.1	1.9	21.0	-2.1	24.4	-13.3	19.3
m	Al-O	340.7	Q2	2.1	2.7	1.0	-1.8	-1.8	-3.1	-3.2	3.8	2.7	30.3	-13.5	16.6	-20.8	25.6
n	Al-O	339.1	Q2	3.8	2.7	0.9	-1.9	-1.9	-3.3	-3.2	3.8	2.7	30.3	-13.9	22.7	-44.5	25.5
o	Al-O	348.0	Q2	-5.2	2.0	0.6	-1.6	-1.6	-2.7	-3.2	2.9	2.0	31.4	-11.0	31.5	-63.9	26.8
p	Si-O	349.0	Q2	-8.5	-0.2	-2.8	-6.6	-6.6	-8.2	-10.1	1.0	-0.7	22.5	-19.5	24.0	-39.0	17.0
q	Si-O	353.2	Q4	-3.3	1.3	0.4	-1.2	-1.2	-2.0	-3.4	2.0	1.6	19.3	-10.5	23.7	-21.5	16.9
r	Si-O	348.1	Q2	-6.9	0.6	-2.1	-6.2	-6.2	-8.1	-9.8	2.0	0.1	20.3	-20.2	27.6	-62.9	15.0
s	C-Si	397.0	Q3	-1.1	-11.9	-7.2	-2.5	-3.6	-2.5	-2.5	-11.8	-10.0	-27.8	23.1	-1.1	-1.1	-21.2
t	O-Si	342.7	Q3	-3.8	-6.7	-5.1	-3.8	-3.6	-3.8	-3.8	-6.5	-5.9	-20.1	3.9	-3.8	-3.8	-16.7
u	O-Si	348.0	Q3	-9.1	-11.0	-4.4	1.1	2.2	3.2	3.2	-9.8	-7.1	-22.5	31.2	-70.8	-9.1	-14.4
v	O-Si	353.2	Q5	-4.4	-5.9	-2.3	0.8	2.0	2.6	5.1	-5.4	-4.1	-16.9	24.2	-23.1	18.8	-12.2
w	O-Si	354.8	Q5	-6.0	-5.6	-1.7	1.7	3.0	3.7	9.0	-5.2	-3.8	-13.8	27.5	-19.3	33.0	-9.1
x	C-S	395.8	Q3	1.4	6.8	5.0	3.3	3.3	2.8	2.8	6.8	5.9	16.3	-3.7	-16.1	1.4	12.8
y	S-S	390.3	Q4	-4.2	0.0	-1.7	-2.9	-2.9	-3.3	-3.3	-0.4	-1.0	-0.9	-9.7	-21.8	-4.2	-2.4
MUE				5.4	4.7	3.6	3.8	4.0	4.8	5.6	5.2	4.4	19.4	15.8	20.9	24.4	16.6

are sensitive to the link atom. In most cases, the H link atom gives larger proton affinities than the full QM calculations, while the F link atom gives smaller proton affinities. As the link atom is directly connected to the QM region, it can greatly change the electronic structure of the QM region. The Q1-H and Q1-F bonds cannot adequately reproduce the properties of the original Q1-M1 bond. Electronegativity can be used as a qualitative criterion to decide which atom is better to be used as the link atom. For example, when the

M1 atom is oxygen or nitrogen, the F link atom gives better results than the H link atom.

Moreover, compared with the charges on the other MM atoms, the charge on the link atom is closest to the QM region and greatly affects the electrostatic potential on the active site. König et al.⁴⁷ have compared different charge schemes to treat the boundary and found that all the balanced methods give, on average, comparable errors in proton affinities and deprotonation energies. Here, we also found

Table 6. Parameters of Pseudopotentials for the Tuned F Link Atoms

case	bond	CPS	balanced SEE	balanced RC	balanced RC2	Amber-1	balanced RC3	Amber-2	balanced RCD	balanced Shift
a	C–C	0.80	0.95	0.95	0.85	0.85	0.85	0.85	1.00	1.00
b	C–C	0.65	0.65	0.65	0.60	0.60	0.55	0.55	0.70	0.65
c	C–C	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.70	0.70
d	N–C	1.40	1.20	1.25	1.40	1.35	1.40	1.40	1.15	1.20
e	O–C	1.35	1.35	1.45	1.50	1.50	1.55	1.55	1.35	1.35
f	S–C	1.05	0.75	0.85	1.05	1.00	1.05	1.10	0.70	0.75
g	C–N	−0.20	−0.25	−0.15	0.05	0.00	0.10	0.15	−0.35	−0.35
h	O–N	1.00	0.35	0.60	0.85	0.85	0.95	0.95	0.30	0.45
i	C–O	−0.40	−0.30	−0.30	−0.25	−0.25	−0.25	−0.25	−0.30	−0.30
j	C–O	−0.25	−0.50	−0.45	−0.30	−0.30	−0.25	−0.25	−0.55	−0.45
k	C–O	−0.25	−0.50	−0.40	−0.25	−0.25	−0.20	−0.20	−0.55	−0.45
l	C–O	−0.45	−0.30	−0.30	−0.25	−0.25	−0.25	−0.25	−0.35	−0.30
m	Al–O	−0.15	−0.30	−0.20	−0.05	−0.05	−0.05	−0.05	−0.30	−0.25
n	Al–O	−0.10	−0.40	−0.20	−0.10	−0.10	−0.10	−0.10	−0.25	−0.15
o	Al–O	−0.15	−0.25	−0.20	−0.10	−0.10	−0.05	−0.05	−0.30	−0.25
p	Si–O	0.00	−0.25	−0.15	0.05	0.05	0.10	0.10	−0.30	−0.25
q	Si–O	−0.30	−0.45	−0.35	−0.25	−0.25	−0.20	−0.20	−0.45	−0.35
r	Si–O	0.00	−0.25	−0.15	0.05	0.05	0.05	0.10	−0.30	−0.15
s	C–Si	0.70	1.10	0.80	0.70	0.75	0.70	0.70	0.90	0.85
t	O–Si	0.60	0.80	0.60	0.60	0.60	0.60	0.60	0.70	0.60
u	O–Si	0.60	1.40	0.70	0.50	0.50	0.50	0.50	0.90	0.70
v	O–Si	0.65	1.40	0.80	0.65	0.65	0.60	0.55	1.00	0.80
w	O–Si	0.65	1.40	0.80	0.65	0.65	0.60	0.55	1.00	0.80
x	C–S	0.45	0.30	0.40	0.40	0.40	0.40	0.40	0.35	0.40
y	S–S	0.35	0.15	0.30	0.35	0.35	0.35	0.35	0.25	0.30

that in the proton affinity calculations, all balanced charge schemes give similar mean unsigned errors (MUEs) of 25 molecules in the test suite. As the link atom affects the charge distribution near the QM/MM boundary, it is possible that the errors brought by the different charge schemes and by the link atom are of the same order of magnitude. In the next section, we will first tune the link atom to make the total charge of the QM subsystem right and then compare different charge schemes.

The position of the active site for each molecule is also listed in the Tables 4 and 5. We found that even when the active site is far from the boundary, the error is still large if the boundary is not well treated. For example, in molecule **v**, the active site is at the Q5 atom but the errors range from −5.9 to 10.4 kcal/mol using different link atoms and different balanced charge schemes. Therefore, increasing the distance of the active site from the boundary cannot completely remove the error due to the link atom. When a polar bond is cut, a tuned link atom and balanced charge scheme should be used.

4.C. Tuned F Link Atom. We used the protocol presented in section 2.B to tune the F link atoms. The final parameters *C* of the pseudopotentials for the tuned F atoms are shown in Table 6. These parameters reflect the differences among various kinds of bonds. The same type of bond shows similar parameters even in different molecules. For example, in the balanced RC scheme, the parameter for the pseudopotential is 0.65–1.45 for a carbon boundary atom, −0.15–0.60 for a nitrogen boundary atom, −0.45 to −0.15 for an oxygen boundary atom, 0.60–0.80 for a silicon boundary atom, and 0.30–0.40 for a sulfur boundary atom. These results are consistent with the electronegativities of the atoms. Also, we found that when the M1 atom is less electronegative than the Q1 atom, the pseudopotential for the tuned F atom is larger. For example, the pseudopotential

needed for an O–N bond (0.60) is much larger than that needed for a C–N bond (−0.15).

The tuned F atom was used as the link atom to calculate the proton affinities, and the results are shown in Table 7. By examination of the mean unsigned error (MUE) of the 25 molecules, one finds that the tuned F link atom gives smaller errors than the H link atom (Table 5) or the F link atom (Table 6) in *all* the charge schemes. This indicates that the accurate treatment of the boundary is very important, especially after the total charge is conserved. The tuned F link atom makes the total charge of the QM region right, and it avoids the artifacts that can be introduced by the use of a link atom.

The balanced RC scheme gives the best results, and the MUE is only 1.3 kcal/mol. The good performance of the balanced RC scheme can be explained as follows. In order to correctly handle the electrostatic interactions between the QM region and the MM region in the QM/MM calculations, it is important to have an accurate charge distribution of the MM region. When we move the boundary charges to avoid overpolarization, the charge distribution of the MM region can be greatly changed if the redistributed charges are moved far from the boundary. In the balanced RC method, the redistributed charges are moved to the midpoints of the M1–M2 bonds, and compared to other charge schemes, they introduce smaller changes to the charge distributions in the MM regions. In the balanced RC2, RC3, and Amber-2 schemes, the redistributed charges are placed farther and farther from the boundary region and the MUE of the proton affinities increases. When we use balanced RC3, in which the charges are redistributed to M2 and M3 atoms, the error is approximately equal to that in the capped primary system (CPS). However, when the bond is nonpolar, such as a C–C bond, the redistributed charge is relatively small and different charge schemes give similar errors.

Table 7. Full-QM Proton Affinities (PA, in kcal/mol), QM/MM Signed Errors (in kcal/mol), and Mean Unsigned Errors (MUE, in kcal/mol) Averaged over 25 cases Using the Tuned F Link Atoms

case	bond	PA	CPS	balanced SEE	balanced RC	balanced RC2	Amber-1	balanced RC3	Amber-2	balanced RCD	balanced Shift
a	C–C	362.3	2.6	–1.5	–0.8	0.0	–0.5	0.0	0.0	–2.0	–1.7
b	C–C	406.6	–6.2	–1.0	–0.2	0.6	–0.2	0.8	1.2	–1.1	–1.1
c	C–C	402.6	–2.0	1.6	2.0	3.1	2.8	3.5	3.5	1.5	1.6
d	N–C	376.7	0.1	1.7	1.0	–0.4	0.5	–1.3	–1.9	2.4	2.7
e	O–C	398.6	–0.2	0.1	–0.6	–1.4	–0.6	–1.7	–2.5	0.8	0.7
f	S–C	394.7	–0.8	1.1	0.3	–0.7	–0.2	–2.3	–2.5	1.5	1.2
g	C–N	355.2	2.8	4.0	2.5	–1.2	2.3	–5.6	–9.0	6.1	4.8
h	O–N	400.0	–3.4	4.6	2.8	–1.2	–1.3	–2.6	–2.6	6.3	5.3
i	C–O	394.9	0.6	1.0	1.1	0.4	0.4	–0.1	–0.5	2.0	1.8
j	C–O	401.0	–4.3	1.6	0.0	–2.5	–2.5	–3.6	–4.5	2.6	2.1
k	C–O	366.7	–1.6	2.0	0.2	–3.1	–3.1	–4.7	–5.7	3.4	2.5
l	C–O	398.3	–0.3	0.6	0.3	–0.1	–0.1	–0.4	–0.6	0.7	0.8
m	Al–O	340.7	1.7	1.8	0.3	–2.0	–2.0	–3.3	–3.3	2.8	1.8
n	Al–O	339.1	3.4	1.5	0.3	–2.3	–2.3	–3.6	–3.5	2.9	2.1
o	Al–O	348.0	–5.6	1.2	0.0	–1.9	–1.9	–2.8	–3.4	1.9	1.2
p	Si–O	349.0	–8.6	–1.2	–3.4	–6.5	–6.5	–7.9	–9.7	–0.2	–1.8
q	Si–O	353.2	–3.2	1.6	0.5	–1.1	–1.1	–1.9	–3.3	2.1	1.7
r	Si–O	348.1	–7.0	–0.3	–2.7	–6.1	–6.1	–7.9	–9.5	0.8	–0.6
s	C–Si	397.0	3.5	–5.0	–2.2	2.1	1.3	2.1	2.1	–6.5	–4.8
t	O–Si	342.7	–0.5	–2.2	–1.9	–0.5	–0.3	–0.5	–0.5	–2.7	–2.6
u	O–Si	348.0	–5.8	–2.7	–0.6	3.8	4.8	5.9	5.9	–5.1	–3.4
v	O–Si	353.2	–2.2	–0.7	0.4	3.1	4.2	4.7	6.9	–2.1	–1.4
w	O–Si	354.8	–3.8	–0.4	1.0	3.9	5.2	5.7	10.9	–1.8	–1.1
x	C–S	395.8	4.3	8.8	7.7	5.8	5.8	5.3	5.3	9.2	8.6
y	S–S	390.3	–1.9	1.0	0.4	–0.6	–0.6	–1.0	–1.0	1.3	1.0
	MUE		3.1	2.0	1.3	2.2	2.3	3.2	4.0	2.8	2.3

In all three tables of mean unsigned errors (Tables 4, 5, and 7), the balanced RC scheme is better than the balanced RCD scheme, whereas previously⁴² the RCD scheme performed slightly better. Since the testing is more thorough in the present paper, we now believe that the simpler RC scheme is to be preferred to the RCD one.

From the above results, we conclude that both a good charge scheme and an appropriate treatment of the link atom are needed to accurately treat the boundary. The balanced RC method with a tuned F atom gives the best results among all the methods.

However, there are still some problems. For example, we found that the error for the C–S bond is still quite large and there is no obvious reason for this error.

4.D. Tuning the F Link Atom Based on Other Methods.

Having established that we can obtain better results with a tuned fluorine atom, we should note that in future work one could also consider other ways to do the tuning. For example, the tuned parameter could be tuned to make the proton affinity or a particular reaction energy come out right at representative geometries of the reactant or product. There would be three advantages of this approach: (i) the proton affinity, unlike the partial charges, is a physical observable that can be calculated straightforwardly; (ii) the proton affinity depends on both the initial and final states, so one does not have to make a decision whether to tune the pseudopotential in the reactant state or the product state; (iii) there is additional flexibility in that one could tune the calculated proton affinity or reaction energy either (a) to a calculation in which a portion of the MM system (e.g., the M1, M2, and M3 atoms or the functional groups that contain them) is treated quantum mechanically but at the same level

as is to be used for the QM portion of the QM/MM calculations or (b) to experiments or high-level calculations.

5. Concluding Remarks

In QM/MM studies it is often inevitable to cut covalent bonds between the QM and MM parts. The present study addresses QM/MM boundary treatments in a systematic manner. All the ingredients considered are well-known in the literature from our own work and that of others (relevant references are cited above), in particular, the use of link atoms, the balancing and redistribution of charges close to the QM/MM boundary, and the tuning of the properties of the link atoms by suitable calibration (e.g., via a pseudopotential). In the present article, we present a new method of tuning and we combine it with the other ideas just mentioned in a new way to yield a method called the balanced redistributed charge (BRC) scheme with tuned fluorine link atom. If an acronym is needed to save space, this could be labeled the TBRC (tuned and balanced redistributed charge) method.

We apply the new method to calculate the proton affinities of 25 diverse compounds (at fixed geometries), and we compare its performance to that of other boundary charge and link atom schemes for treating the QM/MM boundary with regard to their ability to reproduce the quantum mechanical reference values. The balanced redistributed charge scheme with tuned fluorine link atoms outperforms the other treatments for the chosen validation suite of proton affinities, as shown in Tables 4, 5, and 7. Some of the methods listed in Table 4 are typical methods currently in use. For example, as mentioned in section 2.A, the Z1 scheme is the default in CHARMM,⁴⁷ the Z2 scheme is the default in both the *Gaussian 03*⁴⁸ and *Gaussian 09*⁴⁹

packages, and the Amber-2 method is the default in AMBER, version 10.⁴⁶

One important finding is that the errors are generally larger for treatments without charge balancing, whereas the choice of the actual charge redistribution scheme is less crucial, but not insignificant. From Tables 4 and 5, we see that the mean unsigned error (MUE) in the unbalanced schemes ranges from 15 to 24 kcal/mol, much worse than even the CPS method, which has no MM region (just a capped quantum subsystem) and which gives an MUE of 5 kcal/mol. The importance of balancing, which was previously emphasized by others,^{33,43–45} has a dramatic effect, reducing the MUE to 4–7 kcal/mol.

A key physical element of the new scheme is that we tune the link atom to try to make the charge on the primary subsystem (the atoms of the capped primary system but excluding the cap) be the same as it would be in a quantum mechanical calculation on the entire system. In the past there has been much more emphasis on redistributing the MM charge near the boundary than on the charge on the primary subsystem. If, however, the charge on the uncapped portion of the capped quantum mechanical system is inaccurate, then no treatment of the boundary can restore the correct physics. This motivation for tuning is substantiated by the finding that the use of a tuned fluorine link atom further reduces the mean unsigned error in all eight balanced charge schemes, leading to mean unsigned errors of 1–4 kcal/mol. In six of these eight schemes, the MUE is smaller than that for a CPS with a tuned F atom. All these eight schemes, give smaller MUEs than CPS capped with a hydrogen atom. The MUE of 1.3 kcal/mol for the combination of the balanced RC scheme for boundary charges and the tuned fluorine atom is particularly encouraging in light of the difficulty of the test set. In fact, the average proton affinity in all 25 cases of the test set is 373 kcal/mol, and the range of quantum mechanical reference values is 67.5 kcal/mol (from 339.1 to 406.6 kcal/mol). The mean unsigned error of 1.3 kcal/mol for the recommended new methods is only 1.9% of the range.

We conclude that two elements are very important in the energy calculations: balancing the charges of the MM region and tuning the link atom. A general rule (defined for all single bonds in the whole periodic table) is provided to tune the link atom when different types of single bonds are cut at the boundary. We can calculate accurate proton affinities after correctly handling the QM/MM boundary.

In the present work, motivated by interest in a variety of catalytic and redox systems, we intentionally chose a very difficult property, proton affinities, we considered QM/MM boundaries both close and far from the site of protonation, and we created a test set that is much more diverse than previous test sets used for QM/MM methods, in particular in that it includes boundaries that cut very polar bonds, including some in which neither atom is a carbon. Some of the methods found to be inadequate for this demanding test set will perform better for properties that are less sensitive than proton affinities to electrostatic potentials in the quantum subsystem or for cases where only a carbon–carbon bond is broken. However, simply enlarging the quantum system enough to move the QM/MM boundary far from the site of

reaction, when affordable, is not sufficient to guarantee good accuracy. Furthermore, in simulations of complex systems, it is often desirable to use a method that has been validated for even the most challenging problems.

In future work, we will examine the problem of geometry optimization using the new QM/MM scheme. In the present article, the valence and van der Waals terms that involve interactions between the QM subsystem and MM subsystem cancel out, but we will need a protocol for defining them when we begin to optimize geometries or consider dynamics.

Acknowledgment. We thank Masahiro Higashi, Hai Lin, Manjeera Mantina, and Jingjing Zheng for helpful discussions. This work was supported in part by the National Science Foundation under Grant No. CHE07-04974 and by the Office of Naval Research under Award Number N00014-05-1-0538.

References

- (1) Truhlar, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 16824.
- (2) Gao, J. *Rev. Comp. Chem.* **1996**, *7*, 119.
- (3) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
- (4) Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; Neuman Institute for Computing: Jülich, Germany, 2000; Vol. 3, p 285.
- (5) Nicoll, R. M.; Hindle, S. A.; MacKenzie, G.; Hillier, I. H.; Burton, N. A. *Theor. Chem. Acc.* **2001**, *106*, 105.
- (6) Colombo, M. C.; Guidoni, L.; Laio, A.; Magistrato, A.; Maurer, P.; Piana, S.; Röhrig, U.; Spiegel, K.; Sulpizi, M.; VandeVondele, J.; et al. *Chimia* **2002**, *56*, 13.
- (7) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
- (8) Shurki, A.; Warshel, A. In *Protein Simulations*; Vol. 66; Academic Press Inc.: San Diego, CA, 2003; p 249.
- (9) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schafer, A.; Lennartz, C. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1.
- (10) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.
- (11) Hammes-Schiffer, S. *Curr. Opin. Struct. Biol.* **2004**, *14*, 192.
- (12) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
- (13) Moret, M.-E.; Tapavicza, E.; Guidoni, L.; Röhrig, U.; Sulpizi, M.; Tavernelli, I.; Rothlisberger, U. *Chimia* **2005**, *59*, 493.
- (14) Mulholland, A. J. *Drug Discovery Today* **2005**, *10*, 1393.
- (15) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689.
- (16) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
- (17) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; et al. *J. Phys. Chem. B* **2006**, *110*, 6458.
- (18) Zhang, Y. *Theor. Chem. Acc.* **2006**, *116*, 43.

- (19) Senn, H. M.; Thiel, W. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182.
- (20) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
- (21) Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573.
- (22) Sousa, S. F.; Ramos, M. J. In *Computational Proteomics 2008*; Ramos, M. J., Ed.; Transworld Research Network: Kerala, India, 2008; p 101.
- (23) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.
- (24) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
- (25) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
- (26) Théry, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.
- (27) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
- (28) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.
- (29) Redondo, A.; Goddard, W. A.; Swarts, C. A.; McGill, T. C. *J. Vac. Sci. Technol.* **1981**, *19*, 498.
- (30) Koga, N.; Morokuma, K. *Chem. Phys. Lett.* **1990**, *172*, 243.
- (31) Nasluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2228.
- (32) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
- (33) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
- (34) Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.
- (35) Parks, J. M.; Hu, H.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 154106.
- (36) Alary, F.; Poteau, R.; Heully, J. L.; Barthelat, J. C.; Daudey, J. P. *Theor. Chem. Acc.* **2000**, *104*, 174.
- (37) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.
- (38) Moon, S.; Christiansen, P. A.; DiLabio, G. A. *J. Chem. Phys.* **2004**, *120*, 9080.
- (39) DiLabio, G. A.; Wolkow, R. A.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 044708.
- (40) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113.
- (41) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
- (42) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.
- (43) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019.
- (44) Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79.
- (45) Das, D.; Eurenus, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.
- (46) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; and Kollman, P. A.; *AMBER 10*; University of California: San Francisco, CA, 2008.
- (47) König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.
- (48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., 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*, revision D. 01; Gaussian, Inc.: Wallingford, CT, 2004.
- (49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (50) Pacios, L. F.; Christiansen, P. A. *J. Chem. Phys.* **1985**, *82*, 2664.
- (51) Lin, H.; Zhang, Y.; Truhlar, D. G. *QMMM*, version 1.3.5; University of Minnesota: Minneapolis, MN, 2007.
- (52) Ponder, J. W. *TINKER*, version 4.2; Washington University: St. Louis, MO, 2004.
- (53) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (54) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
- (55) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046.
- (56) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
- (57) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.: THEOCHEM* **1999**, *461*, 1.

CT900366M