

YinYang Atom: A Simple Combined *ab Initio* Quantum Mechanical Molecular Mechanical Model

Yihan Shao and Jing Kong*

Q-Chem Inc., 5001 Baum Boulevard, Pittsburgh, Pennsylvania 15213

Received: November 6, 2006; In Final Form: February 23, 2007

A simple interface is proposed for combined quantum mechanical (QM) molecular mechanical (MM) calculations for the systems where the QM and MM regions are connected through covalent bonds. Within this model, the atom that connects the two regions, called YinYang atom here, serves as an ordinary MM atom to other MM atoms and as a hydrogen-like atom to other QM atoms. Only one new empirical parameter is introduced to adjust the length of the connecting bond and is calibrated with the molecule propanol. This model is tested with the computation of equilibrium geometries and protonation energies for dozens of molecules. Special attention is paid on the influence of MM point charges on optimized geometry and protonation energy, and it is found that it is important to maintain local charge-neutrality in the MM region in order for the accurate calculation of the protonation and deprotonation energies. Overall the simple YinYang atom model yields comparable results to some other QM/MM models.

1. Motivation

When simulating complicated biochemical reactions, it is often necessary for one to combine quantum mechanical (QM) methods^{1–4} with molecular mechanical (MM) force fields,^{5–9} as QM methods are prohibitively expensive and MM models cannot always provide enough accuracy or simply lack the capability of treating the problem. At the heart of combined QM/MM models is a division of a molecular system of interest into a QM region and a MM region. Atoms in the QM region directly participate in bond formation and breaking and therefore warrant a quantum mechanical treatment. Atoms in the MM region, on the other hand, affect the reaction by providing an embedding electrostatic potential to the QM atoms.

Many literature exist describing different QM/MM approaches, and a much incomplete list is provided here.^{10–41} QM/MM modeling is expected to become even more widely used in the years to come, thanks to the latest efficiency improvements of QM calculations.⁴² Within the Q-Chem software package,^{43,44} for example, Coulomb techniques (CFMM,⁴⁵ Coulomb engine,⁴⁶ fast Fourier transform,⁴⁷ and resolution of identity⁴⁸), exchange-correlation techniques (IncDFT,⁴⁹ multi-resolution,^{50,51} SG-0 grid⁵²), and parallelization techniques^{53,54} are combined together to allow high-performance density functional theory (DFT) calculations on PC clusters.

In many biochemical reactions, the QM and MM regions are connected together through covalent bonds. A major concern, in the course of developing QM/MM models, is then how to handle these connecting bonds. Since the hybrid of those two fundamentally different methodologies is purely empirical, a variety of models have been reported. These models seem to fall into two broad categories: one is to make modifications to the MM algorithm while making little changes on the QM side, and the other is the opposite, concentrating the changes on the QM side.

The most popular approach to handle the QM/MM interface is to insert link atoms to simulate the covalent bonds connecting QM and MM regions, and requires no change to the existing

QM formalisms for most cases.^{11,12} The link atom is a hydrogen atom most of the time, as the connecting bond is usually a carbon–carbon one, and is constrained along the original covalent bonds at a proper distance. In perhaps the simplest link-atom model, ONIOM,¹⁷ the QM atoms and the link atom are not subjected to any embedding potential from atoms in the MM region. A more recent “learn on the fly” model^{37,38} was developed for molecular dynamics simulations, which takes into account partially the embedding potential through a buffer region.

In more sophisticated link-atom models, QM atoms (and sometimes the link atoms) are readily subjected to an electrostatic embedding potential from most if not all MM atoms. However, this electrostatic potential must be constructed with caution because the hydrogen link atom lies too close to the MM atoms in the immediate vicinity of the connecting bond. Available options for handling the electrostatic potential include: (a) the electrostatic interaction between the link atom and all MM atoms are neglected in the semiempirical QM formalism;^{11,12} (b) the electrostatic potential from the first MM neutral group is excluded;¹⁸ (c) the electrostatic potential from the first MM group is excluded, and nearby MM charges (further away from the interface) are modified to maintain charge-neutrality;²⁸ (d) the charge on the MM-end atom of a connecting bond is redistributed to the midpoints of related MM bonds;⁴⁰ (e) MM charges near the interface are smoothed with Gaussian functions;^{28,29,32} (f) link atoms have also been introduced to the MM region.²⁹ It should also be noted that different research groups might be applying significantly different electrostatic potentials on the QM region when they employ the “link atom model” on the same molecular system, making it sometimes difficult to compare those link atom models.

On the other hand, many attempts have been made to simulate the connecting bond directly based on QM principles so that the use of the fictitious link atom can be avoided. One way of such a simulation is to use some type of orbital function. They include the following: (a) The first is the hybrid orbital model from Warshel and Levitt.¹⁰ A single hybrid orbital on the MM-

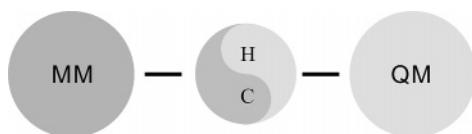


Figure 1. QM/MM interface with a YinYang atom

end atom of a connecting bond (pointing along the connecting bond to the nearest-neighbor QM atom) is included. (b) The second is the local self-consistent field (LSCF) model from Rivail et al.^{13–16} Like Warshel's model, a hybrid orbital on the MM-end atom is also included. Further this orbital is assigned a fixed fractional occupation number. (c) Third, there is the generalized hybrid orbital (GHO) model from Gao et al.^{20–22} A sp^3 MM-end atom would contribute three hybrid orbitals, which adds to QM electronic density, and a fourth hybrid orbital (pointing to its nearest-neighbor QM atom) as an extra basis function for the QM calculation. (d) Finally there is the frozen bond orbital model from Friesner et al.^{23,24} The connecting covalent bond is described with a frozen bond orbital.

In the second kind of link-atom-free QM/MM models, the connecting covalent bond is formed by making the atom at the MM end of the bond a QM atom through effective potentials: (a) Zhang and Yang et al.^{23,24} replaced the MM-end atom with a seven valence-electron atom with optimized effective core potentials and customized s, p basis functions; (b) Christiansen et al.³⁰ developed quantum capping potential for MM host carbon atoms; (c) Poteau et al.³⁴ developed effective group potential for different functional groups; (d) Yasuda and Yamaki developed one-electron effective potentials for methyl group;³⁶ (e) Lilienfeld et al. variationally optimized effective atom centered potentials to describe the methyl group in acetic acid;³⁹ (f) Slavicek and Martinez developed effective potentials for ground-state and excited calculations.⁴¹

The aim of this work is to design a QM/MM model that is conceptually simple and easy to implement at the same time. For the sake of simplicity, our model shall (1) involve no link atoms and (2) involve no geometry constraints. Arbitrary geometry constraints can sometimes lead to extra imaginary frequencies, thus complicating the search for equilibrium geometries and transition states. The model will also (3) involve no localized bond orbitals nor core potentials. We attempted the use of bond orbitals (in unpublished work), but they alone do not yield accurate length for the connecting covalent bonds (and nearby QM bonds), plus it would require extra work for formulating analytical derivatives. (4) The model will use as few new parameters as possible.

In our model, the atom at the MM end of the connecting covalent bond (usually a carbon) is made to have two characters. It is essentially a hydrogen atom to the rest of QM atoms, and a regular MM carbon atom to the rest of MM atoms. We call it a YinYang atom because of this double character as shown in Figure 1. Standard carbon parameters in a force field are used for the YinYang atom as part of the MM region. Standard hydrogen basis functions are used for the YinYang atom as part of the QM region, with the nuclear charge modified according to the MM charge value to maintain charge-neutrality.

YinYang model resembles link atom models in that a simple hydrogen atom is employed to form the covalent bond. It is also quite similar to the second kind of link-atom-free approach discussed above (or Antes and Thiel's "adjusted connection atoms" model,¹⁹ where semiempirical parameters for the MM-end atom are adjusted) except that the QM side of the YinYang atom does not involve any complicated effective potentials. In

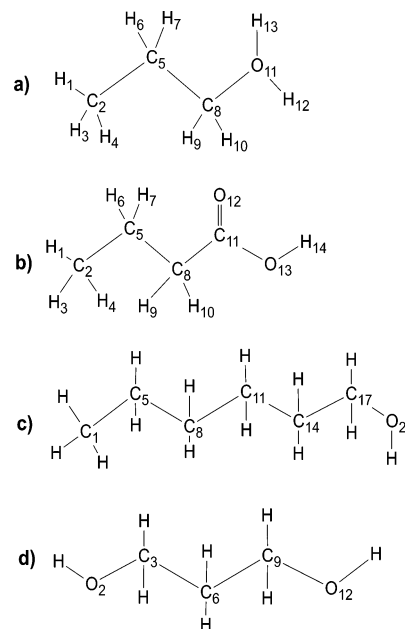


Figure 2. Propanol, butanoic acid, hexanol, and propan-1,2-diol

choosing the simple hydrogen atom, we note that the link-atom models work equally well as the link-atom-free models numerically.^{28,55}

The rest of the paper is organized as follows. Details and variations of the YinYang model will be described in the next section, and its performance will be evaluated in section 3. A summary will be provided in section 4. Special attention will be paid to the effect of the embedding MM potential on the electronic structure of the QM region. The MM force fields are carefully constructed for the simulation of long range point charge interactions. We will show that adjusting those parameters will have quite significant effect on the QM region.

2. Computational Model

2.1. Model Molecule: Propanol. The basics of the YinYang model has been described in the above introduction. To illustrate, we will use propanol as an example. This molecule, shown in Figure 2a, will also serve as a test case for further adjustment of the model. The geometry of propanol is optimized with density functional theory, specifically B3LYP functional,^{56–58} 6-31G* basis set,⁵⁹ and SG-1 grid⁶⁰ using the Q-Chem 3.0 software package.⁴⁴ Some key bond lengths and bond angles are listed in the second column of Table 1. We shall aim at reproducing these geometry parameters with our QM/MM models.

In Table 1, we also list the optimized geometries with the CHARMM27 force field^{5,8} (as implemented within Q-Chem 3.0), where the largest difference (wrt. B3LYP/6-31G* results) in bond lengths is 0.018 Å, and the difference in bond angles is usually less than 1.5° (note C5–C8–O11 has a difference of –4.12°). Results with other DFT functionals are also presented: with BLYP,^{57,58} there is a difference of up to 0.017 Å in bond lengths and up to 0.83° in bond angles; with EDF1,⁶¹ there is a difference of up to 0.004 Å in bond lengths and up to 0.67° in bond angles. All these differences, especially the ones from CHARMM27, will serve as a measurement of the quality of the YinYang QM/MM model.

2.2. The Basic YinYang Model. Suppose that we wish to treat CH_2OH of the propanol as the QM region, and the rest (the CH_3CH_2 group) as the MM region to be described with the CHARMM27 force field.⁸

TABLE 1: Bond Lengths (in Ångströms) and Bond Angles (in Degrees) for the Optimized Structure of Propanol (see Figure 2a) with B3LYP, CHARMM27, BLYP, and EDF1^a

	B3LYP	CHARMM27	BLYP	EDF1
C2–C5	1.532	1.532 (−0.000)	1.544 (0.012)	1.533 (0.001)
C5–H6	1.101	1.113 (0.013)	1.108 (0.008)	1.104 (0.003)
C5–H7	1.097	1.115 (0.018)	1.105 (0.007)	1.100 (0.003)
C5–C8	1.530	1.528 (−0.002)	1.542 (0.012)	1.534 (0.004)
C8–H9	1.096	1.114 (0.018)	1.103 (0.008)	1.099 (0.004)
C8–H10	1.104	1.113 (0.010)	1.112 (0.008)	1.108 (0.004)
C8–O11	1.423	1.421 (−0.002)	1.440 (0.017)	1.427 (0.004)
O11–H12	0.970	0.961 (−0.009)	0.981 (0.011)	0.974 (0.004)
max diff		0.018	0.017	0.004
rms diff		0.011	0.011	0.004
C2–C5–H6	109.55	109.17 (−0.38)	109.50 (−0.05)	109.41 (−0.14)
C2–C5–H7	110.50	109.19 (−1.30)	110.46 (−0.03)	110.37 (−0.13)
C2–C5–C8	112.76	113.29 (0.53)	112.90 (0.14)	113.20 (0.44)
H6–C5–H7	106.49	107.42 (0.93)	106.50 (0.02)	106.36 (−0.13)
H6–C5–C8	108.94	109.21 (0.27)	108.86 (−0.08)	108.93 (−0.01)
H7–C5–C8	108.40	108.39 (−0.02)	108.39 (−0.01)	108.33 (−0.07)
C5–C8–H9	110.04	109.70 (−0.33)	110.13 (0.09)	110.00 (−0.03)
C5–C8–H10	109.85	110.78 (0.93)	109.82 (−0.03)	109.81 (−0.04)
C5–C8–O11	113.12	109.00 (−4.12)	113.21 (0.09)	113.45 (0.34)
H9–C8–H10	106.99	108.29 (1.30)	107.10 (0.10)	106.86 (−0.13)
H9–C8–O11	105.38	108.90 (3.52)	105.02 (−0.35)	105.10 (−0.27)
H10–C8–O11	111.20	110.14 (−1.06)	111.29 (0.09)	111.31 (0.11)
C8–O11–H12	107.36	106.92 (−0.44)	106.53 (−0.83)	106.69 (−0.67)
max diff		−4.12	−0.83	−0.67
rms diff		1.67	0.26	0.27

^a The values in parentheses are differences with respect to B3LYP results in the second column.

With the YinYang model, C5 in Figure 2a then becomes a YinYang atom. The total energy for such a system is

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} \quad (1)$$

The calculation of E_{QM} involves the CH_2OH group plus the YinYang atom. The nuclear charge on the YinYang atom is 0.82, which is a unit charge (for a hydrogen atom) plus -0.18 , the CHARMM27 charge for methylene carbon. The basis functions for the YinYang atom are chosen to be the same ones for a hydrogen atom. This QM system has 18 electrons, counting the pair for the C5–C8 bond, and is subject to 6 “external point charges”: all hydrogen atoms (H1, H3, H4, H6, H7) with a charge of 0.09 and C2 with a charge of -0.27 .

The MM part of the calculation includes the CH_3CH_2 group. Following the conventional MM algorithms, E_{MM} contains several components:⁵ bond energy, angle energy, torsion energy, Urey–Bradley energy, Coulomb energy, and vdw energy. When any of those terms involve a QM atom, the following rules, based on the rules of the MM force field, apply:

(1) For bond, angle, Urey–Bradley, and vdw energies, at least one atom is a MM atom or YinYang atom.

(2) For torsion energies, at least one center atom is a MM atom or YinYang atom.

(3) For Coulomb energies, both atoms must be MM atoms.

The evaluation of analytical energy gradient is straightforward for the total energy as defined above, where the QM forces on the “external point charges” (MM atoms) are calculated the normal way when the point charges on the MM atoms are treated as nuclear ones.

As a test, an QM/MM calculation as described above is carried out with the QM at the level of B3LYP/6-31G* and the results are listed in Table 2. As one can see most of the differences in bond length and bond angle are comparable to the differences between CHARMM27 and B3LYP in Table 1. The lone exception is the bond length of the connecting bond between C5 and C8, which is 0.129 Å too short. The shortening

TABLE 2: Bond Lengths (in Ångströms) and Bond Angles (in Degrees) for the Optimized Structure of Propanol with YinYang QM/MM Models^a

	basic model	adjusted model
C2–C5	1.538 (0.005)	1.533 (0.001)
C5–H6	1.122 (0.021)	1.116 (0.016)
C5–H7	1.123 (0.026)	1.118 (0.020)
C5–C8	1.401 (−0.129)	1.531 (0.001)
C8–H9	1.098 (0.002)	1.093 (−0.003)
C8–H10	1.106 (0.002)	1.099 (−0.004))
C8–O11	1.423 (−0.000)	1.412 (−0.011))
O11–H12	0.968 (−0.002)	0.968 (−0.002)
max diff	−0.129	0.020
rms diff	0.047	0.010
C2–C5–H6	108.02 (−1.54)	109.21 (−0.34)
C2–C5–H7	108.08 (−2.42)	109.26 (−1.23)
C2–C5–C8	114.57 (1.82)	114.73 (1.97)
H6–C5–H7	106.65 (0.17)	107.72 (1.23)
H6–C5–C8	109.73 (0.79)	108.00 (−0.94)
H7–C5–C8	109.48 (1.08)	107.70 (−0.70)
C5–C8–H9	109.06 (−0.97)	107.38 (−2.66)
C5–C8–H10	109.41 (−0.44)	107.65 (−2.20))
C5–C8–O11	112.40 (−0.72)	112.31 (−0.81))
H9–C8–H10	107.93 (0.94)	109.42 (2.43)
H9–C8–O11	106.00 (0.62)	107.01 (1.63)
H10–C8–O11	111.86 (0.66)	112.91 (1.71)
C8–O11–H12	107.47 (0.12)	107.68 (0.32)
max diff	−2.42	−2.66
rms diff	1.13	1.58

^a Errors are computed against B3LYP/6-31G* results and are included in parentheses.

of this does not come as a total surprise, because C5, the YinYang atom, is represented in the QM system as a hydrogen-like atom (but with a net nuclear charge of 0.82) instead of a carbon atom. A C–H bond is shorter than a C–C single bond.

One way to correct this is to simulate the effective potential of a carbon atom in this environment.^{23,24,30,34,36,41} Instead, we elect to overcome this deficiency (connecting bond shortening) with a simple classical approach: adding an extra (repulsive) Coulomb energy term.

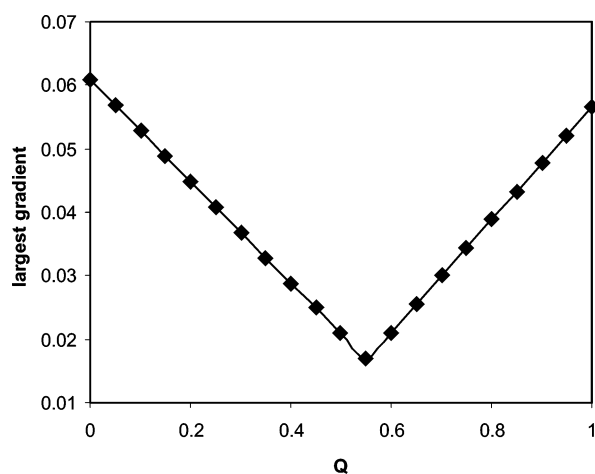
2.3. Adjustment: Connecting Bond Length Corrected Through an Extra Coulomb Energy Term. In order to lengthen the connecting bond, an extra classical Coulomb repulsion, (Q/r) , was added between the atoms that makes up the connecting bond, namely the YinYang atom and the QM atom connected to it, with r being the length of the connecting bond. In order to find an optimal value for Q , we examine the energy gradient with this model at fully optimized quantum geometry (second column in Table 1) at different values of Q . The gradient of the basic YinYang model, i.e., when Q is zero, is tabulated in Table 3. Only the two components (x and y) in the plane of C2–C5–C8 are listed, since the out-of-plane forces are zero due to symmetry. The largest forces are found on atoms C5 and C8 where the components of the gradient are larger than 0.01, while the force components on all other atoms are all much smaller. Our goal is to minimize them with the adjustment of Q .

When we add (Q/r) to the energy, only forces on C5 and C8 are going to be affected. The change in the forces with Q is shown in Figure 3. When Q is between 0 and 0.55, the largest absolute value of the gradient (y component of the force on C5) goes down linearly from 0.061 to 0.17. At $Q = 0.55$, the force on C5 is $(-0.001, -0.017, -0.003)$ and the force on C8 is $(-0.017, -0.008, -0.001)$. When Q is greater than 0.55, the x -directional force on C8 becomes the largest component and increases linearly with Q . Therefore, the optimal value for Q is 0.55.

TABLE 3: Energy Gradient with the YinYang Model (No Coulomb Energy Correction) at Full-Optimized QM Geometry^a

atom	gradient	
	x	y
H1	0.005	0.007
C2	-0.001	0.003
H3	-0.001	-0.003
H4	-0.001	-0.005
C5	-0.050	-0.061
H6	-0.001	0.005
H7	0.000	0.004
C8	0.032	0.036
H9	0.006	0.005
H10	0.005	0.006
O11	0.006	0.004
H12	0.000	-0.001

^a The backbone of propanol, C2–C5–C8–O11, lies roughly in the xy plane, with C2 on the -x end.

**Figure 3.** Largest gradient component vs Q (in a basic YinYang QM/MM calculation).

The optimized geometry with this adjustment ($Q = 0.55$) is listed as the third column in Table 2. Here we observe a very significant improvement in bond length: C5–C8 now falls within 0.001 Å from its full QM length. Meanwhile, there is an increase in the error of bond angles. Most noticeably, C5–C8–H9 and C5–C8–H10 now become even smaller, with errors greater than -2° . With our models, these angles would resemble H–C–H angles, which are usually smaller than C–C–H angles (see Tables 1 and 2). With C5, the YinYang atom, being pushed further away from the QM region with the Coulomb energy correction, the C5–C8 connecting bond becomes even less repulsive, which therefore leads to even smaller values for C5–C8–H angles. This difference, however, is smaller than the maximum difference of this bond angle

between B3LYP and CHARMM27 calculations. This adjusted model will be used as the YinYang model in the rest of this study.

3. Results and Discussions

In this section, we will present results from QM/MM calculations with the YinYang model on some representative molecules and compare them with the results of other QM/MM models from the literature. The protonation energies and the structural differences of the neutral and the protonated/deprotonated species between the full B3LYP/6-31G* QM method and YinYang model are summarized in Tables 4–7. Molecules in Table 4 contain hydroxyl group, amine group, or carboxyl group in the QM region, and an ethyl group makes up the MM region. Molecules in Table 5 have the even smaller methyl group as the MM region. In Table 6, we present results for longer-chain alcohols and for molecules with hydroxyl group or amine group in the MM region. In all those calculations, the YinYang atom is a carbon atom, and is denoted as C_y . All the other QM atoms are denoted in boldface. Most of these molecules were studied with other QM/MM models in the following references:

(a) Zhang used pseudo-bond calculations²⁴ where B3LYP/6-31G* is applied to the QM region.

(b) Nicoll et al. used a link-atom model.²⁸ The QM region is handled with Hartree–Fock/6-31G*, and the MM region with AMBER force field.

(c) Brooks et al. used a double-link-atom model (DLA) with a Gaussian blur width of 1.0 Å (for protonation) or 1.7 Å (for deprotonation).²⁹ Some of their results with the B3LYP/6-31G or B3LYP/6-31G* method applied to the QM region will be mentioned.

(d) DiLabio et al. used a QM/MM model with capping potential.³⁰ Hartree–Fock/6-31G* is applied on the QM region.

(e) Amara and Field obtained QM/MM results.³² Their $LA_g(4, 0)$ and $LA_g(4, 3)$ results with the HF/6-31G* method applied to the QM region are of particular interest to us. It is noted that the OPLS-AA force field is applied in their work.

(f) Pu, Gao, and Truhlar obtained results with their GHO interface.²² We will focus on their B3LYP/6-31G* results with scaling.

(g) Lin and Truhlar obtained results with ESP charges and an RCD scheme.⁴⁰ Note that their QM method is HF/MIDI!

In the following subsections, the structures, and the energetics of each molecule and their charged species will be discussed in detail, with the focus on the difference between QM/MM and full QM calculations.

3.1. $C_3H_7O^-$, C_3H_7OH , and $C_3H_7OH_2^+$. One can see from the results of YinYang QM/MM calculations listed in Table 4 for these species that the geometries and the relative energetics agree quite well with full QM results. The best agreement is

TABLE 4: Errors in the Optimized Geometry (Bond Lengths in Ångstroms and Bond Angles in Degrees) and Protonation Energies for Some Model Compounds^a

molecule	bond length diff		bond angle diff		protonation energy			
	max	rms	max	rms	QM	YinYang	error	literature
$CH_3C_yH_2CH_2O^-$	0.020	0.014	2.83	1.18	-398.42	-399.96	-1.54	-1.87, -2.30, ³² -3.40 ²²
$CH_3C_yH_2CH_2OH$	0.020	0.010	-2.66	1.58	-196.12	-196.96	-0.84	9.64, 8.73 ³²
$CH_3C_yH_2CH_2OH_2^+$	-0.039	0.020	5.40	2.44				
$CH_3C_yH_2CH_2NH^-$	0.030	0.018	2.98	1.43	-425.03	-425.78	-0.75	-3.37, -3.51, ³² -4.2 ²²
$CH_3C_yH_2CH_2NH_2$	0.016	0.009	2.45	1.55	-231.30	-233.21	-1.90	6.51, 7.43, ³² 6.0 ²²
$CH_3C_yH_2CH_2NH_3^+$	0.025	0.013	3.69	1.75				
$CH_3C_yH_2CH_2COO^-$	0.021	0.011	-5.27	2.86	-363.30	-372.30	-9.00	0.64, 0.24, ³² -1.1 ²²
$CH_3C_yH_2CH_2COOH$	0.025	0.013	-4.19	2.32				

^a C_y denotes the carbon atom which becomes the YinYang atom.

TABLE 5: Errors in the Optimized Geometry (Bond Lengths in Ångstroms and Bond Angles in Degrees) and Protonation Energies for Some Model Compounds^a

molecule	bond length diff		bond angle diff		protonation energy			
	max	rms	max	rms	QM	YinYang	error	literature
$C_{\gamma}H_3CH_2O^-$	-0.027	0.016	-4.42	1.80	-399.59	-398.07	1.52	1.9, ²⁴ 4.45, -1.84, ³² -2.0, ²² -8.6 ⁴⁰
$C_{\gamma}H_3CH_2OH$	0.017	0.011	-2.42	1.48	-194.64	-196.48	-1.84	12.2, 7.00 ³²
$C_{\gamma}H_3CH_2OH_2^+$	-0.038	0.020	-4.09	2.15				
$C_{\gamma}H_3CH_2NH^-$	0.039	0.021	-5.03	2.10	-425.90	-423.70	2.20	3.73, -1.94, ³² -2.6 ²²
$C_{\gamma}H_3CH_2NH_2$	-0.015	0.010	2.41	1.62	-230.09	-232.65	-2.56	1.1, ²⁴ 0.7, ²⁹ 9.95, -4.90, ³² 3.9, ²² 0.8 ⁴⁰
$C_{\gamma}H_3CH_2NH_3^+$	0.024	0.013	-3.23	1.50				
$C_{\gamma}H_3CH_2COO^-$	0.017	0.009	-5.23	2.48	-357.64	-365.55	-7.91	-3.8, ²⁴ -3.3, ²⁹ 4.06, 0.26, ³² -0.6, ²² -3.6 ⁴⁰
$C_{\gamma}H_3CH_2COOH$	-0.022	0.013	-3.60	1.89				
$C_{\gamma}H_3CH_2S^-$	-0.020	0.010	-2.64	1.57	-367.84	-371.92	-4.08	-0.7, ²⁴ -3.0, ²⁹ -1.0, ²² -4.1 ⁴⁰
$C_{\gamma}H_3CH_2SH$	0.020	0.013	-2.66	1.55				

^a C_{γ} denotes the carbon atom which becomes the YinYang atom.

TABLE 6: Errors in the Optimized Geometry (Bond Lengths in Ångstroms and Bond Angles in Degrees) and Protonation Energies for Some Model Compounds^a

molecule	bond length diff		bond angle diff		protonation energy			
	max	rms	max	rms	QM	YinYang	error	literature
$CH_3C_{\gamma}H_2(CH_2)_2O^-$	-0.026	0.011	-3.29	1.97	-397.94	-402.04	-4.10	-3.2, ²⁹ 0.36, -0.02, ³² -2.0 ²²
$CH_3C_{\gamma}H_2(CH_2)_2OH$	0.019	0.010	-2.83	1.53	-197.14	-203.62	-6.48	0.5 ²⁹ , 3.90, 3.56 ³²
$CH_3C_{\gamma}H_2(CH_2)_2OH_2^+$	0.071	0.026	-3.93	1.95				
$CH_3C_{\gamma}H_2(CH_2)_4O^-$	0.016	0.006	-2.46	1.14	-394.35	-396.96	-2.61	
$CH_3C_{\gamma}H_2(CH_2)_4OH$	0.019	0.007	-2.41	1.09	-198.75	-200.65	-1.90	
$CH_3C_{\gamma}H_2(CH_2)_4OH_2^+$	0.021	0.008	-2.64	1.09				
$HOCH_2C_{\gamma}H_2CH_2O^-$	-0.020	0.013	2.38	1.13	-396.32	-396.00	0.32	
$HOCH_2C_{\gamma}H_2CH_2OH$	0.021	0.011	-2.63	1.47	-194.97	-194.16	0.81	4.4 ²⁹
$HOCH_2C_{\gamma}H_2CH_2OH_2^+$	-0.037	0.019	-4.33	2.15				
$HOCH_2C_{\gamma}H_2CH_2NH^-$	-0.025	0.015	2.65	1.36	-422.99	-422.36	0.63	
$HOCH_2C_{\gamma}H_2CH_2NH_2$	0.016	0.010	2.57	1.27	-230.42	-230.65	-0.23	2.5 ²⁹
$HOCH_2C_{\gamma}H_2CH_2NH_3^+$	-0.036	0.021	4.85	1.99				
$H_2NCH_2C_{\gamma}H_2CH_2NH^-$	-0.025	0.015	2.57	1.43	-423.82	-427.76	-3.94	
$H_2NCH_2C_{\gamma}H_2CH_2NH_2$	0.018	0.008	-2.46	1.57	-232.09	-236.87	-4.78	4.6 ²⁹
$H_2NCH_2C_{\gamma}H_2CH_2NH_3^+$	0.026	0.014	-3.61	1.67				

^a C_{γ} denotes the carbon atom which becomes the YinYang atom.

TABLE 7: Errors in the Optimized Geometry (Bond Lengths in Å and Bond Angles in Degrees) and Protonation Energies for Two Amino Acids^a

molecule	bond length diff		bond angle diff		protonation energy			
	max	rms	max	rms	QM	YinYang	error	literature
Ser^-	0.087	0.048	-7.40	3.56	-373.15	-377.80	-4.65	11.90, ²⁸ 0.29 ³²
$SerH_{\gamma}$	0.046	0.024	-6.90	2.90				
His	-0.044	0.015	-2.71	1.20	-244.51	-239.66	4.85	-5.03, ²⁸ -4.81 ³²
$HisH_{\epsilon}^+$	-0.060	0.017	-2.90	1.14				
His	-0.044	0.015	-2.96	1.30	-248.19	-242.01	6.18	1.20-2.60, ³⁰ 4.14 ³²
$HisH_{\epsilon}^+$	-0.053	0.016	-4.59	1.81				

^a For structures, see Figure 6. MM regions made up of α carbon and its hydrogen, amine group, and aldehyde (or carboxyl) group. The YinYang atom is an α carbon in our calculations.

found for neutral propanol, where the largest errors in bond length actually come from the MM region (C5-H6 and C5-H7, see Table 2), and the source of these errors can be traced back to the use of CHARMM27 force field (see Table 1). The interface C5-C8 bond has an error of 0.001 Å, and the C8-O11 bond has an error of -0.011 Å. When the system becomes charged, one observes larger errors in the geometry. For $C_3H_7O^-$, C5-C8 is 0.02 Å too long, and C8-O11 is 0.018 Å too short. For $C_3H_7OH_2^+$, C5-C8 is 0.006 Å too long, and C8-O11 is 0.039 Å too short. The more significant error in the C-O bond length for the cationic species can be related to the fact that a fully optimized C-O bond is 1.522 Å in $HCH_2OH_2^+$ (which our QM region resembles) and a much longer 1.571 Å in $C_3H_7OH_2^+$. Apparently, the embedding MM electrostatic potential provides some correction to the bond length (0.01 Å), but it could not account for the majority of the difference.

The energy differences between these neutral and charged species (see Table 4) from YinYang QM/MM calculations are fairly accurate when compared to full QM results: for propanol, its deprotonation energy comes with an error of 1.54 kcal/mol (a 0.4% error), and its protonation energy comes with an error of -0.84 kcal/mol (a 0.4% error). The results are compared favorably with those from the link-atom model in ref 32, where the same division of QM and MM region led to errors of 1.87 and 2.30 kcal/mol in the deprotonation energy, and 9.64 and 8.73 kcal/mol for the protonation energy. With the GHO interface,²² the error in the deprotonation energy is 3.4 kcal/mol.

In order to understand how the molecular structure and the protonation energy are affected by the embedding potential in the MM region, we have examined how the bond length and the protonation energy are affected by an external charge in

Appendices A and B. It is found that (a) an “external” point charge affects the length of an individual bond mainly through some kind of charge–dipole interaction and (b) an “external” point charge affects the protonation energy mainly through charge–charge interaction. As charge–charge interaction scales as $1/r$, the protonation energy can be very sensitive to charge values on close-by MM atoms. For example, if the charge on C2 in propanol (see Figure 2a) increases by 0.1, the protonation energy from YinYang QM/MM calculations would increase approximately by 10 kcal/mol (i.e., the energy gap between neutral C_3H_7OH and cationic $C_3H_7OH_2^+$ narrows down by 10 kcal/mol). Therefore, the accurate predictions of protonation and deprotonation energies seem to indicate that the long range electrostatic interaction in this group of structures is quite small, since all of the functional groups in the CHARMM27 force field are charge-neutral. On the other hand, the shortening of the polarized C–O bond indicates that some atomic charges in the MM region should be adjusted when the molecule is charged.

3.2. $C_3H_7NH^-$, $C_3H_7NH_2$, and $C_3H_7NH_3^+$. Results for propylamine and its protonated or deprotonated species are also summarized in Table 4. For the neutral propylamine, $C_3H_7NH_2$, the polarized C–N bond is shorter by 0.012 Å, while other bonds in the QM region have more accurate lengths. The C–N bond is 0.024 Å too short for $C_3H_7NH^-$, and 0.012 Å too short for $C_3H_7NH_3^+$.

The protonation energies from YinYang QM/MM calculations are fairly accurate: the error is 0.75 kcal/mol (0.2%) for the deprotonation energy and –1.90 kcal/mol (0.8%) for the protonation energy. This compares favorably with the errors (3.37–7.43 kcal/mol) reported in ref 32. In ref 22, the errors were 4.2 and 6.0 kcal/mol.

3.3. $C_3H_7COO^-$ and C_3H_7COOH . As shown in Table 4, the errors in the bond lengths in butanoic acid, C_3H_7COOH , and $C_3H_7COO^-$ are very similar to those for the previous two sets of alcohol or amine molecules. However, the deprotonation energy for C_3H_7COOH comes with a 9.0 kcal/mol (2.5%) error, which is significantly larger than errors for the previous two sets of molecules and the errors (0.24 or 0.64 kcal/mol) reported in ref 32 and –1.1 kcal/mol in ref 22.

We note that the carboxyl group is strongly electronegative, which can attract some electronic charge (roughly 0.05 unit of charge) away from C5, the YinYang atom, in a full-QM calculation. On the basis of the analysis in Appendices A and B, which are also summarized in subsection 3.1, the MM charge value on the YinYang atom can have a significant effect on the protonation energy. If one reduces the MM charge on C5, the YinYang atom, from –0.18 to –0.13, then the net charge on C5 increases from 0.82 to 0.87. With this change, one can indeed reduce the error in the protonation energy to 1.65 kcal/mol.

In general, we do not advocate such an arbitrary change in MM charge values. Alternatively, we can simply increase the size of the QM region. In the case of butanoic acid, we can choose the QM region as CH_2CH_2COOH , so that C2 becomes the YinYang atom (see Figure 2b). Then the error in the protonation energy is reduced to –2.38 kcal/mol.

3.4. $C_2H_5O^-$, C_2H_5OH , and $C_2H_5OH_2^+$. These structures test the limit of the YinYang QM/MM calculation as only three hydrogen atoms in the methyl group are considered as pure “MM” atoms (with point charges of 0.09), and the remaining carbon atom in the methyl group becomes the YinYang atom with the nuclear charge adjusted to 0.73 in the QM calculation to maintain charge neutrality. The same rule applies to molecules in the next three subsections.

The errors in the C–C single bond lengths are 0.028 Å (anion), 0.008 Å (neutral ethanol), 0.020 Å (cation), and the errors in the C–O single bonds are –0.021 Å (anion), –0.006 Å (neutral), and –0.038 Å (cation). The rest of the geometrical errors are much smaller. The energy errors for C_2H_5OH are tabulated in Table 5: –1.52 kcal/mol (deprotonation, 0.4%) and –1.84 kcal/mol (protonation, 0.9%).

In ref 24, similar errors in the bond length were reported for $C_2H_5O^-$ and C_2H_5OH , except for a much smaller error of –0.011 Å for the C–C bond in $C_2H_5O^-$. Their error in the deprotonation energy is –1.9 kcal/mol. In ref 32, errors of –4.45 and 1.84 kcal/mol were reported for the deprotonation energy, and 12.2 and 7.0 kcal/mol for the protonation energy. The error in the deprotonation energy was 2.0 kcal/mol in ref 22 and 8.6 kcal/mol (with ESP/RCD/HF/MIDI!) in ref 40.

3.5. $C_2H_5NH^-$, $C_2H_5NH_2$, and $C_2H_5NH_3^+$. The errors in the C–C single bond lengths are 0.039 Å (anion), 0.008 Å (neutral ethylamine), and 0.007 Å (cation), and the errors in the C–N single bonds are –0.029 Å (anion), –0.015 Å (neutral), and –0.014 Å (cation). For $C_2H_5NH_2$, the energy errors are –2.20 kcal/mol (0.5%) in the deprotonation energy and –2.56 kcal/mol (1.1%) in the protonation energy (see Table 5).

In ref 24, similar errors are reported for bond lengths (C–C, C–N, and N–H) for $C_2H_5NH_2$ and $C_2H_5NH_3^+$ and a more accurate protonation energy for $C_2H_5NH_2$ (error: 1.1 kcal/mol). In ref 29, errors in protonation energy are 3.4 kcal/mol with EXGR and 0.7 kcal/mol with DLA (with B3LYP/6-31G). In ref 32, the errors are –3.73 and 1.94 kcal/mol in the deprotonation energy and 9.95 and –4.90 kcal/mol for the protonation energy. In ref 22, the errors are 2.6 kcal/mol in the deprotonation energy and 3.9 kcal/mol in the protonation energy. In ref 40, the error in the protonation energy is 0.8 kcal/mol (with ESP/RCD/HF/MIDI!).

3.6. $C_2H_5COO^-$ and C_2H_5COOH . For propanoic acid, C_2H_5COOH , the errors in the bond lengths are as follows: –0.012 Å (C1–C5), –0.009 Å (C5–C8), +0.002 Å (C8–O9), 0 (C8–O10), and –0.001 Å (O10–H11). For $C_2H_5COO^-$, the errors are as follows: –0.005 Å (C1–C5), –0.015 Å (C5–C8), +0.004 Å (C8–O9), and 0 Å (C8–O10). Compared to results in ref 24, our bond lengths for C_2H_5COOH come with larger errors, while the errors in geometry for $C_2H_5COO^-$ are similar.

Our protonation energy for $C_2H_5COO^-$ has an error of –7.91 kcal/mol (2.2%, see Table 5). In comparison, the error is –3.8 kcal/mol in ref 24, where this protonation energy was in their training set, –3.3 kcal/mol with DLA in ref 29 and 4.06 and 0.26 kcal/mol in ref 32, and –0.6 kcal/mol in ref 22 and –3.6 kcal/mol in ref 32 (with ESP/RCD/HF/MIDI!). Again, we can improve the accuracy for the deprotonation energy by increasing the charge on the YinYang atoms, but this is somewhat arbitrary.

3.7. $C_2H_5S^-$ and C_2H_5SH . For ethanethiol, C_2H_5SH , the C–S bond is 0.018 Å too short, while the C–S bond is 0.020 Å too short in $C_2H_5S^-$ with the YinYang atom model. In comparison, the C–S bond is 0.016 Å too long in C_2H_5SH and –0.025 Å too short in $C_2H_5S^-$ in ref 24.

The error in the deprotonation energy from YinYang QM/MM calculations is 4.08 kcal/mol (1.1%) (see Table 5), which is larger than 0.7 kcal/mol in ref 24 and 3.0 kcal/mol in ref 29 and 1.0 kcal/mol in ref 22 and 4.1 kcal/mol in ref 40 (with ESP/RCD/HF/MIDI!).

3.8. $C_4H_9O^-$, C_4H_9OH , and $C_4H_9OH_2^+$. For 1-butanol, C_4H_9OH , the MM region in our YinYang QM/MM calculations consists of the methyl group and neighboring CH_2 group, whereas the QM region consists of two remaining CH_2 units

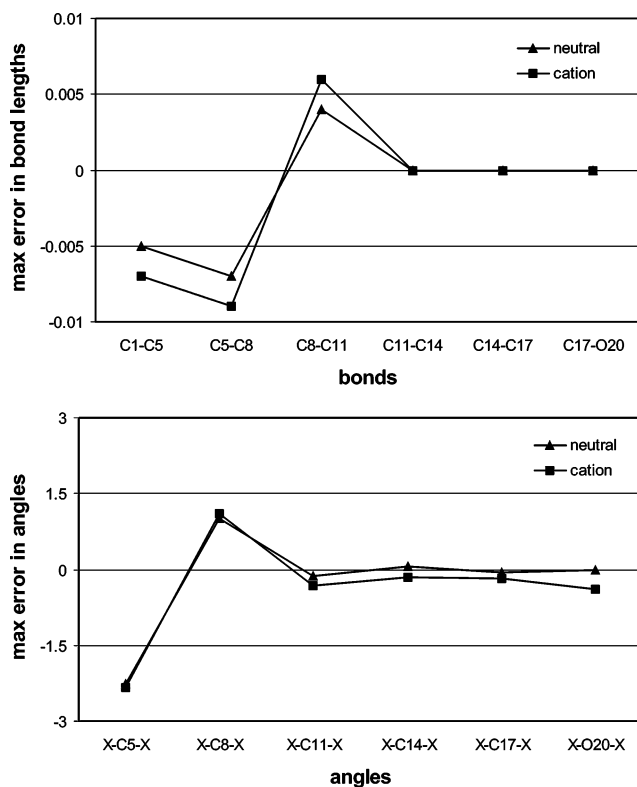


Figure 4. Error decay in geometry parameters for 1-hexanol. YinYang QM/MM calculation were carried with the methyl group being the MM region. C1 is the YinYang atom (see Figure 2c).

and the hydroxyl group. While the geometry for $C_4H_9O^-$ and C_4H_9OH has roughly the same accuracy as smaller molecules (see Table 6), the C–O bond in $C_4H_9OH_2^+$ is 0.071 Å too long, and we still do not understand why this happens.

The error in the deprotonation energy for 1-butanol from YinYang atom calculations is 4.10 kcal/mol (1.0%), while smaller errors were reported in the literature: a 3.2 kcal/mol error in ref 29 and errors of -0.36 or 0.02 kcal/mol in ref 32. For protonation energy, our error is -6.48 kcal/mol (2.5%), and the error is 0.5 kcal/mol in ref 29 and 3.90 or 3.56 kcal/mol in ref 32. The error in the deprotonation energy was 2.0 kcal/mol with the same QM/MM partitioning in ref 22.

We note that, while the errors in the protonation energies become larger for butanol than for propanol and ethanol (see Tables 4 and 5), it is not a general trend that the error increases with the size of QM region. For 1-butanol, the errors in the deprotonation energies are found to be 4.3 kcal/mol (QM = CH_2OH), 4.1 kcal/mol (QM = CH_2CH_2OH), and 2.86 kcal/mol (QM = $CH_2CH_2CH_2OH$), meanwhile the errors in the protonation energies are -0.99 kcal/mol (QM = CH_2OH), -6.48 kcal/mol (QM = CH_2CH_2OH), and -3.18 kcal/mol (QM = $CH_2CH_2CH_2OH$).

3.9. $C_6H_{13}O^-$, $C_6H_{13}OH$, and $C_6H_{13}OH_2^+$. For 1-hexanol, $C_6H_{13}OH$, and its cationic and anionic derivatives, several partitioning scheme were applied in our YinYang QM/MM calculations. In the first scheme, the MM region consists of the methyl group and the neighboring CH_2 group, just like the case of 1-butanol in the previous subsection. So the QM region is made up of four remaining CH_2 groups and the hydroxyl group. With this partitioning, the errors in the geometry and in the relative energies are listed in Table 6, and these errors are very similar to those of ethanol and propanol.

The errors in geometry parameters, such as bond lengths and bond angles, are expected to decay away from the QM/MM

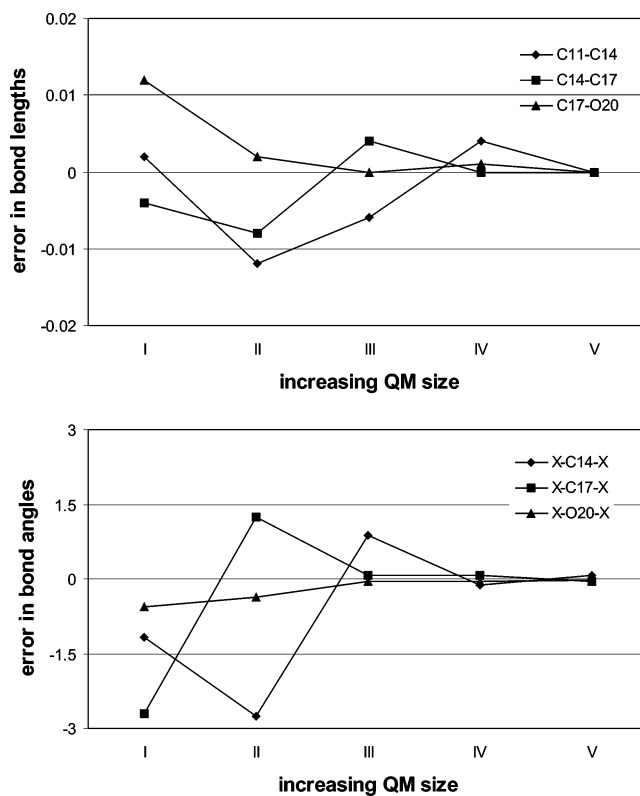


Figure 5. Error decay in geometry parameters for 1-hexanol. YinYang QM/MM calculation were carried with an increasing size of QM region. In model I, the last CH_2 group and OH group make up the QM region. In model II, one more CH_2 group is included in the QM region. By model V, all CH_2 groups are included in the QM region.

interface. In order to confirm that, we first adopt a partitioning where the MM region consists only of the methyl group (in which the carbon atom becomes the YinYang atom). As shown in the upper panel of Figure 4, the errors in bond lengths goes down as the bonds locate farther from the QM/MM interface. The last three bonds, C11–C14, C14–C17, and C17–O20 (see Figure 2c) are all located at least three bonds away from the YinYang atom (C1), and the errors in their lengths are negligibly small. The same also applies to the bond angles, which are shown in the lower panel of Figure 4.

In another series of YinYang QM/MM calculations, the size of the QM region is systematically increased as shown in Figure 5. Model I contains only one CH_2 group, in addition to the hydroxyl group, as the QM region. The size of QM region is increased by one CH_2 group at a time from model II to model V, which contains all five CH_2 groups. The errors for some bonds or angles near the hydroxyl group are plotted in Figure 5. Clearly, these bond lengths and bond angles become more and more accurate as they are moved away from the YinYang atom. At the same time, the errors in the protonation energies are: 0.54 kcal/mol (model I), -1.69 kcal/mol (model II), -3.25 kcal/mol (model III), -1.90 kcal/mol (model IV), and -1.83 kcal/mol (model V), which shows that the protonation energy do not always become worse with larger QM regions.

3.10. $HOC_3H_6O^-$, HOC_3H_6OH , and $HOC_3H_6OH_2^+$. In this and next two subsections, we will examine the effect of having polar groups in the MM region. In this subsection, we start with propanol, and replace one of its methyl hydrogens with a hydroxyl group (see Table 6). The QM region still consists of one hydroxyl group and the neighboring CH_2 group. The MM region contains two other CH_2 groups and the MM-end hydroxyl group.

Within the CHARMM27 force field, the hydroxyl oxygen atom has a charge of -0.66 , and the hydroxyl hydrogen atom has a charge of 0.43 , so the net charge on hydroxyl group is -0.23 . If this charge is not neutralized, the protonation energy will be overestimated by more than 24 kcal/mol. To offset this charge, the charge of the neighboring carbon (C3 in Figure 2d) can be increased from -0.18 to $+0.05$. As mentioned in Appendix B, a $+0.1$ charge at the location of C3 would narrow the energy gap down by approximately 10 kcal/mol. So the proposed charge increase of 0.23 can bring the protonation energy much closer to its correct value. Indeed, as shown in Table 6, with the charge on C3 increased to 0.05 , the energy errors are -0.32 kcal/mol (0.1%) for the deprotonation energy, and 0.81 kcal/mol (0.4%) for the protonation energy. In ref 29, the error in the protonation energy was 4.4 kcal/mol.

3.11. $\text{HOC}_3\text{H}_6\text{NH}^-$, $\text{HOC}_3\text{H}_6\text{NH}_2$, and $\text{HOC}_3\text{H}_6\text{NH}_3^+$. Here we start from propylamine, and replace one of the methyl hydrogen with hydroxyl group. For the same reason mentioned in the last subsection, we need to increase the charge of the MM carbon atom next to hydroxyl group from -0.18 to $+0.05$. Then the energy errors from YinYang atom model calculations are -0.63 kcal/mol (0.1%) for the deprotonation energy and -0.23 kcal/mol (0.1%) for the protonation energy. With the double-link atom model in ref 29, the error in protonation energy was 2.5 kcal/mol.

3.12. $\text{H}_2\text{NC}_3\text{H}_6\text{NH}^-$, $\text{H}_2\text{NC}_3\text{H}_6\text{NH}_2$, and $\text{H}_2\text{NC}_3\text{H}_6\text{NH}_3^+$. Here we also start from propylamine and replace one of the methyl hydrogens with an amine group. The CHARMM27 charges are -0.62 for amine nitrogen and $+0.31$ amine hydrogen, so the amine group (NH_2) is charge-neutral. YinYang QM/MM calculations yield errors of 3.94 kcal/mol (1.0%) in the deprotonation energy, and -4.78 kcal/mol (2.0%) in the protonation energy. Both energy differences are overestimated, which can be related to the fact that the amine nitrogen tends to attract a small amount of charge (about one-tenth of an electron at B3LYP/6-31G* level) away from its neighboring carbon.

3.13. Serine. In this and next sections, we will perform YinYang QM/MM calculations on two amino acids, serine and histidine.

For serine, the MM region consists of the aldehyde group, α carbon and its hydrogen, and the amine group (see Figure 6). The charges values used here are: aldehyde carbon 0.42 , aldehyde oxygen -0.51 , aldehyde hydrogen 0.09 , α carbon 0.09 , hydrogen bound to α carbon 0.07 , amine nitrogen -0.82 , and amine hydrogen 0.33 . So overall the aldehyde group is charge-neutral, α carbon, and its hydrogen has a net charge of 0.16 , which is offset by the amine group. α carbon is the YinYang atom, so its charge is 1.09 in the QM calculation.

From YinYang QM/MM calculations, the deprotonation energy of serine (removal of γ -hydrogen, see Figure 6) comes with an error of 4.65 kcal/mol (1.2%, see Table 7). Link-atom type QM/MM calculations were reported for serine, and the error in the deprotonation energy was 11.9 kcal/mol in ref 28 and -0.29 kcal/mol in ref 32.

3.14. Histidine. For histidine, we performed YinYang calculations with either aldehyde (CHO) group or carboxyl (COOH) group.

If the structure is terminated with aldehyde group, we use the same MM region as in the serine calculation, and therefore the same set of MM charge values. The protonation energy (for adding one proton next to ϵ -nitrogen, see the middle panel in Figure 6) comes with an error of 4.85 kcal/mol (2.0%). Link-

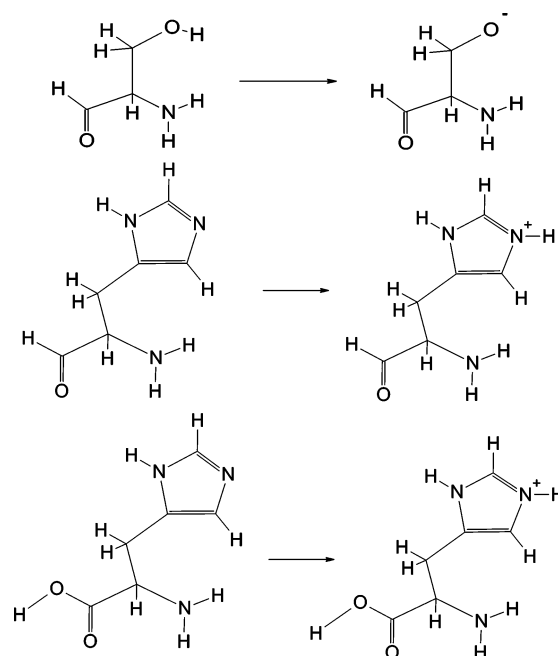


Figure 6. Amino acids

atom type calculations yielded an error of -5.03 kcal/mol in ref 28 and -4.81 kcal/mol in ref 32.

If the structure is terminated with carboxyl group (see the bottom panel in Figure 6), we can still use the same MM charge values as before for α carbon and its hydrogen, and the amine group. For the carboxyl group, the charge values are carbon 0.51 , carbonyl oxygen -0.43 , hydroxyl oxygen -0.51 , and hydrogen 0.43 . With this set of MM charges, the protonation energy comes with an error of 6.18 kcal/mol (2.5%, see Table 7). In ref 30, the error ranges from 1.2 to 2.6 kcal/mol with capped potential model. In ref 32, the error was 4.14 kcal/mol.

Before we proceed to the conclusion, we would like to point out there are limitations as to how small the QM/MM errors can be. As an example, we have performed all-electron DFT calculations on both histidine and protonated histidine (terminated with carboxyl groups) with B3LYP functional and a mixed basis set. 6-31G* basis functions were assigned to “QM” atoms and 3-21G basis functions to “MM” atoms, which were MM atoms in YinYang QM/MM calculations above and are now actually QM atoms. This mixed-basis full-QM calculation yielded geometry with quality similar to that of the YinYang QM/MM calculations, and a protonation energy with an error of -4.38 kcal/mol. As one can see, this level of error is similar to that of YinYang and other QM/MM models.

4. Summary

We have proposed and tested a very simple QM/MM model, in which an atom at the interface between QM and MM regions is made to have double characters: it is like a hydrogen atom to the QM region and a regular MM atom to the MM region. Only one new empirical parameter is introduced for an added classical Coulomb repulsion term to obtain accurate bond length for the connecting bond between QM and MM regions.

The new model is tested for a series of molecules and their charged species. Comparisons are made on the geometry parameters and protonation/deprotonation energies with respect to the full-QM results.

For linear-chain alcohols and amines, this model yields reasonable geometries for neutral species. Its becomes slightly

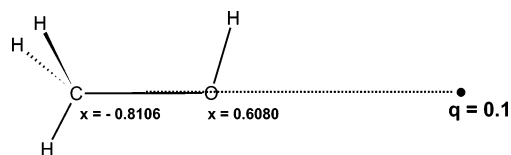


Figure 7. Model system for studying the influence of MM charge values on optimized bond lengths. The external charge is moved along x -axis.

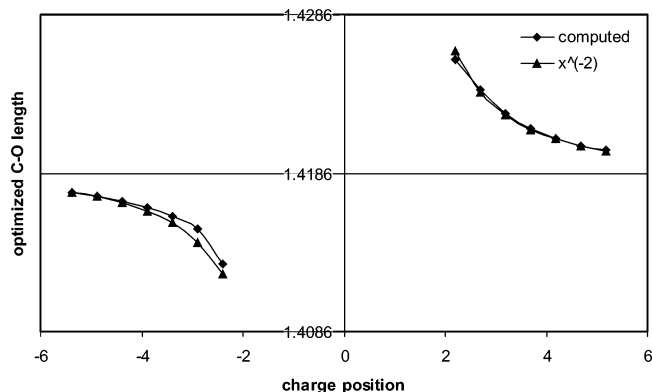


Figure 8. Optimized C–O lengths for methanol under the electrostatic potential of an external point charge. x^{-2} curves are fitted ones: $1.4186 + 0.0368 * SIGN(x)/x^2$.

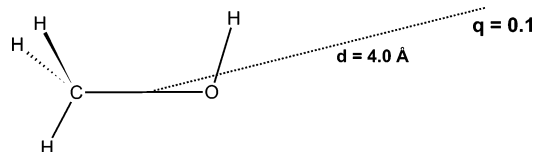


Figure 9. Model system for studying the influence of MM charge values on optimized bond lengths. The external point charge lies in the C–O–H plane and at 4 Å away from the charge center of the C–O bond.

less satisfactory for charged species (worst case for protonated butanol). For these molecules, the protonation energy usually falls within 3 kcal/mol from full-QM results (again with the exception of butanol). For linear-chain carboxyl acids, the protonation energy comes with larger errors, which is related to the assignment of MM charge values. For amino acids, the protonation energy has an error around 4–6 kcal/mol. Overall, those errors are comparable to other QM/MM models which either are based on link-atoms or contain more parameters. Furthermore, the errors decrease as the bond or the protonated functional group in the QM region is moved away from the QM/MM interface.

It is suggested that one should assign MM charge values with extra caution, because these charges perturb bond lengths through some kind of charge–dipole interaction, and even more importantly, perturb the protonation energy through charge–charge interaction. For linear-chain molecules with polar groups attached at the MM end or for amino acids it was found to be important to keep the overall MM region charge-neutral.

Acknowledgment. The authors acknowledge financial support from the National Institutes of Health through SBIR Grant (2R44GM065617). Dr. Jay Ponder's TINKER molecular simulation package served as a practical guide to force fields in the course of this work. Y.S. thanks Drs. Bernard Brooks, Marek Freindorf, Teresa Head-Gordon, Lee Woodcock, and Yingkai Zhang for helpful discussions. The authors also want to thank

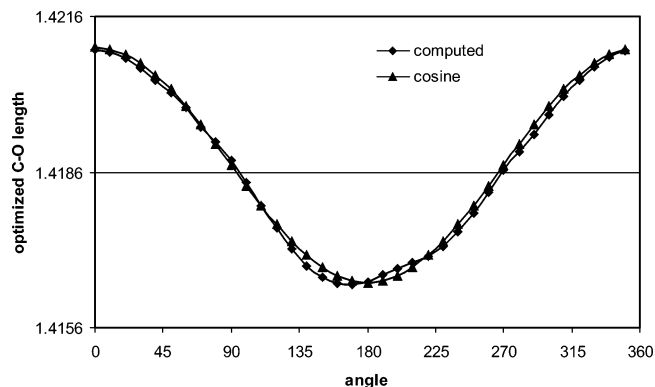


Figure 10. Optimized C–O lengths for methanol under the electrostatic potential of an external point charge. “cosine” curve is a fitted one: $1.4186 + 0.0023 \cos(\Theta)$. When the angle is zero, the charge is placed on the C–O bond axis (to the right of oxygen).

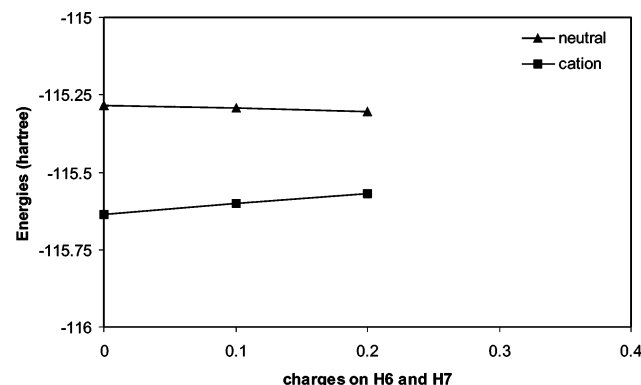
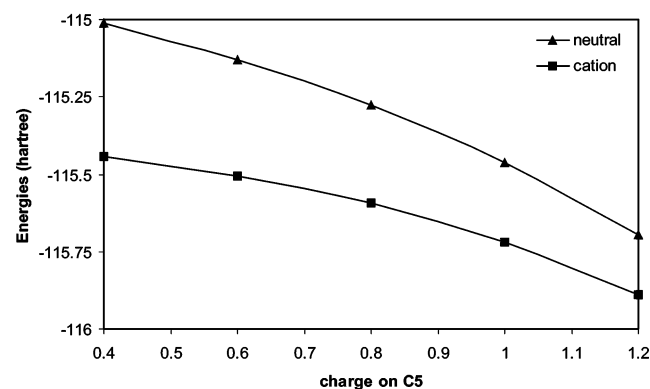
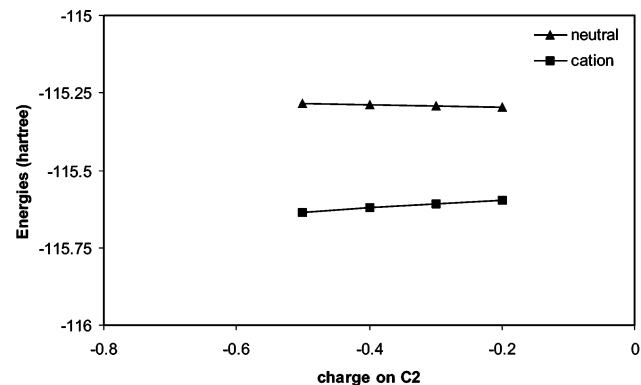


Figure 11. YinYang QM/MM energies for propanol (“neutral”) and protonated propanol (“cation”) at fixed B3LYP/6-31G* geometries. Charges on C2, or C5, or H6 and H7 are varied, so that the whole system has net charge.

Drs. Shawn Brown, Zhengting Gan, Emil Proynov and Nick Russ for technical assistance.

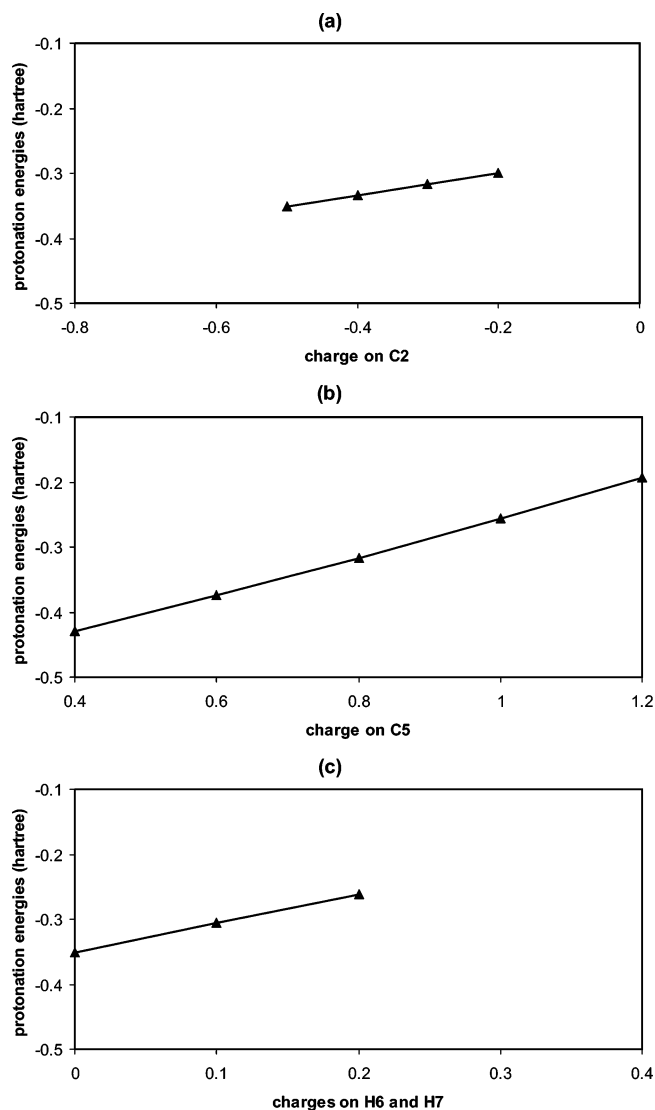


Figure 12. YinYang QM/MM protonation energies for propanol at fixed B3LYP/6-31G* geometries. Charges on C2, or C5, or H6 and H7 are varied, so that the whole system has net charge.

Appendix A. Influence of Point Charges on Optimized Bond Lengths

In QM/MM calculations, a QM model system is often subjected to the electrostatic potential from MM point charges (with the exception of the ONIOM model). Here we want to use methanol as a model system to examine the influence of point charges on optimized bond lengths.

In Figure 7, methanol is perturbed with a point charge (charge value: +0.1) along the C–O bond axis (x -axis). When $x = 0$, the point charge is located at the charge center of the C–O bond. For each charge position (x), the geometry of methanol is fully optimized at B3LYP/6-31G* level. In Figure 8, the optimized C–O bond length is plotted against the charge position, which clearly suggests a R^{-2} dependence for the change in bond length.

In a separate calculation (see Figure 9), the external charge is restricted in the C–O–H plane with a fixed distance 4 Å away from the charge center. The change in the C–O bond length (see Figure 10) clearly mimics a cosine function. This, together with the R^{-2} dependence, suggests to us that an external point charge interacts with an individual covalent bond through some kind of charge–dipole interaction. But it is not clear to

us at this time how to define a dipole moment for a specific covalent bond.

Appendix B. Influence of Point Charges on Protonation Energies

In this section, we are going to use propanol again as a model system to examine the influence of MM point charges on protonation energies. We start from YinYang QM/MM calculations on propanol mentioned in section 2 and vary the MM charge on MM atoms such as C2, C5, and H6. In Figure 11, for example, the charge on C2 is varied from -0.5 to -0.2 (standard CHARMM27 value is -0.27), while the charges on other atoms remain unchanged, so the molecule has a net charge.

The slope of these curves are the values of the electrostatic potential (from the model HCH_2OH QM region) at those atomic positions. For C5, the electrostatic potential (and therefore the slopes) are negative for both neutral propanol and protonated propanol, which is necessary to retain the positively charged nucleus. For neutral propanol, the electrostatic potential is still negative at C2, H6, and H7, while for protonated propanol, the electrostatic potential becomes positive at C2, H6, and H7, reflecting the net +1 charge of the model $HCH_2OH_2^+$ compound.

The protonation energies, which are the differences between the neutral and cationic curves in Figure 11 are shown in Figure 12. So the protonation energy is found to increase linearly with the charges on C2, C5, H6, and H7. The slope is 0.17 for C2, 0.29 for C5, and 0.44 for H6 and H7. If one defines d as the distance from an atom to the charge center of the C–O bond (which is also roughly the charge center for protonated propanol within B3LYP/6-31G*), then the values for $1/d$ are: 0.17 for C2, 0.28 for C5, and 0.22 for H6 and H7. So we reach the conclusion that external MM point charges affect the protonation energy roughly through classical charge–charge interaction.

Finally it is noted that Nicoll et al.²⁸ also studied the effect of point charge values on the energy difference in the context of local self-consistent field (LSCF).

References and Notes

- (1) Hehre, W.; Radom, L.; Pople, J. A.; Schleyer, P. v. R. *Ab initio molecular orbital theory*; Wiley: New York, 1986.
- (2) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford Univ. Press: New York, 1989.
- (3) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry: Introduction to advanced electronic structure theory*; McGraw-Hill Publishing Company: New York, 1993.
- (4) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: New York, 2000.
- (5) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (6) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (7) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (8) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (9) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27–85.
- (10) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (11) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718–730.
- (12) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (13) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269–282.
- (14) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100–106.

- (15) Monard, G.; Loos, M.; Théry, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153–159.
- (16) Ferré, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610–624.
- (17) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- (18) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.
- (19) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290–9295.
- (20) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
- (21) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632–650.
- (22) Pu, J.; Gao, J.; Truhlar, D. G. *ChemPhysChem* **2005**, *6*, 1853–1865.
- (23) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46–54.
- (24) Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.
- (25) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468–1494.
- (26) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *Chem. Phys. Lett.* **2000**, *321*, 113–120.
- (27) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656–6665.
- (28) Nicoll, R. M.; Hindle, S. A.; MacKenzie, G.; Hillier, I. H.; Burton, N. A. *Theor. Chem. Acc.* **2001**, *106*, 105–112.
- (29) Das, D.; Eurenus, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534–10547.
- (30) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578–9584.
- (31) Ferré, N.; Olivucci, M. *J. Mol. Struct. THEOCHEM* **2003**, *632*, 71–82.
- (32) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43–52.
- (33) Loferer, M. J.; Loeffler, H. H.; Liedl, K. R. *J. Comput. Chem.* **2003**, *24*, 1240–1249.
- (34) Bessac, F.; Alary, F.; Carissan, Y.; Heully, J.-L.; Daudey, J.-P.; Poteau, R. *J. Mol. Struct. THEOCHEM* **2003**, *632*, 43–59.
- (35) 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.; Schäfer, A.; Lennartz, C. *J. Mol. Struct. THEOCHEM* **2003**, *632*, 1–28.
- (36) Yasuda, K.; Yamaki, D. *J. Chem. Phys.* **2004**, *121*, 3964–3972.
- (37) Csányi, G.; Albaret, T.; Payne, M. C.; De Vita, A. *Phys. Rev. Lett.* **2004**, *93*, 175503.
- (38) Csányi, G.; Albaret, T.; Moras, G.; Payne, M. C.; De Vita, A. *J. Phys.: Condens. Matter* **2005**, *17*, R691–R703.
- (39) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113.
- (40) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991–4004.
- (41) Slaviček, P.; Martínez, T. J. *J. Chem. Phys.* **2006**, *124*, 084107.
- (42) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- (43) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, C. D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Dachsels, H.; Zhang, W. M.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C. P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532–1548.
- (44) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsels, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W.; Schaefer, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.*, **2006**, *8*, 3172–3191.
- (45) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268–278.
- (46) Shao, Y.; Head-Gordon, M. *Chem. Phys. Lett.* **2000**, *323*, 425–433.
- (47) Fusti-Molnar, L.; Kong, J. *J. Chem. Phys.* **2005**, *122*, 074108.
- (48) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. *Proc. Nat. Acad. Sci. U.S.A.* **2005**, *102*, 6692–6697.
- (49) Brown, S. T.; Kong, J. *Chem. Phys. Lett.* **2005**, *408*, 395–402.
- (50) Brown, S. T.; Fusti-Molnar, L.; Kong, J. *Chem. Phys. Lett.* **2006**, *418*, 490–495.
- (51) Kong, J.; Brown, S. T.; Fusti-Molnar, L. *J. Chem. Phys.* **2006**, *124*, 094109.
- (52) Chien, S. H.; Gill, P. M. W. *J. Comput. Chem.* **2006**, *27*, 730–739.
- (53) Furlani, T. R.; Kong, J.; Gill, P. M. W. *Comput. Phys. Commun.* **2000**, *128*, 170–177.
- (54) Korambath, P. P.; Kong, J.; Furlani, T. R.; Head-Gordon, M. *Mol. Phys.* **2002**, *100*, 1755–1761.
- (55) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- (56) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (57) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (58) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (59) Harihara, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (60) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. *Chem. Phys. Lett.* **1993**, *209*, 506–512.
- (61) Adamson, R. D.; Gill, P. M. W.; Pople, J. A. *Chem. Phys. Lett.* **1998**, *284*, 6–11.