PCCP

View Online
Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 10556–10564

www.rsc.org/pccp

PAPER

Geometry optimization using tuned and balanced redistributed charge schemes for combined quantum mechanical and molecular mechanical calculations†

Bo Wang and Donald G. Truhlar*

Received 11th December 2010, Accepted 22nd February 2011
DOI: 10.1039/c0cp02850a

We performed geometry optimizations using the tuned and balanced redistributed charge algorithms to treat the QM–MM boundary in combined quantum mechanical and molecular mechanical (QM/MM) methods. In the tuned and balanced redistributed charge (TBRC) scheme, the QM boundary atom is terminated by a tuned F link atom, and the charge of the MM boundary atom is properly adjusted to conserve the total charge of the entire QM/MM system; then the adjusted MM boundary charge is moved evenly to the midpoints of the bonds between the MM boundary atom and its neighboring MM atoms. In the tuned and balanced redistributed charge-2 (TBRC2) scheme, the adjusted MM boundary charge is moved evenly to all MM atoms that are attached to the MM boundary atom. A new option, namely charge smearing, has been added to the TBRC scheme, yielding the tuned and balanced smeared redistributed charge (TBSRC) scheme. In the new scheme, the redistributed charges near the QM–MM boundary are smeared to make the electrostatic interactions between the QM region and the redistributed charges more realistic. The TBRC2 scheme and new TBSRC scheme have been tested for various kinds of bonds at a QM–MM boundary, including C–C, C–N, C–O, O–C, N–C, C–S, S–S, S–C, C–Si, and O–N bonds. Charge smearing is necessary if the redistributed charges are close to the QM region, as in the TBSRC scheme, but not if the redistributed charge is farther from the QM region, as in the TBRC2 scheme. We found that QM/MM results using either the TBRC2 scheme or the TBSRC scheme agree well with full QM results; the mean unsigned error (MUE) of the QM/MM deprotonation energy is 1.6 kcal/mol in both cases, and the MUE of QM/MM optimized bond lengths over the three bonds closest to the QM–MM boundary, with errors averaged over the protonated forms and unprotonated forms, is 0.015 Å for TBRC2 and 0.021 Å for TBSRC. The improvements in the new scheme are essential for QM–MM boundaries that pass through a polar bond, but even for boundaries that pass through C–C bonds, the improvement can be quite significant.

1. Introduction

Multiscale modeling[1–6] is a method of choice for the study of chemical and physical processes of complex and large systems, such as practical catalysts and biomolecules. A key element is that a small-scale primary system is treated at a higher level than a large-scale secondary system, and there may even be a hierarchy of levels employed for more than two scales.

Combined quantum mechanical and molecular mechanical (QM/MM) methods are multiscale approaches that can be applied to study chemical reactions in large systems.[7–21] In QM/MM methods, a small region is treated by quantum mechanics, and the remaining part is treated by molecular mechanics. This method can be especially useful for simulations of condensed-phase systems, *e.g.*, biomolecular processes and solid-state chemistry. For example, adsorption and chemical reactions in zeolites can be studied using QM/MM methods.[22]

One important issue in QM/MM methods is how to deal with the QM–MM boundary when it passes through a covalent bond. Link atoms,[7,9,10] generalized hybrid orbitals or other localized orbitals,[23–25] and pseudobonds or effective potentials[26–31] have been used to saturate the dangling valences at the edges of the QM region. To treat various kinds of covalent bonds, especially polar covalent bonds, being cut at the QM–MM boundary, we have developed tuned and balanced redistributed charge methods in a previous study; that article[32] will be called paper I. In these methods a pseudo atom tuned to reproduce

Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455-0431
† Electronic supplementary information (ESI) available: Bond lengths and deprotonation energies for each molecule studied in this paper using the QM and QM/MM schemes. See DOI: 10.1039/c0cp02850a

the electronic structure of the QM region is used as the link atom, and the charge on the MM boundary atom is adjusted to conserve the total charge of the system[33] and is redistributed[34,35] to other MM atoms. For example, in the tuned and balanced redistributed charge (TBRC) scheme, the adjusted (balanced) charge on the MM boundary atom is distributed to the midpoints of bonds between an MM boundary atom and its neighboring MM atoms. It is found that tuning the link atoms and adjusting (balancing) the MM charges can greatly improve the results for proton affinities in single-point calculations (*i.e.*, calculations at fixed geometries), and the TBRC scheme gives the best results.[32]

In the present study, we formulate the QM/MM total energy expression as a function of geometry for systems with QM–MM boundaries that cut bonds, and we carry out QM/MM geometry optimizations. We also add a new option to the tuned and balanced redistributed charge methods, namely charge smearing. When this option is employed, the redistributed charges are not represented by point charges, but by smeared charges. Das *et al.*[36] and Amara and Field[37] have represented the charge distributions of MM atoms close to a QM region by Gaussian functions rather than point charges. For boundaries that cut a C–C bond, they showed that when the nearby MM charges are properly delocalized, the description of electrostatics nears the QM–MM boundary is improved. We have developed[38] a particularly convenient way to delocalize the outermost portion of an atom's charge in a Slater-type orbital[39] (STO). In the present work, we use a similar scheme to delocalize the redistributed charges near a cut covalent bond.

A key element to be examined is whether the tuned and balanced redistributed charge schemes improve the accuracy enough to optimize geometries realistically and calculate reaction energies accurately with optimized geometries. This is important not only for geometry optimization *per se* but for molecular dynamics; a method that yields incorrect geometries when both lengths and bond angles are unconstrained is not suitable for molecular dynamics. Here we test the tuned and balanced redistributed charge schemes for this capability using both the point charges and smeared charges for the balanced redistributed charges. In section 2, we will present all the ingredients of the QM/MM methods used in the present study. In section 3, we will present the test suite and implementation details. Section 4 gives the analysis of the calculations. Section 5 gives an overall comparison of the performance of all boundary charge schemes. Section 6 summarizes the main conclusions.

2. Methods

In this section, we will first review the tuned and balanced redistributed charge schemes proposed in paper I.[32] Then we will present the QM/MM energy expression, which is based on an earlier[35] formulation. This is followed by a description of the placement of the link atom and the smearing of the redistributed charges.

In order to describe the schemes, we label the atoms according to "tiers".[12,32,35] In particular, the MM boundary atoms (*i.e.*, MM atoms covalently bonded to QM atoms) are denoted as M1 atoms; and MM atoms directly bonded to M1 atoms are denoted as M2 atoms. M3 atoms are the third-tier MM atoms, *i.e.*, those bonded to M2 atoms. The QM boundary atoms (*i.e.*, QM atoms covalently bonded to MM atoms) are denoted as Q1 atoms; and the QM atoms directly bonded to Q1 atoms are labeled Q2 atoms.

2.1 Tuned and balanced redistributed charge schemes

In paper I,[32] we introduced tuned and balanced redistributed charge schemes to treat the QM–MM boundary. The balancing consists in adjusting the MM point charge on the M1 atom to conserve the total charge of the entire QM/MM system. In the BRC scheme, this adjusted charge is evenly redistributed to the midpoints of the M1–M2 bonds. The adjusted M1 charge can also be placed in other positions. In particular, when the adjusted M1 charge is evenly redistributed to all M2 atoms, we call the method BRC2, and when the adjusted M1 charge is evenly redistributed to all M2 and M3 atoms, we call the method BRC3. In the balanced redistributed charge and dipole (BRCD) scheme, we double the redistributed charges that are placed at the midpoints of the M1–M2 bonds, and we adjust the charges on M2 atoms to conserve the total charge of the entire QM/MM system.[32] To test the electrostatic effects of the MM charges on the QM region, we also test a $Z\infty$ scheme, in which the electrostatic interactions between the QM and MM regions are completely neglected.

In tuned methods, the link atom is a tuned F atom, which is an atom that has an adjustable pseudopotential centered at its nucleus. The pseudopotential is given by

$$U(r) = C\exp[-(r/r_T)^2] \quad (1)$$

where C is the tuning parameter, and $r_T = 1a_0$ (where a_0 is Bohr radius). The pseudopotential is tuned to make the sum of the partial charges of the uncapped portion of the QM subsystem equal a target value. The partial charges are computed by Mulliken analysis with a 6-31G* basis set when the M1 atom is from the second period (Li through F) and with an STO-3G basis set otherwise. The tuning process has been used successfully[32] to treat polar bonds between the QM and MM subsystems. A detailed description of the tuning process can be found in paper I. We combine the tuned F link scheme with the BRC, BRC2, BRC3, and BRCD schemes, yielding the TBRC, TBRC2, TBRC3, and TBRCD schemes, respectively.

In paper I we tuned the fluorine link atoms on an entire system model (ESM) that includes three tiers of MM atoms, with the third tier capped. Background charges in the ESM (*i.e.*, MM charges and redistributed charges) are present in the tuning process. In the present study, we found that tuning the F link atoms with and without background charges gives similar tuning parameters. Tuning without background charges is more straightforward, so in the present work, we tuned the F link atom without any background charges present, which, in the language of paper I, corresponds to tuning the capped primary system (CPS) without any MM charges. A consequence of this simplification that is important for the present work is that the tuning parameters are

Table 1 The contraction coefficients and exponents[a]

	C_i	α_i
1	0.021221	0.065110
2	0.131906	0.158088
3	0.238573	0.407099
4	0.241818	1.185060
5	0.184113	4.235920
6	0.121984	23.10300

[a] From ref. 40.

independent of the charge models and smearing widths that are used below to treat MM charges and boundary charges.

2.2 Total energy expression

The QM region is also called the primary subsystem (PS), the MM region is called the secondary subsystem (SS), and the whole system is called the entire system (ES). The CPS is the primary system (PS) capped by the link atom. CPS** denotes that the CPS is embedded in the adjusted electrostatic field of the secondary subsystem (or the unadjusted one in the straight electrostatic embedding (SEE) method), and SS* denotes the secondary subsystem with the adjusted M1 charge in the balanced charge methods or with the original M1 charge in the unbalanced charge methods. No charge redistribution is made for SS*. The QM/MM energy is[35]

$$E = E(\mathrm{QM,CPS^{**}}) + [E(\mathrm{val;ES}) - E(\mathrm{val;CPS})] + E(\mathrm{Coul;SS^{*}}) + [E(\mathrm{vdW;ES}) - E(\mathrm{vdW;CPS})] \quad (2)$$

where E(QM,CPS**) is the quantum mechanical energy of the QM system in the presence of (original or adjusted) electrostatic field of the secondary subsystem, the two differences, E(val;ES) − E(val;CPS) and E(vdW;ES) − E(vdW;CPS), are the MM energy differences between the entire system and the capped primary system for the valence interactions and van der Waals interactions, and E(Coul;SS*) is the Coulomb energy of the secondary subsystem with the original or adjusted M1 charge. Therefore, we extend the energy expression formulated in the RC and RCD article[35] to the new balanced charge schemes, such as BRC and BRC2. The only difference between the BRC scheme and the original RC scheme[35] is that the adjusted (balanced) M1 charge, rather than the original M1 charge, is used for the calculations of E(QM,CPS**) and E(Coul;SS*). In the special case where the total charge of the MM region is neutral, the adjusted M1 charge equals the original M1 charge, and these two formulations are same.

2.3 Placement of the link atom

The tuned F link atom is not at its equilibrium position in the QM/MM calculation unless geometry optimization is performed. In the current study, we used the same approach as that used in *Amber 10*[33] to place the link atom (which is denoted as L); in this method the Q1–L bond length is fixed during the QM/MM optimization. The link atom is placed along the bond vector joining Q1 and M1, and the position of the link atom is defined as:

$$\mathbf{r}_{\mathrm{L}} = \mathbf{r}_{\mathrm{Q1}} + d_{\mathrm{Q1-L}} \frac{\mathbf{r}_{\mathrm{M1}} - \mathbf{r}_{\mathrm{Q1}}}{|\mathbf{r}_{\mathrm{M1}} - \mathbf{r}_{\mathrm{Q1}}|} \quad (3)$$

where $\mathbf{r}_{\mathrm{L}}$, $\mathbf{r}_{\mathrm{Q1}}$, and $\mathbf{r}_{\mathrm{M1}}$ are the positions of the link atom, the Q1 atom, and the M1 atom, respectively; and $d_{\mathrm{Q1-L}}$ is the fixed bond length of Q1–L, which is assigned as the standard Q1–L bond length in whatever force field is used for the MM calculations.

2.4 Smearing the redistributed charges

For the option of charge smearing, we placed the redistributed charge q_{MM} in a normalized Slater type orbital (STO)

$$\varphi = \exp(-r/r_0)/(\pi r_0^3)^{1/2} \quad (4)$$

where r is the distance of the charge density from its center, and r_0 is the smearing width. Then the charge density of MM charge q_{MM} is expressed as

$$\rho_{\mathrm{MM}}(r) = q_{\mathrm{MM}} \exp(-2r/r_0)/(\pi r_0^3) \quad (5)$$

We calculated the electrostatic potential generated by the smeared charge and derived the effective charge as

$$q^*_{\mathrm{MM}} = q_{\mathrm{MM}} - q_{\mathrm{MM}}\left(1+\frac{r}{r_0}\right)\exp(-2r/r_0) \quad (6)$$

An explanation of the effective charge concept and a detailed derivation can be found in our previous paper about charge penetration.[38] The only difference here is that we delocalized the outermost electron density in the study of charge penetration,[38] while we delocalize the total redistributed charges in the present study. Because in most programs, the pseudopotentials are expressed as Gaussian types of functions, we used 6 Gaussian functions to fit the Slater function, as is shown in eqn (7), and the contraction coefficients C_i and exponents α_i are listed in Table 1.[40]

$$\exp(-\lambda r) = \sum_{i=1}^{6} C_i \exp(-\alpha_i \lambda^2 r^2) \quad (7)$$

The reason we use this particular way to smear the MM charges is that these effective charges can be easily implemented as pseudopotentials in standard QM programs.

In the present study, we only smear the redistributed charges, while point charges are still used for all other MM charges. Combining the smeared redistributed charge (SRC) scheme with the TBRC, TBRC2, TBRC3, and TBRCD schemes yields the TBSRC, TBSRC2, TBSRC3, and TBSRCD schemes.

3. Computational details

All computations were carried out by using the M06-2X density functional method[41,42] as the QM method. The MMFF94 force field[43] was used for the E(val) and E(vdW) terms of eqn (2). For the MM charges, M06-2X/6-31G(d)/CM4M[44] charges are used because they seem to be reasonable choices to reproduce the electrostatic potentials generated by MM atoms, and they are more accurate for the buried atoms in the systems studied here than are the CHELPG charges. The CM4M charges are derived from the protonated molecules, and–as in the usual procedure in universal MM force fields–are assumed to be the same in the unprotonated molecules as in the protonated ones. The MM parameters for

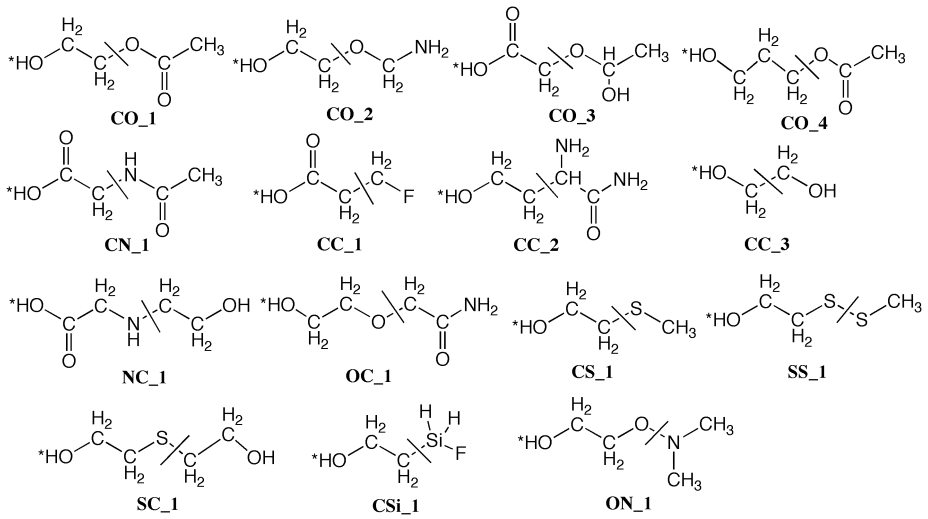

Fig. 1 Test Suite. The asterisk * denotes the deprotonation site. The QM region is on the left of the cut bond, and the MM region is on its right.

the tuned F atom are taken to be same as that for the ordinary F atom.

All QM/MM calculations are carried out using our own QMMM program,[45] which is based on a locally modified module[46] of *Gaussian03*[47] and a modified TINKER[48] program. M06-2X/6-31G(d)/CM4M[44] charges are derived from a locally modified module[49] of *Gaussian03*.[47] In the current study, three basis sets are tested, in particular 6-31G(d),[50–52] def2-TZVP,[53] and MG3S.[54] The 6-31G(d) basis set is used for all calculations in sections 4.1–4.3, and the other basis sets are considered in section 4.4.

Both the protonated and unprotonated molecules are optimized using both full QM and combined QM/MM methods. We compared the QM/MM deprotonation energies and geometries to the QM results for a test suite that contains 15 molecules with 10 kinds of single bonds being cut, in particular C–C, C–N, C–O, O–C, N–C, C–S, S–S, S–C, C–Si, and O–N bonds. The protonated forms of molecules in the test suite are shown in Fig. 1. Because MMFF94 does not include all MM parameters that are required to treat aluminosilicate clusters and we need to examine the validity of available aluminosilicate force fields in a QM/MM context before using them to test the new methods discussed here, we exclude the aluminosilicate clusters that were tested in paper I.

For the geometries, we considered the Q1–M1, M1–M2, and Q1–Q2 bond distances for the QM and QM/MM optimized structures. If there is more than one M1–M2 or Q1–Q2 bond, only the bond not involving a hydrogen atom is counted. For M1–M2 bonds in molecule **ON_1**, there are two N–C bonds and only the longer one is included in the error analysis.

4. Results and discussion

In this section, we first present the tuning parameters used for the tuned F link atom in the present study. This is followed by the QM/MM results using H link atoms and tuned F link atoms with the redistributed charges described by point charges. Then we show the QM/MM results with the redistributed charges being smeared. The smearing widths of the redistributed charges have been optimized for several tuned and balanced redistributed charge schemes. Finally we tested the transferability of the tuning parameters and smearing widths in the TBRC2 and TBSRC schemes to other basis sets.

4.1 Tuning parameters for the tuned F link atom

We performed the tuning process for each protonated molecule in the test suite. Table 2 shows the parameters C of the tuned F link atoms in eqn (1) for all molecules. Because the tuned F link atoms are tuned without the background charges (as explained in Sect. 2.1), the same parameters are used for all charge models and all smearing widths.

Note that the parameters we used here are different from the ones we used in paper I, in which the parameters are tuned in the presence of three tiers of MM charges.[32] Also, the Q1–L bond length is taken in the present study as the standard bond length of Q1–F in the MMFF94 force field, which also differs from paper I. Different placements of the link atom cause some differences in the values of the tuning parameters, but the differences are relatively small. Nevertheless, all results in the present article are recalculated in the new way explained above.

4.2 Redistributed charges as point charges

An H atom has been used as the link atom in most previous QM/MM methods, so we used both H atoms and tuned F atoms as link atoms for the QM/MM calculations. We carried out the QM/MM optimization using the BRC, BRC2, BRC3, BRCD, TBRC, TBRC2, TBRC3, and TBRCD schemes. To make a comparison, we also show the results using the Z∞ scheme. In the Z∞ scheme, all MM charges are zeroed for the calculations of E(QM,CPS**) in eqn (2), and other terms in eqn (2) are evaluated in the same way as in the BRC scheme.

Table 2 Parameters of pseudopotentials for the tuned F link atoms

Molecule	CO_1	CO_2	CO_3	CO_4	CN_1	CC_1	CC_2	CC_3
Parameter	−0.40	−0.15	−0.10	−0.40	−0.05	0.90	0.70	0.75
Molecule	NC_1	OC_1	CS_1	SS_1	SC_1	CSi_1	ON_1	
Parameter	1.40	1.35	0.45	0.40	1.00	0.70	1.05	

Table 3 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM bond lengths (Å) and deprotonation energies (kcal/mol) using H link atoms and tuned F link atoms with point charge representation of the redistributed charge

Link atom	Charge scheme	bond length[a] (Å)		deprotonation energy (kcal/mol)	
		MSE	MUE	MSE	MUE
H atom	Z∞	0.003	0.014	8.48	8.58
	BRC	−0.002	0.016	8.30	8.30
	BRC2	0.004	0.014	7.27	7.41
	BRC3	0.004	0.014	6.22	6.55
	BRCD	−0.010	0.021	9.20	9.20
Tuned F atom	Z∞	−0.002	0.016	1.24	3.59
	TBRC2	−0.002	0.015	0.48	1.65
	TBRC3	−0.002	0.015	−0.55	2.44

[a] Each mean error in bond length is an average over 90 values: 15 molecules, each protonated and unprotonated, and three bond distances for each protonated or unprotonated molecule.

Table 3 shows the results for various charge schemes using H link atoms and tuned F link atoms. When H atoms are used as link atoms, all schemes, including the BRC, BRC2, BRC3, and BRCD schemes, give quite accurate geometries. This indicates that in these redistributed charge schemes with H link atoms, the redistributed charges are already moved far enough from the QM–MM boundary to avoid overpolarization. However, the mean unsigned error (MUE) of deprotonation energies is 6.6–9.2 kcal/mol for various schemes (Z∞, BRC, BRC2, BRC3, and BRCD) employing H link atoms.

When tuned F atoms are used as link atoms, the MUE of deprotonation energies drop to 1.6–3.6 kcal/mol for the Z∞, TBRC2, TBRC3 schemes, with the best performance from the TBRC2 scheme. At the same time, the geometries are also well reproduced in QM/MM optimizations. However, for the TBRC and TBRCD schemes, we found that nearly half of the molecules are severely distorted and are not converged in QM/MM optimizations, and we do not list their errors. One possible reason for the unphysical behavior is that in these two schemes, the redistributed charges located at the midpoints of the M1–M2 bonds are close to the tuned F link atom and may overpolarize the QM region. Hence the TBRC method as originally formulated overestimates induction energies; nevertheless, the errors are largely cancelled in computing relative energies by single-point calculations, and so we obtained good relative energies in the previous study. However, the strong interactions between the redistributed charges and the link atom distort the structures in QM/MM geometry optimizations.

4.3 Redistributed charges as smeared charges

To overcome the problem due to the strong electrostatic interactions between the QM region and the redistributed charges in the TBRC and TBRCD schemes, we used smeared charges to represent the redistributed charges, leading to the TBSRC and TBSRCD schemes. It has been found that smearing the MM charges close to the QM region reduces the large electrostatic interactions near the boundary.[36] We also tested the TBSRC and TBSRC3 schemes, in which the smeared redistributed charges (SRC) are combined, to study the effects of charge smearing, though the TBRC2 and TBRC3 schemes give reasonable optimized geometries.

We first tested the TBSRC and TBSRC2 schemes plus two redistributed charge schemes in which the positions of the redistributed charges are varied. The parameter F is defined as the ratio

$$F = \frac{R(\mathrm{M1-RC})}{R(\mathrm{M1-M2})} \quad (8)$$

where R(M1−RC) is the distance from M1 atom to the redistributed charge, and R(M1−M2) is the distance from M1 atom to M2 atom. The F value is 0.5 for TBSRC, and it is 1.0 for TBSRC2. The other two schemes have F values of 0.3 and 0.7.

In Tables 4 and 5, we show the MSE and MUE of the bond lengths and deprotonation energies using tuned F link atoms with various smearing widths and with four different positions for the redistributed charges. We found that the QM/MM optimization is successful in nearly all cases except one. Table 4 shows that when the smearing width increases, the MUE of the bond lengths decreases. This confirms that smearing the redistributed charges can improve the electrostatics nears the QM–MM boundary and give better optimized geometries. The smearing width needs to be large enough to avoid the distortion of QM/MM structures. If F equals 0.3 or 0.5, the smearing width needs to be greater or equal 1.0 Å. If F equals 0.7, the smearing width needs to be greater or equal to 0.5. If F equals 1.0, there is no need to smear the charges. As the redistributed charges are moved farther from the QM/MM

Table 4 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM bond lengths (Å) using tuned F link atoms with various positions of the smeared redistributed charge and with various smearing widths r_0 (Å)[a]

F	0.3		0.5		0.7		1.0	
r_0 (Å)	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
0.5	−0.022[b]	0.070[b]	−0.025	0.048	−0.018	0.026	−0.007	0.014
1.0	−0.008	0.028	−0.010	0.021	−0.010	0.017	−0.007	0.014
2.0	−0.008	0.016	−0.005	0.014	−0.005	0.014	−0.005	0.014
3.0	−0.002	0.016	−0.005	0.014	−0.003	0.014	−0.004	0.014

[a] TBSRC for F = 0.5 and TBSRC2 for F = 1.0. The other columns do not have a name but may be considered to be nonstandard variants of TBSRC. [b] When r_0 = 0.5 Å and F = 0.3, optimization of **CO_3** (unprotonated form) is not converged, so it has been excluded from calculations in that case.

Table 5 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM deprotonation energies (kcal/mol) using tuned F link atoms with various positions of the redistributed charge and with various smearing widths r_0 (Å)[a]

F	0.3		0.5		0.7		1.0	
r_0 (Å)	MSE	MUE	MSE	MSE	MSE	MUE	MSE	MUE
0.5	1.50[b]	2.30[b]	1.22	1.93	0.87	1.63	0.35	1.80
1.0	1.16	1.73	0.82	1.62	0.52	1.66	0.09	2.10
2.0	0.10	2.18	−0.05	2.38	−0.26	2.64	−0.55	3.01
3.0	−0.71	3.42	−0.91	3.65	−1.01	3.81	−1.21	4.09

[a] TBSRC for $F = 0.5$ and TBSRC2 for $F = 1.0$. The other columns do not have a name but may be considered to be nonstandard variants of TBSRC. [b] When $r_0 = 0.5$ Å and $F = 0.3$, optimization of **CO_3** (unprotonated form) is not converged, so it has been excluded from calculations in that case.

boundary, the overpolarization problem is less severe, and we can use smaller smearing width to get correct geometries. At the same time, Table 5 shows that when the smearing width becomes larger, the MUE of the deprotonation energy increases in most cases, especially for TBSRC2. In the special case in which the smearing width goes to infinity, QM atoms will not feel the redistributed charges and the total charge of the QM/MM entire system is not conserved, which will cause large errors.[32] The TBSRC scheme (F equals 0.5) with the smearing width of 1.0 Å gives the smallest MUE, 1.62 kcal/mol, for the deprotonation energy.

The QM/MM results using the TBSRC3 and TBSRCD schemes are shown in Table 6. For the TBSRC3 scheme, the MUE of the deprotonation energies increases when we increase the smearing width of the redistributed charges. For the TBRCD scheme, the optimum value of the smearing width is 2.0 Å, with the MUE of deprotonation energies of 1.76 kcal/mol.

From the above results, we conclude that for the TBRC and TBRCD schemes, it is necessary to include the smeared redistributed charge (SRC) scheme to reproduce both the geometries and deprotonation energies. For the TBRC2 and TBRC3 schemes, in which the redistributed charges are moved farther from the QM–MM boundary, charge smearing is not necessary and the schemes without charge smearing give the best results.

We also tested whether smearing the redistributed charges with H link atoms can give good results for this test suite. Table 7 shows the results using H link atoms with various smearing widths. Varying the smearing widths from 0.0 to 5.0 Å can change the MSE of the deprotonation energies by up to 4 kcal/mol, and a smearing width of 3.0 Å gives the smallest MUE of the deprotonation energy of 6.7 kcal/mol. In previous studies,[36,37] it was shown that properly adjusting the smearing width of smeared MM charges can greatly reduce the MUE of protonation and deprotonation energies when C–C bonds being cut in the boundary. However, when polar bonds are cut in the QM–MM boundary, as in the present work and paper I, H link atoms may change the electronic structure of the boundary in the QM region, and smearing the redistributed charges does not correct the error.

4.4 Tests with other basis sets

To test the transferability of the scheme among various QM basis sets, we carried out calculations with two other basis sets, in particular def2-TZVP and MG3S. The M06-2X/6-31G(d)/CM4M charges are again used as MM charges. Two methods selected from Sect. 4.2 and 4.3 have been tested: one is the TBRC2 scheme, and the other one is the TBSRC scheme with a smearing width of 1.0 Å for redistributed charges. The results are shown in Table 8. Although for both the TBRC2 and TBSRC schemes, the MUEs of the deprotonation energies, which are around 2.3 kcal/mol, are slightly larger than were found using the 6-31G(d) basis set, they are still much smaller than their counterparts using H link atoms (H link atoms give MUEs of 6–7 kcal/mol). Therefore we conclude that our scheme is transferable, to some extent, among different basis sets. Moreover, we found that def2-TZVP and MG3S, which are more complete basis sets than 6-31G(d), have significant mean signed errors (MSE) with the deprotonation energies in both cases. The systematic character of the error indicates that further developments may be able to reduce the error for large QM basis sets.

5. Comparison of methods

Table 9 provides a consistent overall comparison of the performance of all considered boundary charge schemes for calculating proton affinities of the 15-molecule test suite considered here. The table has eight numerical columns; the first four refer to the entire-system geometries optimized in paper I.[32] The next four refer to geometries optimized by

Table 6 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM bond lengths (Å) and deprotonation energies (kcal/mol) using the TBSRC3 and TBSRCD schemes with various smearing widths r_0 (Å)

Charge scheme	r_0 (Å)	bond length (Å)		deprotonation energy (kcal/mol)	
		MSE	MUE	MSE	MUE
TBSRC3	0.5	−0.004	0.014	−0.60	2.55
	1.0	−0.004	0.014	−0.71	2.78
	2.0	−0.004	0.014	−1.10	3.50
	3.0	−0.003	0.014	−1.59	4.42
TBSRCD	0.5[a]	−0.042	0.086	1.84	2.85
	1.0	−0.012	0.030	1.58	2.14
	2.0	−0.004	0.014	0.45	1.76
	3.0	−0.005	0.015	−0.61	3.22

[a] When $r_0 = 0.5$ Å in the TBSBCD scheme, optimization of **CO_3** (unprotonated form) is not converged, so it has been excluded from calculations.

Table 7 Mean signed error (MSE) and mean unsigned error (MUE) of QM/MM bond lengths (Å) and deprotonation energies (kcal/mol) using H link atoms with the BRC scheme

r_0 (Å)	bond length (Å)		deprotonation energy (kcal/mol)	
	MSE	MUE	MSE	MUE
0.0	−0.002	0.016	8.30	8.30
1.0	0.001	0.015	7.54	7.63
2.0	0.003	0.014	6.65	6.94
3.0	0.004	0.014	5.83	6.70
5.0	0.004	0.014	4.52	7.49

Table 8 MSE and MUE of the QM/MM deprotonation energies and bond lengths with the def2-TZVP and MG3S basis sets using TBSRC with a smearing width of 1.0 Å and TBRC2 without smearing

charge scheme	basis set	bond length (Å)		deprotonation energy (kcal/mol)	
		MSE	MUE	MSE	MUE
TBSRC	def2-TZVP	−0.005	0.020	−1.16	2.18
	MG3S	−0.005	0.019	−1.73	2.43
TBRC2	def2-TZVP	0.003	0.015	−1.28	2.33
	MG3S	0.002	0.014	−1.72	2.31

Table 9 Mean unsigned errors in proton affinities (MUEs, in kcal/mol)

Method	sites	single-point energies				with optimized geometries			
Link atom		H	F[g]	F	F	H	H	F	F
tuned?[a]		no	no	ESM	CPS	no	no	CPS	CPS
smeared?		no	no	no	no	no	yes	no	yes
no MM		6	5	n.a.[b]	2.3	8	n.a	3.5	n.a
	unbalanced								
SEE		19	17						
Z2		18	17						
RC[c]	M2	16	16						
	balanced								
BSEE		9	5	2.4	3.4	unphys.[d]	8	unphys.	unphys.
shift[c]	M2,M2D	8	5	2.7	3.5				
Amber-1[c]	nearest M2	7	4.6	1.5	1.5				
Amber-2[c]	all	5	6	2.9	3.3				
BRCD, BRSCD[c,e]	M1.5,M2	9	6	3.2	4.2	9	8	unphys.	1.8
BRC3, BRSC3[c,e]	M2,M3	6	5	2.3	2.7	7	6	2.4	no im.[f]
BRC2, BRSC2[c,e]	all M2	6	4.5	1.5	1.7	7	7	1.6	no im.
BRC, BRSC[c,e]	M1.5	7	4.6	1.5	2.0	8	8	unphys.	1.6

[a] ESM denotes that F is tuned using the entire system model explained in paper I; CPS denotes that F is tuned on the capped primary system, i.e., without the MM subsystem. [b] n.a. denotes not applicable. [c] The extra column ("sites") for redistributed methods shows the sites affected by redistribution (see text). [d] unphys. denotes that some or all geometry optimizations lead to unphysical structures and do not converge. [e] The first name given applies when there is no charge smearing, and the second name applies when smearing is used. [f] no im. denotes that smearing leads to no improvement, i.e., the optimum smearing width is 0. [g] In the calculations using the untuned F link atom, we used the CRENBL ECP to substitute two electrons in the core of the F link atom. The same approach was also used in paper I in the calculations on the organic molecules (*i.e.*, the 15 molecules also tested in the present study) that used the untuned F link atom (but the other calculations in paper I for untuned F link atoms included all electrons on F).

QM/MM calculations in the present work. All QM calculations, including the QM parts of the QM/MM calculations are based on M06-2X/6-31G(d). The MM force field is MMFF94 except that we use CM4M partial atomic charges. The first three columns of the table are computed for the present 15-molecule test set from calculations originally carried out for paper I. The next five columns are based on new calculations carried out for the present study.

In order to remind the reader of the differences among the redistributed charge schemes, we added an extra column for those methods to show the sites that are affected by charge redistribution. M1, M2 and M3 sites have already been defined; M1.5 denotes a site halfway between M1 and M2; M2D denotes added dipole sites near M2. The shift method is due to Sherwood *et al.*;[34] it is similar to RCD, but we have preferred RCD because RCD is less complicated. The Amber-1 method[33] is the *adjust_q = 1* method in the *AMBER 10* program,[55] and Amber-2 is the default method in that program. Amber-1 is similar to RC2 but has the disadvantage that the redistributed charge can hop discontinuously among M2 atoms as the M1–M2 bonds vibrate. Amber-2 is similar to RC3, but we consider Amber-2 to be unphysical because it redistributes charges to arbitrarily distant locations. Nevertheless, we show these results for comparison.

The first row of Table 9 shows results with the MM subsystem neglected. This provides a baseline that methods

including the MM subsystem should surpass. The effect tuning at this level is remarkable. The MUE is only 2.3 kcal/mol for single-point calculations and 3.5 kcal/mol for optimized geometries; these results are better than any results (5–19 kcal/mol) obtainable with hydrogen link atoms or untuned fluorine link atoms. This shows that attempts to improve untuned link atom methods by improving the boundary charges are missing the point; the dominant error in untuned link atom methods is the incorrect charge distribution in the QM subsystem itself, not in its interaction with the MM subsystem.

Turning our attention to methods including the MM subsystem, we first consider the unbalanced methods. SEE denotes straight electrostatic embedding (the partial charge on M1 is not redistributed), Z2 denotes the default of the ONIOM method as implemented in the *Gaussian 03*[47] and *Gaussian 09*[56] packages in which the charges on the first two tiers of MM atoms are just set equal to zero. We see error for these methods in the range 17–19 kcal/mol. The RC scheme lowers these errors only to 16 kcal/mol. To do better we must use balanced schemes and tuning schemes.

The third and fourth numerical columns of Table 9 compare the two ways to conduct the tuning. The original method (see paper I) for tuning involved tuning in the presence of MM charges in a three-tier entire system model (ESM). The simpler method, introduced in the present paper, is to tune in the absence of MM charges. On average, tuning in the presence of MM charges in the ESM lowers the MUE by 2.9 kcal/mol, whereas tuning in the CPS lowers the error by 2.5 kcal/mol. These lowerings are close enough to one another that we selected to use the much more straightforward CPS tuning in the rest of the work, which is shown in the last two columns of Table 9.

Finally, we consider methods that employ both balanced charges and tuning. First consider balanced straight electrostatic embedding (BSEE). Here the balanced charges on M1 stay on M1. Table 9 shows that geometry optimization with this method is only successful if we employ smearing with H link atoms (the result shown is for a smearing width of 1.0 Å), and even when the geometry optimization is converged, the error is large. We conclude that some redistribution is required.

Second, the results show that charge smearing is necessary if the redistributed charge is close to the QM region (as in the TBSRC and TBSRCD), but not if it is farther from the QM region (as in the TBRC2 and TBRC3 methods).

Considering the last two columns of the last four rows of Table 9, we see that methods employing the M1.5 site are suitable for geometry optimizations only if charge smearing is employed, but when this is done, the MUE drops to 1.6 kcal/mol. The overall best method though seems to be TBRC2. It is simpler than TBRC or TBRCD in that it does not involve the M1.5 site, and it does not require charge smearing; and it is one of the most accurate methods in Table 9 when tuning is based on the CPS, except perhaps for Amber-1, which we did not pursue because of the discontinuous charge redistribution problem.

6. Conclusions

QM/MM optimizations have been performed using the tuned and balanced redistributed charge schemes. A charge-smearing scheme for the redistributed charges has been introduced in order to make the electrostatic interactions near the QM–MM boundary more realistic. It is found that both QM/MM optimized geometries and QM/MM deprotonation energies calculated with optimized geometries can accurately reproduce full QM results even for boundaries through polar bonds, and there are also significant improvements for boundaries through C–C bonds. Both the TBRC2 scheme and the TBSRC scheme with a smearing width of 1.0 Å give a mean unsigned error of 1.6 kcal/mol for the deprotonation energies, and the QM/MM optimized geometries also agree well with the QM geometries for these two choices. Moreover, comparing the results using H link atoms to those tuned F link atoms, we conclude that it is necessary to tune the link atoms when treating diverse kinds of bonds at the QM–MM boundary. In fact tuning and balancing are found to be more important than the choice of charge redistribution scheme, although the literature devoted to tuning and balancing is small, and that devoted to charge redistribution is large.

Acknowledgements

This work was supported in part by the National Science Foundation under grant no. CHE09-56776.

References

1 G. S. Ayton, W. G. Noid and G. A. Voth, *Curr. Opin. Struct. Biol.*, 2007, **17**, 192.
2 A. Heyden, H. Lin and D. G. Truhlar, *J. Phys. Chem. B*, 2007, **111**, 2231.
3 M. Praprotnik, L. D. Site and K. Kremer, *Annu. Rev. Phys. Chem.*, 2008, **59**, 545.
4 P. Sherwood, B. R. Brooks and M. S. P. Sansom, *Curr. Opin. Struct. Biol.*, 2008, **18**, 630.
5 T. Murtola, A. Bunker, I. Vattulainen, M. Deserno and M. Karttunen, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1869.
6 M. Tafipolsky, S. Amirjalayer and R. Schmid, *Microporous Mesoporous Mater.*, 2010, **129**, 304.
7 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718.
8 J. Gao, *Rev. Comput. Chem.*, 1996, **7**, 119.
9 D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580.
10 I. Antes and W. Thiel, *J. Phys. Chem. A*, 1999, **103**, 9290.
11 N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, *J. Phys. Chem. A*, 2000, **104**, 1720.
12 P. Sherwood, *Modern Methods and Algorithms of Quantum Chemistry*, John von Neumann Institute for Computing, Jülich, 2000, pp. 285.
13 J. Gao and D. G. Truhlar, *Annu. Rev. Phys. Chem.*, 2002, **53**, 467.
14 A. Shurki and A. Warshel, in *Protein Simulations*, Academic Press Inc, San Diego, USA, 2003, vol. 66, pp. 249.
15 R. A. Friesner and V. Guallar, *Annu. Rev. Phys. Chem.*, 2005, **56**, 389.
16 M.-E. Moret, E. Tapavicza, L. Guidoni, U. Röhrig, M. Sulpizi, I. Tavernelli and U. Rothlisberger, *Chimia*, 2005, **59**, 493.
17 T. Vreven and K. Morokuma, *Annu. Rep. Comput. Chem.*, 2006, **2**, 35.
18 H. M. Senn and W. Thiel, *Curr. Opin. Chem. Biol.*, 2007, **11**, 182.
19 H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185.
20 H. Hu and W. Yang, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573.
21 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198.
22 B. Boekfa, S. Choomwattana, P. Khongpracha and J. Limtrakul, *Langmuir*, 2009, **25**, 12990.
23 V. Théry, D. Rinaldi, J. L. Rivail, B. Maigret and G. G. Ferenczy, *J. Comput. Chem.*, 1994, **15**, 269.
24 J. Gao, P. Amara, C. Alhambra and M. J. Field, *J. Phys. Chem. A*, 1998, **102**, 4714.
25 J. Pu, J. Gao and D. G. Truhlar, *J. Phys. Chem. A*, 2004, **108**, 632.
26 N. Koga and K. Morokuma, *Chem. Phys. Lett.*, 1990, **172**, 243.

27 Y. Zhang, T.-S. Lee and W. Yang, *J. Chem. Phys.*, 1999, **110**, 46.
28 F. Alary, R. Poteau, J. L. Heully, J. C. Barthelat and J. P. Daudey, *Theor. Chem. Acc.*, 2000, **104**, 174.
29 G. A. DiLabio, M. M. Hurley and P. A. Christiansen, *J. Chem. Phys.*, 2002, **116**, 9578.
30 O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *J. Chem. Phys.*, 2005, **122**, 014113.
31 Y. Y. Ohnishi, Y. Nakao, H. Sato and S. Sakaki, *J. Phys. Chem. A*, 2008, **112**, 1946.
32 B. Wang and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 359.
33 R. C. Walker, M. F. Crowley and D. A. Case, *J. Comput. Chem.*, 2008, **29**, 1019.
34 P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent and I. H. Hillier, *Faraday Discuss.*, 1997, **106**, 79.
35 H. Lin and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 3991.
36 D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodošček and B. R. Brooks, *J. Chem. Phys.*, 2002, **117**, 10534.
37 P. Amara and M. J. Field, *Theor. Chem. Acc.*, 2003, **109**, 43.
38 B. Wang and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 3330.
39 J. C. Slater, *Phys. Rev.*, 1930, **36**, 57.
40 W. J. Hehre, R. F. Stewart and J. A. Pople, *J. Chem. Phys.*, 1969, **51**, 2657.
41 Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157.
42 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.
43 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490.
44 R. M. Olson, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 2046.
45 H. Lin, Y. Zhang and D. G. Truhlar, *QMMM, version 1.3.6*, University of Minnesota, Minneapolis, 2009.
46 Y. Zhao and D. G. Truhlar, *MN-GFM: Minnesota Gaussian Functional Module, version 3.0*, University of Minnesota, Minneapolis, 2006.
47 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, version D. 01*, Gaussian Inc, Wallingford, CT, 2004.
48 J. W. Ponder, *TINKER, version 5.0*, Washington University, St. Louis, 2009.
49 A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer and D. G. Truhlar, *Minnesota Solvation Database, version 2008*, University of Minnesota, Minneapolis, 2008.
50 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257.
51 J. D. Dill and J. A. Pople, *J. Chem. Phys.*, 1975, **62**, 2921.
52 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654.
53 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys*, 2005, **7**, 3297.
54 B. J. Lynch, Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 1384.
55 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, *AMBER 10*, University of California, San Francisco, 2008.
56 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 09, Version A 02*, Gaussian Inc, Wallingford CT, 2009.

