Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods

Yingkai Zhang

Citation: J. Chem. Phys. **122**, 024114 (2005); doi: 10.1063/1.1834899
View online: http://dx.doi.org/10.1063/1.1834899
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v122/i2
Published by the AIP Publishing LLC.

Improved pseudobonds for combined *ab initio* quantum mechanical/molecular mechanical methods

Yingkai Zhang[a)]
Department of Chemistry, New York University, New York, New York 10003

(Received 14 September 2004; accepted 27 October 2004; published online 23 December 2004)

The pseudobond approach offers a smooth connection at the quantum mechanical/molecular mechanical interface which passes through covalent bonds. It replaces the boundary atom of the environment part with a seven-valence-electron atom to form a pseudobond with the boundary atom of the active part [Y. Zhang, T. S. Lee, and W. Yang, J. Chem. Phys. **110**, 46 (1999)]. In its original formulation, the seven-valence-electron boundary atom has the basis set of fluorine and a parametrized effective core potential. Up to now, only the $C_{ps}(sp^3)-C(sp^3)$ pseudobond has been successfully developed; thus in the case of proteins, it can only be used to cut the protein side chains. Here we employ a different formulation to construct this seven-valence-electron boundary atom, which has its own basis set as well as the effective core potential. We have not only further improved $C_{ps}(sp^3)-C(sp^3)$ pseudobond, but also developed $C_{ps}(sp^3)-C(sp^2$,carbonyl) and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds for the cutting of protein backbones and nucleic acid bases. The basis set and effective core potential for the seven-valence-electron boundary atom are independent of the molecular mechanical force field. Although the parametrization is performed with density functional calculations using hybrid B3LYP exchange-correlation functional, it is found that the same set of parameters is also applicable to Hartree-Fock and MP2 methods, as well as DFT calculations with other exchange-correlation functionals. Tests on a series of molecules yield very good structural, electronic, and energetic results in comparison with the corresponding full *ab initio* quantum mechanical calculations. © *2005 American Institute of Physics.*
[DOI: 10.1063/1.1834899]

I. INTRODUCTION

The primary difficulties encountered in computational studies of chemical reactions in macromolecules or solution stem from the need to describe chemical bond breaking/forming and the large size of the system. High level quantum mechanical methods can provide the electronic detail of chemical reactions, but are limited in application to systems of small size. The combined quantum mechanical and molecular mechanical (QM/MM) method[1–3] extends the realm of quantum mechanical calculations to large systems. In a QM/MM calculation, a small chemically active region is treated by a quantum mechanical method, while the remainder of the system containing a large number of atoms is described by a molecular mechanical force field.

A critical issue underlying the accuracy and applicability of the combined QM/MM methods for studying enzyme reactions is how to describe the QM/MM boundary across covalent bonds.[3–8] For example, the side chain of a glutamine residue participates in the chemical reaction as illustrated in Fig. 1, which needs to be treated by quantum mechanical methods. However, the active part resulted from the cutting of the $C_\alpha-C_\beta$ bond is a radical and has a free valence. The behavior of the radical is clearly much different from the original closed-shell system. Thus, it is not acceptable to simply treat the active part quantum mechanically, while the rest molecular mechanically. Over the years, a number of groups have made efforts to develop solutions for this covalent-bond-cutting boundary problem.

Link atom approach is the most straightforward prescription to this boundary problem.[2,3,5,9–13] In the link atom approach, link atoms, which are generally hydrogen atoms, are inserted to cap the free valence of the active part, except in the HYPERCHEM software where pseudohalogen atoms are used in their semiempirical QM/MM program in order to mimic the effect of the fragments which are removed from the quantum mechanical treatment.[14] The link atoms and the atoms in the active part form the closed-shell QM region, which can be described quantum mechanically, while the rest will be treated molecular mechanically. One main drawback of the link atom approach is the introduction of additional degrees of freedom into the system, which complicates the expression of the energy and force, the geometry optimization, and molecular dynamics simulation. Although a variety of approaches have been made to alleviate these complications within the link-atom framework,[11–13] there is a great deal of interest in the search for approaches without introducing additional atoms into the system.

An alternative approach to describe QM/MM interface across covalent bonds is the use of bonding hybrid orbitals, including the Warshel and Levitt's hybrid orbital method,[1] Rivail's local self-consistent field (LSCF) method,[15–18] Gao's generalized hybrid orbital method,[7,19] Friesner's frozen orbital method,[20,21] and Jensen's hybrid orbital method.[22] In the LSCF method, the localized bonding orbitals, which are

[a)]Electronic mail: yingkai.zhang@nyu.edu

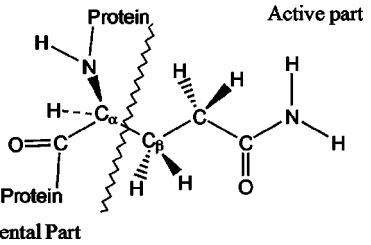

FIG. 1. Illustration of the QM/MM boundary across a covalent bond.

obtained from separate quantum mechanical calculations on small model compounds, are used to cap the free valences of the active part. Due to the use of hybrid orbitals, extensive theoretical formulation and substantial code development are required for the implementation of hybrid orbital methods. Meanwhile, it has been realized that the use of hybrid orbitals alone cannot lead to a satisfactory description of the QM/MM interface and some specific parametrizations are needed.[7,18–21]

A third category of methods for handling the QM/MM boundary problem neither introduces additional atoms into the system nor employs bonding hybrid orbitals, which includes pseudobond method,[23] connection-atom method,[24] quantum capping potential method,[25] effective group potential method,[26] and minimum principle approach.[27] In this category, the pseudobond method developed by Zhang, Lee, and Yang[23] was the first approach developed for *ab initio* QM/MM methods. It uses a seven-valence-electron atom with an effective core potential constructed to replace the boundary atom of the environment part and to form a pseudobond with the boundary atom of the active part. The pseudobond approach offers a smooth connection at the QM/MM interface and does not introduce additional atoms into the system as the link-atom approach. In comparison with hybrid orbital methods, the formalism of the pseudobond approach is simpler and it does not necessitate extensive changes to an existing QM source code. The pseudobond *ab initio* QM/MM approach[23,28] has been demonstrated to be powerful in the study of enzyme reactions.[29–33]

Despite its successes, the pseudobond approach is still much in need of development. In its original formulation,[23] the seven-valence-electron atom has the basis set of fluorine and a parametrized effective core potential. Up to now, only the $C_{ps}(sp^3)-C(sp^3)$ pseudobond has been successfully developed, which limits the applicability of the pseudobond approach. It can be used to cut protein side chains, but not for the cutting of protein backbones and nucleic acid bases. In order to improve the accuracy and applicability of the pseudobond approach, here we have developed a different formulation to construct this seven-valence-electron boundary atom, which has its own basis set as well as the effective core potential. We have developed not only the more accurate $C_{ps}(sp^3)-C(sp^3)$ pseudobond, but also the accurate $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds for the cutting of protein backbones and nucleic acid bases.

II. METHOD

A. Review of the pseudobond idea and the original formulation

The main idea of the pseudobond approach[23] is as follows: we consider that a large molecule is partitioned into two parts, an environment part and an active part, by cutting a covalent σ bond $Y-X$. Y and X refer to boundary atoms of the environment part and the active part, respectively. Instead of using a hydrogen atom to cap the free valence of X atom as in the conventional link atom approach, here a pseudobond $Y_{ps}-X$ is formed by replacing the Y atom with a one-free-valence boundary Y atom (Y_{ps}). The Y_{ps} atom is parametrized to make the $Y_{ps}-X$ pseudobond mimic the original $Y-X$ bond with similar bond length and strength, and also similar effects on the rest of the active part. In the pseudobond approach, the Y_{ps} atom and all atoms in the active part form a well-defined (often closed-shell) QM subsystem which can be treated by quantum mechanical methods. Excluding Y atom, the rest atoms in the environment part form the MM subsystem which will be represented by a molecular mechanical force field.

The $C_{ps}(sp^3)-C(sp^3)$ pseudobond has been successfully developed.[23] In its original formulation, the boundary carbon (C_{ps}) atom has (1) seven valence electrons, (2) nuclear charge seven, (3) an effective core potential, and (4) a basis set of fluorine. Seven valence electrons are just enough to doubly fill three out of the total four valence orbitals and leave the remaining one singly occupied; the C_{ps} atom thus has a free valence to make the pseudobond. Since the effect of core electrons has been included in the effective core potential, there is no core electron needed. Thus the total number of the electrons as well as the nuclear charge for this atom C_{ps} are seven. The following angular momentum dependent formula of the effective core potential[34] has been employed:

$$
\begin{aligned}
&V^{\text{eff}}(r)=V_L^{\text{eff}}(r)+\sum_{l=0}^{L-1}\,[V_l^{\text{eff}}(r)-V_L^{\text{eff}}(r)]\sum_m\,|lm\rangle\langle lm|,\\
&V_L^{\text{eff}}(r)=\frac{a_L e^{-b_L r^2}}{r},\\
&V_l^{\text{eff}}(r)-V_L^{\text{eff}}(r)=a_l e^{-b_l r^2},\quad l=0,1,...,L-1,
\end{aligned}
\tag{1}
$$

where L is the maximum l of the basis set, and a_i and b_i are the fitted parameters. Since the L is 1 for fluorine's 3-21G basis set and 2 for fluorine's 6-31G* basis set, the number of parameters are 4 and 6, respectively, for 3-21G and 6-31G* basis sets.

B. Different formulation to construct pseudobonds

In order to further improve the accuracy and applicability of the pseudobond approach, we have developed a different formulation to construct pseudobonds. *The key difference is that the seven-valence-electron boundary atom has its own basis set instead of the fluorine's.* Here a STO-2G (STO—Slater-type orbital) basis set has been employed for the seven-valence-electron boundary atom, which has four parameters and can be casted into the following form:

TABLE I. Fitted parameters for $C_{ps}(sp^3)-C(sp^3)$, $C_{ps}(sp^3)-C(sp^2$, carbonyl), and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds with 6-31G* basis set. a, and b are the parameters for the effective core potential with the function form as in Eq. (3). α_1, α_2, d_1, and d_2 are the four parameters for the STO-2G basis set as shown in Eq. (2). Atomic units are employed.

Pseudobond	$Cps(sp3)-C(sp3)$	$Cps(sp3)-C(sp2$, carbonyl)	$Cps(sp3)-N(sp3)$
a	5.0	5.0	5.0
b	5.8	6.0	10.0
α_1	1.15	1.00	1.30
α_2	0.20	0.18	0.17
d_1	0.22475	0.11483	0.11436
d_2	0.82747	0.77554	0.82071

$$\phi_s = g_s(\alpha_1, R) + d_1 g_s(\alpha_2, R),$$
$$\phi_p = g_p(\alpha_1, R) + d_2 g_p(\alpha_2, R), \tag{2}$$

where g_s and g_p are normalized s-and p-type Gaussian functions, respectively. α_1, α_2, d_1, and d_2 are the four parameters. For the effective core potential of the seven-valence-electron boundary atom, we use an angular momentum independent formula which has only two parameters a and b,

$$V^{\text{eff}}(r) = a \exp(-br^2)/r. \tag{3}$$

By parametrizing the basis set and the effective core potential of the seven-valence-electron boundary atom, which has six parameters, we have been able to not only significantly improve the $C_{ps}(sp^3)-C(sp^3)$ pseudobond, but also develop accurate $C_{ps}(sp^3)-C(sp^2$,carbonyl) and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds for the cutting of protein backbones and nucleic acid bases for 6-31G* basis set, which means that 6-31G* basis set is used for the rest of QM atoms except the seven-valence-electron boundary atom using its own STO-2G basis set. The developed parameters for the three pseudobonds are listed in Table I.

The overall parametrization procedure is similar to the previous one,[23] which involves two stages: First the parameter space is explored using a training set, then the obtained parameters are tested on a series of molecules which include various changes in the active part with substitutions and net charges. This testing procedure is to increase the transferability of the resulting pseudobond in different chemical environments. For the $C_{ps}(sp^3)-C(sp^3)$ pseudobond, the training set consists of six properties of the ethane:[23] C–C bond length, C–H bond length, C–C–H angle, Mulliken charge on carbon, Mulliken charge on hydrogen, and the bond dissociation energy of C–C bond. Due to the existence of many local minima in the optimization procedure, we employ a combined parameter scan with local minimization procedure to effectively explore the parameter space. Here the parameters a, b, α_1, and α_2 are scanned, while the parameters d_1 and d_2 are optimized such that six properties of pseudobond QM calculations on $C_{ps}-CH_3$ are in accord with the corresponding standard QM calculations for ethane. For the $C_{ps}(sp^3)-C(sp^2$,carbonyl) and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds, the same parametrization procedure is adopted, and the training set comprises the properties of $CH_3CONHCH_3$:

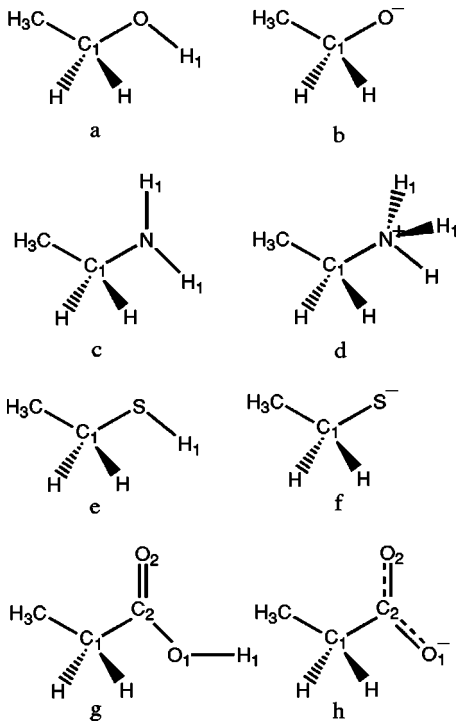

FIG. 2. Illustration of the eight molecules used for testing $C_{ps}(sp^3)-C(sp^3)$ pseudobond parameters.

TABLE II. Test results of the bond lengths (in angstrom) for eight molecules in Fig. 2. "Standard" refers to the standard B3LYP(6-31G*) calculations of the full system, "previous" refers to the pseudobond B3LYP(6-31G*) calculations with the previously developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond, and "current" refers to the pseudobond B3LYP(6-31G*) calculations with the recently developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond. SD is the standard deviation between pseudobond B3LYP(6-31G*) results and standard B3LYP(6-31G*) calculation results.

Molecule	Bond	Standard	Previous[a]	Current
a	C–C1	1.520	1.511	1.513
a	C1–O	1.426	1.412	1.419
a	O–H1	0.969	0.971	0.972
b	C–C1	1.572	1.559	1.561
b	C–O	1.312	1.289	1.289
c	C–C1	1.534	1.529	1.533
c	C1–N	1.467	1.437	1.441
c	N–H1	1.020	1.019	1.020
d	C–C1	1.519	1.497	1.501
d	C1–N	1.533	1.547	1.557
d	N–H1	1.028	1.028	1.028
e	C–C1	1.527	1.523	1.527
e	C1–S	1.849	1.847	1.865
e	S–H1	1.351	1.355	1.354
f	C–C1	1.535	1.536	1.543
f	C1–S	1.841	1.795	1.816
g	C–C1	1.527	1.522	1.527
g	C1–C2	1.514	1.515	1.509
g	C2–O2	1.211	1.210	1.211
g	C2–O1	1.358	1.358	1.357
g	O1–H1	0.976	0.976	0.976
h	C–C1	1.529	1.534	1.538
h	C1–C2	1.579	1.551	1.551
h	C2–O2	1.259	1.257	1.260
h	C1–O1	1.256	1.258	1.258
SD			0.015	0.013

[a]Reference 23.

TABLE III. Test results of the bond angles (degree) for eight molecules in Fig. 2. ''Standard'' refers to the standard B3LYP(6-31G*) calculations of the full system, ''previous'' refers to the pseudobond B3LYP(6-31G*) calculations with the previously developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond, and ''current'' refers to the pseudobond B3LYP(6-31G*) calculations with the recently developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond. SD is the standard deviation between pseudobond B3LYP(6-31G*) results and standard B3LYP(6-31G*) calculation results.

Molecule	Angle	Standard	Previous[a]	Current
a	C–C1–O	107.7	108.4	108.2
a	C1–O–H1	107.8	108.8	108.4
b	C–C1–O	114.3	115.1	114.7
c	C–C1–N	116.0	114.3	116.3
c	C1–N–H1	109.4	110.4	110.9
d	C–C1–N	110.5	109.6	112.4
d	C1–N–H1	111.1	110.3	111.1
e	C–C1–S	109.4	105.8	109.2
e	C1–S–H1	97.0	98.9	98.0
f	C–C1–S	113.7	108.1	112.6
g	C–C1–C2	112.8	111.0	114.0
g	C1–C2–O2	126.1	125.5	125.4
g	C1–C2–O1	111.4	111.6	111.9
g	C2–O1–H1	105.9	106.1	105.8
g	O2–C2–O1	122.4	122.9	122.7
h	C–C1–C2	114.0	111.7	114.5
h	C1–C2–O2	115.6	115.9	115.7
h	C1–C2–O1	114.8	114.3	114.7
h	O1–C2–O2	129.6	129.8	129.7
SD			1.9	0.8

[a]Reference 23.

TABLE IV. Test results of the Mulliken atomic charges for eight molecules in Fig. 2. ''Standard'' refers to the standard B3LYP(6-31G*) calculations of the full system, ''previous'' refers to the pseudobond B3LYP(6-31G*) calculations with the previously developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond, and ''current'' refers to the pseudobond B3LYP(6-31G*) calculations with the recently developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond. SD is the standard deviation between pseudobond B3LYP(6-31G*) results and standard B3LYP(6-31G*) calculation results.

Molecule	Atom	Standard	Previous[a]	Current
a	H1	0.39	0.39	0.39
a	O	−0.61	−0.62	−0.61
b	O	−0.74	−0.63	−0.66
c	N	−0.70	−0.67	−0.67
c	H1	0.29	0.30	0.29
d	N	−0.73	−0.75	−0.73
d	H1	0.43	0.43	0.43
e	S	−0.09	−0.12	−0.12
e	H1	0.09	0.08	0.08
f	S	−0.75	−0.69	−0.76
g	C2	0.58	0.50	0.53
g	O2	−0.46	−0.43	−0.43
g	O1	−0.57	−0.57	−0.57
g	H1	0.41	0.41	0.41
h	C2	0.53	0.45	0.45
h	O2	−0.64	−0.61	−0.62
h	O1	−0.65	−0.60	−0.62
SD			0.05	0.03

[a]Reference 23.

bond lengths, angles, Mulliken charges, $C(sp^3)-C(sp^2$,carbonyl) bond strength, or $C(sp^3)-N(sp^3)$ bond strength. The QM method used in the parametrization is the hybrid density functional method B3LYP.[35–37] All calculations are performed with GAUSSIAN98 program.[38] It should be noted that no programming is needed to perform pseudobond QM calculations since the formulas for the effective core potential and the basis set used here have already been programmed in GAUSSIAN programs. Since only QM calculations are employed in the parametrization and tests, the developed pseudobonds are independent of molecular mechanical force fields.

III. TESTS

We have tested the $C_{ps}(sp^3)-C(sp^3)$ pseudobond parameters in Table I on a series of molecules as shown in Fig. 2. For each molecule in pseudobond QM calculations, the methyl group is replaced by the boundary carbon to form the pseudobond with C_1. The test results of bond lengths, bond angles, atomic Mulliken charges, and the energy differences between pairs of molecules with B3LYP(6-31G*) calculations are listed in Tables II, III, IV, and V, respectively. Since this is the same test set as the previously developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond,[23] the results with the previously developed pseudobond have also been presented for comparison. We would like to point out that for 6-31G* basis set, the previously developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond[23] also has six parameters. The test results in Tables II, III, and IV clearly demonstrate that the $C_{ps}(sp^3)-C(sp^3)$ pseudobond has been further improved compared to the previous one.[23] The standard deviations (SD) between the pseudobond QM calculations and the corresponding standard B3LYP(6-31G*) calculations for bond lengths, angles, and atomic Mulliken charge are 0.013 Å, 0.8°, and 0.03 for the recently developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond, compared to 0.015 Å, 1.9°, and 0.05 for the previously developed pseudobond. Here the SD in angles has been reduced by more than a half. Table V presents the energy differences for these four pairs of molecules, which are the deprotonation energies for molecules CH_3CH_2OH, $CH_3CH_2NH_3^+$, CH_3CH_2SH, and CH_3CH_2COOH, respectively. We should note that this is a quite stringent test since the pseudobond is only one or two bonds away from the reaction bond X–H, X=O, N, S, O. For this test, we can see that the performance of the recently

TABLE V. Energy difference (kcal/mol) for four pairs of molecules in Fig. 2. ''Standard'' refers to the standard B3LYP(6-31G*) calculations of the full system, ''previous'' refers to the pseudobond B3LYP(6-31G*) calculations with the previously developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond, and ''current'' refers to the pseudobond B3LYP(6-31G*) calculations with the recently developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond. MAD refers the mean absolute deviation between pseudobond B3LYP(6-31G*) results and standard B3LYP(6-31G*) calculation results.

	$a\rightarrow b$	$d\rightarrow c$	$e\rightarrow f$	$g\rightarrow h$	MAD
Standard	−399.3	−230.1	−367.3	−364.9	
Previous	−385.3	−225.3	−359.8	−364.4	6.7
Current	−397.4	−229.0	−368.0	−368.7	1.9

TABLE VI. Test results of $C_{ps}(sp^3)-C(sp^3)$ pseudobond parameters in Table I for eight molecules in Fig. 2 with different QM methods (B3LYP, BLYP, PW91, HF, MP2) and 6-31G* basis set. Bond (25), angle (19), charge (17), and ΔE (4) refer to the standard deviation (SD) for 25 bond lengths (Å), SD for 19 angles (degree), SD for 17 atomic Mulliken charges, and the mean absolute deviation for 4 deprotonation energies (kcal/mol) between pseudobond QM results and the corresponding standard QM calculation results with the same QM method, respectively.

Method	Bond (25) (Å)	Angle (19) (deg)	Charge (17)	ΔE (4) (kcal/mol)
B3LYP	0.013	0.8	0.03	1.9
BLYP	0.017	0.9	0.04	1.8
PW91	0.015	0.9	0.04	2.2
HF	0.015	0.7	0.02	3.0
MP2	0.011	1.1	0.02	2.9

developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond is significantly better than the previous one, which reduces the mean absolute error from 6.7 kcal/mol to 1.9 kcal/mol.

In order to test whether the developed $C_{ps}(sp^3)-C(sp^3)$ pseudobond is applicable to other QM methods besides the B3LYP method used for the parametrization, we have performed the tests with different QM methods, including Hartree-Fock (HF), MP2, BLYP,[36,37] and PW91[39] methods, and compared to the corresponding full QM calculation results. The results in Table VI indicate that although the pseudobond is parametrized with B3LYP method, it also performs very well with other QM methods, including HF, MP2, BLYP, and PW91. For example, the standard deviation with MP2 in bond lengths, angles, atomic Mulliken charge are 0.012 Å, 1.3°, and 0.02, respectively, and the mean absolute error of the deprotonation energies is 2.9 kcal/mol.

Besides the further improvement of the $C_{ps}(sp^3)-C(sp^3)$ pseudobond, we have developed accurate $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds which can be used to cut the protein backbones and

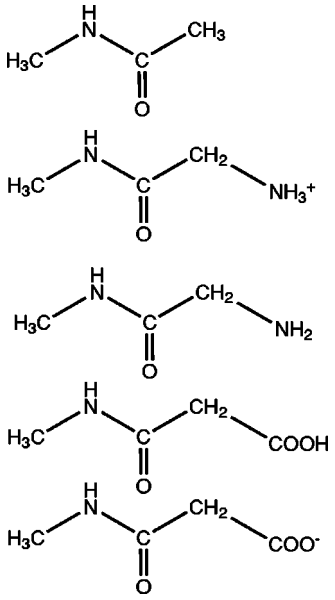

FIG. 4. Illustration of the five molecules used for testing $C_{ps}(sp^3)-N(sp^3)$ pseudobond parameters. The molecules mimic the protein backbones with different substitutions and net charges. In the pseudobond calculations, the methyl group bonded to the nitrogen atom is replaced with C_{ps} boundary atom with $C_{ps}(sp^3)-N(sp^3)$ pseudobond parameters.

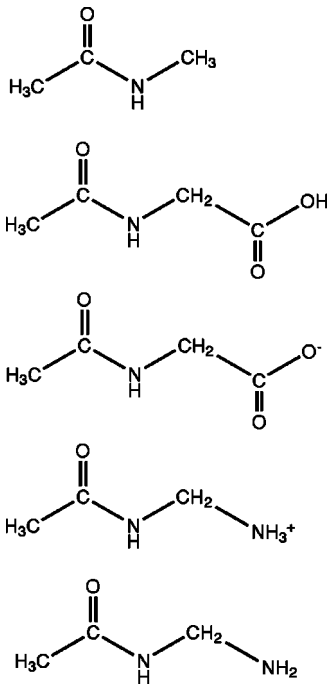

FIG. 3. Illustration of the five molecules used for testing $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ pseudobond parameters. The molecules mimic the protein backbones with different substitutions and net charges. In the pseudobond calculations, the methyl group bonded to the carbonyl carbon is replaced with C_{ps} boundary atom with $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ pseudobond parameters.

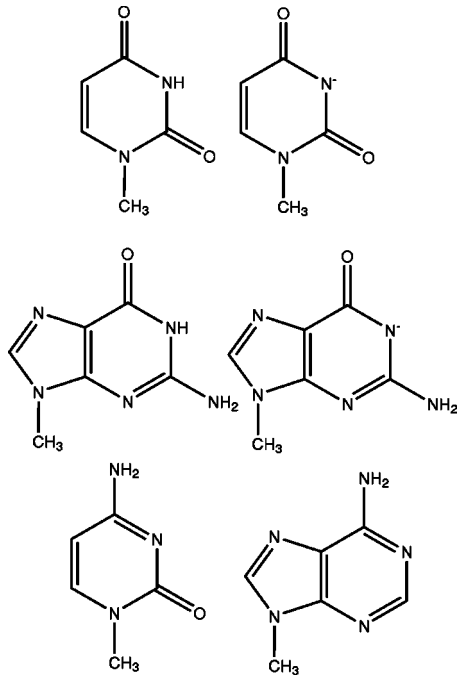

FIG. 5. Illustration of the six molecules used for testing $C_{ps}(sp^3)-N(sp^3)$ pseudobond parameters. The molecules mimic nucleic acid bases. In the pseudobond calculations, the methyl group bonded to the nitrogen atom is replaced with C_{ps} boundary atom with $C_{ps}(sp^3)-N(sp^3)$ pseudobond parameters.

TABLE VII. Test results of $C_{ps}(sp^3)-C(sp^2$, carbonyl) pseudobond parameters in Table I for five molecules in Fig. 3 with different QM methods (B3LYP, BLYP, PW91, HF, MP2) and 6-31G* basis set. Bond (35), angle (40), charge (40), and ΔE (2) refer to the standard deviation (SD) for 35 bond lengths (Å), SD for 40 angles (degree), SD for 40 atomic Mulliken charges, and the mean absolute deviation for 2 deprotonation energies (kcal/mol) between pseudobond QM results and the corresponding standard QM calculation results with the same QM method, respectively.

Method	Bond (35) (Å)	Angle (40) (deg)	Charge (40)	ΔE (2) (kcal/mol)
B3LYP	0.009	1.3	0.05	1.8
BLYP	0.012	1.3	0.06	1.8
PW91	0.013	1.8	0.06	1.9
HF	0.007	1.0	0.02	1.7
MP2	0.006	0.9	0.02	0.4

TABLE IX. Test results of $C_{ps}(sp^3)-N(sp^3)$ pseudobond parameters in Table I for six molecules in Fig. 5 with different QM methods (B3LYP, BLYP, PW91, HF, MP2) and 6-31G* basis set. Bond (77), angle (92), charge (72), and ΔE (2) refer to the standard deviation (SD) for 77 bond lengths (Å), SD for 92 angles (degree), SD for 72 atomic Mulliken charges, and the mean absolute deviation for 2 deprotonation energies (kcal/mol) between pseudobond QM results and the corresponding standard QM calculation results with the same QM method, respectively.

Method	Bond (77) (Å)	Angle (92) (deg)	Charge (72)	ΔE (2) (kcal/mol)
B3LYP	0.010	1.1	0.05	1.1
BLYP	0.011	1.1	0.06	0.8
PW91	0.011	1.2	0.05	1.1
HF	0.008	0.6	0.04	2.2
MP2	0.010	0.7	0.03	3.5

nucleic acid bases. The $C_{ps}(sp^3)-C(sp^2$,carbonyl) pseudobond is tested with five molecules (see Fig. 3) which mimic the protein backbones with different substitutions and net charges. We have also tested $C_{ps}(sp^3)-N(sp^3)$ pseudobond parameters on molecules as shown in Fig. 4 which mimic protein backbones and on molecules as shown in Fig. 5 which mimic nucleic acid bases. Five different QM methods with 6-31G* basis set have been employed for tests, including B3LYP, HF, MP2, BLYP, and PW91. The SD for bond lengths, angles, atomic Mulliken charge, and mean absolute error for deprotonation energies are presented in Tables VII, VIII, and IX. We can see that the developed pseudobond parameters perform very well. For B3LYP method, the errors in bond lengths, angles, atomic Mulliken charge, and deprotonation energies are 0.008 Å, 1.3°, 0.05, and 1.8 kcal/mol for molecules in Fig. 3; 0.010 Å, 1.3°, 0.02, and 0.7 kcal/mol for molecules in Fig. 4; and 0.009 Å, 1.1°, 0.05, and 1.1 kcal/mol for molecules in Fig. 5.

IV. SUMMARY

In order to further improve the applicability and accuracy of the pseudobond *ab initio* QM/MM approach, we have employed a different formulation to develop the pseudobonds. In this formulation, the seven-valence-electron boundary atom, which is used to form the pseudobond, has its own basis set instead of using the basis set of fluorine atom. Here we use an angular-momentum-independent effective core potential and a STO-2G basis set for the boundary atom which has a total of six parameters. By parametrizing both the effective core potential and the basis set, we have not only further improved the $C_{ps}(sp^3)-C(sp^3)$ pseudobond, but also developed accurate $C_{ps}(sp^3)-C(sp^2$,carbonyl) and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds for the cutting of protein backbones and nucleic acid bases with the 6-31G* basis set. The developed pseudobonds are independent of the molecular mechanical force field. Although the parametrization is performed with density functional calculations using hybrid B3LYP exchange-correlation functional, it is found that the same set of parameters is also applicable to Hartree-Fock and MP2 methods, as well as DFT calculations with other exchange-correlation functionals. Tests on a series of molecules yield very good structural, electronic, and energetic results in comparison with the corresponding full *ab initio* quantum mechanical calculations. Meanwhile, we would like to point out that these developed pseudobond parameters are semiempirical in nature because they are parametrized against molecular properties with a limited set of molecules. It would be ideal to design the boundary atom to mimic atomic properties of the corresponding atom instead of molecular properties. Work along this line is in progress.

TABLE VIII. Test results of $C_{ps}(sp^3)-N(sp^3)$ pseudobond parameters in Table I for five molecules in Fig. 4 with different QM methods (B3LYP, BLYP, PW91, HF, MP2) and 6-31G* basis set. Bond (41), angle (42), charge (50), and ΔE (2) refer to the standard deviation (SD) for 41 bond lengths (Å), SD for 42 angles (degree), SD for 50 atomic Mulliken charges, and the mean absolute deviation for 2 deprotonation energies (kcal/mol) between pseudobond QM results and the corresponding standard QM calculation results with the same QM method, respectively.

Method	Bond (41) (Å)	Angle (42) (deg)	Charge (50)	ΔE (2) (kcal/mol)
B3LYP	0.010	1.3	0.02	0.7
BLYP	0.013	1.5	0.02	0.3
PW91	0.013	1.4	0.02	0.4
HF	0.010	1.1	0.05	1.8
MP2	0.009	1.2	0.04	2.5

ACKNOWLEDGMENT

The support from National Science Foundation (CHE-0420870) is gratefully acknowledged.

[1] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[2] U. C. Singh and P. Kollman, J. Comput. Chem. **7**, 718 (1986).
[3] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[4] J. Gao, in *Review in Computational Chemistry*, Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials, Vol. 7 (VCH, New York, 1995), pp. 119–185.
[5] K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoscek, Int. J. Quantum Chem. **60**, 1189 (1996).
[6] D. Bakowies and W. Thiel, J. Phys. Chem. **100**, 10580 (1996).
[7] J. Gao, P. Amara, C. Alhambra, and M. J. Field, J. Phys. Chem. A **102**, 4714 (1998).
[8] M. J. Field, J. Comput. Chem. **23**, 48 (2002).
[9] F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).
[10] P. D. Lyne, M. Hodoscek, and M. Karplus, J. Phys. Chem. A **103**, 3462 (1999).
[11] D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodoscek, and B. R. Brooks, J. Chem. Phys. **117**, 10534 (2002).
[12] P. Amara and M. J. Field, Theor. Chem. Acc. **109**, 43 (2003).

[13] N. Ferre and M. Olivucci, J. Mol. Struct.: THEOCHEM **632**, 71 (2003).
[14] HYPERCHEM *Users Manual, Computational Chemistry* (Hypercube, Inc., Waterloo, Ontario, Canada, 1994).
[15] V. Thery, D. Rinaldi, and J.-L. Rivail, J. Comput. Chem. **15**, 269 (1994).
[16] G. Monard, M. Loos, V. Thery, K. Baka, and J.-L. Rivail, Int. J. Quantum Chem. **58**, 153 (1996).
[17] X. Assfeld and J.-L. Rivail, Chem. Phys. Lett. **263**, 100 (1996).
[18] N. Ferre, X. Assfeld, and J. L. Rivail, J. Comput. Chem. **23**, 610 (2002).
[19] J. Z. Pu, J. L. Gao, and D. G. Truhlar, J. Phys. Chem. A **108**, 632 (2004).
[20] D. M. Philipp and R. A. Friesner, J. Comput. Chem. **20**, 1468 (1999).
[21] R. B. Murphy, D. M. Philipp, and R. A. Friesner, Chem. Phys. Lett. **321**, 113 (2000).
[22] V. Kairys and J. H. Jensen, J. Phys. Chem. A **104**, 6656 (2000).
[23] Y. Zhang, T. S. Lee, and W. Yang, J. Chem. Phys. **110**, 46 (1999).
[24] I. Antes and W. Thiel, J. Phys. Chem. A **103**, 9290 (1999).
[25] G. A. Dilabio, M. M. Hurley, and P. A. Christiansen, J. Chem. Phys. **116**, 9578 (2002).
[26] R. Poteau, I. Ortega, F. Alary, A. R. Solis, J. C. Barthelat, and J. P. Daudey, J. Phys. Chem. A **105**, 198 (2001).
[27] K. Yasuda and D. Yamaki, J. Chem. Phys. **121**, 3964 (2004).
[28] Y. Zhang, H. Liu, and W. Yang, J. Chem. Phys. **112**, 3483 (2000).
[29] H. Liu, Y. Zhang, and W. Yang, J. Am. Chem. Soc. **122**, 6560 (2000).
[30] Y. Zhang, J. Kua, and J. A. McCammon, J. Am. Chem. Soc. **124**, 10572 (2002).
[31] Y. Zhang, J. Kua, and J. A. McCammon, J. Phys. Chem. B **107**, 4459 (2003).
[32] G. A. Cisneros, H. Liu, Y. Zhang, and W. Yang, J. Am. Chem. Soc. **125**, 10384 (2003).
[33] G. A. Cisneros, M. Wang, P. Silinski, M. C. Fitzgerald, and W. Yang, Biochemistry **43**, 6885 (2004).
[34] W. J. Stevens, H. Basch, and M. Krauss, J. Chem. Phys. **81**, 6026 (1984).
[35] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[36] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[37] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[38] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN98, Revision A.11.4 (Gaussian, Inc., Pittsburgh, PA, 2002).
[39] J. P. Perdew, in *Electronic Structure of Solids 91*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).