A pseudobond approach to combining quantum mechanical and molecular mechanical methods

Yingkai Zhang, Tai-Sung Lee, and Weitao Yang

Additional information on J. Chem. Phys.

A pseudobond approach to combining quantum mechanical and molecular mechanical methods

Yingkai Zhang, Tai-Sung Lee,[a] and Weitao Yang
Department of Chemistry, Duke University, Durham, North Carolina 27708

A major challenge for combined quantum mechanical and molecular mechanical methods (QM/MM) to study large molecules is how to treat the QM/MM boundary that bisects some covalent bonds. Here a pseudobond approach has been developed to solve this problem for *ab initio* QM/MM calculations: a one-free-valence atom with an effective core potential is constructed to replace the boundary atom of the environment part and to form a pseudobond with the boundary atom of the active part. This pseudobond, which is described only by the QM method, is designed to mimic the original bond with similar bond length and strength, and similar effects on the rest of the active part. With this pseudobond approach, some well-known deficiencies of the link atom approach have been circumvented and a well-defined potential energy surface of the whole QM/MM system has been provided. The construction of the effective core potential for the pseudobond is independent of the molecular mechanical force field and the same effective core potential is applicable to both Hartree–Fock and density functional methods. 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. © *1999 American Institute of Physics.* [S0021-9606(99)30201-4]

I. INTRODUCTION

Despite the development in computer technology and advances in computational methods, the study of chemical reactions involving large molecules such as enzyme reactions still remains a challenge for computational chemistry. An accurate description of such complex processes, which involve chemical bond formation and breaking, requires quantum mechanics (QM) methods to explicitly consider the electrons.[1,2] However, conventional quantum mechanical methods are so computationally intensive that they are limited to small molecules and are not applicable to such large systems.

One way to approach the problem of large systems is the linear scaling quantum mechanics methods, sometime denoted as $O(N)$ methods. Linear scaling means that the requirement of computer resources, including CPU time and memory, scales linearly with the size of the system. The divide-and-conquer (DAC) approach proposed by Yang was the first linear-scaling method for QM calculations.[3] Subsequently, a great deal of effort has been devoted to the development of linear-scaling algorithms in recent years.[4–18] See a recent review by Yang and Perez-Jorda.[19] With the implementation of the density matrix version of the DAC approach[20] in semiempirical QM methods,[21] handling molecules with several thousand atoms has become possible on a conventional workstation. Recently we have successfully used this DAC approach to study enzyme reactions, and structure and properties of proteins and DNAs.[22–24] Applications of the DAC method have also been made by other groups.[25,26] However, it is still not quite feasible to carry out

similar large scale calculations with *ab initio* methods such as Hartree–Fock (HF) and density functional theory (DFT) methods, even though the algorithms can be linear scaling.

Another more common approach to studying enzyme reactions is to combine quantum mechanical methods with molecular mechanical force fields (QM/MM).[27,28] Warshel and Levitt laid out the basic algorithms of QM/MM approaches in their seminal paper.[29] In the last two decades, there has been much development of such methods by a number of groups.[30–40] In the process of chemical reactions, only a very small number of atoms participate in the bond forming or breaking processes; many other atoms in the system generally serve as a steric and electrostatic environment to influence the properties and reactivities of the reaction active part. Therefore, the small reaction active part can be treated by the quantum mechanical method, while the rest with numerous particles can be described by molecular mechanical methods. Such a combined QM and MM approach can take advantage of the applicability and accuracy of the *ab initio* QM methods for chemical reactions and of the computational efficiency of the MM calculations. The computational scaling of QM/MM method should be between $O(1)$ and $O(N)$.

While the QM/MM method seems to be quite promising, there still remain some obstacles for the QM/MM approach to achieve wide applicability. One of the major challenges is how to treat the QM/MM boundary when the division between QM and MM regions occurs across covalent chemical bonds for a large molecule.[28,31,36,37,39] Until now, the most popular prescription of treating this boundary problem is the link atom approach, whether the QM methods are *ab initio*[30,35,36] or semiempirical.[31,37] Link atoms, which are generally hydrogen atoms, are added to the MM side of the broken covalent bond to satisfy the valency of the QM sys-

[a] Present address: Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446.

tem, except that 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.[41] The link atoms are treated as normal atoms in the QM calculations while the broken bond is still treated by the MM force field. This approach seems to be simple but it has some severe problems. We can see that these artificial link atoms are additional to the system and it introduces some double counting of the interactions. There are quite a few attempts to reduce the effect of the link atoms, but these corrections themselves are arbitrary. Usually they cannot remove the energy and force contributions consistently.[37] Therefore it is not surprising that QM/MM results strongly depend on the placement and treatment of the link atoms.[36] Overall, introducing link atoms not only introduces some artificial effects on the systems, but also can make the QM/MM method lack a unique and consistent definition of the total energy and forces for the QM/MM system.

Since the deficiency of the link atom method is obvious, there is a great deal of interest in the search for better approaches. In the spirit of Warshel and Levitt's hybrid orbital approach,[29] Rivail et al.[34,42,43] have developed the local self-consistent field (LSCF) method which caps the free valences with strictly localized bond orbitals. These strictly localized bond orbitals are obtained from separate quantum mechanical calculations on small model compounds and are assumed to be transferable to large protein molecules. Revail's LSCF approach has been implemented with the semiempirical QM method[34,42] and *ab initio* Hartree–Fock method.[43] LSCF force calculations with the *ab initio* Hartree–Fock method have not been reported. Recently based on Revail's approach, Gao et al.[39] developed a generalized hybrid orbital (GHO) method at the semiempirical level. In the GHO approach, Gao et al. made the localized bond orbitals more transferable by modifying those semiempirical parameters of the boundary atoms.

The development of *ab initio* QM combined with MM is somewhat behind the advances of semiempirical QM combined with MM. For most *ab initio* QM/MM implementations,[30,35,36] hydrogen atoms are usually used as link atoms, although the placement and treatment of link atoms, and the description of the electronic interactions between the QM region and the MM region are quite diverse. To our best knowledge, until now no *ab initio* QM/MM approach without using hydrogen as link atoms has been developed to be capable of performing geometry optimizations or molecular dynamics when the QM/MM boundary occurs across covalent bonds.

Since *ab initio* QM methods like HF and DFT methods are generally more reliable than semiempirical QM methods in describing molecular properties and chemical reactions, we are motivated to develop a pseudobond approach to treat the QM/MM boundary that bisects covalent bonds for the combined *ab initio* quantum mechanical and molecular mechanical methods. This approach aims to overcome major deficiencies of the conventional link atoms approach so that it can provide a well defined *ab initio* QM/MM potential energy surface for modeling large molecules.

II. METHOD

The main idea of our approach is as follows: We consider that a large molecule is partitioned into two parts, an active part and an environment part, by cutting a covalent σ bond $X-Y$. X and Y refer to boundary atoms of the active part and the environment part, respectively. Instead of using a hydrogen atom to cap the free valence of an X atom as in the conventional link atom approach, here a pseudobond $X-Y_{ps}$ is formed by replacing the Y atom with a one-free-valence boundary Y atom (Y_{ps}) which has a parametrized effective core potential. By designing the effective core potential of Y_{ps}, this pseudobond $X-Y_{ps}$ is made to mimic the original $X-Y$ bond with similar bond length and strength, and also similar effects on the rest of the active part.

With the introduction of this pseudobond, 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 the Y atom, the rest of the atoms in the environment part forms the MM subsystem which will be represented by a molecular mechanical force field. The total energy of this QM/MM system can be written as follows:

$$E_{total} = E_{qm}(QM) + E_{mm}(MM) + E_{qm/mm}(QM/MM). \quad (1)$$

where $E_{qm}(QM)$ is the quantum mechanical energy of the QM subsystem, and $E_{mm}(MM)$ the standard molecular mechanical interactions involving exclusively atoms in the MM subsystem.

The QM/MM interaction between the QM subsystem and the MM subsystem can be divided into three terms: electrostatic contribution, van der Waals contribution and MM bonded interaction, as in the following equation:

$$E_{qm/mm}(QM/MM) = E_{electrostatics}(QM/MM)$$
$$+ E_{vdW}(QM/MM)$$
$$+ E_{MM-bonded}(QM/MM), \quad (2)$$

where the MM bonded interaction $E_{MM-bonded}(QM/MM)$ refers to the MM bond, angle and dihedral energy terms, which involve terms with at least one atom from the MM subsystem and one from the QM subsystem. $E_{MM-bonded}(QM/MM)$, $E_{vdW}(QM/MM)$ as well as $E_{mm}(MM)$ are calculated with a MM force field.

In QM calculations, the sum of $E_{qm}(QM)$ and $E_{electrostatics}(QM/MM)$ are calculated as the eigenenergy of an effective Hamiltonian H_{eff} (often ground state); namely,

$$E_{qm}(QM) + E_{electrostatics}(QM/MM) = \langle \Psi | H_{eff} | \Psi \rangle. \quad (3)$$

The QM calculations are performed normally with approximate *ab initio* methods such as HF or DFT. The effective Hamiltonian describes a total number of electrons N_{eff}, which is the sum of the electrons of the active part and the *valence* electrons from all the boundary atoms $Y_{ps}s$. These electrons move in the one-electron potential generated from the nuclei of the atoms in the QM subsystem with charge $\{Z_\alpha\}$, the point change of the MM atoms $\{q_\beta\}$, and the effective core potential (pseudopotential) of the boundary atoms $\{V_{Y_{ps}}^{eff}\}$. The effective Hamiltonian is
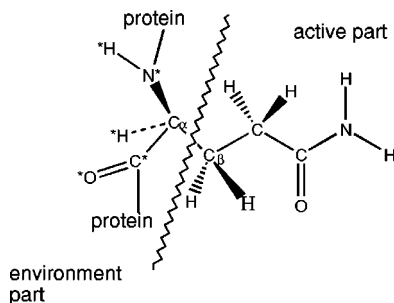
FIG. 1. Illustration of the QM/MM study of a protein, in which the side chain of Gln residue is chosen as the active part. In the pseudobond approach, the C_α is replaced by a one-free-valence boundary carbon atom (C_{ps}) to form a C_{ps}–C_β pseudobond mimicking the original C_α–C_β bond.

$$H_{eff} = -\frac{1}{2}\sum_i^{N_{eff}} \nabla_i^2 + \sum_{i\neq j}^{N_{eff}} \frac{1}{r_{ij}} - \sum_i^{N_{eff}} \sum_{\alpha\in QM} \frac{Z_\alpha}{r_{\alpha i}}$$

$$-\sum_i^{N_{eff}} \sum_{\beta\in MM} \frac{q_\beta}{r_{\beta_i}} + \sum_i^{N_{eff}} \sum_{\gamma\in Y_{ps}} V_{Y_{ps}}^{eff}(r_{i\gamma})$$

$$+\sum_{\alpha_1\neq\alpha_2\in QM} \frac{Z_{\alpha_1}*Z_{\alpha_2}}{r_{\alpha_1\alpha_2}} + \sum_{\alpha\in QM,\beta\in MM} \frac{Z_\alpha^* q_\beta}{r_{\alpha\beta}}, \quad (4)$$

where the last term describes the electrostatic interaction between the nuclei of the QM atoms and the point charges of the MM atoms. The foregoing is a *general* description of our pseudobond approach to the combining QM/MM methods. In the following sections, we address the specific issues on how to handle the MM point charges and the effective core potentials, $\{q_\beta\}$ and $\{V_{Y_{ps}}^{eff}\}$, the remaining two sets of unspecified variables.

A. Electrostatics

There are two issues in the treatment of the MM point charges which need to be addressed: (1) how to achieve the charge consistency—the sum of the partial charges in the MM subsystem plus the charge of the QM subsystem should be equal to the charge of the whole system, (2) in the QM calculation how to reduce the spurious large electrostatics interactions generated from some MM point charges that are very close to the QM subsystem?

Here we resolve these two problems by putting zero point charge on atoms directly bonded to the QM subsystem and some atoms within two bonds of the QM subsystem. This approach takes advantage of the charge pattern of the MM force field. In most MM force fields, such as Charmm[44] and Amber,[45] the MM charges were selected to yield ''groups'' of unit charges. For example considering a protein in Fig. 1, the side chain of Gln residue is the active part and the C_α is the boundary atom of the environment part. In most MM force fields, the amino acid residue has a unit charge, such as -1 for Glu and 0 for Gln. Here we let the QM subsystem have this unit charge, and set zero charge on the rest of the atoms (labeled by an asterisk in Fig. 1) of this amino acid residue. Thus we achieve the charge consistency and also avoid the spurious large electrostatic interactions.

B. Pseudobond with the effective core potential

We describe our approach to constructing this pseudobond for combined *ab initio* QM with MM calculations, which is the key element of present work. Here *ab initio* quantum mechanics methods include both the Hartree–Fock method (HF) and density functional methods (DFT). As the cutting of the sp^3 carbon-carbon bond is most likely to occur in practical QM/MM calculations for enzymes and proteins, we outline our procedures to construct a one-free-valence boundary carbon atom (C_{ps}) to form this pseudobond, and we report the parameters for two very popular basis sets, a small split valence 3-21G basis set and a medium basis set with polarization functions 6-31G*.[1] The procedure outlined herein can also be used to derive the parameters for other kinds of bonds and other basis sets.

Each boundary carbon (C_{ps}) atom has (1) seven valence electrons, (2) nuclear charge seven, and (3) an effective core potential. 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. This choice of seven electrons for such a boundary atom has been used before in the HYPERCHEM software,[41] where a pseudohalogen was used in their combined semiempirical QM and MM program. 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 C_{ps} atom with the effective core potential (or pseudopotential) is here designed to form a pseudobond (C_{ps}–C) with the boundary carbon atom of the active part. This pseudobond mimics the original sp^3 σ carbon–carbon bond. *It should have similar bond length and strength, and similar effects on the active part of the system.* The basis set used for this boundary carbon (C_{ps}) atom is that of the fluorine. The following function form of this effective core potential is found to be successful:

$$V^{eff}(r) = V_L^{eff}(r) + \sum_{l=0}^{L-1} [V_l^{eff}(r) - V_L^{eff}(r)]\sum_m |lm\rangle\langle lm|,$$

$$V_L^{eff}(r) = \frac{a_L * e^{-b_L * r^2}}{r}, \quad (5)$$

$$V_l^{eff}(r) - V_L^{eff}(r) = a_l * e^{-b_l * r^2}, \quad l=0,1,...,L-1,$$

where L is the maximum l of the basis set, a_i and b_i are the fitted parameters. There are four parameters for the 3-21G and six for the 6-31G* basis set. The above formula has the same form as those effective core potentials, such as CEP,[46] programmed in the GAUSSIAN 94 program.[47] Therefore, there is no additional programming effort needed for incorporating the C_{ps} atom.

In our parametrization, the *ab initio* QM method is the most widely used hybrid density functional method B3LYP.[48–50] The training set is comprised of six properties of the ethane: 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 the C–C bond. The
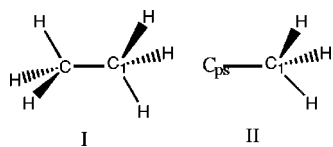
FIG. 2. Illustration for ethane (I) and C_{ps}–CH_3 (II). C_{ps} is the one-free-valence boundary carbon atom with the effective core potential.

ethane molecule and the corresponding C_{ps}–CH_3 are shown in Fig. 2. The parameters of the effective core potentials are optimized such that the six properties of QM calculations on C_{ps}–CH_3 are in accord with the corresponding QM calculations for ethane. Using a combining Monte Carlo sampling with local minimization algorithm[51,52] as our fitting method to efficiently sample the parameter space, a few sets of parameters are obtained which all performed quite well for this training set. Since this pseudobond needs to be transferrable in different chemical environments and ethane itself is quite neutral, these sets of parameters are tested for four additional different systems: C_{ps}–CH_2OH, C_{ps}–CH_2O^{-1}, C_{ps}–CH_2NH_2 and C_{ps}–$CH_2NH_3^+$ to see whether they can reproduce the corresponding QM results on CH_3CH_2OH, $CH_3CH_2O^-$, $CH_3CH_2NH_2$ and $CH_3CH_2NH_3^+$. The set of parameters with the best performance is chosen as the final set. This testing procedure greatly increases the transferability of the resulting pseudobond.

The final parameters for both 3-21G and 6-31G* basis sets are presented in Table I and the results for ethane are shown in Table II and Fig 3. We can see that these two sets of parameters perform quite well not only with the B3LYP method, but also with the HF method and BLYP method, although they were fitted only with the B3LYP method. This shows that these parameters are to some extent independent of the *ab initio* quantum mechanical methods, and they are applicable to both Hartree–Fock and density functional methods, including HF, BLYP and B3LYP. Since in our parametrization, we do not couple with any MM force field, *the design of the effective core potential is independent of any force field*. Therefore, the resulting parameters depend on the QM basis set used, not on the approximate QM methods to some extent, and not on the molecular mechanical force fields. This force-field-free design is particularly appealing for the generality of the methods and the ease of the parametrization.

III. TESTS

We have implemented this pseudobond *ab initio* QM/MM approach based on the GAUSSIAN 94 program[47] and the TINKER program.[53] Here we test our approach on a series of eight molecules as in Fig. 4. Both QM/MM and full *ab initio* calculations are carried out. MM interactions are described by the Charmm22 force field.[44] As for *ab initio* quantum mechanics methods, HF, BLYP and B3LYP methods are used with both basis sets, 3-21G and 6-31G*. For all these eight molecules in QM/MM calculations, the division of the active part and the environment part is achieved by cutting the C–C_1 bond. The C atom of the methyl group is replaced by the boundary carbon to form the pseudobond

with C_1. Three hydrogen atoms of the methyl group comprise the MM subsystem. The rest of the atoms in the molecule belong to the QM subsystem. Since the atoms in the MM subsystem are directly bonded to the QM subsystem, they are all set to have zero partial charges.

A. Structural and electronic properties

The test results of bond lengths, bond angles and atomic Mulliken charges are listed in Tables III–VIII. Note, in particular, that while the eight molecules here include many changes in the active part with substitutions and net charges, the designed pseudobond performs well—it responds to the changes and produces similar effects on the active part when compared to the original C–C_1 bond.

We can see that the standard deviation (SD) between the pseudobond QM/MM approach and the corresponding full QM approach is quite small: the SD for bond lengths is from 0.01 Å for HF (6-31G*) to 0.023 Å for BLYP (3-21G); the SD for bond angles is from 1.3° for HF (3-21G) to 2.0° for BLYP (6-31G*); and the SDs for atomic Mulliken charges are all no more than 0.05. Actually, these standard deviations are similar or even smaller than the deviation between two QM methods with the same basis sets or the same method with different basis sets. For example, for HF (3-21G) and B3LYP (3-21G), their standard deviations for all the data presented in Tables III–V are 0.018 Å for bond lengths, 1.5° for bond angles and 0.14 for Mulliken charges. And for HF

TABLE I. Fitted parameters of the effective core potential for the one-free-valence boundary carbon (C_{ps}) to form the sp^3 C_{ps}–C pseudobond. The function form of the effective core potential is as in Eq. (5).

Basis set	Potential	a_i	b_i
3-21G	V_1	4.362 53	8.516 59
3-21G	$V_0 - V_1$	29.008 29	36.092 43
6-31G*	V_2	0.498 52	2.219 55
6-31G*	$V_0 - V_2$	29.999 04	25.001 20
6-31G*	$V_1 - V_2$	0.402 15	3.568 05

TABLE II. Computed bond lengths (Å), bond angle (°) and atomic Mulliken charges of ethane (I) and of C_{ps}–CH_3 (II) in Fig. 2 with the fitted effective core potential for C_{ps} using various QM methods and basis sets.

| QM method	Basis set	Molecule	C–C_1	C_1–H	C–C_1–H	$q(C_1)$	$q(H)$
HF	3-21G	I	1.542	1.084	110.8	−0.60	0.20
HF	3-21G	II	1.556	1.089	111.9	−0.49	0.16
BLYP	3-21G	I	1.556	1.102	110.8	−0.53	0.18
BLYP	3-21G	II	1.556	1.115	113.1	−0.44	0.13
B3LYP	3-21G	I	1.544	1.095	110.9	−0.56	0.19
B3LYP	3-21G	II	1.551	1.105	112.8	−0.46	0.15
HF	6-31G*	I	1.528	1.086	111.2	−0.48	0.16
HF	6-31G*	II	1.534	1.089	111.5	−0.43	0.12
BLYP	6-31G*	I	1.541	1.104	111.4	−0.39	0.13
BLYP	6-31G*	II	1.545	1.110	111.3	−0.35	0.11
B3LYP	6-31G*	I	1.531	1.096	111.3	−0.43	0.14
B3LYP	6-31G*	II	1.536	1.102	111.2	−0.38	0.12
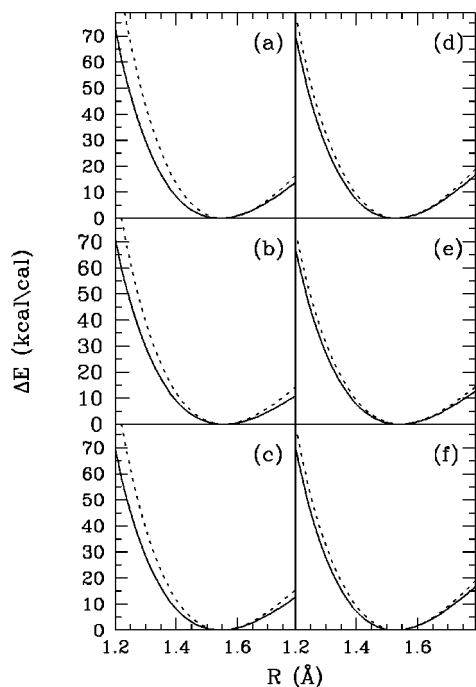
FIG. 3. The C–C bond curve in ethane (solid line) and C_{ps}–C bond curve (dash line) in C_{ps}–CH_3 with the various methods and basis sets: (a) HF (3-21G), (b) BLYP (3-21G), (c) B3LYP (3-21G), (d) HF (6-31G*), (e) BLYP (6-31G*), (f) B3LYP (6-31G*). The curve is obtained by changing the C–C or C_{ps}–C bond length with all other geometry elements fixed as in the optimized geometry.

TABLE III. Test results of the bond lengths (Å) for eight molecules in Fig. 4 with a 3-21G basis set. SD refers to the standard deviation between QM/MM results and QM results with the same QM method.

Mol	Bond	HF	HF/MM	BLYP	BLYP/MM	B3LYP	B3LYP/MM
a	C–C1	1.524	1.527	1.541	1.527	1.529	1.522
a	C1–O	1.444	1.435	1.479	1.474	1.462	1.454
a	O–H1	0.966	0.966	1.010	1.011	0.996	0.996
b	C–C1	1.583	1.572	1.626	1.590	1.606	1.581
b	C–O	1.344	1.308	1.331	1.312	1.328	1.304
c	C–C1	1.543	1.544	1.561	1.548	1.548	1.542
c	C1–N	1.472	1.452	1.501	1.475	1.483	1.457
c	N–H1	1.005	1.004	1.038	1.037	1.026	1.024
d	C–C1	1.527	1.521	1.537	1.509	1.528	1.509
d	C1–N	1.560	1.580	1.586	1.643	1.567	1.606
d	N–H1	1.018	1.017	1.044	1.043	1.035	1.036
e	C–C1	1.532	1.538	1.543	1.527	1.532	1.526
e	C1–S	1.901	1.924	1.946	2.009	1.921	1.964
e	S–H1	1.353	1.353	1.390	1.392	1.376	1.378
f	C–C1	1.540	1.554	1.556	1.557	1.544	1.552
f	C1–S	1.900	1.888	1.921	1.913	1.904	1.891
g	C–C1	1.531	1.541	1.547	1.540	1.535	1.535
g	C1–C2	1.501	1.502	1.527	1.536	1.513	1.519
g	C2–O2	1.203	1.201	1.237	1.235	1.225	1.223
g	C2–O1	1.359	1.360	1.406	1.407	1.383	1.385
g	O1–H1	0.969	0.969	1.012	1.013	0.998	0.998
h	C–C1	1.531	1.544	1.549	1.551	1.537	1.544
h	C1–C2	1.574	1.560	1.630	1.597	1.603	1.579
h	C2–O2	1.251	1.248	1.293	1.284	1.278	1.270
h	C2–O1	1.249	1.253	1.281	1.290	1.269	1.275
SD			0.012		0.023		0.017

(3-21G) and HF (6-31G*), their SD for bond lengths is 0.028 Å, for bond angle is 1.7° and for Mulliken charges is 0.07. These show that this pseudobond QM/MM approach can perform reasonably well in describing the structures and electronic properties of the active part of the system.
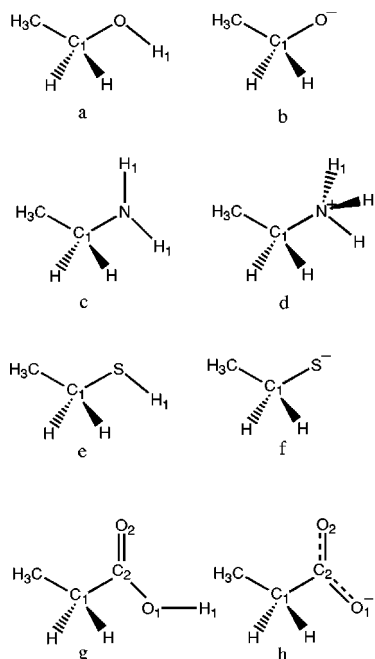

FIG. 4. Illustration for eight molecules: (a) CH_3CH_2OH, (b) $CH_3CH_2O^-$, (c) $CH_3CH_2NH_2$, (d) $CH_3CH_2NH_3^+$, (e) CH_3CH_2SH, (f) $CH_3CH_2S^-$, (g) CH_3CH_2COOH, (h) $CH_3CH_2COO^-$.

TABLE IV. Test results of the bond lengths (Å) for eight molecules in Fig. 4 with 6-31G* basis set. SD refers to the standard deviation between QM/MM results and QM results with the same QM method.

Mol	Bond	HF	HF/MM	BLYP	BLYP/MM	B3LYP	B3LYP/MM
a	C–C1	1.516	1.510	1.530	1.518	1.520	1.511
a	C1–O	1.405	1.401	1.442	1.428	1.426	1.412
a	O–H1	0.946	0.948	0.980	0.983	0.969	0.971
b	C–C1	1.556	1.550	1.588	1.567	1.572	1.559
b	C–O	1.312	1.284	1.319	1.300	1.312	1.289
c	C–C1	1.529	1.526	1.546	1.537	1.534	1.529
c	C1–N	1.455	1.442	1.482	1.448	1.467	1.437
c	N–H1	1.002	1.003	1.030	1.028	1.020	1.019
d	C–C1	1.518	1.504	1.528	1.501	1.519	1.497
d	C1–N	1.519	1.528	1.555	1.576	1.533	1.547
d	N–H1	1.011	1.011	1.036	1.036	1.028	1.028
e	C–C1	1.525	1.525	1.537	1.530	1.527	1.523
e	C1–S	1.828	1.828	1.873	1.872	1.849	1.847
e	S–H1	1.326	1.329	1.364	1.369	1.351	1.355
f	C–C1	1.531	1.532	1.547	1.544	1.535	1.536
f	C1–S	1.829	1.800	1.858	1.814	1.841	1.795
g	C–C1	1.524	1.522	1.538	1.530	1.527	1.522
g	C1–C2	1.507	1.507	1.526	1.528	1.514	1.515
g	C2–O2	1.187	1.188	1.223	1.223	1.211	1.210
g	C2–O1	1.333	1.334	1.379	1.376	1.358	1.358
g	O1–H1	0.952	0.953	0.987	0.988	0.976	0.976
h	C–C1	1.525	1.530	1.541	1.543	1.529	1.534
h	C1–C2	1.557	1.546	1.599	1.560	1.579	1.551
h	C2–O2	1.234	1.235	1.274	1.271	1.259	1.257
h	C1–O1	1.234	1.236	1.269	1.273	1.256	1.258
SD			0.010		0.017		0.015

TABLE V. Test results of the bond angles (°) for eight molecules in Fig. 4 with 3-21G basis set. SD refers to the standard deviation between QM/MM results and QM results with the same QM method.

Mol	Angle	HF	HF/MM	BLYP	BLYP/MM	B3LYP	B3LYP/MM
a	C–C1–O	106.2	108.3	105.4	107.3	105.6	107.7
a	C1–O–H1	110.9	111.2	106.4	107.0	107.6	108.3
b	C–C1–O	113.4	116.7	115.2	117.4	114.8	117.2
c	C–C1–N	114.4	115.7	115.5	116.5	115.1	116.5
c	C1–N–H1	113.4	114.6	109.9	111.4	111.4	112.9
d	C–C1–N	109.5	109.5	109.9	109.8	109.8	109.6
d	C1–N–H1	110.3	109.9	110.0	109.2	110.0	109.4
e	C–C1–S	108.3	106.2	108.0	106.1	108.1	106.0
e	C1–S–H1	98.3	99.9	95.9	98.6	96.5	99.0
f	C–C1–S	112.1	112.1	111.8	112.9	111.8	112.6
g	C–C1–C2	111.4	110.3	110.9	110.4	110.9	110.3
g	C1–C2–O2	127.0	127.9	127.3	128.5	127.0	128.3
g	C1–C2–O1	110.9	109.8	109.8	108.4	110.3	108.8
g	C2–O1–H1	111.8	112.0	106.6	106.8	108.1	108.3
g	O2–C2–O1	122.2	122.3	122.9	123.1	122.8	122.9
h	C–C1–C2	111.9	112.2	110.5	113.1	110.6	112.6
h	C1–C2–O2	115.6	116.8	114.9	117.2	115.1	116.9
h	C1–C2–O1	114.7	113.3	114.9	111.8	115.0	112.4
h	O1–C2–O2	129.7	130.0	130.2	130.9	129.9	130.7
SD			1.3		1.7		1.6

B. Energetics

We have also calculated the energy differences for these four pairs of molecules. We should note that this is a quite stringent test since the pseudobond is only one bond away from the reaction bond X–H, X=O,N,S,O. In order to further test this effect, we also consider the energy difference between $CH_3CH_2CH_2OH$ and $CH_3CH_2CH_2O^-$, for which the pseudobond is two bonds away from the reaction bond O–H in QM/MM calculations, as illustrated in Fig. 5. From Table IX, we can see that for all methods with 3-21G basis set and HF with 6-31G* basis set, the mean absolute deviation between QM/MM and the corresponding QM results is less than 3.5 kcal/mol. The largest deviation is no more than 5.6 kcal/mol. For B3LYP and BLYP with 6-31G* basis set, the deviation between QM/MM and QM is somewhat larger, especially for the deprotonation energy of ethanol. However, in the case of propanol, where the QM/MM boundary is two bonds away from the O–H, the results are much improved in comparison with the ethanol for B3LYP and BLYP with 6-31G* basis set. The deviations between QM/MM and the

TABLE VI. Test results of the bond angles (°) for eight molecules in Fig. 4 with 6-31G* basis set. SD refers to the standard deviation between QM/MM results and QM results with the same QM method.

Mol	Angle	HF	HF/MM	BLYP	BLYP/MM	B3LYP	B3LYP/MM
a	C–C1–O	108.0	109.0	107.6	108.1	107.7	108.4
a	C1–O–H1	109.6	109.7	107.1	108.3	107.8	108.8
b	C–C1–O	113.7	114.7	114.6	115.3	114.3	115.1
c	C–C1–N	115.5	114.8	116.3	114.4	116.0	114.3
c	C1–N–H1	110.6	110.4	108.7	110.2	109.4	110.4
d	C–C1–N	110.3	110.1	110.7	109.5	110.5	109.6
d	C1–N–H1	111.4	110.8	111.1	110.1	111.1	110.3
e	C–C1–S	109.7	107.4	109.4	105.3	109.4	105.8
e	C1–S–H1	98.0	99.4	96.7	98.9	97.0	98.9
f	C–C1–S	113.9	109.0	113.9	108.5	113.7	108.1
g	C–C1–C2	113.0	111.8	112.9	110.7	112.8	111.0
g	C1–C2–O2	126.1	125.8	126.4	125.4	126.1	125.5
g	C1–C2–O1	111.7	112.0	111.2	111.5	111.4	111.6
g	C2–O1–H1	108.1	108.0	105.1	105.3	105.9	106.1
g	O2–C2–O1	122.2	122.2	122.4	123.1	122.4	122.9
h	C–C1–C2	115.2	112.9	113.9	111.7	114.0	111.7
h	C1–C2–O2	116.3	116.0	115.4	116.2	115.6	115.9
h	C1–C2–O1	114.2	114.5	114.9	114.4	114.8	114.3
h	O1–C2–O2	129.5	129.4	129.7	129.4	129.6	129.8
SD			1.5		2.0		1.9

TABLE VII. Test results of the Mulliken atomic charges for eight molecules in Fig. 4 with 3-21G basis set. SD refers to the standard deviation between QM/MM results and QM results with the same QM method.

Mol	Atom	HF	HF/MM	BLYP	BLYP/MM	B3LYP	B3LYP/MM
a	H1	0.37	0.37	0.31	0.30	0.33	0.32
a	O	−0.68	−0.71	−0.50	−0.52	−0.55	−0.57
b	O	−0.86	−0.87	−0.64	−0.63	−0.70	−0.69
c	N	−0.78	−0.82	−0.59	−0.60	−0.65	−0.67
c	H1	0.29	0.29	0.24	0.24	0.26	0.26
d	N	−0.83	−0.85	−0.63	−0.64	−0.69	−0.70
d	H1	0.44	0.44	0.40	0.38	0.41	0.40
e	S	0.02	−0.05	−0.01	−0.11	0.00	−0.08
e	H1	0.06	0.05	0.06	0.05	0.07	0.05
f	S	−0.65	−0.69	−0.65	−0.68	−0.65	−0.67
g	C2	0.87	0.78	0.56	0.49	0.64	0.56
g	O2	−0.61	−0.61	−0.44	−0.44	−0.48	−0.48
g	O1	−0.72	−0.72	−0.50	−0.50	−0.56	−0.56
g	H1	0.41	0.41	0.34	0.34	0.36	0.36
h	C2	0.85	0.76	0.51	0.44	0.58	0.51
h	O2	−0.79	−0.80	−0.58	−0.61	−0.63	−0.66
h	O1	−0.80	−0.81	−0.60	−0.61	−0.65	−0.65
SD			0.04		0.04		0.04

corresponding QM for deprotonation energy of propanol is only 1.3 kcal/mol for B3LYP with a 6-31G* basis set and 4.5 kcal/mol for BLYP with a 6-31G* basis set.

C. QM/MM calculations of succinamide

To further test our pseudobond approach, we perform a QM/MM study of the succinamide. As illustrated in Fig. 6, the division of the active part and the environment part is achieved by cutting the C–C bond. With the replacement of the C of the environment part with a boundary carbon (C_{ps}), the C_{ps} and the active part are treated as a QM subsystem, and the rest is treated classically. Here we use B3LYP with a 6-31G* basis set as the QM method and Charmm22[44] as the

TABLE VIII. Test results of the Mulliken atomic charges for eight molecules in Fig. 4 with 6-31G* basis set. SD refers to the standard deviation between QM/MM results and QM results with the same QM method.

Mol	Atom	HF	HF/MM	BLYP	BLYP/MM	B3LYP	B3LYP/MM
a	H1	0.44	0.44	0.37	0.37	0.39	0.39
a	O	−0.74	−0.78	−0.57	−0.57	−0.61	−0.62
b	O	−0.90	−0.85	−0.68	−0.57	−0.74	−0.63
c	N	−0.83	−0.85	−0.65	−0.60	−0.70	−0.67
c	H1	0.33	0.33	0.27	0.28	0.29	0.30
d	N	−0.85	−0.90	−0.68	−0.69	−0.73	−0.75
d	H1	0.46	0.47	0.42	0.41	0.43	0.43
e	S	−0.07	−0.11	−0.09	−0.13	−0.09	-0.12
e	H1	0.09	0.08	0.08	0.07	0.09	0.08
f	S	−0.76	−0.77	−0.75	−0.68	−0.75	−0.69
g	C2	0.75	0.74	0.53	0.44	0.58	0.50
g	O2	−0.56	−0.58	−0.43	−0.38	−0.46	−0.43
g	O1	−0.70	−0.74	−0.53	−0.52	−0.57	−0.57
g	H1	0.46	0.47	0.38	0.39	0.41	0.41
h	C2	0.74	0.72	0.48	0.39	0.53	0.45
h	O2	−0.76	−0.78	−0.60	−0.55	−0.64	−0.61
h	O1	−0.77	−0.78	−0.61	−0.55	−0.65	−0.60
SD			0.03		0.05		0.05
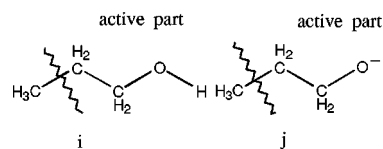
FIG. 5. Illustration for (i) $CH_3CH_2CH_2OH$ and (j) $CH_3CH_2CH_2O^-$.

MM force field. For those atoms in the MM subsystem, only N, H_1 and H_2 have MM partial charges, −0.62 for N and 0.31 for H_1 and H_2. The QM subsystem has no net charge. Thus the sum of the charge for the QM subsystem and the MM subsystem is zero, the same as the whole system. We have performed QM, MM and QM/MM calculations for this system, and listed the geometry and atomic Mulliken charge results in Table X. We can see that those QM/MM geometry elements described quantum mechanically are similar to the full QM calculations, and those described classically are similar to the full MM calculation results. For charges, QM/MM results also agree well with the corresponding QM results.

IV. SUMMARY

In this paper, a pseudobond approach for combined *ab initio* quantum mechanical and molecular mechanical calculations (QM/MM) has been developed for the treatment of the QM/MM boundary across covalent bonds. The main features of the method are as follows: (1) One-free-valence boundary atoms (Y_{ps}) with an effective core potential are used to terminate the electronic wave function of the active part, instead of the more commonly used link atoms. (2) The pseudobond formed between a Y_{ps} atom and a boundary atom of the active part mimics the original covalent bond with similar bond length and strength, and similar effects on the rest of the active part. (3) The effective core potential for the boundary atoms (Y_{ps}) depends on the basis set used in the QM calculations, but not on the MM force fields. The same effective core potential is applicable to both Hartree–

TABLE IX. Energy difference (kcal/mol) for five pairs of molecules in Figs. 4 and 5 with various methods and basis sets. MAD refers to the mean absolute deviation between the QM/MM results and the corresponding QM results.

| Basis set	Method	a→b	d→c	e→f	g→h	i→j	MAD
3-21G	HF	−418.8	−239.9	−355.5	−375.6	−417.3	
3-21G	HF/MM	−415.6	−239.2	−358.7	−380.1	−420.3	2.9
3-21G	BLYP	−406.6	−240.1	−363.4	−371.9	−404.8	
3-21G	BLYP/MM	−402.0	−238.5	−366.3	−376.7	−407.1	3.2
3-21G	B3LYP	−410.2	−240.0	−361.6	−372.4	−408.5	
3-21G	B3LYP/MM	−404.6	−237.9	−363.9	−377.0	−410.9	3.4
6-31G*	HF	−405.5	−231.5	−366.1	−366.7	−404.4	
6-31G*	HF/MM	−403.6	−234.3	−367.9	−371.5	−409.0	3.2
6-31G*	BLYP	−396.4	−229.8	−367.5	−364.3	−395.1	
6-31G*	BLYP/MM	−380.7	−223.4	−358.3	−361.6	−390.6	7.7
6-31G*	B3LYP	−399.3	−230.1	−367.3	−364.9	−398.0	
6-31G*	B3LYP/MM	−385.3	−225.3	−359.8	−364.4	−396.7	5.6
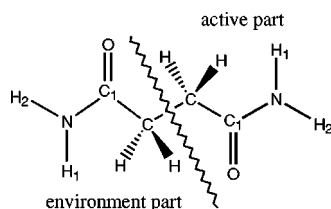
FIG. 6. Illustration of succinamide.

Fock and density functional methods, including HF, BLYP and B3LYP. (4) The quantum mechanical calculations are straightforward and the total energy and forces are also well defined.

In comparison with the conventional link atom approach, there are no additional atoms added and no double counting of the interactions. The boundary region between the active part and the environment part is explicitly and well treated without any energy and force corrections or constraints. Thus

a unified and consistent definition of the energy and force for the whole QM/MM system has been provided by this pseudobond approach.

The tests on a series of molecules have yielded very good structural, electronic and energetic results in comparison with the corresponding full quantum mechanical calculations. To our best knowledge, this is the first approach without using hydrogen as link atoms to provide a well-defined potential energy surface for geometry optimizations at the level of *ab initio* QM/MM calculations when the division of the active part and environment part involves the cutting of covalent bonds. This pseudobond QM/MM approach should be widely applicable to study enzyme reactions.

ACKNOWLEDGMENTS

TABLE X. Geometries and atomic Mulliken charges for succinamide calculated by B3LYP, B3LYP/MM and MM methods. For B3LYP and B3LYP/MM, the 6-31G* basis set is used. The MM force field used is Charmm22.

		B3LYP/MM		
	B3LYP	QM[a]	MM[b]	MM
Bond length Å				
C–C	1.524	1.527		1.567
C–H	1.097	1.105	1.114	1.111
C–C1	1.526	1.531	1.539	1.536
C1–O	1.224	1.222	1.238	1.230
C1–N	1.368	1.371	1.372	1.355
N–H1	1.008	1.009	1.002	0.994
N–H2	1.010	1.010	1.001	0.996
Bond angle (°)				
H1–N–H2	118.9	118.7	118.8	123.5
H1–N–C1	122.7	122.5	121.8	120.0
H2–N–C1	118.4	118.8	119.5	116.5
N–C1–O	122.1	122.4	122.1	120.1
N–C1–C	114.5	114.6	113.5	114.8
O–C1–C	123.4	122.9	124.3	125.1
H–C H	105.3	106.8	107.8	108.3
C1–C–C	112.0	111.3	115.4	114.7
H–C–C	110.1	111.2	108.8	108.7
Dihedral angle (°)				
H1,N,C1,O	180.0	180.0	180.0	180.0
N,C1,C,C	180.0	180.0	180.0	180.0
Atomic Mulliken charge				
H1	0.34	0.33		
H2	0.34	0.34		
N	−0.75	−0.74		
C1	0.60	0.50		
O	−0.51	−0.46		
C	−0.36	−0.32		
H	0.18	0.12		

[a]Refers to those atoms, bonds, bond angles and dihedral angles involving atoms exclusively of the QM subsystem, which are described by the QM method.

[b]Refers to those bonds, bond angles and dihedral angles involving at least one atom of the MM subsystem, which are described by the MM method.

[1] W. Hehre, L. Radom, P. Schleyer, and J. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).
[2] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford Univesity Press, New York, 1989).
[3] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[4] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1992).
[5] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994).
[6] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).
[7] M. S. Daw, Phys. Rev. B **47**, 10895 (1993).
[8] E. Hernández and M. J. Gillan, Phys. Rev. B **51**, 10157 (1994).
[9] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[10] J. M. Millam and G. E. Scuseria, J. Chem. Phys. **106**, 5569 (1997).
[11] A. D. Daniels, J. M. Willam, and G. E. Scuseria, J. Chem. Phys. **107**, 425 (1997).
[12] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[13] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **52**, 1640 (1995).
[14] P. Ordejón, D. Drabold, M. Grumbach, and R. M. Martin, Phys. Rev. B **48**, 14646 (1993).
[15] P. Ordejón, E. Artacho, and J. M. Soler, Phys. Rev. B **53**, R10441 (1996).
[16] W. Hierse and E. B. Stechel, Phys. Rev. B **50**, 17811 (1994).
[17] D. A. Drabold and O. F. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).
[18] J. J. P. Stewart, Int. J. Quantum Chem. **58**, 133 (1996).
[19] W. Yang and J. M. Pérez-Jordá, in *Encyclopedia of Computational Chemistry*, edited by P. Schleyer (Wiley, New York, in press).
[20] W. Yang and T. S. Lee, J. Chem. Phys. **103**, 5674 (1995).
[21] T.-S. Lee, D. York, and W. Yang, J. Chem. Phys. **105**, 2744 (1996).
[22] J. P. Lewis, C. W. Carter, Jr., J. Herman, W. Pan, T. S. Lee, and W. Yang, J. Am. Chem. Soc. **120**, 5407 (1998).
[23] D. York, T. S. Lee, and W. Yang, J. Am. Chem. Soc. **118**, 10940 (1996).
[24] D. York, T. Lee, and W. Yang, Phys. Rev. Lett. **80**, 5011 (1998).
[25] S. L. Dixon and K. M. Merz, Jr., J. Chem. Phys. **104**, 6643 (1996).
[26] R. T. Gallant and A. St-Amant, Chem. Phys. Lett. **256**, 569 (1996).
[27] A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes* (Wiley, New York, 1991).
[28] J. Gao, *Review in Computational Chemistry* (VCH, New York, 1995), Vol. 7, pp. 119–185.
[29] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[30] U. C. Singh and P. Kollman, J. Comput. Chem. **7,** 718 (1986).
[31] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[32] J. Gao and X. Xia, Science **258**, 631 (1992).
[33] R. V. Stanton, D. S. Hartsough, and K. M. Merz, Jr., J. Phys. Chem. **97**, 11868 (1993).

[34] V. Thery, D. Rinaldi, and J.-L. Rivail, J. Comput. Chem. **15**, 269 (1994).

[35] F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).

[36] K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoscek, Int. J. Quantum Chem. **60**, 1189 (1996).

[37] D. Bakowies and W. Thiel, J. Phys. Chem. **100**, 10580 (1996).

[38] I. B. Bersuker, M. K. Leong, J. E. Boggs, and R. S. Pearlman, Int. J. Quantum Chem. **63**, 1051 (1997).

[39] J. Gao, P. Amara, C. Alhambra, and M. J. Field, J. Phys. Chem. A **102**, 4714 (1998).

[40] N. Vaidehi, T. A. Wesolowski, and A. Warshel, J. Chem. Phys. **97**, 4264 (1992).

[41] Hypercube, Inc., *HyperChem Users Manual, Computational Chemistry* (Hypercube, Inc., Waterloo, Ontario, Canada, 1994).

[42] G. Monard, M. Loos, V. Thery, K. Baka, and J.-L. Rivail, Int. J. Quantum Chem. **58**, 153 (1996).

[43] X. Assfeld and J.-L. Rivail, Chem. Phys. Lett. **263**, 100 (1996).

[44] A. D. MacKerell, Jr. *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[45] W. D. Cornell *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).

[46] W. J. Stevens, H. Basch, and M. Krauss, J. Chem. Phys. **81**, 6016 (1984).

[47] M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andes, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, GAUSSIAN 94, *Revision E.2* (Gaussian, Inc., Pittsburgh, PA, 1995).

[48] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[49] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[50] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[51] Z. Li and H. A. Scheraga, Pro. Natl. Acad. Sci. **84**, 6611 (1987).

[52] J. P. K. Doye and D. J. Wales, Phys. Rev. Lett. **80**, 1357 (1998).

[53] J. W. Ponder, TINKER, *Software Tools for Molecular Design, Version 3.4, June 1997*. (The most updated version for the TINKER program can be obtained from J. W. Ponder's World Wide Web site at http://dasher.wustl.edu/tinker).