QM/MM hybrid calculation of biological macromolecules using a new interface program connecting QM and MM engines

2009 J. Phys.: Condens. Matter 21 064234

(http://iopscience.iop.org/0953-8984/21/6/064234)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **21** (2009) 064234 (6pp) doi:10.1088/0953-8984/21/6/064234

QM/MM hybrid calculation of biological macromolecules using a new interface program connecting QM and MM engines

Yohsuke Hagiwara[1,2], **Takehiro Ohta**[2,3] **and Masaru Tateno**[1,2,4]

[1] Graduate School of Pure and Applied Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba Science City, Ibaraki 305-8571, Japan
[2] Center for Computational Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba Science City, Ibaraki 305-8577, Japan
[3] CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi 332-0012, Japan

E-mail: tateno@ccs.tsukuba.ac.jp

Received 19 July 2008, in final form 15 October 2008
Published 20 January 2009
Online at stacks.iop.org/JPhysCM/21/064234

Abstract

An interface program connecting a quantum mechanics (QM) calculation engine, GAMESS, and a molecular mechanics (MM) calculation engine, AMBER, has been developed for QM/MM hybrid calculations. A protein–DNA complex is used as a test system to investigate the following two types of QM/MM schemes. In a 'subtractive' scheme, electrostatic interactions between QM/MM regions are truncated in QM calculations; in an 'additive' scheme, long-range electrostatic interactions within a cut-off distance from QM regions are introduced into one-electron integration terms of a QM Hamiltonian. In these calculations, 338 atoms are assigned as QM atoms using Hartree–Fock (HF)/density functional theory (DFT) hybrid all-electron calculations. By comparing the results of the additive and subtractive schemes, it is found that electronic structures are perturbed significantly by the introduction of MM partial charges surrounding QM regions, suggesting that biological processes occurring in functional sites are modulated by the surrounding structures. This also indicates that the effects of long-range electrostatic interactions involved in the QM Hamiltonian are crucial for accurate descriptions of electronic structures of biological macromolecules.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Biological functions are achieved through interactions of biological macromolecules such as proteins, DNA and RNA. It is difficult to perform theoretical investigations of such interactions based on their electronic structures, because of the following reasons. First, most proteins, for example, consist of more than several thousand atoms, and solvent water molecules contribute to the reaction and/or stabilization/formation of their 3D structures, leading to huge calculation model systems. Second, biological phenomena such as enzymatic reactions and conformational changes are driven by thermal fluctuations, i.e. effects of finite temperatures. Thus, to analyze the function and structure of biological systems, molecular dynamics (MD) calculations using molecular mechanics (MM) potentials have generally been performed, and they play a great role in understanding the structural basis of biological functions. However, there are limitations in such MM-based MD methods when simulating processes such as the formation and/or cleavage of covalent bonds, charge fluctuations due to both conformational and electronic structure changes are crucial for enzymatic reactions. On the other hand, it is not practical to describe entire biological systems using quantum mechanical (QM) methods due to high computational costs. To overcome these difficulties, quantum mechanics/molecular mechanics (QM/MM) calculations [1–5] are utilized, in which the system is divided into QM and MM atoms. Here, QM regions correspond to active sites to be investigated, and are described on the basis of the quantum mechanics theory. MM regions

[4] Author to whom any correspondence should be addressed.

correspond to the remainder of the system, and are described at the molecular mechanics level.

At present, there are some programs available to perform QM/MM calculations. AMBER 9 [6] and CHARMM [7] include subroutines for QM calculations, thereby allowing us to perform QM/MM MD simulations. However, the QM methods used by these programs are limited to semi-empirical schemes. In addition, in some quantum chemical calculation program packages such as Gaussian [8] and GAMESS [9], subroutines for MM calculations are implemented for QM/MM calculations, enabling us to perform calculation by *ab initio* methods for the QM parts. However, since an integrator for MD simulations is not implemented in these packages, free energy calculations coupled with sophisticated algorithms are not available for efficient searches of various molecular states, such as the potential of the mean force (PMF).

To couple the advantages of such conventional packages, we developed a new interface program to connect MM MD (AMBER) and QM (GAMESS) engines, for which the code is, in principle, highly parallelized, and enabled us to perform high-performance QM/MM MD calculations. Although ChemShell [10] and PUPIL [11] were developed as interface programs to couple conventional programs and are now very widely used, this is the first report to describe the combination of AMBER and GAMESS. Our developed interface is written primarily in UNIX shell scripts. To implement complex algorithms, or to perform large scale I/O for storing big files, C programs are used coupled with the modified source code of AMBER and GAMESS.

2. Methodology

2.1. QM/MM scheme

The total Hamiltonian H_{total} of our QM/MM scheme is expressed as follows:

$$\hat{H}_{\text{total}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM1}} + \hat{H}_{\text{QM/MM2}}. \quad (1)$$

Here, H_{QM} is the QM Hamiltonian, and H_{MM} is the classical MM-based Hamiltonian. The MM region is further divided into the MM1 region, which is within a distance r from the center of the QM region, and the MM2 region that includes the remainder of the QM and MM1 regions. $H_{\text{QM/MM1}}$ is the hybrid Hamiltonian that describes the interactions of QM and MM1 atoms; it is defined as follows:

$$\hat{H}_{\text{QM/MM1}} = -\sum_{\text{i,M1}} \frac{q_{\text{m}}}{r_{\text{i,m}}} + \sum_{\text{A,M1}} \frac{Z_{\text{A}} q_{\text{m}}}{r_{\text{A,m}}} + \sum_{\text{A,M1}} \left(\frac{A_{\text{A,M}}}{r^{12}_{\text{A,M}}} - \frac{B_{\text{A,M}}}{r^{6}_{\text{A,M}}} \right). \quad (2)$$

Here, q_{m} denotes the partial charge of atom m included in MM1 region, and Z_{A} is the bare charge (i.e. the effective nuclear charge) of atom A included in the QM region. The subscripts M and A in the summation term denote the numbers of MM1 and QM atoms, respectively. A characteristic feature of our hybrid scheme is that partial charges of MM1 atoms are incorporated into the one-electron integrals corresponding to the first term of equation (2), and therefore, the polarization of the QM region by MM1 atoms can be considered. The second term in equation (2) describes interactions of the nuclei of QM atoms and partial charges, and the third term is the Lennard-Jones (LJ) potential, representing the van der Waals (vdW) energy between the QM and MM1 atoms. The contribution from the QM atoms to the force acting on the MM1 atoms is obtained by integrating the interaction between the partial charges of MM atoms and the electron density at each grid point of the quantum region, as shown in equation (3).

$$F^{\text{QM}}_{\text{m}} = \sum_{j}^{N_{\text{cube}}} \mathbf{r}_{\text{m}j} \frac{q_{\text{m}}}{|r_{\text{m}j}|^3} \text{d}q_j. \quad (3)$$

Here, q_{m} is a partial charge in a MM1 atom, and N_{cube} is the total number of electron density points. Here, $\text{d}q_j$ defines a small volume on the electron density grid, as shown in equation (4).

$$\text{d}q_j = \rho_j \, \text{d}x \, \text{d}y \, \text{d}z. \quad (4)$$

This procedure is more precise than the computation of the interactions of partial charges of MM atoms, and those assigned to QM atoms calculated via procedures of the RESP [12], Mulliken [13] and Löwdin [14] charges, since the polarization of the electron density induced by MM1 atoms considers the forces of MM1 atoms. To save computational time, interactions between QM and MM2 atoms are described at the MM level. In our QM/MM scheme, the LJ potential is incorporated into the QM Hamiltonian H_{QM}, since density functional theory (DFT), which is used for QM calculations in this study, cannot be used in estimating vdW interaction energy. In contrast, when using higher-level *ab initio* calculations, such as the second-order Møller–Plesset perturbation (MP2) calculation and the coupled cluster calculation with singles, doubles and perturbative triples (CCSD(T)), for the QM part, the LJ potential does not need to be included in the QM Hamiltonian.

For the boundary atoms in QM regions, which are usually assigned to sp^3 carbon atoms and form covalent bonds with boundary atoms in MM1 regions, the link atom approach is used to reconstruct a saturated valence structure. Each bond between such boundary atoms in the QM and MM1 regions is 'capped' by a hydrogen atom; this 'link atom' is placed along the bond vector between the QM and MM1 atoms at a distance of ∼1 Å. This new particle is treated as a QM atom, and shares the same pair list for QM/MM interactions as real QM atoms. Although this approach is widely used, it has a problem; when a link atom (L) is attached to a QM atom (Q_1), a physically meaningless interaction appears between the L atom and the MM atom (M_1) that is bonded to the Q_1 atom (see figure 1).

To avoid such illegally close bonded interactions, the partial charge of M_1 is removed from the one-electron integration in the QM Hamiltonian, while it is included in the MM Hamiltonian. By this modification, the equilibrated distance between Q_1 and M_1 is modified from that of the C–C bond to that of the C–H bond, which is shorter and more polarized than the C–C bond.

Using an alternative method for QM/MM calculations, i.e. the 'subtractive' scheme, can avoid this problem. The

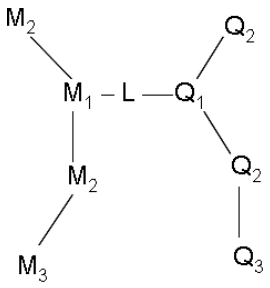

Figure 1. Schematic representation of boundary sites between QM and MM regions. Q_i represents the QM atom i and M_i represents the MM atom i. L represents the link atom, which is attached to Q_1 atom.

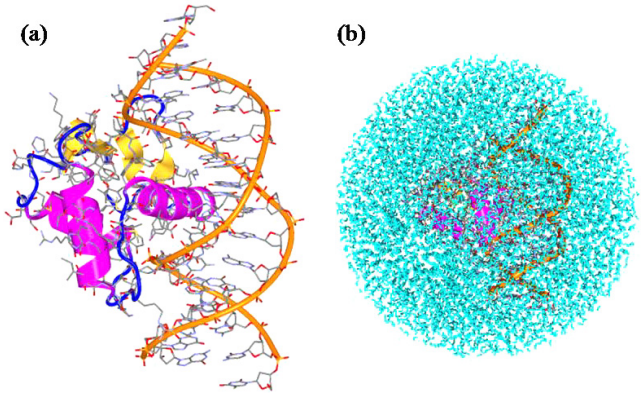

Figure 2. (a) Crystal structure of the complex of PU.1 and the target DNA (PDB ID: 1PUE). (b) Fully solvated structure of the complex of PU.1 and the DNA. A snapshot of the MD simulation is shown.

description of the total energy for this scheme is expressed as follows.

$$E = E(\text{Rayer1} + \text{Link}, \text{QM}) + E(\text{Rayer1} + \text{Rayer2}, \text{MM}) - E(\text{Rayer1} + \text{Link}, \text{MM}), \quad (5)$$

where Rayer1 corresponds to the region to be treated at the QM level, and Rayer2 is outside Rayer1. The first term represents the energy of Rayer1 calculated at the QM level, and the second term is the energy of the entire region (i.e. Rayer1 and Rayer2) calculated at the MM level. The third term corresponds to the correction term used to avoid the following two errors: (i) double counting of the energy of Rayer1 and (ii) contribution of the link atoms to the total energy. Since interactions between Rayer1 and Rayer2 are described at the MM level, influences of Rayer2 on the quantum mechanical properties of Rayer1 are completely ignored.

2.2. Classical MD simulation of the crystal structure of a DNA–protein complex

The coordinates of a crystal structure of PU.1 in a complex with the target DNA was obtained from the Protein Data Bank (PDB) (PDB ID: 1PUE) [15] (figure 2(a)). Hydrogen atoms were added to the crystal structure using the LEAP module of the AMBER 9 package [16]. Then, the protein was fully solvated in spherical water droplets, the radius of which was 40 Å from the center of mass of the complex. The TIP3P water model was used for the solvation (figure 2(b)). Thus, the total atom number of the solvated protein–DNA complex system was 24 414. To obtain energetically favorable configurations of the solvent water molecules, the following procedure was adopted. First, energy minimization was performed by the steepest descent method, where only the solvent water molecules were free from any constraints (the atoms of the protein, DNA and crystallographic water molecules were positionally constrained using the harmonic potential with a force constant of 500 kcal mol^{-1} Å^{-2}). Then, MD simulation was performed for 10 ps at 300 K with positional constraints, which are the same as those used in the minimization phase. Subsequently, energy minimization was performed, in which all atoms in the system were free from positional constraints. All calculations were performed using the Amber 9 program with the parm99 force field parameter. The MD simulation of the solvated system was performed under a constant pressure of 1.013×10^5 Pa, with a periodic boundary condition at 300 K. The temperature and pressure were controlled using the Berendsen algorithm [17]. The SHAKE algorithm was adopted to treat the bonds involving hydrogen [18], and the time step for integration was set as 1 fs. A cut-off distance of 16 Å was used to calculate electrostatic interactions.

2.3. Interface program connecting QM and MM engines

Functions of the software interface are (i) to control the sequence of steps for the QM/MM optimization and MD simulation, and (ii) to exchange information between AMBER and GAMESS. In the first task, with respect to a coordinate set, the input file for GAMESS is generated at each step of the optimization or MD simulation, and a single-point calculation is performed. Each time GAMESS completes a QM calculation, two types of forces can be obtained; the forces on QM atoms derived from QM–QM and QM–MM1 interactions, and the forces on the MM1 atom derived from interactions between their partial charges and the grids of electron density of the QM atoms. With respect to the coordinate set, AMBER calculates (i) forces derived from vdW interactions, which are calculated for all atoms of the system, and (ii) forces on MM1 and MM2 atoms calculated using equation (5), where the electrostatic interactions between the QM and MM1 atoms are removed (the electrostatic interactions can be obtained at the QM level in the GAMESS calculation). After completion of these calculations, the forces calculated by GAMESS and AMBER are combined. At this stage, it is necessary to map the relationship between MM and QM atoms, since the manner of atom labeling varies with software packages. However, the UNIX shell is not suitable for (i) treating large files, including information on forces and the mapping of more than 100 000 atoms, and (ii) complex algorithms such as force merging. To avoid the overhead cost of using the UNIX shell, C programs are used. In fact, even in the case where a system size is much larger, for instance, the number of atoms included in the system equals

∼165 000, the computational time required for the above-mentioned operations is only 3.25 s when using an Intel Core 2 Quad processor (2.83 GHz); thus, it is negligible compared with QM calculations. In the ChemShell, the Tool Command Language (Tcl) and the C programming are used to handle such substantial data, and in the PUPIL, Java is used to address this issue. Thus, it is expected that the overhead costs for the data handling are compatible among those programs and our interface program, while detailed timings were not described in their papers [10, 11]. At the last stage of a cycle in the interface program, coordinates of all atoms are updated using the merged forces in AMBER's MD and optimization routines.

3. Results and discussion

We used a protein–DNA complex as a test system, i.e. a crystal structure of PU.1 in a complex with the target double-stranded DNA molecules. This is a member of the Ets family of proteins, which consists of transcriptional factors that regulate gene expression for biological growth and development. They share a conserved domain composed of around 85 amino acid residues, which binds with the consensus DNA sequence, 5′-GGAA-3′, as the core sequence in the DNA. The crystallographic analyses of PU.1 revealed that two conserved arginines, R232 and R235, participate in the specific recognition of nucleic acid bases of the core sequence. The side chain of R235 forms hydrogen bonds with G8, and the side chain of R232 is positioned between the bases of G9 and A10, which should recognize both of them.

To obtain a completely solvated structure of the complex of PU.1 and the target DNA, we immersed the crystal structure in a solvent box, and performed a classical MD simulation for 10 ps with all atoms of the protein and DNA positionally constrained by a harmonic potential. Then, geometry optimization at the MM level was performed for a final snapshot of the MD simulation. Thus, we obtained the initial structure of the following QM/MM calculations (figure 3).

QM regions include the DNA core sequence, i.e. G8·C26, G9·C25, A10·T24 and A11·T23, and a base pair, G7·C27, which stacks with G8·C26. For the nucleotides in QM regions, their phosphate groups and ribose moieties are not included in the QM regions, with the exception of those of T24, which are recognized by an amino acid residue, K229, of PU.1. With respect to the protein moiety, side chains of the following amino acid residues, which recognize nucleotides involved in QM regions, are also assigned as QM atoms, i.e. the two conserved arginines (R232 and R235), E228 (which recognizes bases of C25 and C26) and K229 (which recognizes the phosphate of T24). In addition, ordered water molecules that form hydrogen bonds with nucleotides and the above-mentioned amino acid residues related to the intermolecular recognition are in the QM regions. Thus, 338 atoms are assigned as QM atoms.

For QM calculations, restricted Hartree–Fock (RHF)/density functional theory (DFT) hybrid all-electron calculations are performed using the B3LYP functional. For the test, we performed two QM/MM calculations, which are referred to as model I and model II. Model I is a simulation using an additive QM/MM scheme, where partial charges of MM atoms are considered in a QM Hamiltonian as one-electron integration; model II is a simulation using a subtractive QM/MM scheme, where interactions of QM atoms and partial charges in MM regions are included only in MM calculations, i.e. the electron density of the QM regions cannot be perturbed by the partial charges of MM atoms. All QM/MM calculations in this study are performed at the self-consistent field (SCF) level. Calculations of geometry optimization are now in progress using the computational models presented here. Furthermore, in this study, our analysis of electronic structures is focused on the effects of the MM regions on some orbitals close to the highest occupied molecular orbitals (HOMOs) in models I and II. They may play a crucial role in such cases as enzymatic reactions in protein–DNA complexes. For the other molecular orbitals (MOs), similar analysis is now in progress.

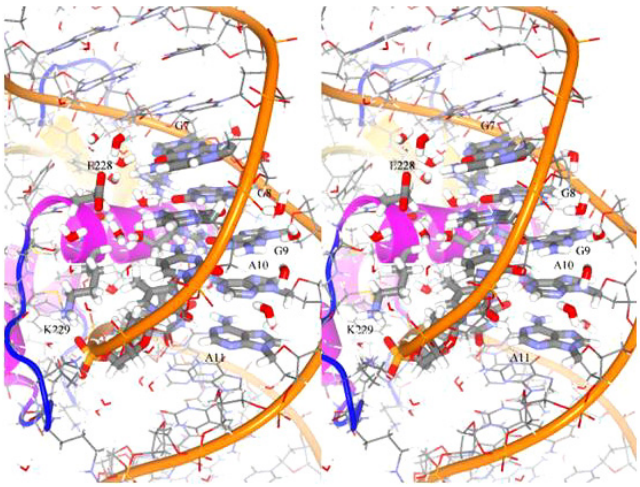

Figure 3. Stereo view of interactions of the core sequence, GGAA, with PU.1. Atoms involved in QM regions are depicted as sticks, and MM atoms as wires.

Figure 4(a) shows a comparison of such orbitals close to HOMOs present in models I and II. In model I, HOMO is localized on the G7 base, while HOMO-3 is located on the two guanine bases, G8 and G9. However, each orbital on G8 and G9 bases in HOMO-3 is similar to HOMO (G7 base); accordingly, these three orbitals located on the guanine bases are equivalent. HOMO-1 and HOMO-2 are localized on the E228 side chain and A11 base, respectively, in model I. On the contrary, in model II, HOMO is not localized on a nucleic acid base, but on the E228 side chain, which is corresponding to HOMO-1 in model I. Similarly, HOMO-1 and HOMO-3 in model II are shifted to HOMO-3 and HOMO, respectively, in model I. In this way, comparison of models I and II shows that the order of some MOs is changed in models I and II; this result shows that the electronic structures of the QM regions are actually perturbed by the MM regions, although HOMO-2 is equivalent in the two models.

Next, we investigated energy differences between the pairs of corresponding MOs belonging to distinct models (figure 4(b)). For example, with respect to HOMO in model I, we first define $\Delta E^{\mathrm{HOMO(I)}}$ and $\Delta E^{\mathrm{HOMO\text{-}3(II)}}$ as the energy

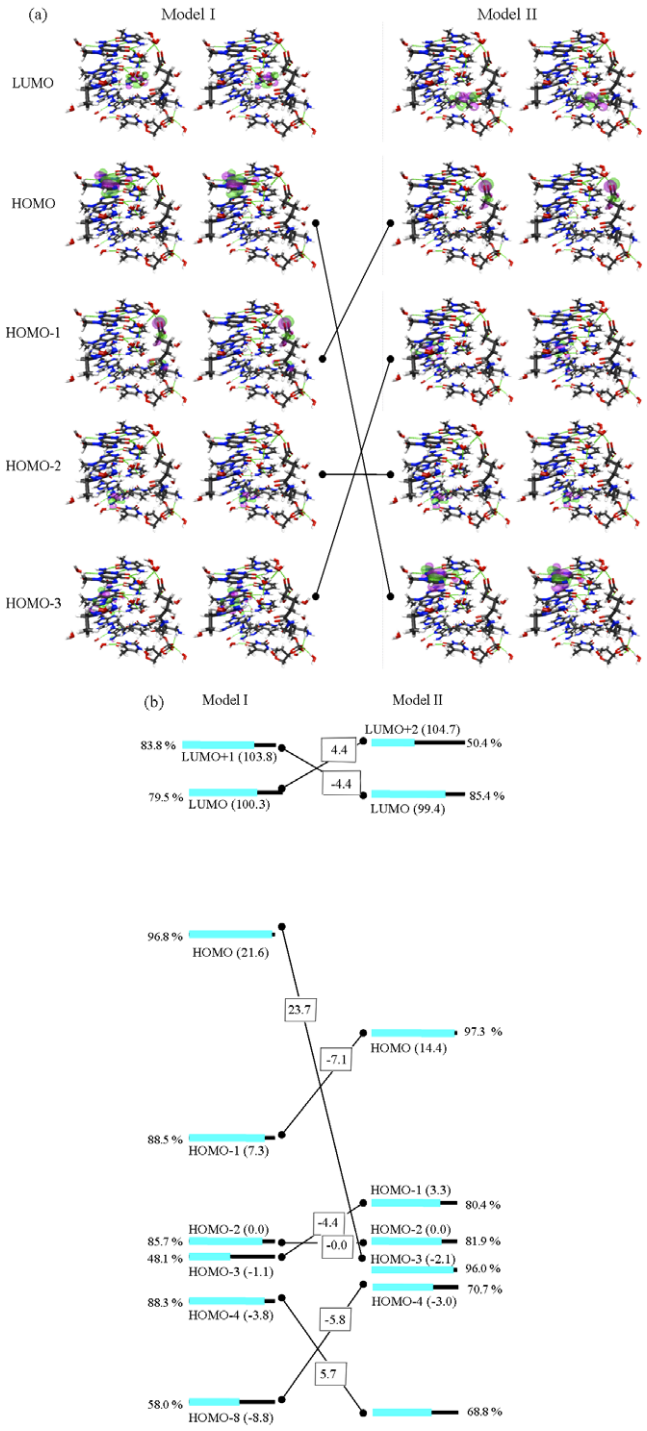

Figure 4. (a) A mapping of MOs obtained using models I and II. A line connects an MO in one model with the corresponding MO in the other model. (b) An energy diagram of the MOs; the values in parentheses are $\Delta E^{\mathrm{MO(model)}}$ in kcal mol^{-1}, e.g. $\Delta E^{\mathrm{HOMO\text{-}1(I)}}$ equals 7.3 kcal mol^{-1}. It should be noted here that $\Delta E^{\mathrm{MO(model)}}$ in each computational model are calculated using HOMO-2 as the reference in each model (see the text). A bold line shows the occupancy equal to the total value of occupancies of atomic orbitals belonging to functional groups with the most dominant contribution to the MO. For example, 96.8% of HOMO in model I is occupied by atomic orbitals of the G7 base; thus, the bold line shows the total value (96.8%) of occupancies of the atomic orbitals that belong to G7, found in HOMO.

differences between HOMO-2 and HOMO in model I and between HOMO-2 and HOMO-3 in model II, respectively. It should be noted here that HOMO-2 is used as the reference for the definition of the energy difference between corresponding MOs in each model, since the orbitals are common in both models. The energy difference between HOMO in model I and HOMO-3 in model II is defined as $\Delta\Delta E^{\mathrm{HOMO(I)}} = \Delta E^{\mathrm{HOMO(I)}} - \Delta E^{\mathrm{HOMO\text{-}3(II)}}$; this value was 23.7 kcal mol^{-1}, indicating that HOMO in model I is significantly shifted in model II (this discussed in detail later in this paper).

In this manner, we obtained values for $\Delta\Delta E^{\mathrm{HOMO\text{-}1(I)}}$ ($=\Delta E^{\mathrm{HOMO\text{-}1(I)}} - \Delta E^{\mathrm{HOMO(II)}}$) and $\Delta\Delta E^{\mathrm{HOMO-3(I)}}$ ($=\Delta E^{\mathrm{HOMO\text{-}3(I)}} - \Delta E^{\mathrm{HOMO\text{-}1(II)}}$) of -7.1 and -4.4 kcal mol^{-1}, respectively. Furthermore, $\Delta\Delta E^{\mathrm{HOMO\text{-}4(I)}}$ ($=\Delta E^{\mathrm{HOMO\text{-}4(I)}} - \Delta E^{\mathrm{HOMO-8(II)}}$) and $\Delta\Delta E^{\mathrm{HOMO8}(I)}$ ($=\Delta E^{\mathrm{HOMO\text{-}8(I)}} - \Delta E^{\mathrm{HOMO\text{-}4(II)}}$) were 5.7 and -5.8 kcal mol^{-1}, respectively. As for LUMOs, $\Delta\Delta E^{\mathrm{LUMO(I)}}$ ($=\Delta E^{\mathrm{LUMO(I)}} - \Delta E^{\mathrm{LUMO+2(II)}}$) and $\Delta\Delta E^{\mathrm{LUMO+1(I)}}$ ($=\Delta E^{\mathrm{LUMO+1(I)}} - \Delta E^{\mathrm{LUMO(II)}}$) were 4.4 and -4.4 kcal mol^{-1}, respectively. Thus, these electronic structures are slightly affected by MM atoms, and thereby, the order of the energy levels is changed. On the contrary, $\Delta\Delta E^{\mathrm{HOMO(I)}}$ ($=\Delta E^{\mathrm{HOMO(I)}} - \Delta E^{\mathrm{HOMO\text{-}3(II)}}$) of 23.7 kcal mol^{-1} is significantly larger, showing that ignoring the MM region has a dramatic influence on HOMO in model I. Since the G7 base is located on the QM and MM boundary, and thus, directly faces the MM regions, the electronic structure of G7 could be more sensitive to polarization by the MM atoms than those of QM atoms located close to the center of the QM region. However, although A11 is also located on the QM and MM boundary, HOMO-2 on A11 does not deviate in the two models (actually, this orbital is used as the reference in each model, as mentioned above). Therefore, the large deviation of the G7 base energy level (HOMO in model I) could be caused by long-range electrostatic interactions, as well as by short-range interactions.

Thus, the long-range electrostatic effects from MM atoms should be included in a QM Hamiltonian to avoid misleading electronic structures. Functional groups in MM regions, which mainly contribute to the polarization of QM regions, are thought to be phosphate groups of DNA, since a phosphate carries a negative formal charge (-1). In addition, unordered bulk water can also contribute to this perturbation, since a water molecule can polarize QM atoms. In this way, it is suggested that biological functions involved in active sites are modulated through long-range electrostatic interactions with the surrounding regions. The factors mentioned here are crucial, in particular, for investigations of catalytic reactions occurring in complexes of proteins and DNA. For instance, in a reaction for repairing damaged bases of DNA, a nucleotide residue is thought to be attacked by an amino acid residue (lysine), to displace the damaged bases [19]. In a methylation reaction of DNA, an adenine base is supposedly activated by an amino acid residue (aspartate), and attacks a methyl group to be attached [20]. Accordingly, to accurately describe such reactions using QM/MM simulations, it would be essential to incorporate the phosphate backbone moieties of nucleic acids in the QM Hamiltonian to interact with QM regions through their partial charges, or to treat them as QM atoms, if computational costs are acceptable.

Recently, on the basis of two distinct QM/MM schemes, i.e. the additive and subtractive schemes, which are similar to those used here, we performed geometry optimization for a Cu-binding active site of azurin, which is assigned as a QM region, using the same QM/MM calculation system described here [21]. We compared the results, and showed that electronic properties obtained using the additive scheme are more consistent with spectroscopic experiments. In addition, we found significant differences in bonds between Cu and the coordinated atoms, particularly for the polarized bonds involved in the Cu coordination. This indicates that partial charges of those surrounding atoms should be incorporated into the QM Hamiltonian for accurate descriptions of the electronic and geometric structures of biological active sites surrounded by atoms involved in polarization effects.

4. Conclusions

In this study, we developed a UNIX shell-based interface program connecting the QM and MM calculation engines, GAMESS and AMBER. Overhead costs of the UNIX shell are avoided using C programs and modifications of the AMBER and GAMESS source code. The GAMESS and AMBER engines used in our system are highly efficient for parallelization; therefore, our system is suitable for simulations of large molecular systems. To test the system, we applied it to a protein–DNA complex, where 338 atoms are treated as QM atoms, and compared the results obtained by the additive and subtractive schemes.

We have found significant differences in electronic structures including boundary orbitals obtained in the two computational models. This discrepancy between the two calculations could be induced by the phosphate groups of the DNA, which are assigned as MM regions in this study, since a phosphate carries a high potency for polarizing QM regions. In addition, unordered bulk water is supposed to contribute to the polarization of QM atoms. It is suggested that biological functions involved in active sites are modulated through long-range electrostatic interactions with the surrounding regions. Therefore, for accurate descriptions of electronic and geometric structures of active sites, partial charges of the surrounding atoms should be incorporated into a QM Hamiltonian. Thus, our interface program can play a critical role as a powerful tool for accurate QM/MM simulations of large systems, such as biological macromolecules.

Acknowledgments

This work is partly supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) under contract Nos 19019003, 19340108 and 20051003. Computations were performed using the computer facilities under the ‘Interdisciplinary Computational Science Program’ at the Center for Computational Sciences, University of Tsukuba and the Computer Center for Agriculture, Forestry, and Fisheries Research, MAFF, Japan.

References

[1] Morokuma K 2007 *Bull. Chem. Soc. Japan* **80** 2247
[2] Lin H and Truhlar D G 2007 *Theor. Chem. Acc.* **117** 185
[3] Senn H M and Thiel W 2007 *Curr. Opin. Chem. Biol.* **11** 182
[4] Bruice T C 2006 *Chem. Rev.* **106** 3119
[5] Murphy R B, Philipp D M and Friesner R A 2000 *J. Comput. Chem.* **21** 1442
[6] Walker R C, Crowley M F and Case D A 2008 *J. Comput. Chem.* **7** 1019
[7] Woodcock H L III, Hodoccek M, Gilbert A T B, Gill P M W, Schaefer H F III and Brooks B R 2007 *J. Comput. Chem.* **28** 1485
[8] Frisch M J *et al* 2004 *Gaussian03* Gaussian, Inc, Wallingford, CT
[9] Schmidt M W *et al* 1993 *J. Comput. Chem.* **14** 1347
[10] Sherwood P *et al* 2003 *J. Mol. Struct. (Theochem)* **632** 1
[11] Torras J, Seabra G M, Deumens E, Trickey S B and Roitberg A E 2008 *J. Comput. Chem.* **29** 1564
[12] Wang J, Cieplak P and Kollman P A 2000 *J. Comput. Chem.* **21** 1049
[13] Mulliken R S 1955 *J. Chem. Phys.* **23** 1833
[14] Löwdin P O 1970 *Adv. Quantum Chem.* **5** 185
[15] Kodandapani R, Pio F, Ni C Z, Piccialli G, Klemsz M, McKercher S, Maki R A and Ely K R 1996 *Nature* **380** 456
[16] Case D A *et al* 2006 *AMBER 9* University of California, San Francisco
[17] Berendsen H J C, Postma J P M, van Gunsteren W F, DiNola A and Haak J R 1984 *J. Chem. Phys.* **81** 3684
[18] Ryckaert J P, Ciccotti G and Berendsen H J C 1977 *J. Comput. Phys.* **23** 327
[19] Verdine G L and Norman D P 2003 *Annu. Rev. Biochem.* **72** 337
[20] Gong W, O’Gara M, Blumental R M and Cheng X 1997 *Nucleic Acids Res.* **25** 2702
[21] Ohta T, Hagiwara Y, Kang J, Nishikawa K, Yamamoto T, Nagao H and Tateno M 2009 *J. Comput. Theor. Nanosci.* at press