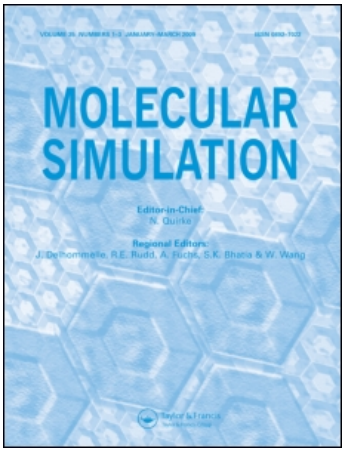

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Ab initio quantum mechanical/molecular mechanical molecular dynamics using multiple-time-scale approach and perturbation theory

Motoyuki Shiga[a]; Masanori Tachikawa[bc]
[a] Center for Computational Science and Engineering, Japan Atomic Energy Agency, Tokyo, Japan [b] Quantum Chemistry Division, Graduate School of Science, Yokohama-city University, Yokohama, Japan [c] PRESTO, Japan Science and Technology Agency, Tokyo, Japan

To cite this Article Shiga, Motoyuki and Tachikawa, Masanori(2007) '*Ab initio* quantum mechanical/molecular mechanical molecular dynamics using multiple-time-scale approach and perturbation theory', Molecular Simulation, 33: 1, 171 — 184
To link to this Article: DOI: 10.1080/08927020601052922
URL: http://dx.doi.org/10.1080/08927020601052922

Molecular Simulation, Vol. 33, Nos. 1–2, 15 January–15 February 2007, 171–184

Taylor & Francis
Taylor & Francis Group

Ab initio quantum mechanical/molecular mechanical molecular dynamics using multiple-time-scale approach and perturbation theory

MOTOYUKI SHIGA†* and MASANORI TACHIKAWA‡¶§

†Center for Computational Science and Engineering, Japan Atomic Energy Agency, 6-9-3, Higashi-Ueno, Taito-ku, Tokyo 110-0015, Japan
‡Quantum Chemistry Division, Graduate School of Science, Yokohama-city University, Seto 22-2, Kanazawa-ku, Yokohama 236-0027, Japan
¶PRESTO, Japan Science and Technology Agency, Tokyo, Japan

(Received July 2006; in final form October 2006)

A new computational method is proposed for *ab initio* quantum-mechanical/molecular-mechanical (QM/MM) molecular dynamics (MD) which is limited to time-independent thermodynamic analysis. The idea is to use the mass scaling method combined with multiple-time-scale (MTS) algorithm and an approximate QM/MM Hamiltonian derived from the first-order Rayleigh–Schrödinger perturbation theory (PT) in which the electronic polarization is neglected as a first approximation. If the polarization effect is not so strong, the correction can also be considered after the simulation run using the weighted sampling method. The advantage and disadvantage of the method is discussed in terms of its computational efficiency and accuracy. As a simple example, we demonstrate an MD simulation of liquid water containing one quantum mechanical (QM) molecule and 255 molecular mechanical (MM) molecules and discuss the advantages in calculating statistical averages such as radial distribution and heat of solution.

Keywords: Molecular dynamics; Water; ab initio; QM/MM

1. Introduction

There has recently been a growing demand for theoretical methods that enable more realistic computer simulations of chemical processes in liquids. One of the promising approaches in this respect is a molecular dynamics (MD) simulation within the *ab initio* quantum-mechanical/molecular-mechanical (QM/MM) description of the system, which consists of a chemically important subsystem (solute) based on *ab initio* quantum mechanical (QM) force field and a surrounding subsystem (solvent) based on molecular mechanical (MM) force field [1–12]. However, the *ab initio* QM/MM calculation has a problem in that it often spends much time in order to take sufficient statistics so as to describe large structural fluctuation of liquid matter. This is simply because the *ab initio* electronic structure calculation of the QM subsystem must be carried out every time step during the MD run. Thus, it has been very hard work to obtain thermodynamic averages using the *ab initio* QM/MM method. For instance, several millions of time steps are needed to quantitatively evaluate the radial distribution of a molecule (QM) solvated in liquid water (MM) and it will spend several weeks or months even if the QM/MM calculation can be done within several seconds per step. To obtain the free energy profiles for a chemical reaction in aqueous solution, the situation is more serious. The QM/MM method should be used in combination with methods such as thermodynamic integration [13], or other methods [14,15] which would be extremely time-consuming unless some kind of approximation is made.

Therefore, it should be very important to consider how the computation can be reduced in the QM/MM method with minimal loss of accuracy. This report suggests a new way to carry out QM/MM calculations efficiently, if we are only interested in *time-independent* thermodynamic properties, such as the radial distribution, heat of solution and free energy curve in a chemical reaction. Thus, time-dependent properties are beyond our scope here.

*Corresponding author. Email: shiga.motoyuki@jaea.go.jp
§Email: tachi@yokohama-cu.ac.jp

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020601052922

In the QM/MM method, the calculation usually spends much more time on the QM forces (the forces acting on QM atoms) rather than on the MM forces (the forces acting on MM atom or site), even when the QM subsystem is much smaller than the MM subsystem. On the other hand, the MD sampling as to the MM configurations is required much more abundantly than the QM configurations, since the configuration space is essentially proportional to the degrees of freedom. Considering this situation, it would be useful to employ multiple-time-scale (MTS) MD [16,17], in which different time scales (step sizes) can be used for the QM and MM forces. The sampling by MTS algorithm will become more powerful when the MM molecules move relatively faster than the QM molecules.

Here, it is important to point out that, as long as the time-independent thermodynamics is concerned, we can control the relative speed of QM and MM molecules by changing their masses. In other words, the MM molecules can be *made* faster in the simulation than in the reality, by scaling the atomic masses of the MM molecules to be artificially smaller than the real ones. Of course, the dynamical aspect of the trajectory is completely destroyed with the mass scaling. However, the finite-temperature configurations would be generated properly along the trajectory (in the same sense as in the Monte-Carlo method). The thermal distribution of configurations has nothing to do with the mass, in principle, because the statistical averages of classical canonical ensemble do not depend on the atomic masses.

The idea to apply the MTS algorithm and mass scaling to the QM/MM calculation has been first suggested by Woo *et al.* [18]. Let us briefly explain this nice idea, which is also employed in this paper. The MTS algorithm consists of a double MD cycle to treat with fast MM molecules and the slow QM molecules. One is an outer cycle which updates the QM force by a long time step and the other is an inner cycle which updates the MM–MM and the QM–MM interaction forces by a short time step. The algorithm will become more efficient than the usual single-time-step algorithm if the inner cycle, especially the QM–MM interaction, can be computed very quickly. Thus, whether the algorithm works successfully or not depends on the type of theoretical model used for the QM–MM coupling.

Woo *et al.* [18] has employed this method in Car–Parrinello based QM/MM calculation for gaseous hydrocarbon molecules by the link atom approach where the QM–MM electrostatic (ES) forces were totally neglected. Therefore, the aim of this paper is to explicitly take account of the QM–MM ES coupling within this scheme, which should be important in wide range of condensed matter such as aqueous solutions. Moreover, both the *ab initio* part (QM and QM–MM interactions) is systematically calculated based on conventional quantum chemical method with a localized Gaussian basis set.

Now, in order to take advantage of the MTS algorithm in QM/MM calculations, we have the following conditions: (a) the forces should be divided properly into contributions from the QM force, the MM force and the QM–MM interaction force and (b) the QM–MM interaction force should be computed much faster than the QM force. Let us consider if these conditions are met when the QM–MM ES coupling is included. Conventionally, there are two types of theoretical model in the QM–MM ES interaction. The first type is the electrostatic embedding (EE) model (hereafter, "QM/MM-EE") which employs the *ab initio* determination of QM–MM ES interaction by one-electron Coulomb integrals. In this model, the MM charges induce the QM electronic polarization and the polarization, in turn, has an influence on the MM forces reflecting the ES interaction between polarized QM electrons and MM charges. The QM–MM interaction can not be extracted simply from the *ab initio* calculation without solving the QM electronic wave function by self-consistent-field (SCF) procedure in the presence of MM atomic charges. As a result, a full *ab initio* calculation must be done each instant of time to obtain the QM–MM interaction and this QM/MM-EE model is not suited for the MTS algorithm. The second type is the mechanical embedding (ME) model (hereafter, "QM/MM-ME"), in which the QM–MM interaction is completely replaced by the MM force field (i.e. the MM*–MM interaction, where the QM molecules are referred to as the "MM*" molecules for convenience). In this model, the computation of the MM force and the MM*–MM interaction can usually be done much faster than the QM force by *ab initio* calculation of isolated QM subsystem. Therefore, as discussed in the next section, the MTS algorithm usually works well in the QM/MM-ME model. However, it is well known that the QM/MM-ME model has a drawback in that the electric charges on QM atoms must be given artificially. (Finding the QM charges might be troublesome if they do not exist in literatures.) As a particular case, the ME model may not be good for chemical reactions since the QM charges are usually fixed to constant values under bond breaking process.

For the reasons above, here we consider a new QM/MM approach so that it is suitable to MTS simulation whereas the QM–MM ES interactions given explicitly by *ab initio* one-electron Coulomb integrals. As we show in the next section, we refer to this scheme as "perturbative embedding" (or "QM/MM-PE" method) since it is derived from the conventional QM/MM Hamiltonian based on first-order Rayleigh–Schrödinger perturbation theory (PT). Physically, this approximation employs the ES interaction between the MM point charges and the QM electron density corresponding to that of the isolated QM molecule at the instantaneous configuration. As the effect of electronic polarization of the QM molecule induced by the surrounding MM charges is considered only beyond the second-order PT, such an effect is neglected within the QM/MM-PE method. However, the method would help a great deal to accelerate MD simulation for combined QM/MM systems, because we do not have to compute the two-electron integrals and the SCF every time step.

In addition, the method has the good performance in parallel computation by dividing the sum of QM–MM interactions (one-electron Coulomb integrals).

In Section 2, the theoretical background of QM/MM-PE method is introduced. In Section 3, we test the validity of the approximation by comparing the results among the QM/MM-PE, the conventional QM/MM and the full QM calculations on a water dimer cluster. To assess the computational efficiency of this method, an MD simulation of liquid water, which is composed of one QM water molecule dissolved in 255 MM water molecules, has been carried out at the room temperature. The water dimer and liquid water systems are studied in detail since they have been used frequently as a validation test of the QM/MM methods [19–24] In addition to the QM/MM-PE MD results, we show a prescription to include the electronic polarization effect using the weighted sampling method. It is shown that the calculated radial distribution functions, heat of solution and molecular dipole moment are found in good agreement with experimental results. These calculations are basically done by Hartree–Fock approximation with 6-31G(d,p) basis set for the QM molecule and a modified version of the flexible simple point charge model [25] (SPC/F2) for the MM molecules.

2. Theory

2.1 The QM/MM-PE method

In the combined QM/MM description, the total Hamiltonian is given by [1–12]

$$H_{\text{total}} = T_{\text{QM}} + T_{\text{MM}} + \hat{H}_{\text{QM}}(\mathbf{r},\mathbf{R}) + V_{\text{MM}}(\mathbf{R}') + \hat{H}^{\text{es}}_{\text{QM-MM}}(\mathbf{r},\mathbf{R},\mathbf{R}') + V^{\text{vdw}}_{\text{QM-MM}}(\mathbf{R},\mathbf{R}'), \quad (1)$$

where the first two terms are the kinetic energy of the QM and MM molecules,

$$T_{\text{QM}} = \sum_{I}^{\text{QM}} \frac{\mathbf{P}_I^2}{2M_I}, \quad T_{\text{MM}} = \sum_{A}^{\text{MM}} \frac{\mathbf{P}_A'^2}{2M_A'}, \quad (2)$$

and the last four terms are the potential energies of the QM subsystem, the MM subsystem and the QM–MM ES interaction, the QM–MM van der Waals (VDW) interaction, respectively. The vectors $\mathbf{R}$ and $\mathbf{P}$ ($\mathbf{R}'$ and $\mathbf{P}'$), respectively give the nuclear (atomic) coordinates and momenta in the QM (MM) subsystem and $\mathbf{r}$ denotes the electron coordinate. We use I and A as the indices of QM and MM atoms, respectively. The MM potential V_{MM} is the sum of MM–MM inter-molecular contribution and MM intra-molecular contribution.

In equation (1), the last four terms are related to the adiabatic potential surface. In the conventional QM/MM-EE method, we solve the Schrödinger equation,

$$\left(\hat{H}_{\text{QM}} + \hat{H}^{\text{es}}_{\text{QM-MM}}\right)|\Psi\rangle = E|\Psi\rangle, \quad (3)$$

where E is the sum of potential energies for the QM subsystem and the QM–MM ES interaction, $E = V^{(0)}_{\text{QM}} + V^{\text{es}}_{\text{QM-MM}}$. Using the SCF procedure in *ab initio* method, the electronic ground state Ψ is solved in the presence of electric field due to the MM atomic charges. In other words, the electronic polarization effect is included in E and Ψ.

Thus, the total potential is given by

$$V_{\text{total}} = E(\mathbf{R},\mathbf{R}') + V_{\text{MM}}(\mathbf{R}') + V^{\text{vdw}}_{\text{QM-MM}}(\mathbf{R},\mathbf{R}'), \quad (4)$$

Here, we emphasize that E is determined by solving Ψ at each instantaneous molecular configuration of the whole system. As E depends not only on the QM configuration $\mathbf{R}$ but also on the MM configuration $\mathbf{R}'$, the SCF calculation is necessary whenever the positions of MM atoms (as well as QM atoms) are updated. In this situation, it should be difficult to improve the computational efficiency of the QM/MM MD even if we use the MTS algorithm.

To circumvent this problem, we apply the Rayleigh–Schrödinger PT to equation (1), regarding the third and fourth terms as the reference Hamiltonian and the fifth and sixth terms as the perturbation. Within the first-order, we have only to solve the zero-th order electronic ground state $\Psi^{(0)}$, which corresponds to that of *isolated* QM molecules in the absence of MM molecules,

$$\hat{H}_{\text{QM}}|\Psi^{(0)}\rangle = V^{(0)}_{\text{QM}}|\Psi^{(0)}\rangle. \quad (5)$$

Using $\Psi^{(0)}$, we can calculate the first-order perturbation term giving the QM–MM interaction

$$V^{\text{es}(1)}_{\text{QM-MM}} = \langle\Psi^{(0)}|\hat{H}^{\text{es}}_{\text{QM-MM}}|\Psi^{(0)}\rangle, \quad (6)$$

and then the total potential becomes

$$V^{\text{PE}}_{\text{total}} = V^{(0)}_{\text{QM}}(\mathbf{R}) + V_{\text{MM}}(\mathbf{R}') + V_{\text{QM-MM}}(\mathbf{R},\mathbf{R}'), \quad (7)$$

where

$$V_{\text{QM-MM}}(\mathbf{R},\mathbf{R}') = V^{\text{es}(1)}_{\text{QM-MM}}(\mathbf{R},\mathbf{R}') + V^{\text{vdw}}_{\text{QM-MM}}(\mathbf{R},\mathbf{R}'). \quad (8)$$

Now, the QM potential $V^{(0)}_{\text{QM}}(\mathbf{R})$, in the first term in equation (7), only depends on the QM configuration $\mathbf{R}$. In other words, total potential is divided so that the QM potential is separated from the parts that are depend on the MM configuration $\mathbf{R}'$, i.e the second and third terms in equation (7). This character plays the key role in the MTS algorithm, as shown below. We note in passing that the total potential in the QM/MM-ME model is the same as equation (7) except that the QM–MM ES interaction is replaced by MM*–MM ES interaction and therefore the QM/MM-ME model is also suitable to the MTS algorithm. From now on, the superscripts such as "(0)" and "(1)" are removed for simplicity.

Equation (6) corresponds to the QM–MM interaction without polarization effect, assuming that the QM electronic wave function is identical to that of the isolated QM system. In other words, the effect of electronic

polarization due to the MM charges is neglected in the QM/MM-PE model. To consider the polarization effect, we must include the higher-order corrections, which would result in coupling between electronic excitations. We will come back to this issue later.

Now, the QM–MM ES interaction term $V^{\mathrm{es}}_{\mathrm{QM-MM}}(\mathbf{R},\mathbf{R}')$ in equation (8) is given by

$$V^{\mathrm{es}}_{\mathrm{QM-MM}}(\mathbf{R},\mathbf{R}') = \sum_{M}^{\text{MM molecules}} \left[\left\{ \sum_{a}^{\text{AO}} \sum_{b}^{\text{AO}} D_{ab}(\mathbf{R}) I_{ab,M} \times(\mathbf{R},\mathbf{R}') \right\} f_{\mathrm{cut}}\left(\left| \mathbf{R}'^{(c)}_{M} - \mathbf{R}^{(c)} \right| \right) \right], \quad (9)$$

where we have defined one-electron reduced density matrix by $D_{ab}(\mathbf{R})$ and the sum of one-electron Coulomb integrals

$$I_{ab,M}(\mathbf{R},\mathbf{R}') = \sum_{A_M}^{\text{MM atoms}} Q'_{\mathrm{A_M}} \int \mathrm{d}\mathbf{r} \chi_a(\mathbf{r},\mathbf{R}) \times \frac{-1}{\left| \mathbf{R}'_{A_M} - \mathbf{r} \right|} \chi_b(\mathbf{r},\mathbf{R}). \quad (10)$$

Here, the indices a and b are used for the atomic orbitals (AOs) χ which is expressed in terms of Gaussian basis functions centered at one of the QM nuclei $\mathbf{R}$. In Hartree–Fock theory, one-electron reduced density matrix is defined by

$$D_{ab}(\mathbf{R}) = \sum_{i}^{\text{MO}} f_i^{\text{occ}} C_{ai}(\mathbf{R}) C^*_{bi}(\mathbf{R}). \quad (11)$$

The indices i and j are used for the molecular orbitals (MOs) ϕ described as the linear combination of atomic orbitals (LCAO) and C_{ai} and f_i^{occ} give the LCAO coefficient and the MO occupation number, respectively. In equation (10), $\mathbf{R}'_{A_M}$ and Q'_{A_M} are the Cartesian coordinate and electric charge of the A-th atom in M-th MM molecule. In equation (9), the switching function $f_{\mathrm{cut}}(R)$ is used to cut off the Coulomb interaction smoothly when the inter-molecular distance R, between the center of mass coordinate for the M-th MM molecule $\mathbf{R}'^{(c)}_{M}$ and the QM subsystem $\mathbf{R}^{(c)}$, is ranged between R_{in} and R_{out} [26]. Here, we have employed $f_{\mathrm{cut}}(R) = 1$ for $R < R_{\mathrm{in}}$, $f_{\mathrm{cut}}(R) = (R - R_{\mathrm{out}})^2(2R + R_{\mathrm{out}} - 3R_{\mathrm{in}})/(R_{\mathrm{out}} - R_{\mathrm{in}})^3$ for $R_{\mathrm{in}} < R < R_{\mathrm{out}}$ and $f_{\mathrm{cut}}(R) = 0$ for $R > R_{\mathrm{out}}$. The switching function is practically very useful for MD simulation, as it smooths the potential energy surface. Although it makes a small shift in the energy, the atomic forces can be evaluated appropriately. With the switching function, the total energy and total linear momentum of the system is always conserved and the total angular momentum is also conserved under free boundary condition. These conservation laws can be proved from the fact that the potential energy V_{total} is a single-valued differentiable function and that V_{total} is invariant under translation or rotation of the total system (Noether's theorem). The derivatives of equation (9) with respect to $\mathbf{R}_I$ and $\mathbf{R}'_A$ can be derived analytically and they are used in the force calculation with respect to QM–MM interaction. We note that the derivatives for switching function can be easily obtained in this way, because it does not include the electron coordinate $\mathbf{r}$. As another advantage of using the switching function, computation time can be reduced by neglecting small QM–MM interactions with respect to the MM molecules that are far enough from the QM subsystem.

The QM/MM method based on PT has some advantages and disadvantages similar to the conventional QM/MM-EE method. First, there is no need to assume the charge parameters for QM atoms, which may change their character when the QM subsystem undergoes an intramolecular rearrangement, especially if the bond breaking/forming processes are involved. Second, we need an *ad hoc* determination of VDW parameters between QM and MM atoms. However, this does not seem to be a serious problem usually in practice, since the VDW interaction is significant mainly at short distances and the use of only one set of VDW parameters is often adequate [12]. Third, the method is not appropriate whenever the electronic polarization of QM subsystem is very strong. This is not just because of the PT in which the isolated QM molecule is employed as the reference. When the QM–MM interaction is so strong and the electronic structure of QM subsystem can be drastically changed by strong polarization, the QM/MM Hamiltonian itself might be problematic since the MM region next to the QM subsystem can also become chemically active. Thus, we could say that the validity of using the PT is partially related to the QM/MM description itself.

In the QM/MM-PE method, we assume that the electronic structure of QM subsystem is frozen to that of the isolated molecule. In addition, we ignore the polarization of MM charges in response to the environment [27–31], since it will result in a more complicated implementation and increased computational cost. Therefore, in the language of inter-molecular forces, the ES forces are taken into account in terms of interaction between the electron distribution of isolate QM electron and MM charges, as in $V_{\mathrm{QM-MM}}$. The induction forces are totally neglected within the QM/MM-PE method. The Pauli exchange repulsion and the dispersion forces are modeled as a classical force field with *ad hoc* VDW parameters.

It may be useful to point out that there are not just good things to include polarization in the QM/MM method, however. One of the disadvantages of QM/MM-EE method is on the instability in the complete basis set limit. As an extreme example, polarization can be overestimated if we add some basis functions near the positively charged MM atoms to which the electrons will be attracted. Apparently, this is an unphysical artifact of in the QM/MM modeling, because the Pauli exchange repulsion between the QM and MM electrons are not considered explicitly. Thus, we must always be careful in the choice of basis set (being not too large!) when we use the QM/MM-EE method. In contrast,

the QM/MM-PE as well as QM/MM-ME methods have a good behavior in the complete basis set limit because the *ab initio* calculation is done only for the isolated QM subsystem.

2.2 The multiple-time-scale method

To make use of MTS method, the propagator (time evolution operator) should be factorized into the components that change slowly and rapidly. Here, we consider the changes in the potential, equation (7), when the QM coordinates $\mathbf{R}$ move slowly and the MM coordinates $\mathbf{R}'$ rapidly. As can be easily expected, the QM potential $V_{\mathrm{QM}}(\mathbf{R})$ shift more slowly than the MM potential $V_{\mathrm{MM}}(\mathbf{R}')$ and the QM–MM interaction $V_{\mathrm{QM-MM}}(\mathbf{R}, \mathbf{R}')$ in this case. Therefore, we can employ a larger step size Δt for the former and a smaller step size δt for the latter. Applying the time-reversible reference system propagator algorithm [16,17] (RESPA) to this case, the propagator is approximated by

$$\exp[i(L_{\mathrm{total}} + L_{\mathrm{bath}})\Delta t] \approx \exp\left(iL_{\mathbf{P}}\frac{\Delta t}{2}\right)\exp(iL_B\Delta t) \times \exp\left(iL_{\mathbf{P}}\frac{\Delta t}{2}\right), \tag{12}$$

where

$$\exp(iL_B\Delta t) = \left[\exp\left(iL_{\mathrm{bath}}\frac{\delta t}{2}\right)\exp\left(iL_{\mathbf{P},\mathbf{P}'}\frac{\delta t}{2}\right)\exp(iL_{\mathbf{R}}\delta t) \times \exp(iL_{\mathbf{R}'}\delta t)\exp\left(iL_{\mathbf{P},\mathbf{P}'}\frac{\delta t}{2}\right)\exp\left(iL_{\mathrm{bath}}\frac{\delta t}{2}\right)\right]^{N_{\mathrm{ref}}}. \tag{13}$$

where the integer N_{ref} is the ratio between the step sizes, $N_{\mathrm{ref}} = \Delta t/\delta t$. In equation (13), L_{total} is the Liouville operator related to Hamiltonian of the total QM/MM system, H_{total} and L_{bath} is the Liouville operator for the bath variables (thermostats). The Liouville operators $L_{\mathbf{R}}$, $L_{\mathbf{P}}$, $L_{\mathbf{R}'}$, $L_{\mathbf{P},\mathbf{P}'}$ are related to the terms T_{QM}, V_{QM}, T_{MM} and $V_{\mathrm{MM}} + V_{\mathrm{QM-MM}}$, respectively.

The MTS algorithm for QM/MM-PE model is shown schematically in figure 1. (The same algorithm in figure 1 can be also used in the QM/MM-ME model simply by setting the QM–MM interaction identical to the MM–MM interaction. In this case, the calculation of one-electron integrals and the density matrix in steps (c) and (j) are unnecessary.) The MD cycle consists of a double MD cycle containing parts A, B and C as shown in the figure. The outer and inner MD cycles are respectively repeated $N_{\mathrm{step}}^{\mathrm{QM}}$ and N_{ref} times, so the number of MD time steps is $N_{\mathrm{step}} = N_{\mathrm{step}}^{\mathrm{QM}} \times N_{\mathrm{ref}}$ in total. In the outer MD cycle (parts A and C), the QM momenta $\mathbf{P}$ are updated according to the gradients $\partial V_{\mathrm{QM}}/\partial\mathbf{R}$ which are obtained from an ordinary *ab initio* calculation of the isolated QM system. In the inner MD cycle (part B), all the QM and MM coordinates and momenta, $(\mathbf{R}, \mathbf{P}, \mathbf{R}', \mathbf{P}')$, are updated according to the gradients $\partial(V_{\mathrm{MM}} + V_{\mathrm{QM-MM}})/\partial\mathbf{R}$ and $\partial(V_{\mathrm{MM}} + V_{\mathrm{QM-MM}})/\partial\mathbf{R}'$, respectively. We note that the $N_{\mathrm{ref}} = 1$ case is exactly the same with the conventional velocity Verlet method except that the density matrix extrapolation is used (see below). The Nosé–Hoover thermostats can also be updated in the inner MD cycle in the appropriate manner [16,17].

In the QM/MM-PE model, the algorithm becomes slightly complicated because it is necessary to compute the density matrix $\mathbf{D}$ in equation (11) and the derivatives $\partial\mathbf{D}/\partial\mathbf{R}$ to obtain the QM–MM interaction $V_{\mathrm{QM-MM}}$ and the gradients $\partial V_{\mathrm{QM-MM}}/\partial\mathbf{R}$ and $\partial V_{\mathrm{QM-MM}}/\partial\mathbf{R}'$. Although they should be evaluated every δt in the inner MD cycle where the QM coordinates $\mathbf{R}$ are updated, it is clearly inconvenient to solve the SCF and coupled-perturbed Hartree–Fock (CPHF) equations to obtain them every δt. So, as a practical way, in the outer MD cycle the SCF and CPHF equations are solved to obtain $\mathbf{D}$ accurately by the interval of Δt, while in the inner MD cycle we make a short-cut by the extrapolation of $\mathbf{D}$ and $\partial\mathbf{D}/\partial\mathbf{R}$ using the fifth-order Gear predictor–corrector algorithm. The implementation is shown in detail in the next section. It is shown that this method works reasonably well as to the energy conservation (although the time reversibility is very slightly destroyed).

At this stage, the parameter N_{ref} can be chosen arbitrarily as long as the step sizes Δt and δt are chosen appropriately. Now, as mentioned in Section 1, we introduce the scaling factor λ with respect to the masses of MM atoms, $\lambda M'$, to make the MD sampling more efficient. When the parameter λ is chosen to be smaller than 1, the motion of MM atoms will be faster than the reality. In this case, we should select a smaller δt (larger N_{ref}) so as to follow the fast motion. As a guide, it may be reasonable to choose $\lambda \approx 1/N_{\mathrm{ref}}^2$ considering an application to simple harmonic oscillator; A reasonable choice is to take the step size $\delta t = \Delta t/N_{\mathrm{ref}}$ so that the ratio between δt and the oscillation period τ does not change with λ, using the fact that τ is proportional to the square root of the mass, $\sqrt{\lambda M'}$. Of course, the best choice of the parameters λ and N_{ref} generally depends on the system of interest. For example, when we consider a small and stiff molecule in liquid, N_{ref} could be taken to be very large, because the sampling on the solvation structures would be much more important than the intra-molecular deformation within the solute. The choice will also depend on the ratio of computational costs in the inner and outer MD cycles. However, the MTS algorithm is theoretically exact in that the parameters λ and N_{ref} does not have influence on the thermodynamic averages as long as the sampling is sufficient.

As can be seen in figure 1, part B is calculated every step, while parts A and C are calculated once in N_{ref} steps. Thus, the calculation time per unit step is given by the sum $t_{\mathrm{total}} = t_{\mathrm{A}}/N_{\mathrm{ref}} + t_{\mathrm{B}} + t_{\mathrm{C}}/N_{\mathrm{ref}}$ where t_{A}, t_{B} and t_{C} are the calculation time of parts A, B and C, respectively. Here, most of the time-consuming calculations are concentrated in the outer MD cycle, particularly in part C. This part involves the calculation of two-electron integrals and the solution of SCF for the QM subsystem, as in the ordinary

THE MD CYCLE

do istep = 1, N_step_QM
  PART A
  do jstep = 1, N_ref
    PART B
  end do
  PART C
end do

PART A

(a) update QM momentum: $\mathbf{P} = \mathbf{P} - \partial V_{\mathrm{QM}}/\partial \mathbf{R} \times \Delta t/2$

PART B

(d) update thermostats ($\times \delta t/2$)
(e) update QM momentum: $\mathbf{P} = \mathbf{P} - \partial (V_{\mathrm{MM}} + V_{\mathrm{QM-MM}})/\partial \mathbf{R} \times \delta t/2$
(f) update MM momentum: $\mathbf{P}' = \mathbf{P}' - \partial (V_{\mathrm{MM}} + V_{\mathrm{QM-MM}})/\partial \mathbf{R}' \times \delta t/2$
(g) update QM position: $\mathbf{R} = \mathbf{R} + \mathbf{P}/M \times \delta t$
(h) update MM position: $\mathbf{R}' = \mathbf{R}' + \mathbf{P}'/(M'\lambda) \times \delta t$
(i) calculation of MM intra-molecular and MM-MM inter-molecular forces
— (i-1) calculate V_{MM} and $\partial V_{\mathrm{MM}}/\partial \mathbf{R}'$
(j) calculation of QM-MM interaction
— (j-1) calculate 1-electron AO integrals of Eq.(5) type (+ QM-MM LJ interaction)
— (j-2) get $\mathbf{D}$ and $\partial \mathbf{D}/\partial \mathbf{R}$ (using extrapolation)
— (j-3) use $\mathbf{D}$ and $\partial \mathbf{D}/\partial \mathbf{R}$ to calculate $V_{\mathrm{QM-MM}}$, $\partial V_{\mathrm{QM-MM}}/\partial \mathbf{R}'$, and $\partial V_{\mathrm{QM-MM}}/\partial \mathbf{R}$
(f) update MM momentum: $\mathbf{P}' = \mathbf{P}' - \partial (V_{\mathrm{MM}} + V_{\mathrm{QM-MM}})/\partial \mathbf{R}' \times \delta t/2$
(e) update QM momentum: $\mathbf{P} = \mathbf{P} - \partial (V_{\mathrm{MM}} + V_{\mathrm{QM-MM}})/\partial \mathbf{R} \times \delta t/2$
(d) update thermostats ($\times \delta t/2$)

PART C

(b) Electronic structure calculation of isolated QM subsystem
— (b-1) calculate 1- and 2-electron AO integrals
— (b-2) solve self-consistent field (electron density matrix $\mathbf{D}$); calculate $\partial \mathbf{D}/\partial \mathbf{R}$
— (b-3) calculate V_{QM} and $\partial V_{\mathrm{QM}}/\partial \mathbf{R}$
(c) get $\mathbf{D}$ and $\partial \mathbf{D}/\partial \mathbf{R}$
(a) update QM momentum: $\mathbf{P} = \mathbf{P} - \partial V_{\mathrm{QM}}/\partial \mathbf{R} \times \Delta t/2$

Figure 1. A MTS algorithm for QM/MM-PE MD.

ab initio method. In addition, the CPHF equation is solved to obtain the derivatives of the density matrix $\mathbf{D}$ (or equivalently, the derivatives of the LCAO coefficients $\mathbf{C}$). So, part C requires the computation about $t_{\mathrm{A}} \propto N_{\mathrm{basis}}^{x}$ where N_{basis} is the number of basis functions employed and $x \approx 4$ in Hartree–Fock theory. The rest of the time-consuming calculation is in part B, especially the QM–MM interaction and MM–MM interaction. Usually, the former one is more expensive since it involves the summation of one-electron Coulomb integrals in equation (9). Then, part B requires the computation about $t_{\mathrm{B}} \propto N_{\mathrm{MM}} N_{\mathrm{basis}}^{2}$ where N_{MM} is the number of MM molecules. However, part B can be computed efficiently using parallel algorithm, because the summation can be divided into the contributions from each MM molecules and they can be calculated individually on different processors. As the number of MM molecules are usually larger than the number of processors N_{procs}, the computation of part B will become $t_{\mathrm{B}} \propto N_{\mathrm{MM}} N_{\mathrm{basis}}^{2}/N_{\mathrm{procs}}$ in the ideal situation. The cost on part A is negligibly small compared with other parts.

From the above discussion, the total time is approximately

$$t_{\mathrm{total}} \propto \max\left(N_{\mathrm{basis}}^{x}/N_{\mathrm{ref}}, N_{\mathrm{MM}} N_{\mathrm{basis}}^{2}/N_{\mathrm{procs}}\right) \qquad (14)$$

in the QM/MM-PE combined with MTS algorithm. This relation implies that by employing a large N_{ref} value, this

method would be more useful as N_{basis} becomes large, i.e. as the size of QM subsystem becomes large.

3. Results and discussion

3.1 Water dimer

As a rough check on the accuracy of QM/MM-PE approximation, the results for the equilibrium (geometry optimized) structure of water dimer are shown in table 1. Here, the binding energy V and the interatomic distances $R(\text{O}\cdots\text{O})$, $R(\text{O}\cdots\text{H}^*)$ and $R(\text{O—H}^*)$ (where H^* is the H atom mediating the hydrogen bond), are listed with respect to the QM/MM-PE, conventional QM/MM-EE, QM/MM-ME, full QM and full MM methods. We also show the results in the QM/MM-EE method with charge scaling where all the MM charges are multiplied by 0.84 in order to correct the overestimation of QM–MM interaction [23]. The abbreviation "a" and "d" indicate which of the QM and MM water molecules is hydrogen acceptor(a) or hydrogen donor(d). We can see that the results of QM/MM-PE method are in reasonable agreement with those of other methods. The difference between the QM/MM-PE method and the conventional QM/MM-EE method is due to the polarization effect. For instance, the QM/MM-PE gives the binding energy smaller by about 7–10% and the O···O distance slightly larger compared with the conventional QM/MM-EE. Both the QM/MM and full MM methods tend to slightly underestimate O···O distance compared with full QM results within Hartree–Fock or the second-order Møller–Plesset PT (MP2).

3.2 Liquid water: the MD method

MD simulation of liquid water has been carried out based on the QM/MM-PE method. The system contained one QM molecule (HF/6-31G(d,p)) and 255 MM molecules (SPC/F2) in a cubic box with the side length 19.7 Å, which corresponds to molar density of 1 g/cm^3 and the periodic boundary condition is applied. The temperature is controlled to 300 K using the massive Nosé–Hoover chain (MNHC) thermostat technique which generates the canonical ensemble for ergodic systems. The thermostats are attached respectively to the QM system and the MM system. According to the minimum image convention, the QM–MM interaction and short-range Lennard–Jones interactions were smoothly cut off between 8.3 and 9.8 Å applying the switching function method. For the MM–MM molecular pairs, the long-range ES interaction was treated by Ewald method. The QM/MM MD simulation was performed in two ways by the single- and MTS methods, $N_{\text{ref}} = 1$ and $N_{\text{ref}} = 100$. After the system has been equilibrated, the former run (run #1) has been done for $N_{\text{step}} = N_{\text{step}}^{\text{QM}} = 40,000$ steps with the step size $\Delta t = \delta t = 0.1$ fs, while the latter run (run #2) for $N_{\text{step}} = 4,000,000$ steps ($N_{\text{step}}^{\text{QM}} = 40,000$ steps) with $\Delta t = 0.1$ fs and $\delta t = 0.001$ fs where the MM masses $\lambda M'$ were scaled by $\lambda = 1/10,000$. In the test calculations, we also performed QM/MM MD simulations with $N_{\text{ref}} = 10$ ($\Delta t = 0.1$ fs and $\delta t = 0.01$ fsusing mass scaling $\lambda = 1/100$).

3.3 Liquid water: test calculations

Before starting the QM/MM simulation, the MTS algorithm has been checked by the "MM/MM" (full MM) simulation. This simulation are done exactly in the same way as run #2, except that the QM molecule is replaced by an MM molecule labeled as MM^*. Figure 2 shows the oxygen–oxygen radial distribution between MM–MM and MM^*–MM molecules. We can see that the agreement of these two distributions is very good. A tiny statistical error in MM^*–MM distribution is found simply because there is only one MM^* molecule in the system. Although it is not shown explicitly in the figure, the conventional MD simulation (for 256 MM molecules with usual single-time-scale velocity Verlet algorithm) gave the same distribution almost without error. This result guarantees the fact that the MTS algorithm with mass scaling generates the correct canonical ensemble of ergodic systems.

Next, we must confirm that the QM/MM MD simulation is running correctly when the MTS algorithm is used. So, we check the numerical accuracy in terms of the energy conservation. In the present case, the QM–MM ES interaction $V_{\text{QM-MM}}^{\text{es}}$ in equation (9) and its MM

Table 1. Binding energy and interatomic distances of water dimer.

Method	*QM/MM*	*ΔV (kcal/mol)*	*R (O—O) (Å)*	*R(O…H*) (Å)*	*R (O—H*) (Å)*
QM/MM-PE	a/d	6.7	2.74	1.78	0.96
	d/a	6.9	2.76	1.74	1.02
QM/MM-ME	a/d	6.2	2.77	1.81	0.96
	d/a	6.7	2.74	1.72	1.02
QM/MM-EE	a/d	7.5	2.70	1.74	0.96
	d/a	7.4	2.73	1.71	1.02
QM/MM-EE (0.84)†	a/d	5.9	2.75	1.79	0.95
	d/a	5.9	2.78	1.76	1.02
QM/QM (HF/6-31G(d,p))		5.5	2.98	2.04	0.95
QM/QM (MP2/6-31G(d,p))		7.0	2.91	1.97	0.97
MM/MM		7.0	2.73	1.71	1.02

† The electric charge is scaled by 0.84 only for QM–MM interaction. See Ref. [23].

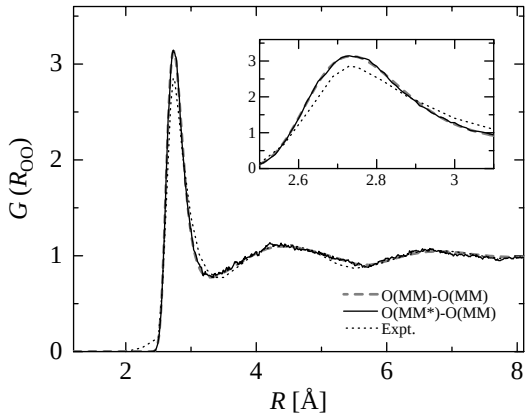

Figure 2. Oxygen–oxygen radial distributions of liquid water by a MTS "MM/MM" MD simulation with mass scaling ($N_{\mathrm{ref}} = 100$, $\lambda = 1/10,000$) The statistical average is taken for $N_{\mathrm{step}} = 4,000,000$ ($N_{\mathrm{step}}^{\mathrm{MM}'} = 40,000$). The solid and dashed lines are the distribution between MM–MM and MM–MM* pairs, respectively. The dotted line is from the X-ray diffraction experiment [34].

gradient $\partial V_{\mathrm{QM-MM}}^{\mathrm{es}}/\partial \mathbf{R}'$ include the density matrix $\mathbf{D}$ in equation (11). The gradient $\partial V_{\mathrm{QM-MM}}^{\mathrm{es}}/\partial \mathbf{R}$ includes the density matrix gradient $\mathbf{D}' \equiv \partial \mathbf{D}/\partial \mathbf{R}$ as well as the density matrix $\mathbf{D}$. Thus, all the matrix elements of $\mathbf{D}$ and $\mathbf{D}'$ have to be extrapolated in the inner MD cycle and therefore the accuracy may depend on the extrapolation method, as introduced in figure 1. Note that the main contribution to QM–MM interaction energy (forces) is from the $V_{\mathrm{QM-MM}}^{\mathrm{es}}$ itself, not the shift in $\mathbf{D}$ ($\mathbf{D}$ and $\mathbf{D}'$) values. However, this slight shift may have subtle effect on the numerical accuracy, since time reversibility is slightly destroyed.

In order to estimate the error, we have tested several treatments on the evaluation of $\mathbf{D}$ and $\mathbf{D}'$. The first approximation (case A) is to assume that $\mathbf{D}$ and $\mathbf{D}'$ are constant values within the inner MD cycle, i.e. they are kept to the latest values updated at the outer MD cycle. This is a reasonable choice because we can anticipate that the density matrix $\mathbf{D} = \mathbf{D}(\mathbf{R})$ will not vary so much during the time of Δt as long as the QM atoms move slowly and the displacements in $\mathbf{R}$ are very small during Δt. For further correction, we have considered the second approximation (case B) in which the differential equation

$$\dot{\mathbf{D}}(t) = \frac{\partial \mathbf{D}}{\partial \mathbf{R}} \cdot \dot{\mathbf{R}}(t) \tag{15}$$

is solved by Gear predictor–corrector algorithm. Using the coefficients $\mathbf{D}(0), \dot{\mathbf{D}}(0), \ldots$ obtained every Δt step, the extrapolated density matrix, $\mathbf{D}_{\mathrm{p}}(t)$, is calculated by the fifth-order Taylor expansion at the previous step [32],

$$\mathbf{D}_{\mathrm{p}}(t) = \mathbf{D}(0) + \dot{\mathbf{D}}(0)t + \ddot{\mathbf{D}}(0)\frac{t^2}{2} + \cdots \quad \text{for } 0 < t < \Delta t. \tag{16}$$

The third approximation (case C) uses a similar extrapolation method for both $\mathbf{D}_{\mathrm{p}}$ and $\mathbf{D}'_{\mathrm{p}}$. In this case, however, it is not simple to solve the differential equation for $\mathbf{D}'$,

$$\dot{\mathbf{D}}'(t) = \mathbf{X}(t), \tag{17}$$

because the second derivatives of $\mathbf{D}$ is included in $\mathbf{X}(t)$. Therefore, we have used an approximation derived from the Gear method

$$\mathbf{X}(t) = \dot{\mathbf{D}}'_{\mathrm{p}}(t) + \frac{\mathbf{D}'(t) - \mathbf{D}'_{\mathrm{p}}(t)}{a_0 \Delta t} \tag{18}$$

where $a_0 = 95/288$ is the zero-order coefficient of the fifth-order Gear predictor–corrector method and $\mathbf{D}'_{\mathrm{p}}(t)$ and $\dot{\mathbf{D}}'_{\mathrm{p}}(t)$ are the predictor values obtained from the Taylor expansion similar to equation (16).

In figure 3, the energy E of the extended system (the QM/MM system + MNHC thermostats) is plotted for cases A, B and C in the QM/MM MD simulation of liquid water. We can see that case C is much improved from cases A and B both on the drift and noise errors. Figure 3 also shows the same plot for a single-time-scale velocity Verlet QM/MM MD simulation which does not use the extrapolation. The small noise error remained in case C is almost the same with that in the conventional velocity Verlet method. The amplitude of the noise error, which is less than 10^{-4} a.u., does not depend on the parameter N_{ref} in the MTS algorithm with the mass scaling under the conditions $\delta t = \Delta t/N_{\mathrm{ref}}$ and $\lambda = 1/N_{\mathrm{ref}}^2$.

We have seen that the extrapolation method (case C) is successful for the QM/MM-PE scheme. However, we can expect that any type of extrapolation method will *not* work well when the QM/MM-EE scheme, which treats the polarization effect, is used in the MTS MD algorithm. In the QM/MM-EE scheme, the density matrix $\mathbf{D}$ will depend the MM charge distribution given by the MM configuration, $\mathbf{D} = \mathbf{D}(\mathbf{R}, \mathbf{R}')$. This means that the values of $\mathbf{D}$ and $\mathbf{D}'$ will change as the MM molecules move. Especially when the MM molecules move rapidly by mass scaling, the extrapolation for $\mathbf{D}$ and $\mathbf{D}'$ would be difficult as the number of inner MD cycle N_{ref} is increased. On the other hand, the extrapolation itself is unnecessary for the QM/MM-ME scheme, because the QM–MM interaction does not depend on the density matrix $\mathbf{D}$.

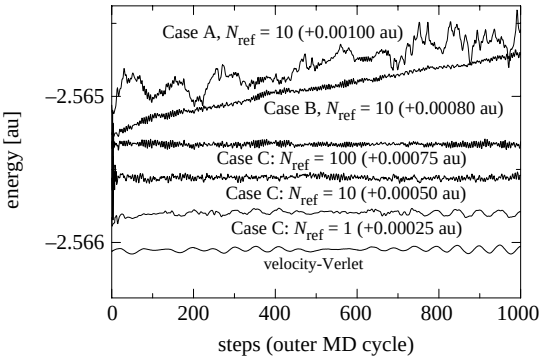

Figure 3. The total energies E in the QM/MM MD simulations as a function of time steps of outer MD cycle ($N_{\mathrm{step}}^{\mathrm{QM}}$). Several cases in extrapolation methods (cases A and B with $N_{\mathrm{ref}} = 10$ and case C with $N_{\mathrm{ref}} = 1, 10, 100$) are compared with the velocity Verlet method.

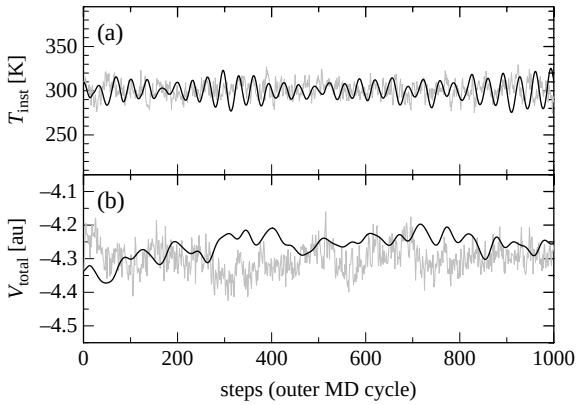

Figure 4. (a) Instantaneous temperature T_{inst} and (b) total potential V_{total} in the QM/MM MD simulations of liquid water as a function of time steps of outer MD cycle (N_{step}^{QM}). The thick and thin lines give the results of single-time scale simulation ($N_{ref} = 1$, run #1) and the MTS simulation ($N_{ref} = 100$, run #2), respectively.

3.4 Liquid water: MD simulation

Now, we move on to the QM/MM MD simulation runs, #1 and #2. First, the instantaneous temperature T_{inst} and the total potential V_{total} are shown in figure 4. We can see that T_{inst} is controlled appropriately to the temperature $T = 300$ K. At this temperature, the potential V_{total} fluctuates by the magnitude of about 0.1 a.u. Comparing figures 3 and 4, we note that the error in the energy conservation is much smaller than the potential fluctuation. The average and the extent of fluctuation is the same between the two simulation runs. However, run #2 moves much faster and it covers larger configuration space for MM molecules.

Figure 5(a)–(c), respectively show the oxygen–hydrogen, hydrogen–hydrogen and oxygen–oxygen radial distributions obtained from the QM/MM MD simulation using the MTS algorithm, run #2. We can see that the QM–MM interatomic distributions are in reasonable agreement with neutron and X-ray diffraction experiments [33,34]. Somewhat surprisingly, the QM–MM distributions are very similar to the MM–MM counterpart for oxygen pair and hydrogen pair, even though the QM and MM molecules are modeled quite differently. However, the distributions

Table 2. Peak positions and intensities in radial distributions of liquid water[†].

	$G(R_{OO})$		$G(R_{OH})$		$G(R_{HH})$	
Peaks	*1st*	*2nd*	*1st*	*2nd*	*1st*	*2nd*
QM/MM-PE	2.8	4.5	1.8‡	3.2‡	2.3	3.8
	3.1	1.1	1.4‡	1.5‡	1.4	1.1
Full MM (SPC/F2)	2.7	4.5	1.7	3.2	2.4	3.9
	3.1	1.1	1.6	1.5	1.4	1.1
Full QM (DFT)[¶]	2.7	4.4				
	3.5	1.5				
Experiment[§]	2.7	4.5	1.8	3.2	2.3	3.8
	2.8	1.1	1.1	1.5	1.3	1.1

[†] Top values are the peak position in (Å), the bottom values are the intensity.
[‡] Average values between the O(QM)—H(MM) and H(QM)—O(MM) radial distributions.
[¶] *Ab initio* DFT calculation within BLYP/TZV2P level at 292 K (Ref. [40]).
[§] Experimental values from Refs. [33,34].

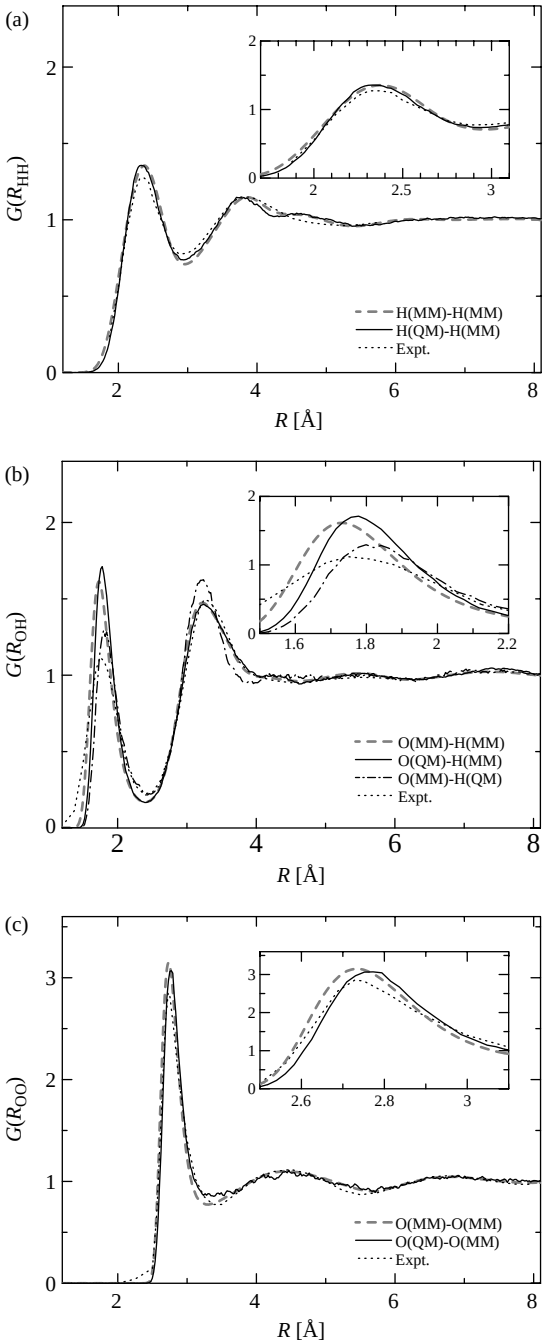

Figure 5. Radial distributions of liquid water obtained from QM/MM MD simulation using MTS method (run #2). The result of (a) hydrogen–hydrogen, (b) oxygen–hydrogen and (c) oxygen–oxygen distributions. The solid line, dotted-dashed line and dashed line are the distribution for QM–MM, MM–QM and MM–MM pairs, respectively. The dotted lines in (a) and (b) are from the neutron diffraction experiment [33] and the dotted line in (c) is from the X-ray diffraction experiment [34].

between O(QM)—H(MM) and between O(MM)—H(QM) slightly differs from each other and also slightly differs from that of O(MM)—H(MM). Compared with the O(MM)—H(MM) distribution, the O(QM)—H(MM) distribution has the first peak a little higher and the second peak with a similar height, while the O(MM)—H(QM) distribution has the first peak a little lower and the second peak a little higher. These results can also be compared

with recent *ab initio* calculations based on density functional theory (DFT) [35–40], but the DFT results are somewhat dependent on the exchange-correlation functional, especially on the peak intensities [39,40]. A comparison of the peak positions and intensities of radial distributions are summarized in table 2.

Owing to the efficiency of MTS algorithm, the distribution profiles for MM*–MM pair in figure 2 and QM–MM pair in figure 5 are found reasonably accurate. To see this more clearly, a QM/MM simulation has been undertaken with the single-time-scale algorithm (run #1). Figure 6 shows the oxygen–oxygen radial distribution obtained from this simulation after the same step length, $N_{\mathrm{step}}^{\mathrm{QM}} = 40,000$. As for QM–MM pair, the peak positions and peak intensities remain unclear compared with figure 5(c). This is apparently due to the shortage of configuration sampling in the single-time-scale algorithm.

We can compare the radial distributions in figure 5 with conventional QM/MM-EE simulations by Tu and Laaksonen [22,23] and Xenides *et al.* [24] However, their data might be only qualitatively correct especially as to the peak intensities, because the statistical error is inevitable within a short simulation run at least in our experience. However, we think that the result would not change drastically since the QM/MM-PE results are already in good agreement with experiment, at least for the radial distributions.

Figure 7 shows the QM intra-molecular bond length distribution obtained from MTS QM/MM MD simulation (run #2). It is interesting that the calculated average bond lengths in the liquid phase water, $R_{\mathrm{OH}} = 0.954\,\mathrm{\AA}$ and $R_{\mathrm{HH}} = 1.524\,\mathrm{\AA}$, are slightly longer than the equilibrium bond length of the gas phase water, $R_{\mathrm{OH}} = 0.943\,\mathrm{\AA}$ and $R_{\mathrm{HH}} = 1.506\,\mathrm{\AA}$. This figure shows that the statistics in this simulation is also sufficient as to the QM intra-molecular degrees of freedom by $N_{\mathrm{step}}^{\mathrm{QM}} = 40,000$ steps ($N_{\mathrm{step}} = 4,000,000$ steps).

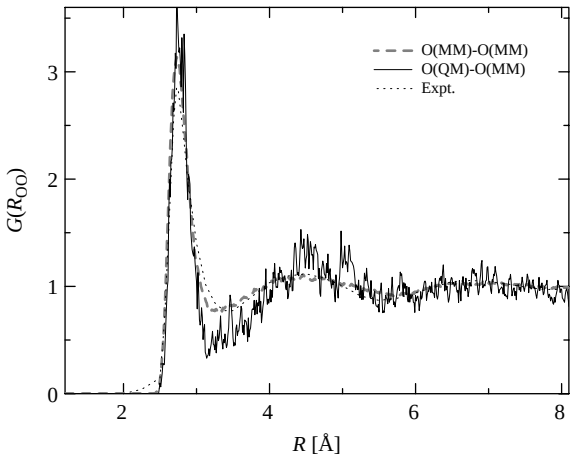

Figure 6. Oxygen–oxygen radial distributions of liquid water obtained from QM/MM MD simulation using single-time-scale method (run #1). The solid and dashed lines are the distribution between QM–MM and MM–MM pairs, respectively. The dotted line is from the X-ray diffraction experiment [34].

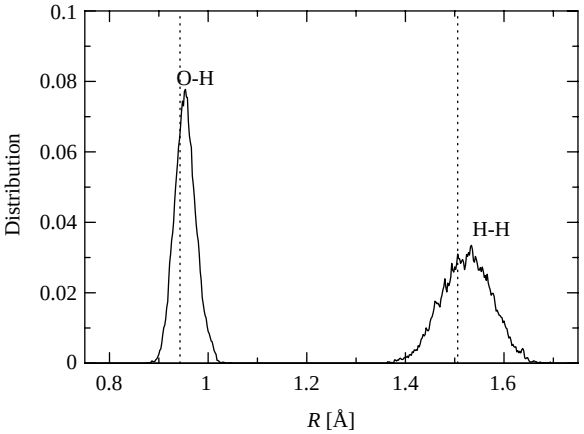

Figure 7. Intramolecular O—H and H—H bond length distributions of the QM molecule, obtained from the QM/MM MD simulation of liquid water (run #2). The dotted lines are the equilibrium bond length of isolated QM water molecule.

In table 3, we list the calculation time in the MD simulation of liquid water. The total time is composed of the parts spent on the calculation of QM, MM and QM–MM potential/forces and others (update of coordinates, etc). For parallel computation, the communications among processors are also necessary to share the information of atomic coordinates, forces and potentials. In the case of the single-time-step method ($N_{\mathrm{ref}} = 1$), the calculation time of QM/MM-PE is about the same as that of the conventional QM/MM-EE method. In the $N_{\mathrm{ref}} = 1$ case, the main difference between the QM/MM-EE and QM/MM-PE is whether the SCF calculation is done including the MM charges or not, but this seems to have little influence on the calculation time in the present system. Here, most of the time is spent on the potential/force calculations, particularly the QM and QM–MM parts. First, using the MTS algorithm, the calculation of the QM part is greatly reduced. Compared with the $N_{\mathrm{ref}} = 1$ case, the calculation time for the QM part is damped to 1% in the $N_{\mathrm{ref}} = 100$ case, because this part is calculated only once in 100 steps. Now, the total time is reduced to less than 1/3. Second, by parallel computation, the calculation of the QM–MM part is greatly reduced. As the parallel efficiency is almost 100%, the calculation time is reduced to about 1/16 using 16 processors. However, as the number of processors N_{procs} increases, the total time tends to depend on the time of communication. As a result, the total time in the $N_{\mathrm{procs}} = 16$ case is 0.088 s/step. This is about 22 times faster than the conventional QM/MM code using single-time-step algorithm.

3.5 Liquid water: weighted sampling

So far, we have neglected the electronic polarization of QM molecule within the QM/MM-PE method. Now, it would be useful discuss how the method can be improved for the polarization effect. For example, it is well known the dipole moment of water is larger in liquid phase than in the gas phase. This is mainly due to the shift in

Table 3. Calculation time per MD step in seconds[‡].

Number of CPUs[¶]	$N_{ref} = 1$ *1*	$N_{ref} = 100$ *1*	*2*	*4*	*8*	*16*
QM/MM-PE						
QM potential and force	1.297	0.013	0.013	0.013	0.013	0.013
MM potential and force	0.101	0.101	0.051	0.026	0.013	0.007
QM–MM potential and force	0.486	0.486	0.237	0.118	0.064	0.041
Communication	0.003	0.003	0.005	0.010	0.019	0.023
Others	0.029	0.024	0.013	0.008	0.005	0.003
Total time	1.915	0.626	0.319	0.175	0.114	0.088
Parallel efficiency[†]			98.2%	89.5%	68.6%	44.6%
QM/MM-EE						
Total time	1.921					

[†] Ratio of total time in single and multiple processors divided by the number of multiple processors.
[‡] The original *ab initio* program code "FVOPT" was linked to the MD program.
[¶] Pentium IV (3.0 GHz) linux PC machine with Gigabit Ethernet is used.

the electronic structure induced by the solvation structure rather than the geometrical change of the QM molecule itself. (This can be also seen in figure 7 that the difference of the average intra-molecular bond length and bond angle is not so large between vapor and liquid water.) There are several ways to include the polarization effect. The first way is to apply the second-order PT, which has been introduced in Section 2. However, implementation of the method will become more complicated, because the matrix elements between electronic excitations (virtual molecular orbitals within the QM isolated subsystem) are necessary. The second way, which may be more convenient, is to first run the QM/MM-PE MD simulation and then do the average calculation by more accurate method (by QM/MM-EE or *ab initio* calculations of clusters including solvent) on the selected configurations randomly sampled from the MD trajectory. As a modification of the latter approach, the third way, which we recommend here, is to use the weighting sampling technique. This is based on a rigorous thermodynamic relation with respect to the canonical ensemble average of quantity A,

$$\langle A \rangle \equiv \frac{\int d\mathbf{\Gamma}\, A(\mathbf{\Gamma}) \exp(-\beta V(\mathbf{\Gamma}))}{\int d\mathbf{\Gamma} \exp(-\beta V(\mathbf{\Gamma}))} = \frac{\langle A \exp(-\beta(V - V_0))\rangle_0}{\langle \exp(-\beta(V - V_0))\rangle_0} \tag{19}$$

where $\mathbf{\Gamma} = (\mathbf{R}, \mathbf{R}')$ is the configuration space. Here, we set V and V_0 to be the potential in QM/MM-PE and the potential in improved accuracy, respectively. The numerator and the denominator in equation (19) are calculated by the average

$$\langle B \rangle_0 = \frac{\int d\mathbf{\Gamma}\, B(\mathbf{\Gamma}) \exp(-\beta V_0(\mathbf{\Gamma}))}{\int d\mathbf{\Gamma} \exp(-\beta V_0(\mathbf{\Gamma}))} \tag{20}$$

using the MD trajectory within QM/MM-PE approximation. In other words, after A, V and V_0 are calculated at molecular configurations sampled randomly from the MD trajectory, $\langle A \rangle$ is calculated as the ratio between the average of $\exp(-\beta(V - V_0))$ and the average of $A \exp(-\beta(V - V_0))$. This method would work if the potential surfaces $V(\mathbf{\Gamma})$ and $V_0(\mathbf{\Gamma})$ are close to each other. The Boltzmann distribution on the $V(\mathbf{\Gamma})$ surface must have large overlap with that on the $V_0(\mathbf{\Gamma})$ surface since the sampling is done based on the latter distribution as in equation (20).

In order to calculate the molecular dipole moment in liquid water by the weighted sampling method, we set V and $A = \mu$ to be the potential energy and dipole moment, respectively of the QM/MM-EE model based on the MP2 PT and Sadlej-pVTZ basis set (MP2/Sadlej-pVTZ)[41], while V_0 is the potential energy used in the MD simulation (QM/MM-PE model based on HF/6-31G(d,p)). Note that the electron correlation and basis set should be selected properly to obtain accurate dipole moment. If fact, the calculated dipole moment of an isolated water molecule (in geometry optimized structure) is 2.11 Debye for HF/6-31G(d,p) and 1.88 Debye for MP2/Sadlej-pVTZ, whereas the experimental dipole moment of an isolated water molecule is 1.855 Debye [42]. In figure 8, the accumulative average $\langle \mu \rangle$ of QM water molecule in liquid phase (the average molecular dipole moment of liquid water at 300 K) is plotted based on the weighted sampling method as a function of randomly sampled configurations from the QM/MM-PE MD simulation. The $\langle \mu \rangle$ value is converged to 2.67 Debye. This is very good agreement with the result 2.64 ± 0.13 by Tu and Laaksonen [23] which

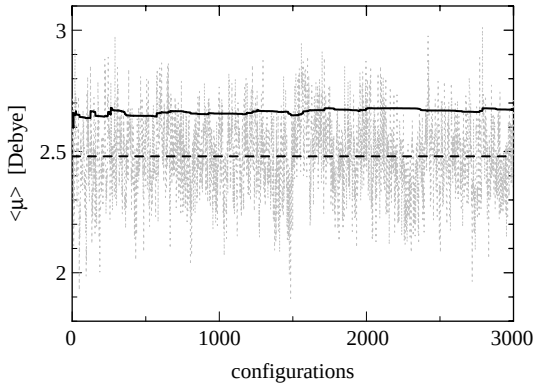

Figure 8. The dipole moments in the weighted sampling (dotted line) and its accumulative average (solid line). The dashed line is the simple average without weighting (the second method, see text).

Table 4. Average molecular dipole moment $\langle\mu\rangle$.

	$\langle\mu\rangle$ *(Debye)*	
Methods	*Vapor*	*Liquid*
Full MM	2.31[†]	
Ab initio	2.15[‡], 1.88[¶]	2.6–3.2[§]
QM/MM-EE[‖]		2.64 ± 0.13
QM/MM-PE + correction[#]	1.88	2.67
Experiment	1.858[**]	2.9(6)[††]

[†] Modified flexible SPC/F.
[‡] HF/6-31G(d,p).
[¶] MP2/Sadlej-pVTZ.
[§] Refs. [37,43–46].
[‖] Ref. [23].
[#] This work.
[**] Ref. [42].
[††] Ref. [47].

employs the *ab initio* QM/MM-EE approach and theoretical results based on *ab initio* calculations (2.6–2.7 Debye) [43–45], while it is somewhat smaller than the values of recent *ab initio* Car–Parrinello MD simulation [37,46] and experiment [47] (2.9–3.2 Debye). It is worthwhile to note that $\langle\mu\rangle_0$ (the second method in the previous paragraph, the simple average of figure 8) is found 2.48 Debye, which is smaller than $\langle\mu\rangle$. This means that the simple average underestimates the polarization effect. The results for the dipole moment is summarized in table 4.

The statistical average of QM–MM interaction energy, $\langle V_{\mathrm{QM-MM}}\rangle$, is approximately equal to the heat of solution (with the negative sign), which is defined by the energy change when the QM solute is embedded into the MM solvent. In a pure liquid, $U = -\langle V_{\mathrm{QM-MM}}\rangle/2$ is approximately the heat of evaporation. In figure 9, we show the statistical convergence of U as a function of MD steps, obtained from the QM/MM simulation run #2. We can see that the value is converged to 12.1 kcal/mol after about $N_{\mathrm{step}} = 2{,}500{,}000$ steps ($N_{\mathrm{step}}^{\mathrm{QM}} = 25{,}000$ steps). We may further correct this value by taking account of electronic polarization because hydrogen bonding between the QM–MM molecules will be a little stronger by the induction force and thus the heat of evaporation will

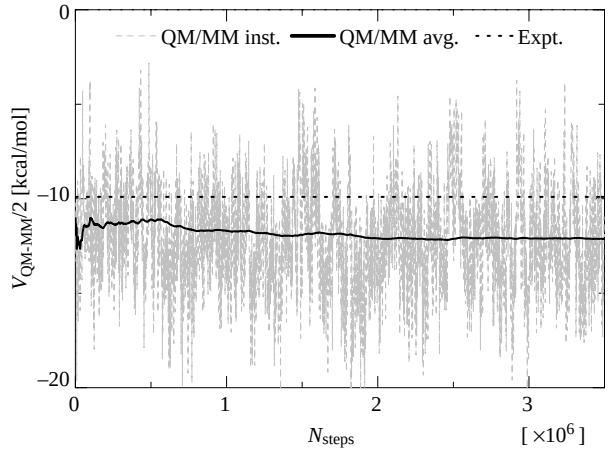

Figure 9. Time series of the QM–MM interaction $V_{\mathrm{QM-MM}}/2$ (dashed line) and its accumulative average, – U (solid line) obtained from the QM/MM MD simulation (run #2). The dotted line is the experimental heat of evaporation of liquid water (Ref. [50]).

Table 5. Heat of evaporation U

Methods	*U (kcal/mol)*
Full MM[†]	11.2
QM/MM-PE	12.1
+ Reweighting correction	13.3
+ Quantum nuclear effect[‡]	11.8
Experiment[¶]	9.9

[†] Ref. [48].
[‡] See the text and Refs. [48,49].
[¶] Ref. [50].

become a little larger. This shift can be evaluated by the weighted sampling method using equation (19), where A is the difference of two potential energies calculated by QM/MM-EE with QM = MP2/Sadlej-pVTZ and QM/MM-PE with QM = HF/6-31G(d,p). The result of this correction was +1.2 kcal/mol. Meanwhile, the heat of evaporation also shifts by quantum nuclear effect. From the previous results for classical MD and quantum path integral MD simulations [48,49], we estimate that the quantum correction is about −1.5 kcal/mol for liquid water. Thus, the overall evaporation energy is estimated to be 11.8 kcal/mol, which is 20% higher than experiment, 9.9 kcal/mol [50]. This data is summarized in table 5.

4. Conclusion

In this report, we have developed a new type of treatment for ES interactions in combined QM/MM systems, "QM/MM-PE", which suits to a MTS MD simulation. This method would be especially useful to investigate thermodynamic properties of liquid phase such as radial distribution and heat of solution. For properties in which the electronic polarization effect is important, such as molecular dipole moment, we recommend the correction by weighted sampling method after the MD simulation.

The method is computationally very efficient especially on parallel computers. Running our program code on the state-of-the-art 16-CPU PC cluster, it took only 0.088 s per MD step in the simulation of liquid water containing one QM molecule (HF/6-31G(d,p)) and 255 MM molecules (SPC/F2). This is about 22 times as fast as the conventional QM/MM simulation within a single CPU machine.

The method shall be promising also in calculating other thermodynamic properties, such as electronic spectra of a molecule dissolved in liquids. The method might also be efficient in the free energy calculation of chemical processes, since such a calculation has often been computational demanding in conventional *ab initio* QM/MM approach. Based on the preliminary tests shown in this report, application studies are now in progress at our laboratory.

Acknowledgements

We thank Dr Y. Shigeta at the University of Tokyo and Dr W. Shinoda at National Institute of Advanced Industrial

Science and Technology for useful discussions. This work was supported in part by Grant-in-Aid for Scientific Research and for the priority area by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] A. Warshel, M. Levitt. Theoretical studies of enzymatic reactions: dielectric electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 227 (1976).

[2] U.C. Singh, P.A. Kollman. A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the $CH_3Cl + Cl^-$ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, **7**, 718 (1986).

[3] M.J. Field, P.A. Bash, M. Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, **11**, 700 (1990).

[4] J. Gao. Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Rev. Comput. Chem.*, **7**, 119 (1996).

[5] K. Morokuma. New challenges in quantum chemistry—quests for accurate calculations for large molecular systems. *Phil. Trans. Roy. Lond. Ser. A*, **360**, 1149 (2002).

[6] G. Monard, K.M. Merz Jr. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Acc. Chem. Res.*, **32**, 904 (1999).

[7] J. Gao, M.A. Thompson (Eds.). *Combined Quantum Mechanical and Molecular Mechanical Methods: ACS Symp. Ser. 712*, American Chemical Society, Washington, DC (1998).

[8] I.H. Hillier. Chemical reactivity studied by hybrid QM/MM methods. *J. Mol. Struct. (Theochem)*, **463**, 45 (1999).

[9] P. Sherwood. *Modern Methods and Algorithms of Quantum Chemistry*, J. Grotendorst (Ed.), Vol. 3, p. 285 NIC-Directors, Princeton (2000).

[10] T. Mordasini, W. Thiel. Combined quantum mechanical and molecular mechanical approaches. *Chimia*, **52**, 288 (1998).

[11] D. Bakowies, W. Thiel. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.*, **100**, 10580 (1996).

[12] H. Lin, D.G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here?, http://comp.chem.umn.edu/truhlar/hilight/QMMMreview05.html.

[13] D. Frenkel, B. Smit. *Understanding Molecular Simulation, From Algorithms to Applications*, 2nd ed., Academic, San Diego (2001).

[14] N. Okuyama-Yoshida, K. Kataoka, M. Nagaoka, T. Yamabe. Structure optimization via free energy grandient method: application to glycine zwitterion in aqueous solution. *J. Chem. Phys.*, **113**, 3519 (2000).

[15] H. Takahashi, N. Matsubayashi, M. Nakahara, T. Nitta. A quantum chemical approach to the free energy calculations in condensed systems: the QM/MM method combined with the theory of energy representation. *J. Chem. Phys.*, **121**, 3989 (2004).

[16] M.E. Tuckerman, G.J. Martyna, B.J. Berne. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, **97**, 1990 (1992).

[17] G.J. Martyna, M.E. Tuckerman, D.J. Tobias, M.L. Klein. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, **87**, 1117 (1996).

[18] T.K. Woo, P. Margl, P.R. Blöchl, T. Ziegler. Sampling phase space by a combined QM/MM *ab initio* Car–Parrinello molecular dynamics method with different (multiple) time steps in the quantum mechanical (qm) and molecular mechanical (mm) domains. *J. Phys. Chem. A*, **106**, 1173 (2002).

[19] R.V. Stanton, L.R. Little, K.M. Merz Jr.. An examination of a Hartree–Fock/molecular mechanical coupled potential. *J. Phys. Chem.*, **99**, 17344 (1995).

[20] R.V. Stanton, D.S. Hartsough, K.M. Merz Jr.. An examination of a density functional/molecular mechanical coupled potential. *J. Comput. Chem.*, **16**, 113 (1995).

[21] I. Tuñón, M.T.C. Martins-Costa, C. Millot, M.F. Ruiz-López, J.L. Rivail. A coupled density-functional-molecular mechanics Monte Carlo simulation method. The water molecule in liquid water. *J. Comput. Chem.*, **17**, 17 (1996).

[22] Y. Tu, A. Laaksonen. On the effect of Lennard–Jones parameters on the quantum mechanical and molecular mechanical coupling in a hybrid molecular dynamics simulation of liquid water. *J. Chem. Phys.*, **111**, 7519 (1999).

[23] Y. Tu, A. Laaksonen. Combined Hartree–Fock quantum mechanical and molecular mechanical molecular dynamics simulations of water at ambient and supercritical conditions. *J. Chem. Phys.*, **113**, 11264 (2000).

[24] D. Xenides, B.R. Randolf, B.M. Rode. Structure and ultrafast dynamics of liquid water: a quantum mechanics/molecular mechanics molecular dynamics simulations study. *J. Chem. Phys.*, **122**, 174506 (2005).

[25] J. Lobaugh, G.A. Voth. A quantum model for water: equilibrium and dynamical properties. *J. Chem. Phys.*, **106**, 2400 (1997).

[26] Unfortunately, the Ewald sum technique for *ab initio* QM/MM simulations using localized Gaussian basis set has not been developed yet to our knowledge, while it is commonly used in semi-empirical QM/MM simulations. See, J. Gao, C. Alhambra. A hybrid semiempirical quantum mechanical and lattice-sum method for electrostatic interactions in fluid simulations. *J. Chem. Phys.*, K. Nam, J. Gao, and D.M. York, An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations, *J. Chem. Theor. Comput.* **1**, 2 (2005). **107**, 1212 (1997).

[27] M.A. Thompson. QM/MMpol: a consistent model for solute/solvent polarization. Application to the aqueous solvation and spectroscopy of formaldehyde, acetaldehyde, and acetone. *J. Phys. Chem.*, **100**, 14492 (1996).

[28] J. Gao. Energy components of aqueous solution: insight from hybrid QM/MM simulations using a polarizable solvent potential. *J. Comput. Chem.*, **18**, 1061 (1997).

[29] R.A. Bryce, R. Buesnel, I.H. Hillier, N.A. Burton. A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. *Chem. Phys. Lett.*, **279**, 367 (1997).

[30] M.S. Gordon, M.A. Freitag, P. Bandyopadhyay, J.H. Jensen, V. Kairys, W.J. Stevens. The effective fragment potential method: a QM-based MM approach to modelling environmental effects in chemistry. *J. Phys. Chem. A*, **103**, 293 (2001).

[31] M. Dupuis, M. Aida, Y. Kawashima, K. Hirao. A polarizable mixed Hamiltonian model of electronic structure for micro-solvated excited states. I. Energy and gradients formulation and application to formaldehyde ($^1A_2$). *J. Chem. Phys.*, **117**, 1242 (2002).

[32] This correspond to "predictor" part in the Gear algorithm.

[33] A.K. Soper. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.*, **258**, 121 (2000).

[34] T. Head-Gordon, G. Hura. Water structure from scattering experiments and simulation. *Chem. Rev.*, **102**, 2651 (2002).

[35] S. Izvekov, G.A. Voth. Car–Parrinello molecular dynamics simulation of liquid water: new results. *J. Chem. Phys.*, **116**, 10372 (2002).

[36] D. Asthagiri, L.R. Pratt, J.D. Kress. Free energy of liquid water on the basis of quasichemical theory and *ab initio* molecular dynamics. *Phys. Rev. E*, **68**, 041505 (2003).

[37] B. Chen, I. Ivanov, M.L. Klein, M. Parrinello. Hydrogen bonding in water. *Phys. Rev. Lett.*, **91**, 215503 (2003).

[38] J.C. Grossman, E. Schwegler, E.W. Draeger, F. Gygi, G. Galli. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *J. Chem. Phys.*, **120**, 300 (2004).

[39] E. Schwegler, J.C. Grossman, F. Gygi, G. Galli. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II. *J. Chem. Phys.*, **121**, 5400 (2004).

[40] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, M. Parrinello. The influence of temperature and density functional models in *ab initio* molecular dynamics simulation of liquid water. *J. Chem. Phys.*, **122**, 014515 (2005).

[41] A.J. Sadlej's. pVTZ basis set was obtained from the extensible computational chemistry environment basis set database, version 02/25/04, Pacific Northwest Laboratory (http://www.emsl.pnl.gov/forms/basisform.html).

[42] S.A. Clough, Y. Beers, G.P. Klein, L.S. Rothman. Dipole moment of water from Stark measurements of H_2O, HDO, and D_2O. *J. Chem. Phys.*, **59**, 2254 (1973).
[43] K. Laasonen, M. Sprik, M. Parrinello, R. Car. '*Ab initio*' liquid water. *J. Chem. Phys.*, **99**, 9080 (1993).
[44] G. Jansen, F. Colonna, J.G. Angyan. Mixed quantum–classical calculations on the water molecule in liquid phase: influence of a polarizable environment on electronic properties. *Int. J. Quantum Chem.*, **58**, 251 (1996).
[45] Y. Tu, A. Laaksonen. The electronic properties of water molecules in water clusters and liquid water. *Chem. Phys. Lett.*, **329**, 283 (2000).
[46] P.L. Silvestrelli, M. Parrinello. Water molecule dipole in the gas and in the liquid phase. *Phys. Rev. Lett.*, **82**, 3308 (1999).
[47] Y.S. Badyal, M.-L. Saboungi, D.L. Price, S.D. Shastri, D.R. Haeffner, A.K. Soper. Electron distribution in water. *J. Chem. Phys.*, **112**, 9206 (2002).
[48] W. Shinoda, M. Shiga. Quantum simulation of the heat capacity of water. *Phys. Rev. E*, **71**, 041204 (2005).
[49] M. Shiga, W. Shinoda. Calculation of heat capacities of light and heavy water by path-integral molecular dynamics. *J. Chem. Phys.*, **123**, 134502 (2005).
[50] G.S. Kell. Density, thermal expansivity, and compressibility of liquid water from 0° to 150° C: correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *J. Chem. Eng. Data*, **20**, 97 (1975).