JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Size-Consistent Multipartitioning QM/MM: A Stable and Efficient Adaptive QM/MM Method

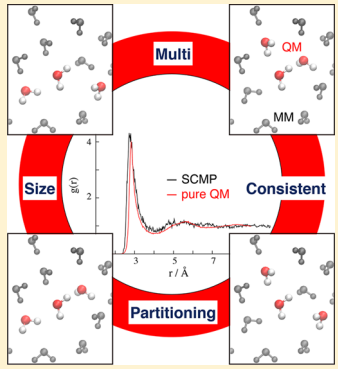
Hiroshi C. Watanabe,*[,†,‡] Tomáš Kubař,[‡] and Marcus Elstner[‡]

[†]Center for Biological Resources and Informatics, Tokyo Institute of Technology, B-62 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501 Japan

[‡]Institute of Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany

Supporting Information

ABSTRACT: We propose a new adaptive QM/MM method, the size-consistent multipartitioning (SCMP) QM/MM scheme, which enables stable and computationally efficient QM/MM simulations. A number of partitionings with identical size of the QM and the MM regions are considered with a new adaptive scheme in order to (1) realize smooth QM/MM switching, (2) introduce a conserved quantity (total energy, Hamiltonian), (3) avoid spurious artificial forces on the QM/MM border, and (4) allow for an efficient parallel implementation. Benchmark simulations performed for "QM water in MM water" show that energy conservation can be significantly improved and the computational efficiency allows treating also larger QM regions, for which previous methods had to face an intractable increase in computer time.



I. INTRODUCTION

Molecular simulations, a subject of constant development throughout the recent decades, are classified into several groups according to the molecular model employed. With quantum mechanics (QM), the electronic structure of a molecular system is considered explicitly, aiming for a transferable and accurate description of interatomic and intermolecular interactions but demanding substantial computational resources. Molecular mechanics (MM) on the other hand, adopts a simplified empirical potential for the interatomic interactions in order to reduce the computational cost but possibly sacrificing some of the transferability and accuracy. According to the observables and phenomena of interest, a proper representation of the molecular system should be selected among the available models.

An appealing way to complement the respective advantages of QM and MM is the hybrid QM/MM approach. Since their introduction in 1976,[1] QM/MM simulations have been instrumental in studies of reaction mechanisms in complex environments, such as in aqueous solutions and biomolecular complexes.[2−10] Since many various QM and MM methods are available, numerous ways to combine them in QM/MM have been proposed as well, considering also the possibility to apply multiple QM methods of different accuracy and computational cost (ONIOM).[11,12] As for the choice of QM approach, semiempirical QM methods (e.g., AM1,[13] PM6,[14] OM2,[15] and DFTB3[16−18]) are attractive compromise options because of the reduced computational cost and largely maintained accuracy in many cases. Indeed, it was demonstrated that semiempirical methods in QM/MM molecular dynamics (MD) simulations extend the available time scales in simulations, making global structural sampling accessible.[19]

Within a QM/MM framework, the molecular system is divided physically into two regions that are treated on different footings—one of them with QM and the other with MM. Apart from the obvious issue of how to deal with the interaction between both regions, additional difficulties arise in MD simulations involving mobile molecules that are free to cross the physical boundary of the QM and MM regions, which will be denoted as spatial and temporal discontinuities. Spatial discontinuities arise because the different methods used, QM and MM, deliver properties of the molecules and thus of the bulk matter. Therefore, solvent in the QM and MM regions will have different properties. As a result, there arises a sort of "phase boundary" at the QM and the MM interface, on which artifacts may occur. This is inevitable in any QM/MM simulation, and a practical way to minimize these artifacts is the introduction of a smooth switching of QM and MM of the molecules near the boundary. Temporal discontinuities occur in MD simulations in which molecules are allowed to move from the QM region to the MM and vice versa. The switches of description of the molecules result in changes of the Hamiltonian in the simulations, leading to discontinuities of energy in the course of the simulations. Even more serious are the discontinuities of forces, which may lead to unstable simulations.

Various MD methods have been developed, which combine QM embedding and a classical force model in a unified scheme free of any boundary region.[20−22] In such approaches, the quantum-chemical information is incorporated into a simple parametrized force model at run time and there are no boundary

Received: July 1, 2014
Published: September 10, 2014

 © 2014 American Chemical Society dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

problems. However, such methods are useful for rather special classes of applications, and most QM/MM simulations of processes taking place in complex environments will still involve a division of the system into QM and MM regions. Note that the discontinuity problems are common to any multiscale simulation scheme, occurring, for example, in hybrid coarse-grained/all-atom MD simulations as well.

To alleviate the discontinuity, several corrective techniques have been proposed, which are known as adaptive QM/MM. The first generation of adaptive QM/MM is mainly designed to improve the spatial continuity. Note that the QM and MM regions in the simulation may still have different properties (depending on the QM and MM methods employed), but the passage from QM to MM will be gradual rather than abrupt. In the hotspot method, a transition zone is introduced in which the forces are switched gradually from those calculated with QM to MM.[23,24] In this method, the number of QM solvents suddenly increases or decreases whenever an MM solvent molecule intrudes into the transition zone or a QM solvent molecule diffuses away from the center of the QM region. Correspondingly, the Hamiltonian changes and discontinuities of the energy and forces appear. In the buffered-forced scheme, these discontinuities are suppressed by the introduction of a marginal QM setup beyond the transition region, in combination with a hysteresis switching of the QM region.[25] Stability of MD simulation is achieved; however, there is no total energy expression (Hamiltonian) from which equations of motion can be derived; thus, there is no conserved quantity in the course of simulation. In the ONIOM-XS method,[26] the Hamiltonian is introduced by combining two types of ONIOM energy function with different numbers of QM molecules. However, this works fine only with one solvent molecule in the buffer region, and discontinuities emerge again otherwise.

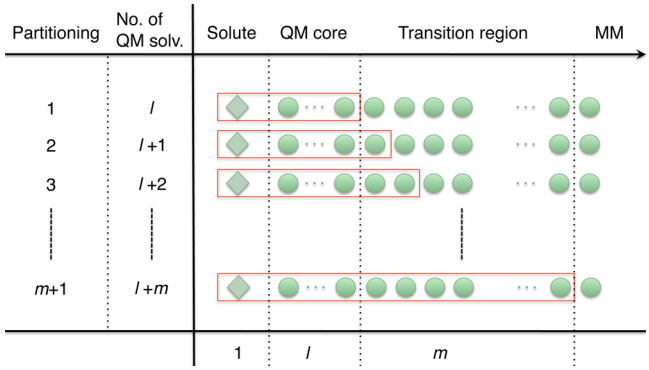
The first generation methods still suffer from the temporal discontinuity, and this issue is tackled by adaptive multipartitioning schemes proposed in the second generation. These techniques introduce several partitionings of the molecular system into QM and MM regions (of generally different sizes). A partitioning denotes a particular division of the molecular system into a QM and a MM region. Clearly, there are a multitude of choices to construct such partitionings. Then, QM/MM calculations are performed independently for each of the partitionings, and the resulting energy and forces on atoms are combined to give the total energy and forces to be considered in the MD simulation. The name "adaptive partitioning" (AP) denotes the concept that the partitionings are allowed to change during the simulations. The specific design of an AP scheme is arbitrary as long as the requirements of continuous energy and forces are satisfied. Thus, several different types of AP schemes have been proposed, such as Permuted AP,[27] Sorted AP (SAP),[27] and Difference-based Adaptive Solvation (DAS).[28,29]

Common to all of these AP schemes each QM/MM, partitioning contains a different number of QM solvent molecules. To this end, two regions are defined in the molecular system: the QM core region that is included entirely in the QM region in each partitioning and the transition region that contains solvent molecules included in the QM regions in some but not all of the QM/MM partitionings. All of the other molecules in the system are considered as MM in all partitionings. QM/MM calculations are conducted for each partitioning, and the resulting energy and forces are combined with an AP scheme to realize a gradual switching of the QM and MM characters of the solvent molecules over the transition region. The second generation of the adaptive QM/MM methods succeeded to define an energy function and remove the discontinuity as shown in MM/MM simulations.[27,28] Unfortunately, in the case of QM/MM simulations, however, they still suffer from substantial energy drifts, although the situation is much improved compared to the first generation methods.[29,30]

The current AP schemes require that the number of QM/MM partitionings (thus the number of QM/MM calculations performed) is larger than the number of solvent molecules in the transition region (see Scheme 1). Since the volume of the transition region increases with the square of the radius of the QM core region, the number of required partitionings increases steeply with the QM core radius as well (See Figure 1). Thus, the computational effort grows steeply for middle-sized QM core regions. On the other hand, a considerable advantage of multipartitioning methods is the possibility of trivial parallelization, because the calculations for the individual partitionings are entirely independent of each other. However, the distribution of CPU loads is not ideal, because partitionings with large QM regions will request much longer computational times than those with smaller QM zones, and compute nodes being assigned smaller QM zones will idle for considerable amounts of time. Thus, the practicability and applicability of AP QM/MM methods designed in this way are limited.

In order to improve adaptive QM/MM methods, we propose a size-consistent multipartitioning (SCMP) QM/MM scheme

Scheme 1. Multiple Partitionings[a]



[a]See text for explanation.

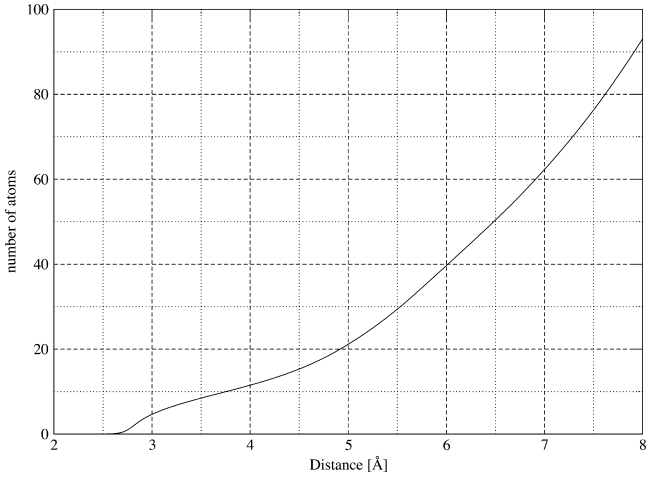
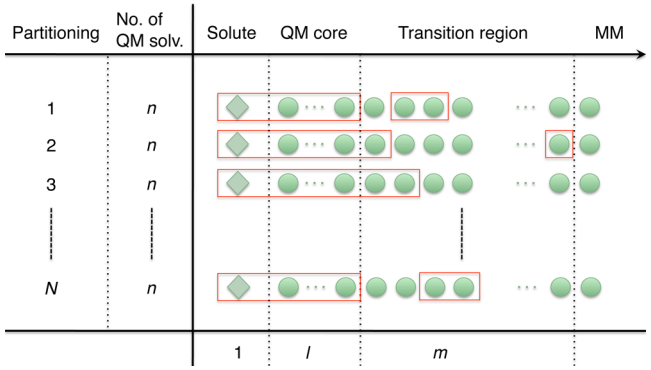


Figure 1. Integrated oxygen−oxygen radial distribution function for the system of 256 water molecules in a pure QM (DFTB3) simulation.

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

Scheme 2. Size-Consistent Multipartitioning Method



where all the partitionings contain the same number of QM and MM molecules (see Scheme 2). We demonstrate that the present method yields a conserved Hamiltonian as well as continuous forces, and it outperforms the previous multipartitioning methods concerning the applicability for larger QM core regions as well as conserved energy independent of the number of partitionings.

In section II, we present the partitioned potential energy derivation for each single partitioning based on Born–Oppenheimer approximation. In section III, we introduce a new AP scheme with a weight function to combine forces in each partitioning to calculate effective forces and total energy. In section IV, we describe the partitioning updating scheme, and in section V, the computational details. In section VI, we demonstrate the application to bulk water.

II. FORCE AND SIZE-CONSISTENT MULTIPARTITIONING QM/MM HAMILTONIAN

Here, we introduce a size-consistent multipartitioning QM/MM scheme where N different partitionings are defined such that each partitioning involves the same number of QM solvent molecules. The effective force $\boldsymbol{F}_j^{\text{eff}}$ acting on the j-th atom is obtained as weighted average

$$\mathbf{F}_j^{\text{eff}}(\mathbf{r}) = \sum_n^N \sigma^{(n)}(\mathbf{r})\cdot\mathbf{f}_j^{(n)}(\mathbf{r}) \tag{1}$$

of forces $\boldsymbol{f}_j^{(n)}$ calculated as negative gradients of QM/MM energy $V^{(n)}(\mathbf{r})$ in the n-th partitioning and $\sigma^{(n)}(\mathbf{r})$ is a normalized weight function as described in the next section. Then, an effective potential energy $V^{\text{eff}}(\mathbf{r})$ is given as an integral

$$\begin{aligned} V^{\text{eff}}(\mathbf{r}) &= -\int \mathbf{F}^{\text{eff}}(\mathbf{r})\cdot\mathrm{d}\mathbf{r} \\ &= -\sum_n^N \int \sigma^{(n)}(\mathbf{r})\mathbf{f}^{(n)}(\mathbf{r})\cdot\mathrm{d}\mathbf{r} \end{aligned} \tag{2}$$

Since $V^{\text{eff}}(\boldsymbol{r})$ is difficult to calculate directly, it is transformed into a discrete form in the temporal domain

$$\begin{aligned} V^{\text{eff}}(\mathbf{r}:\, t=T) &= \sum_n^N \int_{t=0}^T \sigma^{(n)}(\nabla V^{(n)}(\mathbf{r}))\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)\mathrm{d}t \\ &= \sum_n^N \int_{t=0}^T \sigma^{(n)}\left(\frac{\partial V^{(n)}(t)}{\partial t}\right)\mathrm{d}t \\ &= \sum_n^N \sum_{i=0}^s \sigma^{(n)}(t_i)\cdot\Delta V^{(n)}(t_i) \end{aligned} \tag{3}$$

where the index i runs over the simulation steps and s is the total number of steps. (Then, the total simulation time is $T = s\Delta t$, where Δt is the time step).

Alternatively, $V^{\text{eff}}(\boldsymbol{r})$ can also be written using partial integration of eq 3 as

$$\begin{aligned} V^{\text{eff}}(\mathbf{r}:\, t=T) &= \sum_n^N \sigma^{(n)}V^{(n)} - \sum_n^N \int_{t=0}^T V^{(n)}\left(\frac{\partial\sigma^{(n)}(t)}{\partial t}\right)\mathrm{d}t \\ &= \sum_n^N \sigma^{(n)}V^{(n)} - \sum_n^N \sum_{i=0}^s V^{(n)}(t_i)\cdot\Delta\sigma^{(n)}(t_i) \\ &= V^{\text{ad}} + V^{bk} \end{aligned} \tag{4}$$

and it consists of two contributions: The first term is a weighted average of the potential energies in the individual partitioning and is called the adaptive potential. The second term is a bookkeeping term, as introduced by Bulo et al.[28]

For the sake of fair comparison with previous literature,[28−30] we employed eq 4 in the following application and algorithm.

Therefore, the Hamiltonian for MD simulation in the NVE ensemble is given as

$$\begin{aligned} H &= K(\mathbf{p}) + V^{\text{eff}}(\mathbf{r}) \\ &= \sum_i \frac{p_i^2}{2m_i} + V^{\text{ad}} + V^{\text{bk}} \end{aligned} \tag{5}$$

III. WEIGHT FUNCTION

To derive a computationally efficient scheme, the weight function $\sigma^{(n)}(\mathbf{r})$ has to be designed such that large deformations of the QM region as well as intrusion of MM molecules into the QM core region are avoided; that is, a reasonably compact shape of the QM region has to be maintained. This is achieved by imposing two conditions on $\sigma^{(n)}$: (1) $\sigma^{(n)} = 0$ whenever the QM region in the n-th partitioning is fragmented or deformed, and (2) $\sigma^{(n)} = 0$ when the QM region in the n-th partitioning has the most compact form. A QM region is considered to be the most compact when only those solvent molecules in the transition region are included, which are the closest to the QM center. Although the latter condition may seem to be counterintuitive at the first sight, the combination of these conditions makes it possible to replace a partitioning with a fragmented QM region by one with the most compact QM region without any discontinuity in forces and energy appearing. After being inserted into the ensemble, the most compact partitioning will gradually be deformed, and thereby, its weight will gradually increase in the course of the dynamics following.

Let us consider N different partitionings (with different QM and MM regions). L and M are the indices of QM and MM solvent molecules in the QM/MM partitionings, and the n-th partitioning contains molecule subsets $S_{\text{QM}}^{(n)}$ and $S_{\text{MM}}^{(n)}$. To each solvent molecule, we assign two transition profiles $\Lambda^{(n)}{}_{\text{in}}$, $\Lambda^{(n)}{}_{\text{out}}$ defined as

$$\Lambda_{\text{out}}^{(n)}(\mathbf{r}_i) = \begin{cases} \lambda_{\text{QM}}^{(n)}(\mathbf{r}_i:\, s_{\text{QM}}^{\text{out}},\, t_{\text{QM}}^{\text{out}}) & \text{if} \quad i \in S_{\text{QM}}^{(n)} \\ \lambda_{\text{MM}}^{(n)}(\mathbf{r}_i:\, s_{\text{MM}}^{\text{out}},\, t_{\text{MM}}^{\text{out}}) & \text{if} \quad i \in S_{\text{MM}}^{(n)} \end{cases} \tag{6}$$

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

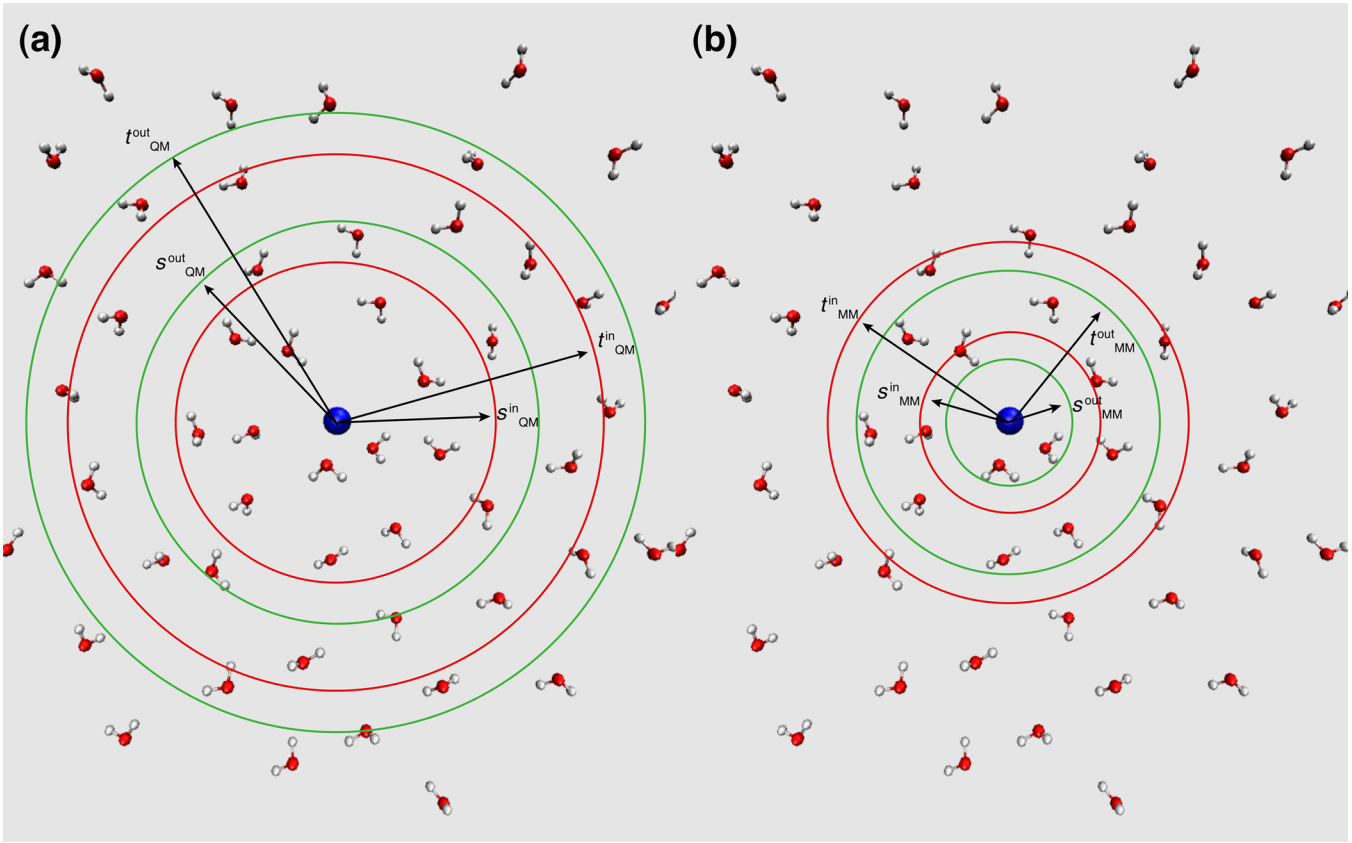


Figure 2. Schematic illustration of the transition region for (a) the fade-out function and (b) the fade-in function.

$$\Lambda^{(n)}_{in}(\mathbf{r}_i) = \begin{cases} \lambda^{(n)}_{QM}(\mathbf{r}_i; s^{in}_{QM}, t^{in}_{QM}) & \text{if} \quad i \in S^{(n)}_{QM} \\ \lambda^{(n)}_{MM}(\mathbf{r}_i; s^{in}_{MM}, t^{in}_{MM}) & \text{if} \quad i \in S^{(n)}_{MM} \end{cases} \tag{7}$$

where λ is a progress function written with parameters s and t ($s < t$),

$$\lambda_{QM}(\mathbf{r}_i; s_{QM}, t_{QM}) = \begin{cases} 1 & \text{if} \quad r_i \le s_{QM} \\ \dfrac{(t_{QM} - r_i)^2(2r_i + t_{QM} - 3s_{QM})}{(t_{QM} - s_{QM})^3} & \text{if} \quad s_{QM} \le r_i \le t_{QM} \\ 0 & \text{if} \quad t_{QM} \le r_i \end{cases} \tag{8}$$

$$\lambda_{MM}(\mathbf{r}_i; s_{MM}, t_{MM}) = \begin{cases} 0 & \text{if} \quad r_i \le s_{MM} \\ \dfrac{(r_i - s_{MM})^2(3t_{MM} - 2r_i - s_{MM})}{(t_{MM} - s_{MM})^3} & \text{if} \quad s_{MM} \le r_i \le t_{MM} \\ 1 & \text{if} \quad t_{MM} \le r_i \end{cases} \tag{9}$$

$\lambda_{QM}(\mathbf{r}_i)$ stands for the progress function of distance r_i between the center of the QM core and the i-th solvent molecule, which is determined by two parameters s and t ($s < t$). λ_{QM} decreases from 1 down to 0 if the molecule diffuses from the center of QM region across the transition region defined by s and t. Note that each QM solvent molecule has two $\lambda_{QM}(\mathbf{r}_i)$ values due to the parameter sets (s^{out}_{QM}, t^{out}_{QM}) and (s^{in}_{QM}, t^{in}_{QM}) (Figure 2a). In contrast, $\lambda_{MM}(\mathbf{r}_i)$ increases up to 1 over the transition region, being determined by (s^{out}_{MM}, t^{out}_{MM}) or (s^{in}_{MM}, t^{in}_{MM}) (Figure 2b).

Put in a simplified way, the motivation for such a setup is to reduce the MM character of molecules approaching the QM core, starting at the distance t_{MM}, and to reduce the QM character of molecules leaving the QM core, starting at s_{QM}. Typically, t_{MM} will be close to or equal s_{QM}. While these parameters give the distance where the transition profiles set on, t_{QM} or s_{MM} determine the point where the respective transition profiles terminates.

Next, we introduce fade-out and fade-in functions, which are given by products of transition profiles as follows:

$$O^{(n)}_{QM}(\mathbf{r}) = \prod_{i \in S^{(n)}_{QM}} \Lambda^{(n)}_{out}(\mathbf{r}_i) \qquad I^{(n)}_{QM}(\mathbf{r}) = 1 - \prod_{i \in S^{(n)}_{QM}} \Lambda^{(n)}_{in}(\mathbf{r}_i)$$
$$O^{(n)}_{MM}(\mathbf{r}) = \prod_{i \in S^{(n)}_{MM}} \Lambda^{(n)}_{out}(\mathbf{r}_i) \qquad I^{(n)}_{MM}(\mathbf{r}) = 1 - \prod_{i \in S^{(n)}_{MM}} \Lambda^{(n)}_{in}(\mathbf{r}_i) \tag{10}$$

$O^{(n)}_{QM}$ and $O^{(n)}_{MM}$ are fade-out functions for the QM molecules and the MM molecules in the n-th partitioning, respectively. Note that $O^{(n)}_{QM} = 0$ if a QM solvent molecule in the n-th partitioning diffuses far away from the QM region center and $O^{(n)}_{MM} = 0$ if a MM solvent molecule is approaching the QM center. Thus, the fade-out functions account for the first condition described above. Similarly, the fade-in functions $I^{(n)}_{QM}$ and $I^{(n)}_{MM}$ realize the second requirement: the partitioning with the most compact QM region possesses a weight of zero and is available to replace a strongly fragmented partitioning with zero weight, retaining the continuity of energy and forces during this exchange.

Then, the weight of the n-th partitioning is defined as

$$\sigma^{(n)}(\mathbf{r}) = \frac{O^{(n)}_{QM}(\mathbf{r}) \cdot O^{(n)}_{MM}(\mathbf{r}) \cdot I^{(n)}_{QM}(\mathbf{r}) \cdot I^{(n)}_{MM}(\mathbf{r})}{\sum_n^N O^{(n)}_{QM}(\mathbf{r}) \cdot O^{(n)}_{MM}(\mathbf{r}) \cdot I^{(n)}_{QM}(\mathbf{r}) \cdot I^{(n)}_{MM}(\mathbf{r})} \tag{11}$$

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

and the set of weights is normalized trivially:

$$\sum_{n}^{N} \sigma^{(n)}(\mathbf{r}) = 1 \tag{12}$$

Since each partitioning in the SCMP calculation contains the same number of QM and MM molecules, all partitionings span an identical phase space. So, the absolute values of potential energies $V^{(n)}$ are comparable in the sense that they differ only by the difference of interaction energies due to different configurations (structures), in contrast to the previous adaptive QM/MM where the absolute energies in the respective partitionings are largely different due to the solvent molecules in the QM and MM representations. In the present study, the partitioning ensemble is generated randomly once at the start of the simulation and is equilibrated with an update scheme as described in the next section, which leads to a convergence of total energy correspondingly, even if the potential energies vary systematically due to the selection of partitionings. The properties of the ensemble, its convergence, and the conservation of the SCMP total energy are discussed in the Results section.

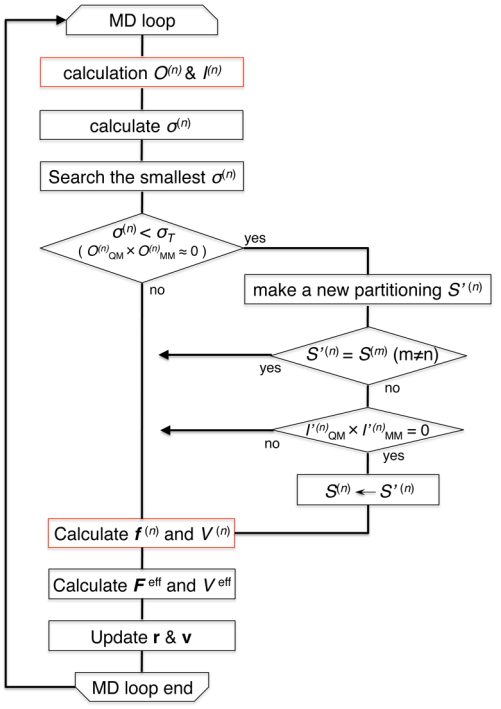
IV. UPDATING ALGORITHM

Since forces and energy calculations for each partition are independent of each other and all of the QM regions have the same size, the multipartitioning QM/MM simulations are ideal for parallel computing where each QM/MM partition is assigned to one processing node. Contrary to the conventional multiscale partitioning methods, which require a specific number of partitionings larger than that of the solvent molecules in the transition region,[27,28] SCMP is more flexible regarding the number of partitionings. Still, a shortage of partitionings can cause artifacts around the QM/MM border.

In the SCMP method, out of the vast number of possible partitionings, only several partitionings can be involved in the calculation. In order to make efficient use of the available computational resources, any partitioning that exhibits a small statistical weight $\sigma^{(n)}$ should be promptly replaced by another. To this end, we introduce a threshold value σ_T and implement the following procedure: A partitioning with a fragmented QM region with $\sigma^{(n)} < \sigma_T$ is replaced by another partitioning with a compact QM region with $\sigma^{(n)} < \sigma_T$. Among the various possible replacement algorithms, we have adopted the simplest one, namely a minimum update: a partitioning with $\sigma^{(n)} < \sigma_T$ is replaced by the partitioning with the most compact QM region, which consists of the nearest solvent molecules to the QM center. When all QM solvent molecules are within s^{in}_{QM} from the QM center, or all MM solvent molecules are further away than s^{in}_{MM}, the fade-in functions $I^{(n)}_{QM}$ and $I^{(n)}_{MM}$ of the corresponding partitioning vanishes, which leads to zero weight.

As shown in Scheme 3, we employ the following procedure in one MD step: (1) calculation of the fade-in and fade-out functions with eq 6−10 on respective computational nodes, (2) calculation of the weights $\sigma^{(n)}$ with eq 11 on the master node, (3a) search for a partitioning with a fragmented QM region, for which $\sigma^{(n)} < \sigma_T$ and $O^{(n)}_{QM}O^{(n)}_{MM} \approx 0$, (3b) check whether the most compact partitioning is available (not yet included in the set of considered partitionings), (3c) substitution of the fragmented partitioning by the most compact one, (4) calculation of forces $f_i^{(n)}$ and energy $V^{(n)}$ for each partitioning on the respective computational node, (5) calculation of the effective potential and of the effective forces with eq 1 and 4, (6) coordinate update.

Scheme 3. Algorithm of Size-Consistent Multipartitioning QM/MM Simulation[a]



[a]Red-lined box can be computed on respective node.

V. COMPUTATIONAL DETAILS

Size-Consistent Multipartitioning QM/MM. We implemented DFTB3 and the SCMP method in a local version of the GROMACS 4.5.5 package.[31] QM/MM calculations are parallelized with MPI such that the calculation of each QM/MM partitioning is assigned to a separate MPI node (CPU core), and the results of the calculations are collected on the master node, in each step of the MD simulation.

For benchmark, we perform "QM water in MM water" simulations over 20 ps with a time step of 0.5 fs. The QM waters are treated by DFTB3[16−18] with the 3ob parameter set,[32] and MM waters are represented with the flexible TIP3P model.[33] The electric potential induced by the charges of MM atoms on the QM atoms is obtained with the smooth particle−mesh Ewald method as implemented in GROMACS with a switching function for electrostatic interaction.[34] van der Waals interactions are shifted starting at 14 Å to vanish at the cutoff distance of 16 Å. The Nosé−Hoover thermostat[35,36] is adopted for temperature control with a reference temperature of 300 K and a time constant of 0.1 ps.

The transition profiles in the SCMP method are estimated based on oxygen atom coordinates in water molecules. We list the used transition parameter sets (s^{in}_{QM}, t^{in}_{QM}, s^{out}_{QM}, t^{out}_{QM}, s^{in}_{MM}, t^{in}_{MM}, s^{out}_{MM}, t^{out}_{MM}) in Table 1. The initial partitioning set is always generated so that two-thirds of the QM water molecules are those nearest to the QM center in all partitionings; the remaining

Table 1. Transition Parameters [Å]

	s^{in}_{QM}	t^{in}_{QM}	s^{out}_{QM}	t^{out}_{QM}	s^{in}_{MM}	t^{in}_{MM}	s^{out}_{MM}	t^{out}_{MM}
[15, 48]	4.5	7.5	4.5	7.5	2.5	4.5	2.5	4.5
[19, 48]	5.4	8.4	5.4	8.4	2.2	5.2	2.2	5.2
[23, *n*]	5.5	8.5	5.5	8.5	2.5	5.5	2.5	5.5
[31, *n*]	5.8	8.8	5.5	8.5	2.9	5.9	2.5	5.5

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

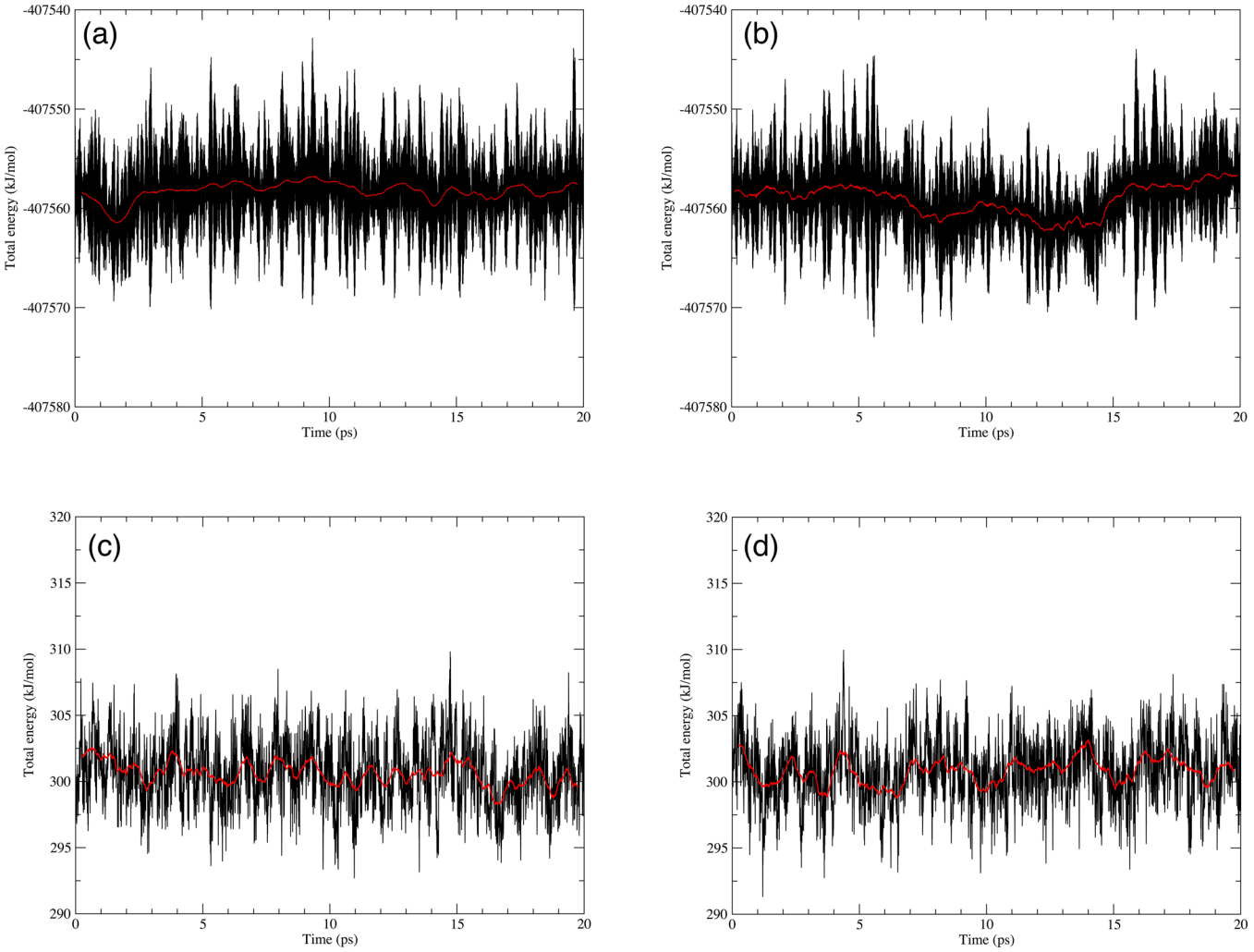


Figure 3. Total energy in the NVT SCMP simulations (a) [31, 36] and (b) [31, 42]. Temperature in the NVE SCMP simulations (c) [23, 28] and (d) [23, 36]. Black and red lines represent instantaneous values of data and averages over 1000 time steps, respectively.

QM molecules are selected randomly from the molecules that are present in the range of distances between s_{QM}^{in} and t_{QM}^{out}. A threshold of $\sigma_T = 0.00001$ is applied for the partitioning update; while this value has not been optimized, it will turn out that good energy conservation will be achieved, proving the value of σ_T sufficiently small.

Pure QM Simulations. For reference, a pure QM simulation was conducted with the DFTB+ package;[37] much the same as in the QM/MM simulations, the DFTB3 method with the 3ob parameters set was used. The system of 256 water molecules enclosed in a periodic cubic box was simulated for 20 ps with a time step of 0.5 fs. The Nosé thermostat and the Berendsen barostat[38] were employed to maintain constant temperature of 300 K and constant pressure of 1 bar.

VI. RESULTS

Energy Conservation. For the sake of convenience, we introduce a designation $[q, p]$ for a SCMP simulation with q QM solvent molecules and p partitionings. As described above, the SCMP Hamiltonian in eq 5 may depend on the partitioning ensemble, which is equilibrated in the initial phase of every simulation through the update algorithm. Here, we analyze the time course of the total energy in SCMP simulations of the "QM water in MM water" system containing 2177 water molecules.

Figure 3a demonstrates that some energy relaxation takes place in the first 3 ps in the NVT ensemble with the [31, 36] configuration, and the total energy remains constant for the rest of the simulation. On the other hand, the [31, 42] simulation does not show a clear relaxation, and there seems to be weak energy oscillation with a period of ca. 4 ps (see Figure 3b) instead. Note that in both of these simulations, after ca. 400 partitioning update steps, neither discontinuity nor monotonic drift of energy is observed. Correspondingly, the SCMP simulation in the NVE ensemble keeps temperature the initial value of ca. 300 K (see Figure 3c and d) in contrast to the conventional adaptive QM/MM methods that show temperature drift of more than 80 K/ps in the other adaptive QM/MM simulations with DFTB/TIP3P.[30] This is a major improvement relative to the second generation of the adaptive QM/MM simulations.[27,29,30]

For further analysis of the SCMP stability, we carried out simulations with various numbers of QM solvent molecules and various numbers of partitionings, and the results are summarized in Table 2. Several significant features of the SCMP simulations as observed:

(1) The total energy in the SCMP simulations is conserved better than in the second generation of the adaptive QM/MM. Note that the energy drift in the SCMP simulation, which is an order of magnitude smaller than in the

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

Table 2. Energy Drifts Per Picosecond[a]

method	no. of exchangeable QM waters	no. of partitionings	no. of partitioning update [ps^{-1}]	energy drift [kJ/mol·ps]	avg. energy [kJ/mol]
conventional QM/MM	0[c]	1		8.7×10^{-3}	-77491 ± 2
	15[c]	1		0.12	-236925 ± 0.3
	19[c]	1		0.12	-279623 ± 0.4
SCMP	15	48	19.4	0.34	-236883 ± 3
	19	36	10.0	0.14	-279563 ± 5
		48	20.3	0.36	-279558 ± 3
	23	32	17.5	8.5×10^{-3}	-346718 ± 3
		36	15.1	-0.21	-346723 ± 3
		48	21.5	-0.13	-346717 ± 3
	31	32	14.1	-1.9×10^{-3}	-407566 ± 4
		36	20.2	3.4×10^{-2}	-407558 ± 4
		42	24.8	1.4×10^{-2}	-407559 ± 4
		48	21.0	0.32	-407561 ± 5
multiscale partitioning QM/MM[b]	17	3		2.11	no data
	21	17		4.13	no data
	29	17		2.52	no data

[a]The total energies are fitted by a linear function in order to estimate an energy drift from the start to the end point of the simulation. [b]Number-adaptive difference-based solvation multiscale QM/MM simulation[29] combining the semiempirical PM3 method and the SPC water model. [c]The number of water molecule in QM region surrounding a water in the QM center.

second-generation AP simulations, is temporary and concentrated to the initial relaxation phase as seen in Figure 3a, as opposed to the second-generation AP where the energy was reported to increase monotonically.[27−30]

(2) No energy drift is observed on occasions that the partitioning updates take place. This implies that the update threshold σ_T used in the present study is small enough not to endanger the energy conservation. The energy drift observed may be attributed to the relaxation process.

(3) Energy drifts is independent of the size of the QM region, as seen at the comparison of the $[n, 48]$ simulations (n = 15, 19, 23, and 31 solvent molecules in the QM region, respectively).

(4) Similarly, the comparison of the groups of simulations $[23, n]$ (n = 32, 36, and 48 partitionings) and $[31, n]$ (n = 32, 36, 42, and 48 partitionings) suggests that the energy drift does not depend on the number of partitionings in the SCMP simulations.

(5) Apart from the possible drift of energy, we also present the averaged absolute values of energy in Table 2. These values are constant among the groups of simulations with the same numbers of QM molecules: $[19, n]$ (n = 36 and 42 partitionings), $[23, n]$ (n = 32, 36, and 48 partitionings) and $[31, n]$ (n = 32, 36, 42, and 48 partitionings). This consistence of energy should make it possible to combine the SCMP approach with other advanced simulation techniques, such as replica exchange simulations.[39] Still, what may be surprising at the first sight is that the total energies in the SCMP simulations are slightly larger than in conventional QM/MM simulations with the same size of the QM region. (Here, we are comparing, for example, conventional QM/MM with 15 QM molecules and the [15, 48] SCMP simulation.) Note that the conventional simulation in the present study starts with the most compact QM region, while the SCMP simulation features partitionings with more spatially expanded QM regions. Since the compactness of the QM region may affect the potential energy, as described above, this difference of energy between SCMP and conventional QM/MM is understandable.

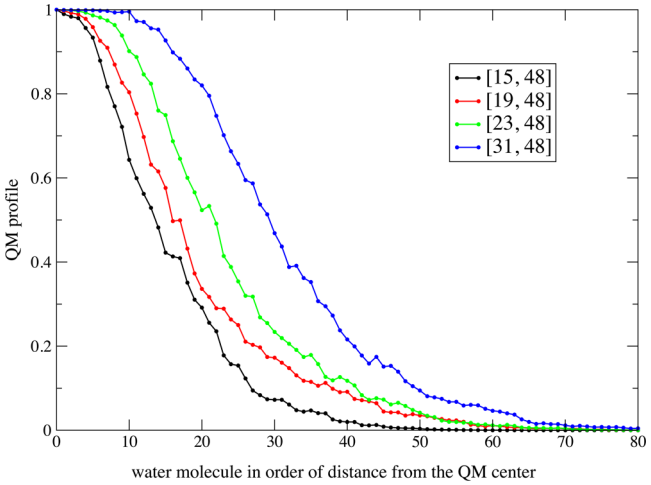


Figure 4. QM profiles of water molecules observed in the different SCMP QM/MM simulations: [15, 48] (black), [19, 48] (red), [23, 48] (green), and [31, 48] (blue).

QM Profile. In order to describe the effective size of the QM region in an SCMP simulation, we introduce the (instantaneous) QM profile $\omega_t(j)$ for the j-th exchangeable solvent molecule in time t as

$$\omega_t(j) = \sum_n^N \sigma_t^{(n)}(\mathbf{r}) \cdot \delta^{(n)}(j) \tag{13}$$

where

$$\delta^{(n)}(j) = \begin{cases} 1 & \text{if} \quad j \in S_{\mathrm{QM}}^{(n)} \\ 0 & \text{if} \quad j \in S_{\mathrm{MM}}^{(n)} \end{cases} \tag{14}$$

Then, the time-averaged QM profile $\omega(j)$ is defined as

$$\omega(j) = \frac{1}{T} \sum_{t=1}^{T} \omega_t(j) \tag{15}$$

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

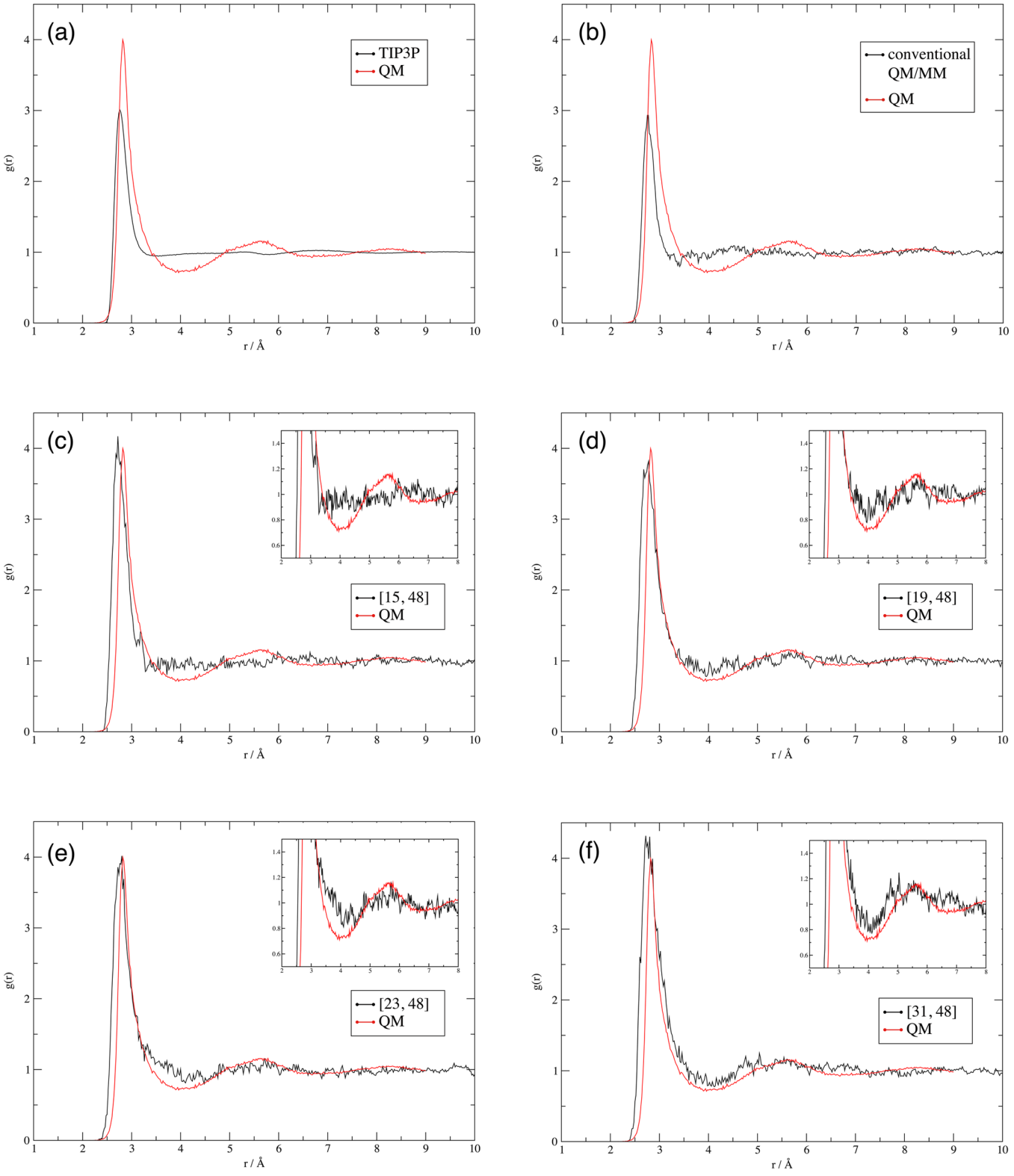


Figure 5. Molecular structure of water obtained from MD simulations performed with different approaches: (a) O−O RDF of all oxygen atoms in TIP3P water. (b) Oc-O RDF (Oc is central oxygen) in a conventional QM/MM simulation where one water molecules is QM and the remaining solvent is MM. (c−f) Oc-O RDF in different SCMP simulations: (c) [15, 48], (d) [19, 48], (e) [23, 48], (f) [31, 48]. Black lines are the data from each respective simulation; red line is a reference RDF of all oxygen atoms obtained from a "pure QM" (DFTB3) simulation.

So, a molecule with $\omega(j) = 1$ always belongs to the QM region in all partitionings, while $\omega(j) = 0$ indicates a purely MM molecule. Figure 4 illustrates the QM profiles of water molecules in the $[n, 48]$ simulations (n = 15, 19, 23, and 48 QM water molecules). Evidently, all of the SCMP simulations realize a smooth transition between the QM and the MM character of the molecules. For instance, in the [31 48] simulation, $\omega(j)$ is larger than zero for all water molecules up to the 98th closest to the

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

center of the QM region. This is the longest QM/MM switching range considered in any adaptive QM/MM method so far.

Since not only the interaction of the solute with the first solvation shell but also the second and the third shell may affect the solvation, the solvent molecules in these shells should be treated at the QM level in switching framework as well. In a second-generation AP method, however, this would require massive computational resources to be deployed because the required number partitionings is larger than the number of solvent molecules in transition region; in the above case, this would be 99 partitionings in the DAS technique.[28] Thus, the application of the SCMP simulation scheme may reduce the requirements for computational resources significantly.

Note that the instantaneous QM profiles are not necessarily monotonically decreasing functions in the SCMP method, while most of the second-generation AP methods ensure a smooth QM-MM switching at each MD step (although they are limited to short-range switching).[27,28] Thus, to verify the continuities on the QM/MM border in the SCMP simulations, we investigate the structure of water as described by the radial distribution function in the next section.

Oxygen−Oxygen Radial Distribution Function (O−O RDF). The RDF of oxygen atoms of liquid water obtained from QM (DFTB3) and MM (TIP3P) simulations are presented in Figure 5a. The DFTB3 water shows a peak value of 4.0 at 2.8 Å and a second peak appearing at 5.7 Å. The performance of DFTB models for water, including DFTB3/3OB as used here, was discussed previously,[40] and it was found that improvements of the DFTB model are necessary. This, however, is not the scope of this work, and the RDF of water obtained with the pure QM simulation serves merely as a reference for the evaluation of the performance of the SCMP method. Note that any other QM method can be implemented in principle: DFTB3 is used here only because of its computational efficiency. On the other hand, the MM water model TIP3P yields a peak value of 3.0 and misses the second peak altogether.

Practically, an adaptive QM/MM method is designed to reproduce the pure QM results for short distances and that of the MM model for longer distances, dependent on the choice of the transition radii. Note especially that the result cannot be better (e.g., closer to an experimental observation) than that from the pure QM method.

The RDF obtained with a conventional QM/MM simulation where a QM water molecule is immersed in MM water is presented in Figure 5b. The first peak is as high as that in the pure MM simulation, and this is the expected result because the nonbonded (van der Waals and electrostatic) interactions are determined by the force field terms. Contrary to that, the SCMP simulations reproduce the results from the pure DFTB3 simulations well, see Figures 5c−f. Note that the second peak is distinct in the [23, 48] and [31, 48] simulations, while it is partly captured in [19, 48], and it is missing in the [15, 48] simulation altogether.

According to the integrated O−O RDF from the pure DFTB3 simulation (see Figure 1), the 34th closest water molecule is 5.7 Å away from the center of the QM zone. Thus, in order to reproduce the second peak, it is important that the water molecules up to ca. the 35th closest should have considerable QM character, which may be quantified by the QM profiles. This is in line with the data in Figure 4, where the QM profile shows clear difference in this range among the [n, 48] simulations (n = 15, 19, 23, and 31).

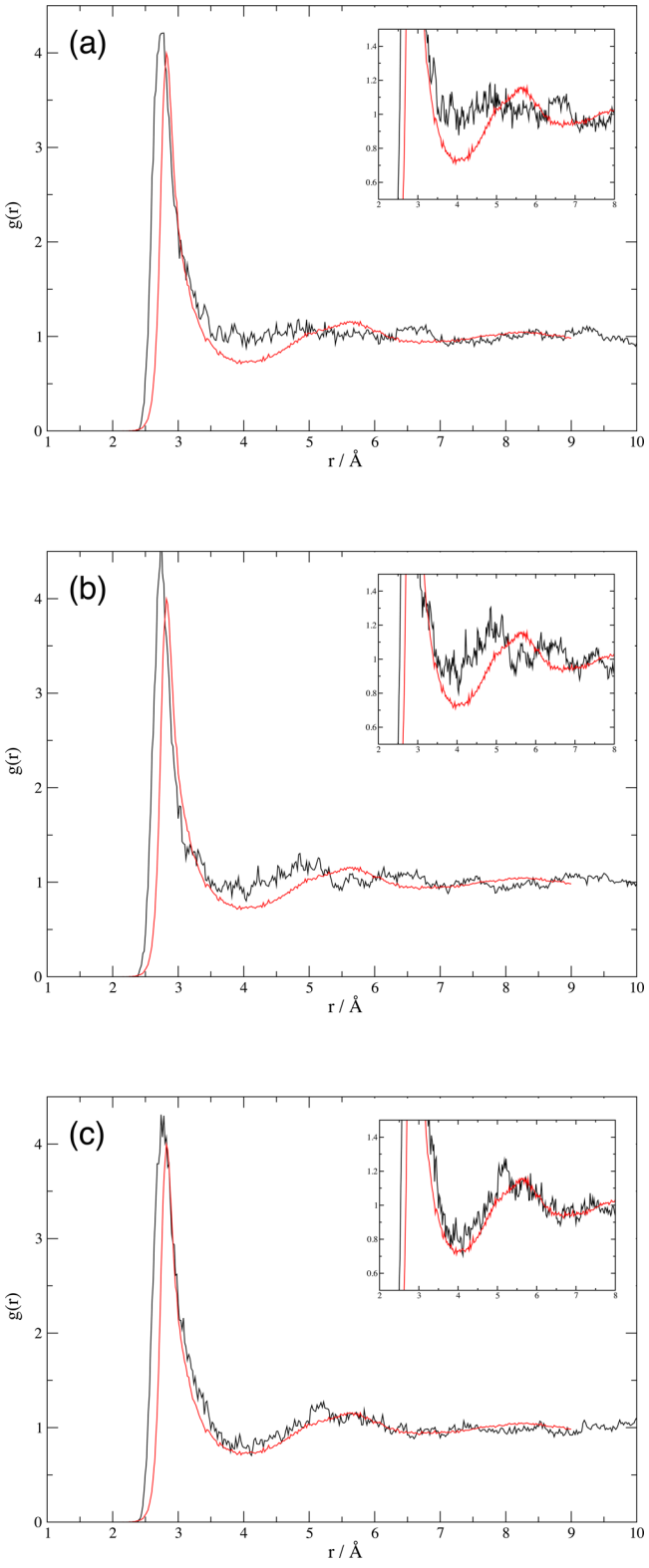


Figure 6. Effect of the number of different partitionings on the molecular structure: Oc-O RDF (Oc is central oxygen) in the SCMP simulations (a) [31, 32], (b) [31, 36], and (c) [31, 42]. Black lines are the data from each respective simulation; red line is a reference RDF of all oxygen atoms obtained from a "pure QM" (DFTB3) simulation.

Note that we did not observe any distorted structure reported in the previous works in the RDFs,[30,41] which is presumably caused by spatial and temporal discontinuities. Therefore, these

dx.doi.org/10.1021/ct5005593 | J. Chem. Theory Comput. 2014, 10, 4242−4252

results indicate the SCMP simulation will provide the most realistic description of solvation so far.

On the Effect of the Number of QM/MM Partitionings. To examine the effect of the number of partitionings considered on the structure of water, additional [31, n] simulations were performed with n = 32, 36, 42. As shown in Figure 6, the second peak in O−O RDF is not observed in the [31, 32] simulation and it is only partly formed in [31, 36], while it is reproduced in the [31, 42] simulation perfectly. The simulations [31, 36] and [31, 42] do not exhibit any pronounced difference in the QM profile distributions, as can be inferred from Figure 7.

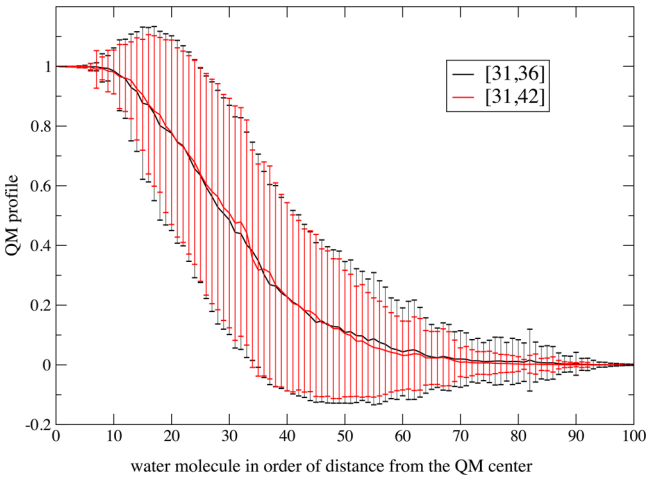


Figure 7. QM profiles of the nearest 100 water molecules observed in the SCMP simulations [31, 36] (black) and [31, 42] (red). (The errors bars represent standard deviation.)

Notably, there is large variance of QM profile in both of the simulations, and the error bars even extend beyond the interval (0,1). Since the QM profile is always between 0 and 1, this result implies a largely asymmetric distribution of the QM profiles.

It was also sought after any patterns in the distribution of QM profiles of the individual solvent molecules. To this end, a correlation matrix was built from the QM profiles; see Figure S1 in the Supporting Information. No apparent correlations between the QM profiles of the molecules were found in the simulations performed in this work.

Another property that may be monitored is the maximum weight $\sigma_{max}(t)$ among the partitioning set at a given time t in the simulation, and $\sigma_{max}(t)$ along the simulations [31, 36] and [31, 42] is presented in Figure 8. Note that a low value of $\sigma_{max}(t)$ implies that many partitionings contribute to the forces on atoms significantly, and the QM/MM switching involves a large number of molecules. In our simulations, $\sigma_{max}(t)$ occasionally took extensive values, more often in the simulation [31, 36] than in [31, 42]. Each of these events occurred after a series of partitioning updates were performed within a short time period. The occurrence of such a dominant partitioning means that the forces on atoms are determined by a single partitioning (the dominant one), and the spatial discontinuity of QM/MM is retained in such a case.

We conclude that the spatial discontinuity is still present in SCMP simulations performed with a smaller number of partitionings. Since the required number of partitionings to be considered will certainly vary depending on the type, number, and switching range of QM molecules, one may need to estimate the number of partitionings on the basis of pilot simulations.

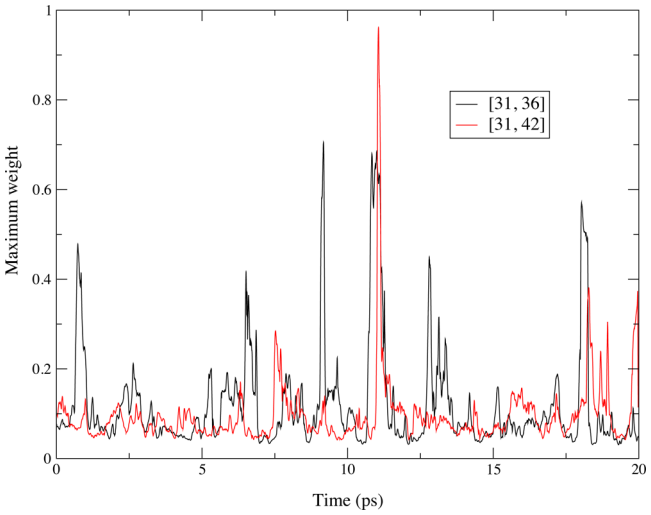


Figure 8. Maximum weight among partitionings in the SCMP simulations [31, 36] (black) and [31, 42] (red). Note that the smallest possible value of $\sigma_{max}(t)$ is $1/p$, where p is the number of partitionings; this corresponds to 0.028 and 0.024 for [31, 36] and [31, 42], respectively.

VII. CONCLUSIONS

We propose a new adaptive QM/MM method, the size-consistent multipartitioning (SCMP) QM/MM, where all partitionings involve the same numbers of QM and MM solvent molecules. Test simulations indicate that the SCMP method is a further development of previous multiscale AP methods in several respects: (1) better energy conservation, (2) consistence of the energy scale among the individual simulations, and (3) better computational efficiency. The application of SCMP to the "QM water in MM water" system revealed that QM calculations involving extended portions of the solvent are required for accurate treatment of the solvation. In addition, the QM profile analysis indicated that more than 30 solvent molecules should have a pronounced QM character in order to reproduce the second peak of the O−O RDF of water.

■ ASSOCIATED CONTENT

Supporting Information

One figure for a correlation matrix of QM profiles obtained from the [31, 32] simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Email: watanabe.h.au@m.titech.ac.jp.

Author Contributions

H.W. designed and coded the size-consistent multipartitioning QM/MM and performed all simulations with support for DFTB theory from M.E. T.K. implemented DFTB3 into GROMACS. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (FOR1279 "Protein Based Photoswitches") and research

fellowships of the Japan Society for the Promotion of Science for Young Scientists to H.C.W. The computations were performed at the computer centers of the Karlsruhe Institute of Technology. H.C.W gratefully acknowledges support from Dr. B. Aradi for DFTB+ usage.

■ ABBREVIATIONS

AP, adaptive partitioning; QM, quantum mechanics; MM, molecular mechanics; MD, molecular dynamics; SCMP, size-consistent multipartitioning; RDF, radial distribution function

■ REFERENCES

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103* (2), 227−249.
(2) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7* (6), 718−730.
(3) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11* (6), 700−733.
(4) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16* (9), 1170−1179.
(5) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.
(6) Gao, J. L.; Ma, S. H.; Major, D. T.; Nam, K.; Pu, J. Z.; Truhlar, D. G. *Chem. Rev.* **2006**, *106* (8), 3188−3209.
(7) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H. B.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G. H.; Xu, D. G.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110* (13), 6458−6469.
(8) Hu, H.; Yang, W. T. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.
(9) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48* (7), 1198−1229.
(10) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113* (5), 1253−1272.
(11) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100* (50), 19357−19363.
(12) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2* (3), 815−826.
(13) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902−3909.
(14) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13* (12), 1173−1213.
(15) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103* (6), 495−506.
(16) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58* (11), 7260−7268.
(17) Yang, Y.; Yu, H. B.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111* (42), 10861−10873.
(18) Gaus, M.; Cui, Q. A.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7* (4), 931−948.
(19) Seabra, G. D.; Walker, R. C.; Roitberg, A. E. *J. Phys. Chem. A* **2009**, *113* (43), 11938−11948.
(20) Csanyi, G.; Albaret, T.; Payne, M. C.; De Vita, A. *Phys. Rev. Lett.* **2004**, *93*, 175503.
(21) Kubar, T.; Elstner, M. *J. R. Soc., Interface* **2013**, *10* (87), 20130415.
(22) Kubar, T.; Elstner, M. *Phys. Chem. Chem. Phys.* **2013**, *15* (16), 5794−5813.
(23) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* **1996**, *211* (1−3), 313−323.
(24) Schwenk, C. F.; Loeffler, H. H.; Rode, B. M. *J. Am. Chem. Soc.* **2003**, *125* (6), 1618−1624.
(25) Bernstein, N.; Varnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csanyi, G. *Phys. Chem. Chem. Phys.* **2012**, *14* (2), 646−656.
(26) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355* (3−4), 257−262.
(27) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111* (9), 2231−2241.
(28) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *J. Chem. Theory Comput.* **2009**, *5* (9), 2212−2221.
(29) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. *Chem. Phys. Lett.* **2012**, *524*, 56−61.
(30) Bulo, R. E.; Michel, C.; Fleurat-Lessard, P.; Sautet, P. *J. Chem. Theory Comput.* **2013**, *9* (12), 5567−5577.
(31) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26* (16), 1701−1718.
(32) Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9* (1), 338−354.
(33) Teleman, O.; Jonsson, B.; Engstrom, S. *Mol. Phys.* **1987**, *60* (1), 193−203.
(34) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(35) Nose, S. *J. Chem. Phys.* **1984**, *81* (1), 511−519.
(36) Hoover, W. G.; Holian, B. L. *Phys. Lett. A* **1996**, *211* (5), 253−257.
(37) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111* (26), 5678−5684.
(38) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.
(39) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151.
(40) Goyal, P.; Qian, H.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2014**, DOI: 10.1021/jp503372v.
(41) Park, K.; Gotz, A. W.; Walker, R. C.; Paesani, F. *J. Chem. Theory Comput.* **2012**, *8* (8), 2868−2877.