JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Toward Hamiltonian Adaptive QM/MM: Accurate Solvent Structures Using Many-Body Potentials

Jelle M. Boereboom,[†] Raffaello Potestio,[‡] Davide Donadio,[¶,‡,§,∥] and Rosa E. Bulo*[,†]

[†]Inorganic Chemistry and Catalysis group, Debye Institute for Nanomaterials Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

[‡]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

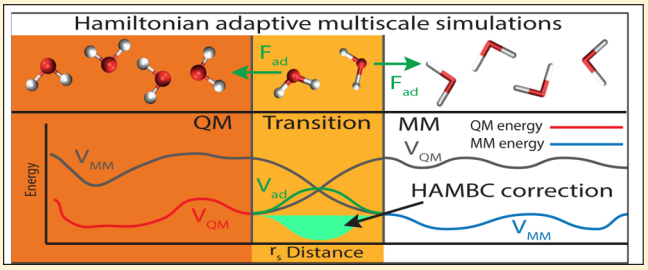
[¶]Department of Chemistry, University of California Davis, One Shields Avenue, Davis, California 95616, United States

[§]Donostia International Physics Center, Paseo Manuel de Lardizabal, 4, E-20018 Donostia-San Sebastian, Spain

[∥]IKERBASQUE, Basque Foundation for Science, E-48011 Bilbao, Spain

Supporting Information

ABSTRACT: Adaptive quantum mechanical (QM)/molecular mechanical (MM) methods enable efficient molecular simulations of chemistry in solution. Reactive subregions are modeled with an accurate QM potential energy expression while the rest of the system is described in a more approximate manner (MM). As solvent molecules diffuse in and out of the reactive region, they are gradually included into (and excluded from) the QM expression. It would be desirable to model such a system with a single adaptive Hamiltonian, but thus far this has resulted in distorted structures at the boundary between the two regions. Solving this long outstanding problem will allow microcanonical adaptive QM/MM simulations that can be used to obtain vibrational spectra and dynamical properties. The difficulty lies in the complex QM potential energy expression, with a many-body expansion that contains higher order terms. Here, we outline a Hamiltonian adaptive multiscale scheme within the framework of many-body potentials. The adaptive expressions are entirely general, and complementary to all standard (nonadaptive) QM/MM embedding schemes available. We demonstrate the merit of our approach on a molecular system defined by two different MM potentials (MM/MM′). For the long-range interactions a numerical scheme is used (particle mesh Ewald), which yields energy expressions that are many-body in nature. Our Hamiltonian approach is the first to provide both energy conservation and the correct solvent structure everywhere in this system.



1. INTRODUCTION

Molecular dynamics (MD) simulations of chemical processes in a complex environment can be significantly accelerated with a dual-resolution approach, modeling the region of interest (active or A-region, Figure 1) at high resolution (HR), while the environment (E-region, Figure 1) is modeled at lower resolution (LR).[1−3] Conventional dual-resolution approaches[4,5] define the active region as a preselected set of atoms. This strategy works well if the molecular system is rigid, but in a solute−solvent system HR solvent molecules readily diffuse away from the active region, to be replaced by LR solvent molecules. Adaptive resolution methods address this issue by dynamically assigning molecules HR or LR character based on their proximity to the active site.[6−18] Generally, this procedure involves a transition region that smoothly connects the active and environment regions (T-region, Figure 1). The solvent molecules in the T-region have partial HR and partial LR character, and the description of each solvent molecule s gradually changes from HR to LR (or vice versa) as it moves across the region. The fraction of HR character changes with the distance r_s from a predefined HR center (Figure 1, blue water molecule).

Adaptive resolution methods fall into two categories. The first category combines complex many-body potential energy expressions, of which the combination of quantum mechanics (QM) and molecular mechanics (MM) is the most common example (QM/MM).[6−13] Other examples include QM/QM, but also MM/MM, since some MM descriptions go beyond two-body terms. A very typical example is particle mesh Ewald (PME), which is often used to accelerate the computation of interactions across periodic boundaries, and cannot be reduced to a sum of pairwise interactions.[19,32] The second category combines potential energy expressions that can be reduced to a simple sum over particle pairs, such as most MM and coarse grained (CG) particle descriptions (MM/CG).[14−18] While the two problems are similar, MM/CG developments cannot always straightforwardly be extended to QM/MM, because the QM configurational potential in the HR region introduces

Received: February 25, 2016
Published: June 22, 2016

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448

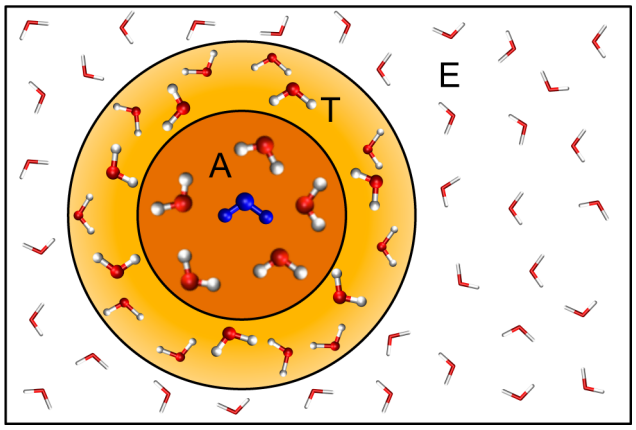


Figure 1. Schematic representation of an adaptive dual-resolution description of water-in-water, partitioned into an A-region [orange], a T-region [yellow], and an E-region [white], around a central blue water molecule. Ball and stick water molecules are HR, and LR molecules are depicted with thick lines. The HR character of a solvent molecule s is determined by its distance r_s from the HR center.

further layers of complexity, stemming from its many-body nature. Contrary to the MM and CG expressions, the QM potential energy contains an integral over all electronic degrees of freedom, and the multi-atom expansion of this integral contains higher order terms.

Currently, there are only few examples of adaptive QM/MM simulations based on a Hamiltonian formalism (energy conserving).[6,8] In all practical situations, the proposed methods failed to provide reliable solvent structures, while alternative non-Hamiltonian schemes performed much better in this respect.[8,20] Similar findings also directed MM/CG developments,[21] until recently, when a Hamiltonian approach was introduced (H-AdResS) that conserves both energy and solvent structure.[17] This approach relies on the MM/CG energy expression consisting of pairwise interactions only, and can therefore not be applied in combination with many-body potentials (e.g., QM potentials). In this work we exploit the concepts and formalism developed in the context Hamiltonian adaptive MM/CG[17] to develop a more general formulation that is applicable to adaptive QM/MM.[6,8,9] We first reformulate the energy and force expressions for the current Hamiltonian adaptive QM/MM approaches[6,8,9] in terms that match the simpler MM/CG expressions. We then derive a novel Hamiltonian scheme that is very general, and can connect any potential in an adaptive manner (e.g., QM/MM, QM/QM, (many-body) MM/MM, or MM/CG). The novel approach is an extension of current adaptive QM/MM formulations, and can be combined with many of the available flavors.[6,8] To achieve the desired generality, the combined expressions are by necessity different from the H-AdResS expressions, and even when used in combination with pair potentials they do not reduce to the H-AdResS expressions.

This paper is organized as follows: In section 2 we introduce the theory behind adaptive dual-resolution simulations. We first compare the latest adaptive MM/CG and QM/MM expressions, and we discuss the H-AdResS MM/CG correction to the Hamiltonian. We then rewrite the adaptive QM/MM (many-body) expressions so that we can define the criteria for a "per-particle" correction analogue to MM/CG. Finally, we discuss a simple correction that has been applied in previous works,[7,8,22] and then introduce our novel Hamiltonian adaptive many-body correction (HAMBC). In section 3, we describe the model system we use to validate our Hamiltonian scheme, and we provide the computational details. In section 4, we present the results of two possible per-particle corrections to the Hamiltonian, and show that the HAMBC correction results in accurate solvent structures, while conserving the total energy.

2. THEORY

The Hamiltonian of any system consists of a global kinetic energy function $T(\mathbf{p})$ (of momentum vector $\mathbf{p}$), and a potential energy function $V(\mathbf{r})$ (of coordinate vector $\mathbf{r}$) describing all particle−particle interactions in the system. In this section we first compare the potential energy expressions $V(\mathbf{r})$ currently used in adaptive MM/CG and QM/MM simulations. Then we explain why these uncorrected expressions produce inaccurate forces on the particles, and we describe the per-particle H-AdResS correction to the MM/CG Hamiltonian. We then deduce the criteria for an analogue per-particle correction to the many-body QM/MM Hamiltonian, and discuss a simple correction that has been previously applied. In the final subsection we derive our novel HAMBC correction from the criteria presented before.

2.1. Adaptive Dual-Resolution Potential Energy Expressions.

The H-AdResS approach combines an MM and a CG potential energy ($V^{\mathrm{MM}}(\mathbf{r})$ and $V^{\mathrm{CG}}(\mathbf{r})$) into a global $V^{\mathrm{MM/CG}}(\mathbf{r})$. Since $V^{\mathrm{MM}}(\mathbf{r})$ and $V^{\mathrm{CG}}(\mathbf{r})$ can both be expressed as a sum of pairwise interaction terms, the combined potential energy can also be written as a sum over particle pairs,

$$V^{\mathrm{MM/CG}}(\mathbf{r}) = \sum_{\alpha<\beta}\left(\frac{\lambda_\alpha+\lambda_\beta}{2}V^{\mathrm{MM}}_{\alpha\beta} + \left(1-\frac{\lambda_\alpha+\lambda_\beta}{2}\right)V^{\mathrm{CG}}_{\alpha\beta}\right) \tag{1}$$

where $V_{\alpha\beta} = V(\mathbf{r}_\alpha - \mathbf{r}_\beta)$ is an interaction potential for the particle pair $\alpha\beta$, and $\lambda_{\alpha/\beta} = \lambda(\mathbf{r}_{\alpha/\beta})$ is a simple continuous function of the distance $r_{\alpha/\beta}$ of particle α/β to the center of the A-region. Together, λ_α and λ_β determine the contribution (or weight) of the MM potential $V^{\mathrm{MM}}_{\alpha\beta}$, and each function the MM (HR) character of the corresponding particle.[17] The function $\lambda_{\alpha/\beta}$ has a value 1 if particle α/β is in the A-region, a value of 0 in the E-region, and a fractional number between 1 and 0 in the T-region.

In contrast, a many-body interaction potential cannot be decomposed into pair contributions, and the dual-resolution potential energy must be expressed in many-body terms. Conventional QM/MM methods partition the molecular system into a set of fully QM solvent molecules and a set of fully MM solvent molecules. Labeling a specific choice of QM/MM partitioning p, with the set of QM molecules named $\mathcal{S}_p$ and the complementary set of MM molecules $\mathcal{S}'_p$, the QM/MM potential energy becomes,

$$E_p(\mathbf{r}) = V^{\mathrm{QM}}_{\mathcal{S}_p} + V^{\mathrm{MM}}_{\mathcal{S}'_p} + V_{\mathrm{int}}(\mathbf{r}) \tag{2}$$

Here $V^{\mathrm{QM}}_{\mathcal{S}_p} = V^{\mathrm{QM}}(\mathbf{r}_{\mathcal{S}_p})$ is the QM potential energy for the subsystem of molecules belonging to the set $\mathcal{S}_p$, and $V_{\mathrm{int}}(\mathbf{r})$ is an interaction energy between the two types of molecules, which can be defined in several different ways.[23] Mechanical embedding and electrostatic embedding are the most common choices for this interaction energy, although many other options are available. The expression in eq 2 is completely general, covering all types of QM/MM embedding.

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448

An adaptive QM/MM simulation must account for the fact that all solvent molecules in the T-region have different partial QM character. This can be achieved by including contributions from different partitions into the adaptive potential energy expression.[6]

$$V^{\mathrm{QM/MM}}(\mathbf{r}) = \sum_{p\in\mathcal{P}} \sigma_p(\mathbf{r}) E_p(\mathbf{r}) \tag{3}$$

Here $\mathcal{P}$ is the set of all possible partitions p (2^n partitions in case of n solvent molecules). The function $\sigma_p(\mathbf{r})$ denotes the contribution (or weight) of partition p, and the sum over all weights equals 1 ($\sum_{p\in\mathcal{P}} \sigma_p(\mathbf{r}) = 1$). Each weight $\sigma_p(\mathbf{r})$ is a function of the coordinates of all atoms. The fractional QM character $\omega_s(\mathbf{r})$ of a solvent molecule with label s is then the sum of weights of the contributing partitions that describe this solvent molecule QM. It is the QM/MM analogue of $\lambda_{\alpha/\beta}$ in MM/CG, and can be written as,

$$\omega_s(\mathbf{r}) = \sum_{p\in\mathcal{P}} \delta_s(\mathcal{S}_p)\sigma_p(\mathbf{r}) \tag{4}$$

where $\delta_s(\mathcal{S}_p)$ is the Dirac measure, which is 1 if $s \in \mathcal{S}_p$ and 0 if $s \in \mathcal{S}'_p$. The concept of multiple partitions is schematically visualized in Figure 2 for the simple case of only two solvent molecules in the T-region. Partitions P_1−P_4 describe all solvent molecules in the A-region as QM and all molecules in the E-region as MM, and should therefore provide all important contributions to $V^{\mathrm{QM/MM}}(\mathbf{r})$. Partitions P_5 and P_6 describe a molecule in the E-region as QM, or a molecule in the A-region as MM, and should therefore not contribute. The general form of the energy expression in eq 3 is used in many of the current adaptive QM/MM methods.[6,8,9] The expressions only differ in the functional form of the weights $\sigma_p(\mathbf{r})$.

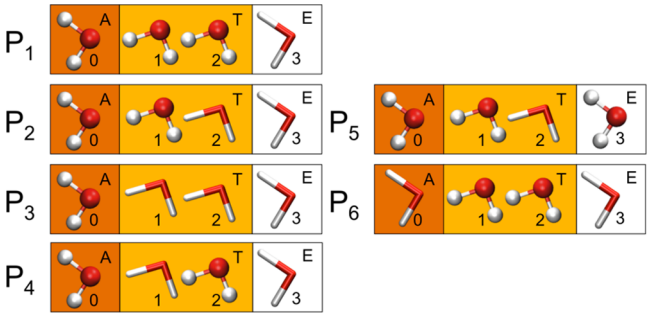


Figure 2. Six possible QM/MM partitions, with QM molecules depicted as ball and stick and MM molecules as thick lines. Partitions P_1−P_4 may all contribute to the adaptive QM/MM energy expression, while by construction P_5 and P_6 do not.

The adaptive QM/MM methods that use the general expression in eq 3 are Permuted Adaptive Partitioning (PAP),[6] Sorted Adaptive Partitioning (SAP),[6] Difference-based Adaptive Solvation (DAS),[8] and Size-Consistent Multi-Partitioning (SCMP).[9] The PAP method defines nonzero weight functions $\sigma_p(\mathbf{r})$ for all QM/MM partitions that describe the A-region molecules QM and the E-region molecules MM (P1−P4, Figure 2). This results in exponential scaling of the computational cost with the number of molecules in the T-region (M). The DAS method reduces the computational cost by assigning zero weight to a (much) larger number of partitions. In principle only contributions from "ordered" partitions (P1−P3, Figure 2) are nonzero, but to ensure continuity of the forces extra partitions are included if two solvent molecules are at a similar distance from the QM center. As a result, DAS scales approximately linear with M provided that the solvent structure in the T-region does not contain regions of extremely high density. Like DAS, the SAP weight-functions are only nonzero for ordered partitions. The SAP computational costs scale linearly with M in all cases, but this is achieved at the cost of the simplicity of the weight-functions. The SAP potential energy surface is very steep in places, and simulations require a small time-step for proper integration of the equations of motion. Finally, the SCMP weight functions are constructed in such a way that the number of contributing partitions is conserved throughout the simulation, and all contributing partitions have the same predetermined number of QM (HR) molecules. These partitions can yield a continuous energy as long as the total number of solvent molecules in the system is sufficiently large. An SCMP simulation on a parallel platform can exhibit nearly perfect linear scaling behavior.

2.2. Transition Forces. The forces derived from the adaptive QM/MM potential energy in eq 3 have the form,

$$\begin{aligned} F_i^{\mathrm{QM/MM}}(\mathbf{r}) &= -\sum_{p\in\mathcal{P}} \sigma_p(\mathbf{r}) \frac{\partial E_p(\mathbf{r})}{\partial x_i} - \sum_{p\in\mathcal{P}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} E_p(\mathbf{r}) \\ &= F_i^{\mathrm{ad}}(\mathbf{r}) + F_i^{\mathrm{tr}}(\mathbf{r}), \end{aligned} \tag{5}$$

where x_i is a component of the vector $\mathbf{r}$. This negative gradient of the potential energy is a sum of two terms: an adaptive force $F_i^{\mathrm{ad}}(\mathbf{r})$ and a transition force $F_i^{\mathrm{tr}}(\mathbf{r})$. The former term is a linear combination of force terms derived from a conventional QM/MM potential energy (eq 2), but the second term (the transition force) contains the gradients of the weights, and causes anomalies in the structure of the solvent.[8,20]

A pragmatic solution to the problem of the distorted structures is to discard the Hamiltonian formalism and simply neglect the offending term (e.g., use only $F_i^{\mathrm{ad}}(\mathbf{r})$ in eq 5 for propagation). It has been demonstrated repeatedly that, while such non-Hamiltonian simulations do not conserve energy (the applied forces do not integrate to a consistent potential energy[21]), they do provide reliable structures.[8,20,22,24] Nonetheless, giving up a Hamiltonian formalism has fundamental and practical consequences.[17] (1) Without a well-defined energy a partition function cannot be defined, and it is difficult to attach formal meaning to average values obtained from a simulation. In practice this disadvantage is somewhat diminished by the empirical observation that molecular structures obtained from these simulations are reliable. (2) The MD simulations require local thermostats with strong coupling to prevent heating, and the reliability of dynamical quantities (diffusion coefficients, time correlation functions) is severely compromised.[22] In practice, the primary added value of a Hamiltonian scheme lies in the possibility for the reliable acquisition of dynamical quantities.

A more rigorous solution is to adjust the Hamiltonian with an extra term in the potential energy expression of eq 3. The gradient of this term should then effectively cancel $F_i^{\mathrm{tr}}(\mathbf{r})$. The H-AdResS approach pursues the MM/CG equivalent of this strategy. The MM/CG forces derived from eq 1 have the form,

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448

$$F_{\alpha i}^{\mathrm{MM/CG}}(\mathbf{r}) = -\sum_{\beta}\left(\frac{\lambda_\alpha+\lambda_\beta}{2}\frac{\partial V_{\alpha\beta}^{\mathrm{MM}}}{\partial x_{i\alpha}} + \left(1-\frac{\lambda_\alpha+\lambda_\beta}{2}\right)\frac{\partial V_{\alpha\beta}^{\mathrm{CG}}}{\partial x_{i\alpha}}\right) - \frac{1}{2}\frac{\partial \lambda_\alpha}{\partial x_{i\alpha}}\sum_{\beta}[V_{\alpha\beta}^{\mathrm{MM}} - V_{\alpha\beta}^{\mathrm{CG}}] \quad (6)$$

where the last term, containing the gradient of the MM/CG weight function λ_α, is the equivalent of $F_i^{\mathrm{tr}}(\mathbf{r})$ in eq 5, and has similar catastrophic effects on the structure of the system. Unlike the QM/MM force expression in eq 5, the simpler MM/CG force expression in eq 6 clearly shows that the transition force scales linearly with the difference between the two types of interaction energy (MM and CG) of particle α with the rest of the system. This energy difference is the potential energy released or absorbed when particle α is converted from CG to MM at fixed conformation of the system,[17] or the vertical difference between two potential energy surfaces at a specific point $\mathbf{r}$. The transition force $F_{\alpha i}^{\mathrm{QM/MM}}(\mathbf{r})$ is a force driving the particle to become either QM or MM. For each particle α the H-AdResS approach applies a correction term to the potential energy expression in eq 1. This term is constructed such that its gradient effectively cancels the vertical energy release for particle α.

2.3. Per-Molecule Correction to a Many-Body Hamiltonian. In this work we show that we can arrive at a similar per-particle correction for the QM/MM potential energy expression in eq 3. The first step is to rewrite eq 5 into a form comparable to eq 6. We define an empty set of QM molecules $\mathcal{S}_0$ associated with the QM/MM potential energy $E_0(\mathbf{r})$ that describes the system fully MM. In eq 3 the sum of the weights $\sigma_p(\mathbf{r})$ equals 1 ($\sum_{p\in\mathcal{P}}\sigma_p(\mathbf{r}) = 1$), so that

$$\sigma_0(\mathbf{r}) = 1 - \sum_{\substack{p\in\mathcal{P}\\ p\neq 0}}\sigma_p(\mathbf{r})$$

We can therefore rewrite eq 3 as

$$V^{\mathrm{QM/MM}}(\mathbf{r}) = \sum_{\substack{p\in\mathcal{P}\\ p\neq 0}}\sigma_p(\mathbf{r})\,E_p(\mathbf{r}) + \left(1-\sum_{\substack{p\in\mathcal{P}\\ p\neq 0}}\sigma_p(\mathbf{r})\right)E_0(\mathbf{r}) \quad (7)$$

and the transition force defined in eq 5 as

$$F_i^{\mathrm{tr}}(\mathbf{r}) = -\sum_{\substack{p\in\mathcal{P}\\ p\neq 0}}\frac{\partial\sigma_p(\mathbf{r})}{\partial x_i}(E_p(\mathbf{r}) - E_0(\mathbf{r})) \quad (8)$$

Equation 8 shows that the transition force $F_i^{\mathrm{tr}}(\mathbf{r})$ is linearly dependent on the difference between the partition energies $E_p(\mathbf{r})$ and the fully MM potential energy $E_0(\mathbf{r})$. This can be seen as the energy release when all the QM solvent molecules in a set $\mathcal{S}_p$ are converted to MM. Analogously to the MM/CG expressions we can simply divide this term into per-molecule contributions, representing the energy released when the specific solvent molecule is converted from QM to MM,

$$\Delta E_p(\mathbf{r}) = E_p(\mathbf{r}) - E_0(\mathbf{r}) = \sum_{s\in\mathcal{S}_p}\Delta\epsilon_s^p(\mathbf{r}) \quad (9)$$

Note that $\Delta E_p(\mathbf{r})$ is different for each global geometry $\mathbf{r}$ and for each partition p, and that eq 9 assumes nothing about the definition of $\Delta\epsilon_s^p(\mathbf{r})$. In fact, there is no unique way to subdivide $\Delta E_p(\mathbf{r})$ into molecular contributions, since these depend on the order in which the molecules are converted to MM.

The intuitive solution is to assume equal contributions from all molecules s in eq 9. Earlier attempts at a Hamiltonian expression went one step further,[6,8] and also assumed similar contributions for each partition p and geometry $\mathbf{r}$, attempting to cancel each $\Delta\epsilon_s^p(\mathbf{r})$ with the same constant C. An appealingly simple choice for C is the difference in QM and MM energies for a single (gas phase) molecule at the geometries $\mathbf{r}_s^{\mathrm{QM}}$ and $\mathbf{r}_s^{\mathrm{MM}}$ optimized with the respective methods ($C = V^{\mathrm{QM}}(\mathbf{r}_s^{\mathrm{QM}}) - V^{\mathrm{MM}}(\mathbf{r}_s^{\mathrm{MM}})$). A corrected potential energy expression of the form,

$$\tilde{V}^{\mathrm{QM/MM}}(\mathbf{r}) = V^{\mathrm{QM/MM}}(\mathbf{r}) - \sum_{p\in\mathcal{P}}\sigma_p(\mathbf{r})\sum_{s\in\mathcal{S}_p}C \quad (10)$$

has the desired corrected gradient,

$$\tilde{F}_i^{\mathrm{QM/MM}}(\mathbf{r}) = F_i^{\mathrm{ad}}(\mathbf{r}) - \sum_{p\in\mathcal{P}}\frac{\partial\sigma_p(\mathbf{r})}{\partial x_i}(E_p(\mathbf{r}) - \sum_{s\in\mathcal{S}_p}C) \quad (11)$$

Substituting eq 8 in eq 11, and knowing that for partition $p = 0$ no correction term applies as $\mathcal{S}_0$ is an empty set, we can write

$$\tilde{F}_i^{\mathrm{QM/MM}}(\mathbf{r}) = F_i^{\mathrm{ad}}(\mathbf{r}) - \sum_{\substack{p\in\mathcal{P}\\ p\neq 0}}\frac{\partial\sigma_p(\mathbf{r})}{\partial x_i}(E_p(\mathbf{r}) - \sum_{s\in\mathcal{S}_p}C - E_0(\mathbf{r})) \quad (12)$$

If C is indeed a good representation of $\Delta\epsilon_s^p(\mathbf{r})$ for each geometry $\mathbf{r}$, QM/MM partition p, and solvent molecule s, then each $E_p(\mathbf{r})$ is corrected to be similar to $E_0(\mathbf{r})$, and the transition force becomes approximately zero. However, it has previously been shown that such a simple correction results in equilibrium solvent structures with a depletion of solvent molecules across the T-region.[7,8,22]

The depletion in the transition region can be explained using insights presented in the H-AdResS paper.[17] The authors suggested that the energy release upon conversion of a particle α from MM to CG ($\sum_\beta[V_{\alpha\beta}^{\mathrm{MM}} - V_{\alpha\beta}^{\mathrm{CG}}]$) depends strongly on the fractional MM (HR) character of $\alpha(\lambda_\alpha)$, which is reflected in its geometry. In our QM/MM simulations the analogue of λ_α is $\omega_s(\mathbf{r})$. For a molecule in the A-region with $\omega_s(\mathbf{r}) = 1$ and energy E_x, the geometry resembles the equilibrium QM geometry. Therefore, the MM potential energy of the molecule in this configuration will be higher than E_x and the energy increase upon converting molecule s from MM to QM will be negative ($\langle\Delta\epsilon_s^p(\mathbf{r})\rangle_{\omega_s=1} < 0$, Figure 3, black versus red line). The reverse

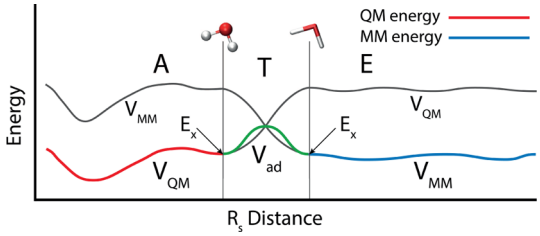


Figure 3. Schematic representation of QM, MM, and adaptive QM/MM potential energies after correction according to eq 10.

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448

argument holds for the energy of a molecule in the E-region, for which $\langle \Delta\epsilon_s^p(\mathbf{r})\rangle_{\omega_s=0} > 0$ (Figure 3, black versus blue line). This already explains why a simple correction using a constant C cannot be sufficient. In the T-region, the molecule has neither the equilibrium QM geometry, nor the equilibrium MM geometry. Because of these nonequilibrium geometries both the QM and the MM energies will have a value higher than E_x. Any combination of a QM and an MM description will therefore raise the energy above E_x, which is reflected in the higher energies of particles in the T-region (Figure 3), and leads to a local depletion of the water density in this region.

2.4. The HAMBC Correction. The above argument implies a trend in the error of the water densities that depends on the distance of the water molecules to the QM center. This distance is directly related to the QM-character of the water molecules (1 at short distances, 0 at large distances). Such a trend complies with insights put forward in ref 17, stating that the energy release upon conversion of a particle α from MM to CG ($\sum_\beta[V_{\alpha\beta}^{\mathrm{MM}} - V_{\alpha\beta}^{\mathrm{CG}}]$, eq 6) depends strongly on its fractional MM character as reflected in its geometry. Similar to ref 17 the final step toward a working Hamiltonian approach is to approximate $\sum_{s\in\mathcal{S}_p}\Delta\epsilon_s^p(\mathbf{r})$ with an expression that depends on the QM character $\omega_s(\mathbf{r})$ of each molecule s. The chosen per-molecule correction term is an (ensemble) average over all coordinates $\mathbf{r}$, partitions p, and solvent molecules s with QM character ω_s: $\langle \Delta\epsilon_s^p(\mathbf{r})\rangle_{\omega_s}$. In the following, we will show that correcting the per-molecule energy release $\Delta\epsilon_s^p(\mathbf{r})$ with this average is equivalent to correcting the potential energy expression $V^{\mathrm{QM/MM}}(\mathbf{r})$ with a function $H^{\mathrm{c}}(\omega_s)$ for each molecule s, as long as $\frac{\mathrm{dH^c}(\omega_s)}{d\omega_s}$ equals $\langle \Delta\epsilon_s^p(\mathbf{r})\rangle_{\omega_s}$. For a system of n solvent molecules,

$$\hat{V}^{\mathrm{QM/MM}}(\mathbf{r}) = V^{\mathrm{QM/MM}}(\mathbf{r}) - \sum_{s=1}^{n} H^{\mathrm{c}}(\omega_s) \tag{13}$$

The forces derived from this expression have the form,

$$\hat{F}_i^{\mathrm{QM/MM}}(\mathbf{r}) = F_i^{\mathrm{ad}}(\mathbf{r}) - \sum_{p\in\mathcal{P}} \frac{\partial\sigma_p(\mathbf{r})}{\partial x_i} E_p(\mathbf{r}) + \sum_{s=1}^{n} \frac{\mathrm{d}H^{\mathrm{c}}(\omega_s)}{\mathrm{d}\omega_s}\frac{\partial\omega_s(\mathbf{r})}{\partial x_i} \tag{14}$$

Since $\omega_s(\mathbf{r})$ is a sum over $\sigma_p(\mathbf{r})$ values for those partitions that describe molecule s QM (eq 4), we can express its gradient to x_i as follows,

$$\sum_{s=1}^{n}\frac{\partial\omega_s(\mathbf{r})}{\partial x_i} = \sum_{s=1}^{n}\sum_{p\in\mathcal{P}}\delta_s(\mathcal{S}_p)\frac{\partial\sigma_p(\mathbf{r})}{\partial x_i} = \sum_{p\in\mathcal{P}}\frac{\partial\sigma_p(\mathbf{r})}{\partial x_i}\sum_{s\in\mathcal{S}_p} 1 \tag{15}$$

Inserting eq 15 into eq 14 we obtain,

$$\begin{aligned}\hat{F}_i^{\mathrm{QM/MM}}(\mathbf{r}) &= F_i^{\mathrm{ad}}(\mathbf{r}) - \sum_{p\in\mathcal{P}}\frac{\partial\sigma_p(\mathbf{r})}{\partial x_i}\left(E_p(\mathbf{r}) - \sum_{s\in\mathcal{S}_p}\frac{\mathrm{d}H^{\mathrm{c}}(\omega_s)}{\mathrm{d}\omega_s}\right)\\ &= F_i^{\mathrm{ad}}(\mathbf{r}) - \sum_{\substack{p\in\mathcal{P}\\ p\neq 0}}\frac{\partial\sigma_p(\mathbf{r})}{\partial x_i}\left(E_p(\mathbf{r}) - \sum_{s\in\mathcal{S}_p}\frac{\mathrm{d}H^{\mathrm{c}}(\omega_s)}{\mathrm{d}\omega_s} - E_0(\mathbf{r})\right)\end{aligned} \tag{16}$$

If $\frac{\mathrm{dH^c}(\omega_s)}{d\omega_s}$ is a good estimate for the $\Delta\epsilon_s^p(\mathbf{r})$ value of a solvent molecule s with QM character $\omega_s(\mathbf{r})$ at any geometry $\mathbf{r}$ accessed during the simulation, then $E_p(\mathbf{r})$ is corrected toward the MM potential energy $E_0(\mathbf{r})$, and the transition force $F_i^{\mathrm{tr}}(\mathbf{r})$ becomes approximately zero.

Since we know that the NVT ensemble of structures is well-reproduced by a non-Hamiltonian simulation using the force expression $F_i^{\mathrm{ad}}(\mathbf{r})$ in eq 5,[24] the ensemble average $\frac{\mathrm{d}H^{\mathrm{c}}(\omega_s)}{\mathrm{d}\omega_s} = \langle\Delta\epsilon_s^p(\mathbf{r})\rangle_{\omega_s}$ can be extracted from such a non-Hamiltonian simulation. The energy differences ($\Delta\epsilon_s^p(\mathbf{r})$) that are obtained from the non-Hamiltonian simulation are the differences in energy between two QM/MM partitions that differ only in the description of one solvent molecule. For example, in Figure 2 the value of $\Delta\epsilon_2^{\mathrm{P1}}(\mathbf{r})$ for solvent molecule 2 and partition P1 is defined as the difference between $E_{\mathrm{P1}}(\mathbf{r})$ and $E_{\mathrm{P2}}(\mathbf{r})$. The correction term in the energy $H^{\mathrm{c}}(\omega_s)$ (eq 15) can be obtained by thermodynamic integration[25] of this average over ω_s,

$$H^{\mathrm{c}}(\omega_s) = \int_0^{\omega_s}\langle\Delta\epsilon_s^p(\mathbf{r})\rangle_{\omega_s'}\,\mathrm{d}\omega_s' \tag{17}$$

The above integral is by definition zero for solvent molecules that are purely MM ($\omega_s = 0$).

3. COMPUTATIONAL DETAILS

We demonstrate the performance of our approach on a test system of water in water, using the adaptive QM/MM weight-functions $\sigma_p(\mathbf{r})$ as formulated in the SAP method.[6] The SAP method has been selected in this work for practical reasons. The number of contributing partitions to the SAP energy expression always equals M+1, while the DAS method,[8] due to the construction of the weight functions, includes more partitions when the solvent structure is extremely distorted in the T-region (e.g., Hamiltonian simulations with a simple correction). Note that for simulations with a homogeneous solvent structure the SAP and DAS methods compute an equal number of energy terms $E_p(\mathbf{r})$ (eq 3). All simulations are performed with FlexMD, a python library that serves as a wrapper around several molecular program packages, each providing the required QM or MM energies and forces. FlexMD itself is distributed with the ADF program package,[26,27] and uses the atomistic simulation environment (ASE)[28] for MD propagation. Our model system is a 30.025 Å periodic simulation box containing 915 water molecules. The two selected potentials for our test simulations are both MM potentials, where we chose the SPC/Fw force-field to describe the central A-region (Figure 1),[29] while the E-region is described with the TIP3P/Fs force-field.[29,30] SPC/Fw and TIP3P/Fs energies and forces are both computed with the NAMD program,[31] and the many-body character of both descriptions is introduced in the computation of the long-range electrostatics, for which PME is used.[32] For the 'QM'/MM interaction $V_{\mathrm{int}}(\mathbf{r})$ (eq 2) we use two different approaches. The first is the IMOMM approach, which equates simple mechanical embedding,[33,34] and effectively means that $V_{\mathrm{int}}(\mathbf{r})$ consists only of TIP3P/Fs interaction terms. The second approach is electrostatic embedding, which differs from the IMOMM approach in the Coulombic component of $V_{\mathrm{int}}(\mathbf{r})$ in eq 2. This expression now consists of pairwise interaction terms between SPC/Fw charges for the $\mathcal{S}_p$ molecules, and TIP3P/Fs charges for the $\mathcal{S}'_p$ molecules. In both cases the active region

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448

(A-region) is a 5.5 Å sphere around a central water molecule, while the transition region (T-region) is a 0.9 Å thick layer around the A-region. We performed an additional set of IMOMM simulations with a smaller A-region of 4.0 Å, and a T-region of 0.9 Å thickness.

The IMOMM system with the large A-region was first equilibrated for 10 ps (for 5 seeds) using the non-Hamiltonian approach (eq 5) at a water density of 1.01 g/mL in the canonical ensemble (NVT). A Langevin thermostat was used with a friction of 82.7 ps^{-1}. The equilibrated structures were used as starting geometries for five non-Hamiltonian simulations of 10 ps each. These simulations also use a Langevin thermostat with a friction of 82.7 ps^{-1}. All five simulations have varying starting velocities, which were randomly generated. The results from these five simulations were combined to obtain the HAMBC correction term $H^c(\omega_s)$. At each time step the QM character of each water molecule in the transition region was determined, and based on this it was assigned to a bin of width 0.1 in ω_s. Subsequently, the (two) partitions were selected that differed only in the description of this water molecule. The energy difference between the two partitions was then extracted, and the value was stored in the bin. When all 50 ps of data was collected, the values in each bin were averaged, and $H^c(\omega_s)$ was computed by numeric integration (eq 17), followed by spline fitting to interpolate the points. The $g(r)$ from the non-Hamiltonian simulations was used as the reference for all other results. The equilibrated structures were also used as starting geometries for 5 ps NVE simulations (0.1 fs time-step) using the Hamiltonian expressions $\tilde{V}^{QM/MM}(\mathbf{r})$ (eq 10) and $\hat{V}^{QM/MM}(\mathbf{r})$ (eq 13). Finally, five 20 ps Hamiltonian NVT simulations were performed (of which the first 10 ps is the equilibration) with the simple correction term (eq 10), using random starting velocities and a 0.5 fs time-step. The same procedure was repeated with the HAMBC correction of eq 13. For all NVT simulations a time-step of 0.5 fs was used, and it was verified that this yielded the same equilibrated structures as the NVE simulations with a smaller time-step.

The entire above procedure was repeated using a smaller A-region (4.0 Å) The only simulations omitted were the simulations using $\tilde{V}^{QM/MM}(\mathbf{r})$ (eq 10) and the NVE simulations. The same was done with an electrostatic embedding setup for the interactions between the two sets of molecules in the partitions.

4. RESULTS

4.1. Mechanical Embedding (IMOMM). As demonstrated previously,[8,20,22,24] the non-Hamiltonian simulations (using only $F_i^{ad}(\mathbf{r})$ to propagate the trajectory) result in a radial distribution ($g(r)$) of water-oxygens around the central oxygen atom O^* that is very similar to the O−O radial distribution of the reference (in this case SPC-Fw and TIP3P-fs[29,30]). The non-Hamiltonian IMOMM simulation with a large A-region produces a first sharp peak at 2.8 Å from the central oxygen atom, and two much shallower peaks at 4.4 and 6.7 Å, respectively (Figure 4, solid red line). We use this result as a reference for the performance of the Hamiltonian simulations.

The corrected Hamiltonian approach using $\tilde{V}^{QM/MM}(\mathbf{r})$ (eq 10) results in an O*−O radial distribution (Figure 4, dotted green line) that deviates strongly from the reference. The main difference is a depletion of the water density in the T-region,

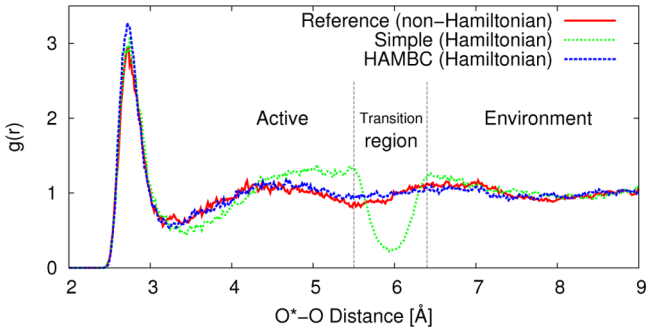


Figure 4. Radial distribution of oxygen around the central "QM" water oxygen atom O* for three different simulations: Non-Hamiltonian [red], Hamiltonian using a simple constant correction [green, dotted], and HAMBC [blue, dashed]. All three simulations use the IMOMM scheme to describe the interactions between the two types of molecules.

which is balanced by an increased density at the edges of the A- and E-regions. The water densities on either side of the T-region are similar, indicating that there is no significant chemical potential difference between a "QM" water molecule in the A-region and an MM water molecule in the E-region. The correct density balance between the A- and E-regions is the result of the correction C, which proves to be a good approximation for the chemical potential difference between a QM and an MM water molecule at their respective equilibrium geometries.

The HAMBC energy correction $H^c(\omega_s)$ obtained from the 50 ps of non-Hamiltonian simulation is depicted in Figure S1 of the Supporting Information. The correction energy converged to within 0.002 kcal/mol in approximately 40 ps (Figure S2, Supporting Information). The blue line in Figure 4 is the O*−O $g(r)$ resulting from a simulation using this energy correction. We extracted the values of the correction force and the transition force for 17 water molecules over a short segment of the simulation (450 fs). We found that the instantaneous value of the transition force oscillates strongly around the (negative of the) correction force, with an average root-mean-square deviation of 3.7 kcal/mol. The average of the transition force over a short time interval of 50 fs deviates very little from the negative correction force with a root-mean-square error of only 0.8 kcal/mol. The resulting $g(r)$ clearly indicates that the average compensation is enough to cancel the effect of the transition forces $F_i^{tr}(\mathbf{r})$ such that the correct solvent structure is obtained everywhere in the solution. In the Supporting Information we plotted the behavior of the transition and correction forces in Figure S3, and the trajectory of the "QM" character of a few selected water molecules in Figure S4.

The merit of our novel Hamiltonian approach is further confirmed with microcanonical (NVE) simulations that reveal no significant drift in total energy (no more than 0.005 kcal mol^{-1} ps^{-1}), which can be seen in Figure 5. This is in stark contrast to the total energy drift in the non-Hamiltonian simulations of 10 kcal mol^{-1} ps^{-1}.

The HAMBC simulation with the smaller A-region yielded very similar structures, which agreed well with the corresponding non-Hamiltonian reference (Figure S5 of the Supporting Information).

4.2. Electrostatic Embedding. In an electrostatic embedding simulation the conversion of a molecule from one description to another involves not only a changing interaction

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448

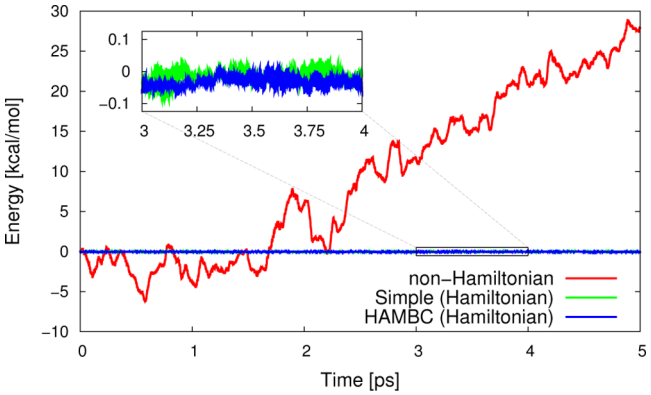


Figure 5. Total energy of the non-Hamiltonian [red], Hamiltonian with simple constant correction [green], and HAMBC [blue] simulations against simulation time.

with the molecules in $\mathcal{S}_p$ (like IMOMM), but also with the molecules in $\mathcal{S}'_p$. We therefore expected these simulations to produce slightly larger values for $\langle \Delta\epsilon_s^p(\mathbf{r}) \rangle_{\omega_s}$ than the IMOMM simulations. Instead, we observe that the "QM" description is slightly destabilized with respect to the MM description compared to the IMOMM result with an average value of 0.16 (±0.02) kcal/mol. We attribute this destabilization to the Coulombic interaction of the water molecules in $\mathcal{S}_p$ with a large reservoir of $\mathcal{S}'_p$ water molecules with charges that are not optimal for this force field. The destabilization is independent of the value ω_s of water molecule s, indicating that the water molecules in the A-region have very little possibility to accommodate the new interactions by adjusting their geometry. The increase in the energies of the individual MM molecules ($\mathcal{S}'_p$) is smaller, because they only experience an unfavorable interaction with a small number of $\mathcal{S}_p$ molecules. In complete analogy with IMOMM simulations the HAMBC correction forces nicely compensate the transition forces, resulting in a $g(r)$ that matches the non-Hamiltonian reference (Figure S7, Supporting Information).

5. CONCLUSIONS

In summary, we propose a new Hamiltonian adaptive dual-resolution approach that combines two many-body potentials and is able to correctly describe the structure of a molecular solution while simultaneously conserving the total energy. The forces from an uncorrected adaptive Hamiltonian consist of two terms, $F_i^{\mathrm{ad}}(\mathbf{r})$ and $F_i^{\mathrm{tr}}(\mathbf{r})$. The former term by itself provides the correct equilibrium solvent structures, but the latter term causes artifacts. A crucial step in our derivation is the separation of the transition force $F_i^{\mathrm{tr}}(\mathbf{r})$ into characteristic contributions from each solvent molecule. We then introduce a per−molecule correction term that is a function of the high-resolution character of the molecule and that cancels the undesirable transition force. When we extract the HAMBC correction term from a non-Hamiltonian adaptive multiscale simulation, the resulting Hamiltonian can be used in simulations that conserve energy and preserve the solvent structure throughout the molecular system. The new approach thus enables micro-canonical simulations that provide meaningful fluctuations and response functions, and can be used to obtain vibrational spectra and other dynamical properties.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00205.

Details about the HAMBC correction energy applied in the mechanical embedding simulation; plot of the $g(r)$ for the simulations with smaller active and smaller transition regions; description of a simulation with a strongly altered potential, and a plot of the resulting correction energy and $g(r)$; description of the electrostatic embedding simulation details, and a plot of the corresponding correction energy and $g(r)$ (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: R.E.Bulo@uu.nl.

Funding

The research was supported by The Netherlands Organization for Scientific Research (NWO) (Vidi 723.012.104). R.P. and D.D. acknowledge financial support under the project SFB-TRR146 of the Deutsche Forschung Gemeinschaft.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank T. Jiang for fruitful discussions.

■ REFERENCES

(1) Gao, J. L.; Xia, X. F. *J. Am. Chem. Soc.* **1993**, *115*, 9667−9675.
(2) Gao, J. L. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(3) Park, K.; Gotz, A. W.; Walker, R. C.; Paesani, F. *J. Chem. Theory Comput.* **2012**, *8*, 2868−2877.
(4) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(5) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(6) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231−2241.
(7) Pezeshki, S.; Lin, H. *J. Comput. Chem.* **2014**, *35*, 1778−1788.
(8) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *J. Chem. Theory Comput.* **2009**, *5*, 2212−2221.
(9) Watanabe, H. C.; Kubař, T.; Elstner, M. *J. Chem. Theory Comput.* **2014**, *10*, 4242−4252.
(10) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* **1996**, *211*, 313−323.
(11) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355*, 257−262.
(12) Bernstein, N.; Varnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csanyi, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 646−656.
(13) Böckmann, M.; Doltsinis, N. L.; Marx, D. *J. Chem. Theory Comput.* **2015**, *11*, 2429−2439.
(14) Praprotnik, M.; Delle Site, L.; Kremer, K. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545.
(15) Heyden, A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 217.
(16) Park, J. H.; Heyden, A. *Mol. Simul.* **2009**, *35*, 962.
(17) Potestio, R.; Fritsch, S.; Español, P.; Delgado-Buscalioni, R.; Kremer, K.; Everaers, R.; Donadio, D. *Phys. Rev. Lett.* **2013**, *110*, 108301.
(18) Praprotnik, M.; Delle Site, L.; Kremer, K. J. *J. Chem. Phys.* **2005**, *123*, 224106.
(19) Ewald, P. P. *Ann. Phys.* **1921**, *369*, 253−287.
(20) Pezeshki, S.; Davis, C.; Heyden, A.; Lin, H. *J. Chem. Theory Comput.* **2014**, *10*, 4765−4776.
(21) Delle Site, L. *Phys. Rev. E* **2007**, *76*, 047701.

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448

(22) Bulo, R. E.; Michel, C.; Fleurat-Lessard, P.; Sautet, P. *J. Chem. Theory Comput.* **2013**, *9*, 5567−5577.
(23) Pezeshki, S.; Lin, H. *Mol. Simul.* **2015**, *41*, 168−189.
(24) Kreis, K.; Donadio, D.; Kremer, K.; Potestio, R. *Europhys. Lett.* **2014**, *108*, 30007.
(25) Kirkwood, J. *J. Chem. Phys.* **1935**, *3*, 300.
(26) *ADF2016, SCM, Theoretical Chemistry*; Vrije Universiteit: Amsterdam, The Netherlands, 2016; http://www.scm.com.
(27) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931−967.
(28) Bahn, S.; Jacobsen, K. W. *Comput. Sci. Eng.* **2002**, *4*, 56−66.
(29) Wu, Y.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, *124*, 024503.
(30) Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **1999**, *111*, 9361−9381.
(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kal, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(32) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(33) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170−1179.
(34) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

DOI: 10.1021/acs.jctc.6b00205
J. Chem. Theory Comput. 2016, 12, 3441−3448