MiMiC: A Novel Framework for Multiscale Modeling in Computational Chemistry

Jógvan Magnus Haugaard Olsen,[*,†] Viacheslav Bolnykh,[‡,¶,§] Simone Meloni,[‖] Emiliano Ippoliti,[§] Martin P. Bircher,[⊥] Paolo Carloni,[§,#,‡] and Ursula Rothlisberger[⊥]

†*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway*

‡*Department of Physics, RWTH Aachen University, 52056 Aachen, Germany*

¶*CaSToRC, The Cyprus Institute, 2121 Aglantzia, Nicosia, Cyprus*

§*Institute for Advanced Simulation (IAS-5) and Institute of Neuroscience and Medicine (INM-9), Forschungszentrum Jülich, 52425 Jülich, Germany*

‖*Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, 00184 Rome, Italy*

⊥*Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

#*Institute for Neuroscience and Medicine (INM-11), Forschungszentrum Jülich, 52425 Jülich, Germany*

E-mail: jogvan.m.olsen@uit.no

Abstract

We present a flexible and efficient framework for multiscale modeling in computational chemistry (MiMiC). It is based on a multiple-program multiple-data (MPMD)

model with loosely coupled programs. Fast data exchange between programs is achieved through the use of MPI intercommunicators. This allows exploiting the existing parallelization strategies used by the coupled programs while maintaining a high degree of flexibility. MiMiC has been used in a new electrostatic embedding quantum mechanics/molecular mechanics (QM/MM) implementation coupling the highly efficient CPMD and GROMACS programs but it can also be extended to use other programs. The framework can also be utilized to extend the partitioning of the system into several domains that can be treated using different models, such as models based on wave function or density functional theory as well as coarse-graining and continuum models. The new QM/MM implementation treats long-range electrostatic QM–MM interactions through the multipoles of the QM subsystem which substantially reduces the computational cost without loss of accuracy compared to an exact treatment. This enables QM/MM molecular dynamics (MD) simulations of very large systems.

1 Introduction

Multiscale simulations have become important tools in many fields of chemistry, physics, and biology. In particular, quantum mechanics / molecular mechanics (QM/MM) approaches are widely used to investigate large and complex molecular systems where quantum effects of a subdomain play a key role.[1–4] Examples include (bio)chemical reactions[5–8] and light–matter interactions[9–11] that require explicit modeling of the electronic structure of the active parts of the total system. The rest of the system is modeled using a molecular mechanics description, in which the effect of changes in the electronic structure is taken into account via an effective potential depending only on the atomic positions. However, depending on the physical or chemical phenomena being studied and the level of accuracy that is needed, different levels of approximation may be employed, ranging from a fully quantum mechanical (QM) description, over atomistic molecular mechanics (MM) and coarse-grained (CG) modeling, to continuum mechanics (CM). Multiscale approaches take advantage of the fact that it is

often sufficient to use the more fine-grained and accurate models on a smaller subsystem of a complete molecular system while the remainder can be modeled using coarser, less accurate, but computationally efficient models.

The implementation of multiscale methods follows different strategies, each with their own advantages and limitations. The addition of multiscale modeling capabilities to existing programs can follow two principal strategies for coupling the methods, which can be categorized in terms of tight and loose coupling approaches. In their extreme versions, a tight coupling strategy has the methods integrated in a single monolithic program with a high degree of interdependency, whereas a loose coupling strategy avoids interdependencies as far as possible by implementing the individual methods in separate programs or by providing either part in a library.

Using the tight coupling strategy has the immediate advantage that it is, in some respects, more straightforward and easier to optimize compared to loose coupling. For example, data exchange between the parts of the program that treat different subsystems becomes simple and efficient. However, it can lead to considerably higher long-term maintenance efforts because of the larger code base. Extending the multiscale capabilities of a tightly coupled implementation, e.g. adding support for new QM methods or MM force fields, can involve substantial implementation effort and often results in duplicated functionality that is already present in other programs. Moreover, multiscale capability is often added to existing programs that are otherwise dedicated to a specific model and thus also optimized with that model in mind.[12–19] This, combined with the fact that the best parallelization strategy for different models can be different, can complicate the parallel optimization of multiscale methods in monolithic programs. The advent of massively parallel computers, with millions of cores, and hybrid architectures (e.g. hybrid CPU/GPU computers) makes it even more challenging to achieve high performance in the tight coupling strategy. Exploiting the evolution of computer architectures within the monolithic tight coupling paradigm may prove particularly challenging in such a context.

Loose coupling has the implicit advantage of providing great flexibility and low maintenance efforts but typically requires more efforts in the planning stage. Using this approach, it is relatively easy to couple independent programs, thus giving access to new features as they become available in a program, potentially without requiring any modifications in the other coupled programs. This substantially accelerates the availability of new methodological and technical developments. Moreover, the increased flexibility also facilitates the development of new multiscale approaches that go beyond QM/MM. One limiting factor, however, is that computational efficiency can be hampered by the communication between the coupled programs. For example, in the file-based approach, in which the communication between programs is handled by writing/reading files, there is computational overhead stemming from file input/output operations and initialization/termination of external programs.[20] This is the general path taken by several popular classical MD packages, such as AMBER,[21] GROMACS,[22] and NAMD,[23] as well as the PUPIL system[24] and ChemShell environment.[25,26] However, AMBER also provides a more efficient interface based on message passing interface (MPI), and ChemShell has the option to directly link certain external programs in order to increase computational efficiency.

In this work, we present a pilot implementation of a new framework for multiscale modeling in computational chemistry (MiMiC), with the goal to combine the best of both worlds: flexibility and efficiency. To this end, we have designed a strategy based on a multiple-program multiple-data (MPMD) model with loose coupling between programs (see illustration in Figure 1). In this strategy, MiMiC couples a main driver, which runs the molecular dynamics (MD) simulation, to a set of external programs, each of which concurrently computes the contributions that are relevant to a specific subsystem using their own optimal parallelization strategy. The aim is to support a wide range of models, such as QM, MM, CG, and CM, and also to provide the necessary functionality to compute the subsystem interaction terms when needed. Efficient communication between programs is achieved through the use of a lightweight communication library (CommLib), which was developed in the

context of this work. CommLib uses MPI intercommunicators and thus gives access to high-speed interconnects, such as InfiniBand and Omni-Path Architecture.

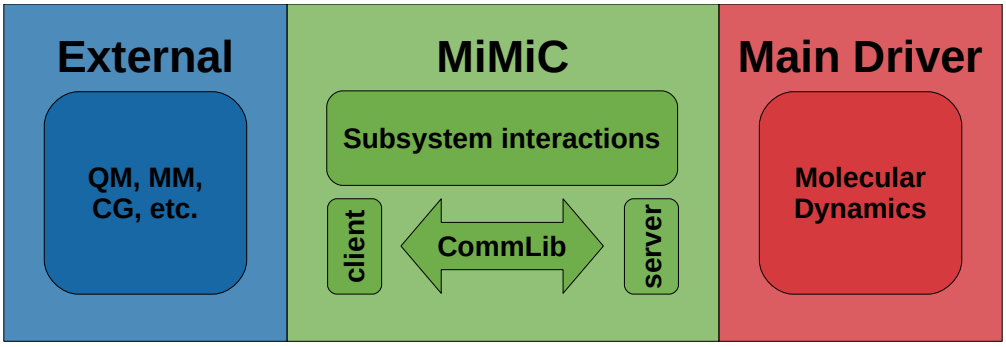


Figure 1: Illustration of the strategy used in the MiMiC framework.

We demonstrate the prowess of the MiMiC framework in a new efficient implementation of electrostatic embedding QM/MM by combining the highly efficient CPMD[27] and GROMACS[28,29] programs. This new QM/MM implementation features a generalized version of the electrostatic coupling scheme by Laio et al.[30] which greatly accelerates the computation of QM–MM interaction terms potentially without any loss of accuracy. In the following, we provide a theoretical description of the QM/MM approach used in this work followed by an overview of the MiMiC framework. Finally, we present results from illustrative applications of the methodology before concluding with an outlook.

2 QM/MM Methodology

In the following, we will consider a system that is composed of two separable subsystems: a QM subsystem, where the electronic and nuclear interactions are described by Kohn–Sham density functional theory (KS-DFT), and an MM subsystem, in which atoms interact according to an MM force field. We use an electrostatic embedding QM/MM formalism in which the KS equations must be solved including the external field that is generated by the MM atoms. Typically, the electrostatic force field parameters consist of point charges but can also include higher-order multipoles. The electron density of the QM subsystem is

thus directly polarized by the atoms in the MM subsystem. The nuclei in the QM domain are propagated as classical point particles by using either the Born–Oppenheimer (BO) or Car–Parrinello (CP)[31] approach. The Hamiltonian for this system can be written as

$$H_{\mathrm{tot}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}},\mathbf{R}^{\mathrm{MM}}) = H_{\mathrm{QM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}) + H_{\mathrm{MM}}(\mathbf{R}^{\mathrm{MM}}) + H_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}},\mathbf{R}^{\mathrm{MM}}) \quad (1)$$

where $\mathbf{R}^{\mathrm{QM}}$ and $\mathbf{R}^{\mathrm{MM}}$ are the positions of the nuclei in the QM subsystem and the atoms in the MM subsystem, respectively, and $\rho^{\mathrm{QM}}(\mathbf{r})$ is the electron density of the QM subsystem. The Hamiltonian of the QM subsystem, $H_{\mathrm{QM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}})$, is the sum of the kinetic energy of the nuclei and the KS energy, i.e. $H_{\mathrm{QM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}) = \sum_{i=1}^{N_{\mathrm{QM}}} \frac{\mathbf{P}_i^2}{2M_i} + E_{\mathrm{KS}}^{\mathrm{QM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}})$. The QM KS energy, $E_{\mathrm{KS}}^{\mathrm{QM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}})$, contains the same terms as conventional KS-DFT, i.e. kinetic, nuclear–electron interaction, Coulomb, and exchange–correlation energy. The total KS energy contains, in addition, the potential from the MM atoms that arises from the QM/MM Hamiltonian which is detailed further below. The Hamiltonian describing the MM subsystem, $H_{\mathrm{MM}}(\mathbf{R}^{\mathrm{MM}})$, is the sum of the kinetic energy of the MM atoms and the force field energy, which depends on the specifics of the force field. Typically, it consists of bonded (b) energy terms (associated with bond stretching, angle bending, torsions, etc.) and nonbonded (nb) energy terms (describing electrostatic and van der Waals interactions), i.e. $H_{\mathrm{MM}}(\mathbf{R}^{\mathrm{MM}}) = \sum_{i=1}^{N_{\mathrm{MM}}} \frac{\mathbf{P}_i^2}{2M_i} + V_{\mathrm{MM}}^{\mathrm{b}}(\mathbf{R}^{\mathrm{MM}}) + V_{\mathrm{MM}}^{\mathrm{nb}}(\mathbf{R}^{\mathrm{MM}})$. The QM/MM Hamiltonian, $H_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}},\mathbf{R}^{\mathrm{MM}})$, describes bonded and nonbonded interactions between the particles of the two subsystems, i.e. $H_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}},\mathbf{R}^{\mathrm{MM}}) = V_{\mathrm{QM/MM}}^{\mathrm{b}}(\mathbf{R}^{\mathrm{QM}},\mathbf{R}^{\mathrm{MM}}) + V_{\mathrm{QM/MM}}^{\mathrm{nb}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}},\mathbf{R}^{\mathrm{MM}})$. The nonbonded term consists of electrostatic interactions, which are fully detailed further below, and van der Waals interactions, which are treated using the same force field that is used for the MM subsystem. The bonded QM–MM interactions are only present when there are covalent bonds that cross the QM–MM interface. In such cases, the undercoordinated QM atoms are saturated by adding boundary atoms that are characterized by monovalent pseudopotentials[32] and contribute a single electron to the QM

subsystem. The only bonded terms that are included are those not described through the QM subsystem. Consider a string of bonded atoms that cross the QM–MM interface, i.e. M_3–M_2–M_1–B–Q_1–Q_2, where M is an MM atom, Q is a QM atom, and B is a boundary atom. Then the following bonded terms are kept: the M_1–B bond, the M_1–B–Q_1 and M_2–M_1–B angles, and the M_1–B–Q_1–Q_2, M_2–M_1–B–Q_1, and M_3–M_2–M_1–B dihedrals.

In this work, we use the Hamiltonian electrostatic coupling scheme developed by Laio et al.[30] where the electrostatic interactions between the QM and MM subsystems are split into a short-range (sr) and a long-range (lr) contribution

$$V^{\mathrm{es}}_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}}) = V^{\mathrm{es,sr}}_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}}) + V^{\mathrm{es,lr}}_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}}) \quad (2)$$

The short-range contribution consists of the interaction between the electron density and nuclear (or core) charges of the QM subsystem and the effective point charges (or multipoles) in the short-range domain. The latter includes all MM atoms within a given distance from the QM subsystem. The long-range interaction is computed using a multipole expansion of the electrostatic potential originating from the QM subsystem. This greatly reduces the computational cost of the evaluation of electrostatic QM–MM interactions for large molecular systems (which can be of the order of 10^5 atoms or more). MiMiC supports a generalized version of the approach by Laio et al. allowing, in principle, any order of multipoles to be used. This can potentially increase the precision of the long-range interactions and thus allow for a reduction of the number of atoms in the short-range domain. The original implementation by Laio et al. was later extended to three coupling regions to achieve a similar reduction, using dynamically generated restrained electrostatic potential (D-RESP) charges to describe interactions with MM atoms in an intermediate domain.[33] This extension is especially useful for elongated QM subsystems which we will address in a further development of our implementation.

In the following, we will use a multi-index notation which facilitates compact and open-

ended expressions of the QM/MM terms. A multi-index, denoted by α, β, γ, etc., is a 3-tuple consisting of three non-negative integers (e.g., $\alpha = (\alpha_x, \alpha_y, \alpha_z)$) that are associated with the x, y, and z Cartesian components, respectively. For instance, $\alpha = (0, 0, 0)$ denotes zeroth order with respect to all Cartesian components, $\alpha = (1, 0, 0)$ is first order with respect to the x component, $\alpha = (0, 0, 2)$ is second order with respect to the z component, etc. The short-range electrostatic QM–MM interaction energy can then be written as

$$V^{\mathrm{es,sr}}_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}}) = \sum_{i=1}^{N^{\mathrm{sr}}_{\mathrm{MM}}} \sum_{|\alpha|=0}^{A_i} \frac{(-1)^{|\alpha|}}{\alpha!} M_i^{[\alpha]} \left(\int T^{[\alpha]}_{\mathrm{mod}}(\mathbf{R}^{\mathrm{MM}}_i, \mathbf{r}) \rho^{\mathrm{QM}}(\mathbf{r})\, \mathrm{d}\mathbf{r} + \sum_{j=1}^{N_{\mathrm{QM}}} T^{[\alpha]}_{\mathrm{mod}}(\mathbf{R}^{\mathrm{MM}}_i, \mathbf{R}^{\mathrm{QM}}_j) Z^{\mathrm{QM}}_j \right) \tag{3}$$

where $M_i^{[\alpha]}$ is a component of the multipole associated with the ith MM atom, Z_j^{QM} is the nuclear (or core) charge of the jth QM atom, $\rho^{\mathrm{QM}}(\mathbf{r})$ is the QM electron density, and $T^{[\alpha]}_{\mathrm{mod}}(\mathbf{R}_i, \mathbf{R}_j)$ is a component of a modified interaction tensor, which is defined further below. The Cartesian component of the multipole and interaction tensor is indicated by a multi-index in square brackets. For example, $M^{[0,0,0]}$ is a charge, $M^{[1,0,0]}$ is the x-component of a dipole, $M^{[1,0,1]}$ is the xz-component of a quadrupole, and so on. Moreover, eq 3 introduces the multi-index norm, which is defined as the sum of the elements, i.e. $|\alpha| = \alpha_x + \alpha_y + \alpha_z$, and the multi-index factorial, which is defined by the product of factorials of the elements, i.e. $\alpha! = \alpha_x! \cdot \alpha_y! \cdot \alpha_z!$. The sum over multi-indices includes all multi-indices whose norm is less than or equal to the upper summation limit, so that for $A_i = 1$ it includes $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. The maximum order of the multipoles is thus defined by A_i.

An interaction tensor is generally defined as

$$T^{[\alpha]}(\mathbf{R}_a, \mathbf{R}_b) = \partial^{\alpha}_{\mathbf{R}_b} \frac{1}{|\mathbf{R}_b - \mathbf{R}_a|} \tag{4}$$

where we have used the definition of the multi-index partial derivative, i.e. $\partial^{\alpha}_{\mathbf{R}_b} = \frac{\partial^{|\alpha|}}{\partial R_{b,x}^{\alpha_x} \partial R_{b,y}^{\alpha_y} \partial R_{b,z}^{\alpha_z}}$. To avoid electron spill-out, we use a modified form of the interaction tensor for the short-

range interactions which is defined as[30]

$$T^{[\alpha]}_{\mathrm{mod}}(\mathbf{R}_a, \mathbf{R}_b) = \partial^{\alpha}_{\mathbf{R}_b} \frac{r^4_{c,a} - |\mathbf{R}_b - \mathbf{R}_a|^4}{r^5_{c,a} - |\mathbf{R}_b - \mathbf{R}_a|^5} \tag{5}$$

where $r_{c,a}$ is the covalent radius of the ath atom. This form of the interaction tensor models charge penetration effects and, importantly, avoids $1/r^n$ singularities by ensuring that the potential remains finite as $|\mathbf{R}_b - \mathbf{R}_a| \to 0$.

The interactions between the electrons and nuclei of the QM subsystem and the atoms in the MM subsystem that are farther away than the short-range cutoff distance are approximated using a multipole expansion of the electrostatic potential due to the QM charge density. With this approximation, the long-range electrostatic QM–MM interaction energy can be written as

$$V^{\mathrm{es,lr}}_{\mathrm{QM/MM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}}) = \sum_{i=1}^{N^{\mathrm{lr}}_{\mathrm{MM}}} \sum_{|\alpha|=0}^{A_i} \sum_{|\beta|=0}^{B_{\mathrm{QM}}} \frac{(-1)^{|\alpha+\beta|}}{\alpha!\beta!} M^{[\alpha]}_i T^{[\alpha+\beta]}(\mathbf{R}^{\mathrm{MM}}_i, \bar{\mathbf{R}}^{\mathrm{QM}}) M^{[\beta]}_{\mathrm{QM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}) \tag{6}$$

where $\bar{\mathbf{R}}^{\mathrm{QM}}$ is the origin of the multipole expansion, which in our implementation is the centroid of the QM subsystem (i.e., $\bar{\mathbf{R}}^{\mathrm{QM}} = \sum_{j=1}^{N_{\mathrm{QM}}} \mathbf{R}^{\mathrm{QM}}_j / N_{\mathrm{QM}}$), B_{QM} is the maximum order of the multipole expansion, and the QM multipoles are computed as

$$M^{[\beta]}_{\mathrm{QM}}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}) = \int \rho^{\mathrm{QM}}(\mathbf{r})(\mathbf{r} - \bar{\mathbf{R}}^{\mathrm{QM}})^{\beta} \, \mathrm{d}\mathbf{r} + \sum_{j=1}^{N_{\mathrm{QM}}} Z^{\mathrm{QM}}_j (\mathbf{R}^{\mathrm{QM}}_j - \bar{\mathbf{R}}^{\mathrm{QM}})^{\beta} \tag{7}$$

Equation 6 also introduces multi-index addition that is defined through $\alpha \pm \beta = (\alpha_x \pm \beta_x, \alpha_y \pm \beta_y, \alpha_z \pm \beta_z)$, i.e. the addition/subtraction is performed element-wise.

The polarizing potential due to the multipoles in the MM subsystem that acts on the electrons in the QM subsystem is needed to solve the KS equations. The potential is the functional derivative of the QM–MM interaction energy with respect to the electron density. Since only the electrostatic part of the QM/MM energy depends on the electron density (see

eqs 3 and 6), it is only these terms that are added to the KS equations. For the short-range part, the potential is thus given by

$$v^{\mathrm{es,sr}}_{\mathrm{QM/MM}}(\mathbf{r}) = \sum_{i=1}^{N^{\mathrm{sr}}_{\mathrm{MM}}} \sum_{|\alpha|=0}^{A_i} \frac{(-1)^{|\alpha|}}{\alpha!} M_i^{[\alpha]} T^{[\alpha]}_{\mathrm{mod}}(\mathbf{R}^{\mathrm{MM}}_i, \mathbf{r}) \tag{8}$$

whereas the potential from MM multipoles in the long-range part is

$$v^{\mathrm{es,lr}}_{\mathrm{QM/MM}}(\mathbf{r}) = \sum_{i=1}^{N^{\mathrm{lr}}_{\mathrm{MM}}} \sum_{|\alpha|=0}^{A_i} \sum_{|\beta|=0}^{B_{\mathrm{QM}}} \frac{(-1)^{|\alpha+\beta|}}{\alpha!\beta!} M_i^{[\alpha]} T^{[\alpha+\beta]}(\mathbf{R}^{\mathrm{MM}}_i, \bar{\mathbf{R}}^{\mathrm{QM}})(\mathbf{r} - \bar{\mathbf{R}}^{\mathrm{QM}})^{\beta} \tag{9}$$

The QM–MM interactions also lead to contributions to the forces acting on the QM and MM atoms. The additional forces are given by the derivative of the QM–MM interaction energy with respect to the positions. Here, we will only show the contributions originating from the electrostatic QM–MM interactions. Considering first the contributions from the short-range interactions (eq 3), we can write a component of the force acting on a QM nucleus as

$$F^{\mathrm{es,sr}\ [\gamma]}_{\mathbf{R}^{\mathrm{QM}}_j} = -\sum_{i=1}^{N^{\mathrm{sr}}_{\mathrm{MM}}} \sum_{|\alpha|=0}^{A_i} \frac{(-1)^{|\alpha|}}{\alpha!} M_i^{[\alpha]} T^{[\alpha+\gamma]}_{\mathrm{mod}}(\mathbf{R}^{\mathrm{MM}}_i, \mathbf{R}^{\mathrm{QM}}_j) Z^{\mathrm{QM}}_j \tag{10}$$

where γ can be $(1,0,0)$, $(0,1,0)$, or $(0,0,1)$ corresponding to the x, y, and z component, respectively. The short-range contributions to the force acting on an MM atom are given by

$$\begin{aligned} F^{\mathrm{es,sr}\ [\gamma]}_{\mathbf{R}^{\mathrm{MM}}_i} = -\sum_{|\alpha|=0}^{A_i} \frac{(-1)^{|\alpha+\gamma|}}{\alpha!} M_i^{[\alpha]} \Bigg(& \int T^{[\alpha+\gamma]}_{\mathrm{mod}}(\mathbf{R}^{\mathrm{MM}}_i, \mathbf{r}) \rho^{\mathrm{QM}}(\mathbf{r})\, \mathrm{d}\mathbf{r} \\ & + \sum_{j=1}^{N_{\mathrm{QM}}} T^{[\alpha+\gamma]}_{\mathrm{mod}}(\mathbf{R}^{\mathrm{MM}}_i, \mathbf{R}^{\mathrm{QM}}_j) Z^{\mathrm{QM}}_j \Bigg) \end{aligned} \tag{11}$$

The contributions to the force on a QM nucleus due the long-range interactions (eq 6) are somewhat more complicated because of the dependency on the nuclear positions through the nuclear part of the QM multipoles and the multipole expansion center. Successive application

of the chain rule yields

$$
\begin{aligned}
F^{\mathrm{es,lr}\ [\gamma]}_{\mathbf{R}_j^{\mathrm{QM}}} = &- \sum_{i=1}^{N_{\mathrm{MM}}^{\mathrm{lr}}} \sum_{|\alpha|=0}^{A_i} \sum_{|\beta|=1}^{B_{\mathrm{QM}}} \frac{(-1)^{|\alpha+\beta|}}{\alpha!(\beta-\gamma)!} M_i^{[\alpha]} T^{[\alpha+\beta]}(\mathbf{R}_i^{\mathrm{MM}}, \bar{\mathbf{R}}^{\mathrm{QM}}) Z_j^{\mathrm{QM}}(\mathbf{R}_j^{\mathrm{QM}} - \bar{\mathbf{R}}^{\mathrm{QM}})^{\beta-\gamma} \\
&- \frac{1}{N_{\mathrm{QM}}} \sum_{i=1}^{N_{\mathrm{MM}}^{\mathrm{lr}}} \sum_{|\alpha|=0}^{A_i} \left(\sum_{|\beta|=0}^{B_{\mathrm{QM}}} \frac{(-1)^{|\alpha+\beta|}}{\alpha!\beta!} M_i^{[\alpha]} T^{[\alpha+\beta+\gamma]}(\mathbf{R}_i^{\mathrm{MM}}, \bar{\mathbf{R}}^{\mathrm{QM}}) M_{\mathrm{QM}}^{[\beta]}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}) \right. \\
&\qquad \left. + \sum_{|\beta|=1}^{B_{\mathrm{QM}}} \frac{(-1)^{|\alpha+\beta|}}{\alpha!(\beta-\gamma)!} M_i^{[\alpha]} T^{[\alpha+\beta]}(\mathbf{R}_i^{\mathrm{MM}}, \bar{\mathbf{R}}^{\mathrm{QM}}) M_{\mathrm{QM}}^{[\beta-\gamma]}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}) \right)
\end{aligned}
\tag{12}
$$

where the summation over the β multi-indices is restricted to only include those where $\beta - \gamma$ has elements that are ≥ 0. The long-range contribution to the force on an MM atom is given by

$$
F^{\mathrm{es,lr}\ [\gamma]}_{\mathbf{R}_i^{\mathrm{MM}}} = - \sum_{|\alpha|=0}^{A_i} \sum_{|\beta|=0}^{B_{\mathrm{QM}}} \frac{(-1)^{|\alpha+\beta+\gamma|}}{\alpha!\beta!} M_i^{[\alpha]} T^{[\alpha+\beta+\gamma]}(\mathbf{R}_i^{\mathrm{MM}}, \bar{\mathbf{R}}^{\mathrm{QM}}) M_{\mathrm{QM}}^{[\beta]}[\rho^{\mathrm{QM}}](\mathbf{R}^{\mathrm{QM}}) \tag{13}
$$

The equations are written in an open-ended form and are therefore also valid for higher-order derivatives (i.e., when $|\gamma| > 1$). Finally, there are also indirect force contributions on the QM nuclei (and the electronic degrees of freedom in case of CP dynamics) that stem from the polarization of the electron density due to the electrostatic potentials given in eqs 8 and 9.

The described QM/MM methodology has been implemented using a plane wave (PW) basis where the computationally most expensive part is the calculation of the potential and forces originating from the short-range interactions. These contributions are expensive because they require integration of the electron density over the entire real space grid for each MM atom in the short-range part. The corresponding long-range interactions involve only one integration of the electron density, which is performed when determining the QM multipoles, and can thus lead to a substantially reduced computational cost compared to the

full explicit coupling used for the short-range part. In section 5, we investigate the effect of the long-range coupling with respect to the short-range cutoff distance and the order of the multipole expansion in detail.

3 MiMiC Framework

MiMiC is a multiscale modeling framework aimed at supporting a wide range of models, e.g. QM, MM, CG, and CM, and at coupling, in principle, any number of these models, which can either be used for different spatial regions or within the same subsystem. To achieve this flexibility while at the same time retaining a high degree of computational efficiency, MiMiC uses a strategy based on the MPMD model with a loose coupling between programs. This implies that instead of creating program-specific interfaces, i.e. tight coupling, the communication between programs goes through a well-defined application programming interface (API). Specifically, MiMiC has a two-sided API, where on one side MiMiC interfaces to a host program, which is the main driver evolving the system in time, while on the other side it can connect to multiple external programs in a plugin-like fashion (see illustration in Figure 1). The communication between external programs and MiMiC goes through a separate, lightweight communication library that has a simple C API which avoids major intervention in the source code of the external programs. The intention is to keep the setup and execution of the external programs as close as possible to the original. This is convenient for users who are already familiar with a given program and also allows existing documentation and tutorials to be used. The communication library uses MPI intercommunicators, which gives access to high-speed interconnects, such as InfiniBand and Omni-Path Architecture, but it can be extended to include other protocols without affecting the API. This approach allows us to achieve better use of computational resources by letting the external programs run concurrently with efficient exchange of data and by exploiting the fact that different programs have been optimized for their own specific use case. In this work, we have implemented

the QM/MM method described in the previous section within the MiMiC framework. Here, the QM and MM programs use separately allocated resources where they can apply their own parallelization strategies, including multithread/multiprocess and CPU/GPU-hybrid approaches, and thus obtain the highest possible efficiency for the individual subsystems.

The data structure of the MiMiC framework is general and allows modeling of subsystems at different levels of theory and resolution. A subsystem can contain a set of objects, such as fragments (used to describe, e.g., molecules or amino acid residues) and particles (used to describe, e.g., QM nuclei, MM atoms, or CG beads). Properties, such as electron densities, nuclear charges, point charges, masses, etc., are associated to the appropriate object. MiMiC supports both additive and subtractive subsystem schemes. In an additive scheme, the total system is divided into a number of subsystems in such a way that the total energy can be written as sums of (embedded) subsystem energies and interaction energies. This is the scheme we used for the present QM/MM implementation. A subtractive scheme, on the other hand, uses overlapping subsystems such that the total energy can be written as additions and differences of overlapping subsystem energies, which is the approach used, for example, in the ONIOM method.[34]

The MiMiC framework provides embedding capabilities which are currently limited to electrostatic embedding for QM/MM. The embedding functionality will be extended in future work, e.g., to support polarized embedding and QM/QM-type embedding. The electrostatic embedding supports a generalized version of the electrostatic coupling scheme by Laio et al.[30] in which long-range electrostatic QM–MM interactions are computed using the multipoles of the QM subsystem. MiMiC provides two schemes to partition the MM subsystem into short- and long-range parts, namely fragment-based and atom-based partitioning. The fragment-based scheme uses the distances between the centroids of the fragments in the MM subsystem and the centroid of the QM subsystem while the atom-based scheme uses the distance between the centroid of the QM subsystem and the individual MM atoms. When fragment information is available from the external MM program, e.g. as charge groups or

molecules, it is generally advisable to use the fragment-based partitioning scheme, because this will usually ensure neutral charge of the short- and long-range parts. It is not a requirement that the short- and long-range parts are neutral but it can potentially improve the convergence with respect to cutoff distance and the order of the multipole expansion needed to achieve a given accuracy.

For our first application of the MiMiC framework, we have coupled the CPMD[27] and GROMACS[28,29] programs to enable electrostatic embedding QM/MM-MD. GROMACS was chosen based on its high performance, versatility, and the availability of different force fields, as well as its active development community which ensures long-term support. The use of the CPMD program as the main driver enables MD within the CP and BO variants, with efficient orbital extrapolation schemes to improve the computational efficiency for the latter, using a plane wave/pseudopotential implementation of KS-DFT. Density functional perturbation theory calculations can be performed,[35] and electronically excited states can be accessed within the linear response[36] or real-time[37,38] time-dependent DFT (TDDFT) frameworks as well as the restricted open-shell KS (ROKS) formalism.[39] Nonadiabatic effects[40] can be included either within an Ehrenfest scheme[38] or by virtue of trajectory surface hopping (TSH),[41,42] while nuclear quantum effects can be taken into account within a path integral (PI) formalism.[43] The program is highly parallelized,[17,44] offers access to a large number of state-of-the-art exchange–correlation functionals[45,46] and a computationally efficient treatment of hybrid functionals,[44,47] and allows for long time scale simulations by virtue of efficient multiple time step (MTS) acceleration algorithms.[48] The choice of CPMD therefore allows for optimal flexibility and high efficiency in the description of the QM subsystem.

The workflow of a simulation employing the MiMiC framework is not predetermined but can be adapted to the multiscale implementation being implemented and to the chosen external programs. In the QM/MM implementation of this work, we use the workflow illustrated in Figure 2. Here the CPMD program is the main simulation driver and QM engine, while GROMACS is the MM engine. Both programs are executed independently and

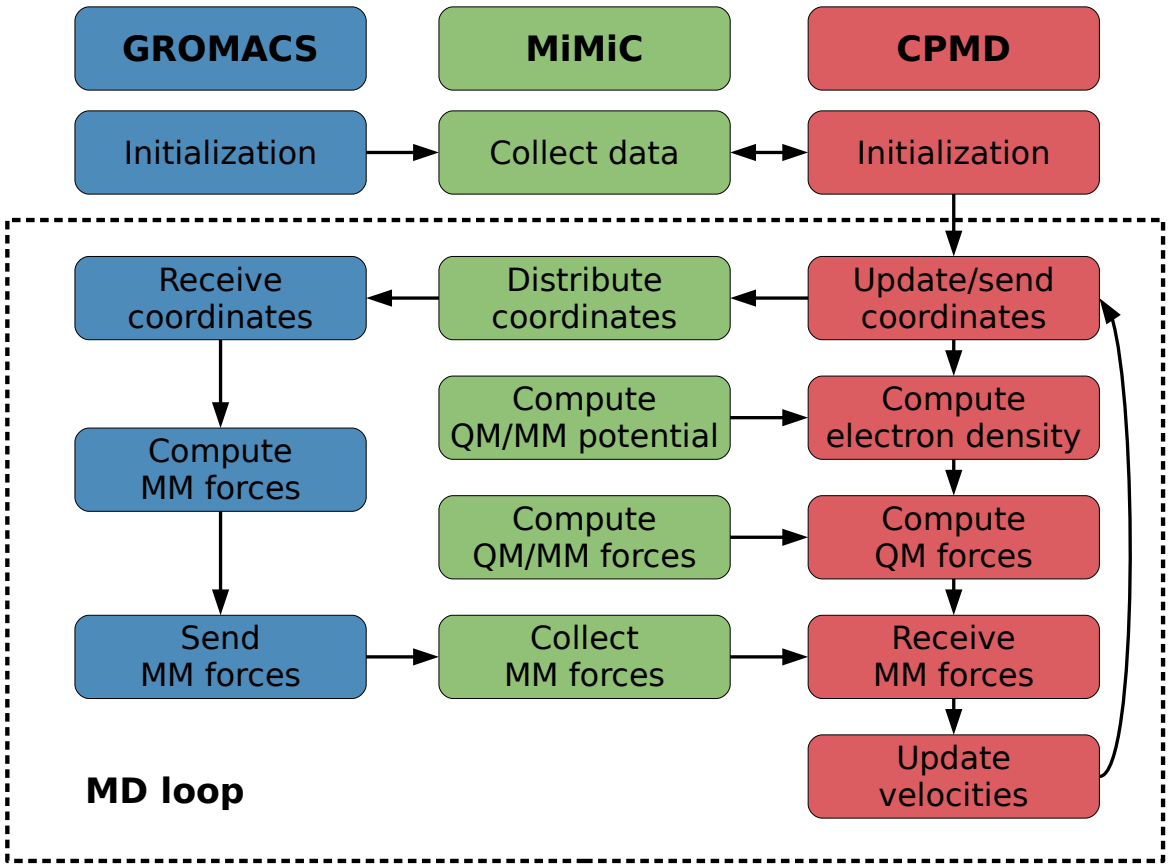


Figure 2: Schematic illustration of a QM/MM-MD workflow using the MiMiC framework where the CPMD program is the main simulation driver and QM engine, GROMACS is the MM engine, and MiMiC handles communication and computes QM/MM contributions.

run simultaneously using MiMiC to communicate and to compute QM/MM contributions. In the initialization phase, the programs read their respective input files, and MiMiC collects data from GROMACS and CPMD and sends the necessary data to CPMD, e.g. coordinates, atom types, etc., to allow setup of data structures. Once CPMD is initialized, the QM/MM-MD loop is entered. The loop consists of several stages. The first one involves the sending of coordinates to GROMACS which then proceeds to compute the MM energy and forces. At the same time as GROMACS computes the MM energy and forces, CPMD computes the corresponding QM contributions subject to the electrostatic potential computed by MiMiC on the QM grid. MiMiC also computes the QM/MM energy and forces. Finally, all force contributions are collected, and CPMD integrates the equations of motion and continues to the next iteration of the MD loop.

A MiMiC-based QM/MM simulation using CPMD and GROMACS runs one instance of each program. GROMACS computes all contributions related to the MM subsystem as well as the classical nonbonded van der Waals QM–MM interactions and, if present, bonded QM–MM interactions. The GROMACS input thus includes the entire system, i.e. the QM subsystem is also described by the MM force field. The atoms belonging to the QM subsystem are added to a special index group within GROMACS which will have the following effect: a) point charges of the atoms belonging to the group are zeroed, b) nonbonded van der Waals interactions between atoms in the group are excluded, and c) bonds between atoms in the group are deleted. This avoids treating interactions twice, i.e. at the MM level *and* at the QM level. In cases where the MM engine does not allow such operations intrinsically, MiMiC can follow a different approach where two instances of the MM engine run simultaneously. The first instance runs an MM calculation of the entire system, while the second instance runs an MM calculation only including atoms belonging to the QM subsystem. The point charges of QM atoms are set to zero in both instances. The energies and forces computed by the second instance are then subtracted from those computed by the first one. This approach has a minor computational overhead since the cost of the additional MM calculation of the

(typically small) QM subsystem is negligible. Moreover, it allows us to use virtually any external MM program in the MiMiC-based QM/MM implementation.

The MiMiC framework will be made freely available under the terms of the GNU Lesser General Public License (LGPL) version 3, or any later version, as published by the Free Software Foundation (FSF). The interface in CPMD will be part of a forthcoming release of CPMD, but patches for the current release can be obtained from the corresponding author, while the interface in GROMACS is part of the 2019 release.[49]

4 Computational Details

We used four different systems of varying size and composition to validate the MiMiC framework for QM/MM-MD and to analyze the performance of the long-range coupling scheme. Three of the systems are solute–solvent systems, of which two are relatively small and one is large. The small systems consist of a water in water solution (Wat(aq)) and an acetone in water solution (Ace(aq)), while the large system is an *n*-butanol dissolved in acetone (BuOH(ace)). The QM subsystem in each case consists of a single solvated molecule, i.e. one water, acetone, or *n*-butanol molecule, respectively, and the MM subsystem is made up of all the solvent molecules, which is 981 water molecules in the Wat(aq) system, 978 water molecules in the Ace(aq) system, and 13,948 acetone molecules in the BuOH(ace) system. In addition to the solute–solvent systems, we also used a zinc-binding GB1 mutant[50,51] (shown in Figure 3) as a representative of more complex and biologically relevant systems and one that also involves bonded QM–MM interactions. The QM subsystem of the GB1 system consists of the zinc ion, together with three coordinating histidine side chains and one coordinated water molecule, while the MM subsystem contains the remainder of the protein and 8,372 water molecules.

The QM/MM-MD simulations were performed using the MiMiC framework, coupling a locally modified version of the CPMD development trunk and a locally modified GROMACS

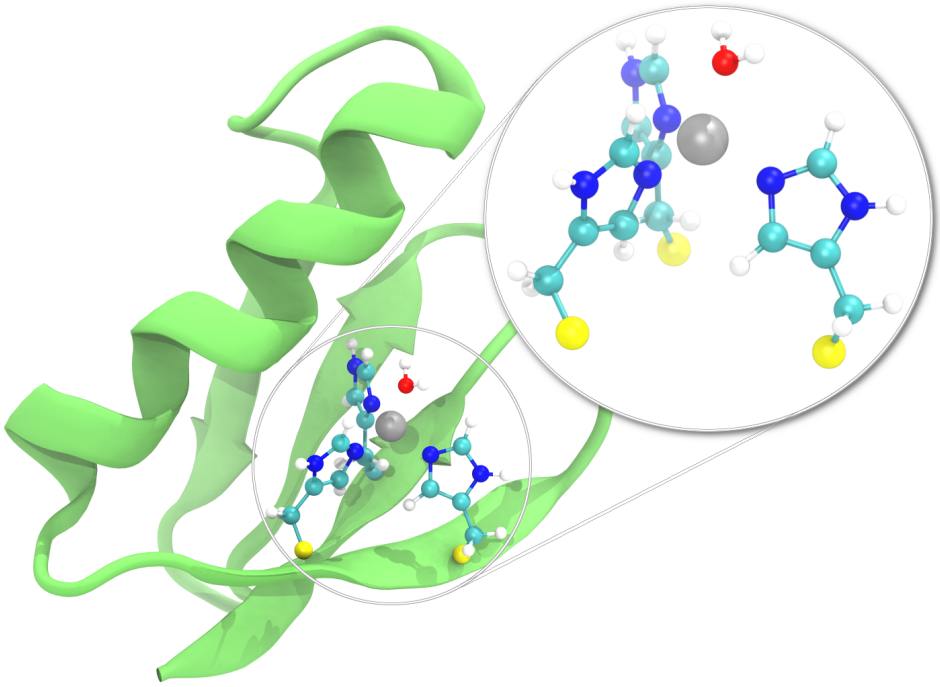


Figure 3: Zinc-binding site of the GB1 mutant. The QM subsystem consists of the atoms shown as ball-and-stick. Boundary atoms are shown in yellow. Bulk water is not shown for clarity.

2019 development version compiled using double precision. All systems were pre-equillibrated through MM-MD simulations before they were equilibrated using QM/MM-MD simulations. We refer to the Supporting Information (SI) for the details of the pre-equillibration procedure.

The first set of QM/MM-MD simulations used the BO approach for the QM subsystem and was performed in the NVT ensemble, running for ∼19.35 ps, corresponding to 40,000 steps using a time step of ∼0.48 fs (20.0 a.u.). The first half is considered part of the equilibration, while the remainder was used for further analysis. While this was adequate for the solute–solvent systems, it was not enough to fully equilibrate the GB1 system. The latter half was used regardless since it does not affect the analyses performed in this work. In these simulations, the short-range coupling was used for all MM atoms in the Wat(aq) and Ace(aq) systems, while the long-range coupling was employed for the BuOH(ace) and GB1 systems. The BuOH(ace) system was partitioned into short- and long-range domains using the fragment-based scheme with a cutoff distance of 32 a.u. and using a fifth-order multipole

expansion for the long-range coupling. For the GB1 system, we used an atom-based partitioning with a cutoff distance of 36 a.u. and a seventh-order multipole expansion for the long-range coupling. In both cases, the list of short- and long-range atoms was updated every 50th step. The temperature was initialized at 300 K and was subsequently maintained with a Nosé–Hoover chain thermostat[52] using a chain length of 4 and a frequency of 3500 cm^{-1} for the Wat(aq) system and 3000 cm^{-1} for the Ace(aq), BuOH(ace), and GB1 systems. In the case of the solute–solvent systems, the MM subsystem was modeled using the optimized potential for liquid simulations all-atom (OPLS/AA) force field,[53] employing the transferable intermolecular potential with three points (TIP3P)[54] water model and topologies of acetone and *n*-butanol that were downloaded from the GROMACS Molecule & Liquid Database (`http://virtualchemistry.org`).[55,56] The MM subsystem of GB1 was modeled using the AMBER ff14SB[57] force field. The bonds of all classical water molecules in the Wat(aq), Ace(aq), and GB1 systems were constrained using the RATTLE algorithm.[58,59] The simulation cell dimensions were set according to the final dimensions obtained from NPT MM-MD simulations in the pre-equillibration procedure (see details in the SI). We thus used cubic cells with lengths of 58.67, 58.66, and 225.8 a.u. for Wat(aq), Ace(aq), and BuOH(ace), respectively, and an orthorhombic cell for GB1 with lengths of 121.3, 120.4, and 121.2 a.u. The QM subsystem was in all cases modeled using the BLYP[60,61] exchange–correlation functional, employing the new driver by Bircher and Rothlisberger,[45] and Troullier–Martins norm-conserving pseudopotentials.[62] For GB1 the bonds between the C_α and C_β of the zinc-coordinating histidines were cut and the C_α's were replaced by boundary atoms described by an analytic monovalent pseudopotential[32] (boundary atoms are shown in yellow in Figure 3). Periodic boundary conditions were applied to the purely classical interactions, employing a 28.3 a.u. (26.5 a.u. for GB1) cutoff on the short-range nonbonded interactions, and using the smooth particle mesh Ewald (SPME)[63] method for the long-range electrostatic interactions. The electrostatic QM–MM interactions were computed within the minimum image convention without periodic repetition. Isolated system conditions for the QM subsystem

were enforced using the Martyna–Tuckerman[64] method to solve Poisson's equation. The QM cell sizes and PW cutoffs were determined based on convergence analyses shown in the SI (see Figs. S1–S3) and taking into consideration the requirements of the Martyna–Tuckerman method. Thus, we chose a cell size of 30 a.u. for the Wat(aq) system and 40 a.u. for the other three systems, and PW cutoffs of 100, 80, 85, and 85 Ry for Wat(aq), Ace(aq), BuOH(ace), and GB1, respectively. The DIIS method[65–67] was used in the self-consistent field (SCF) optimization of the wave function using a convergence threshold of 1×10^{-5} a.u. with respect to the largest component of the electronic gradient. The initial guess of the wave function in each MD step is based on an extrapolation from the five previous steps to accelerate convergence and to improve energy conservation.[68]

To investigate the effect of the long-range coupling on the energy, forces, and dipole moment of the QM subsystem for different short-range cutoffs and multipole orders, we performed an analysis on 10 structures that were extracted from the last half of the simulations (20,000 steps) at regular intervals, i.e. every 2,000th step (~ 0.97 ps). In the case of GB1, we calculated the Hirshfeld partial charge of the zinc ion instead of the dipole moment. The SCF convergence threshold was set to 1.0×10^{-6} a.u. in these single-point calculations.

We also examined the robustness of the QM/MM implementation by monitoring the energy fluctuations of the Wat(aq), Ace(aq), and BuOH(ace) systems during QM/MM-MD simulations in the NVE ensemble. Simulations using BO and CP methods were performed both with and without long-range coupling. For the simulations of Wat(aq) and Ace(aq) that include the long-range coupling, we first performed additional equilibrations for $\sim$2.42 ps (5,000 steps) including the long-range coupling and otherwise using the same settings as described above. Short-range cutoffs of 28, 30, and 32 a.u. were used for Wat(aq), Ace(aq), and BuOH(ace), respectively, while the short-range coupling was used for the entire system in the simulations without long-range coupling. The same settings as described above were used for the QM/MM-MD simulations based on the BO approach except that the convergence threshold was set to 1.0×10^{-6} a.u. and eight previous steps were used in the wave function

extrapolation. For the QM/MM-MD simulations using the CP method, a time step of ∼0.07 fs (3.0 a.u.) was used for Wat(aq) and ∼0.12 fs (5.0 a.u.) for the other two systems. The fictitious electron mass was set to 700 a.u. for all systems except for Wat(aq) where it was set to 800 a.u.

5 Results and Discussion

To validate the new MiMiC-based QM/MM implementation, we performed simulations on four molecular systems of varying size and composition as described in the Computational Details (section 4).

The numerical quality of the implementation is assessed by measuring the fluctuation of the total energy through QM/MM-MD simulations of the solute–solvent systems in the NVE ensemble. Both BO and CP approaches are investigated either using a full short-range coupling for the entire system or including also the long-range coupling with short-range cutoffs of 28, 30, and 32 a.u. for Wat(aq), Ace(aq), and BuOH(ace), respectively. The results are presented in Figure 4 where the fluctuations of the total energy per particle over time are shown relative to the average energy. In all cases, the energy profile displays fluctuations with negligible drift, showing that stable simulations can be performed using the MiMiC-based QM/MM implementation. No appreciable difference is observed between the CP and BO approaches in terms of drift due to the relatively tight SCF convergence threshold of 1.0×10^{-6} a.u. together with the wave function extrapolation used in the BO simulations. However, larger fluctuations are observed when using the BO approach compared to CP in the case of BuOH(ace). This can be attributed to the larger time step used in the BO simulations (0.48 fs) compared to the one used in the CP simulations (0.12 fs). In the case of Wat(aq) and Ace(aq), the energy fluctuations are larger and more irregular compared to those observed for BuOH(ace). This is most likely related to the difference in the size of the systems combined with the use of nonperiodic electrostatic QM–MM interactions.

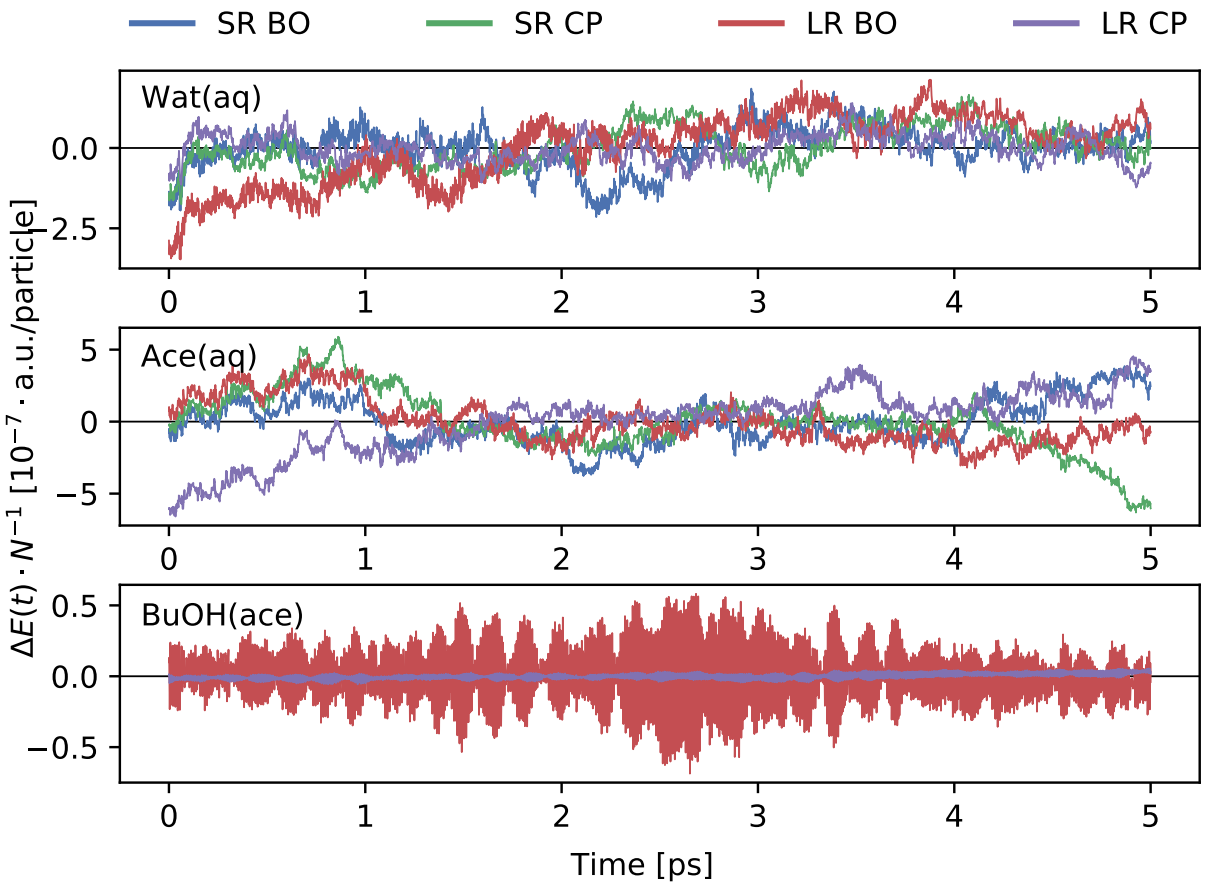
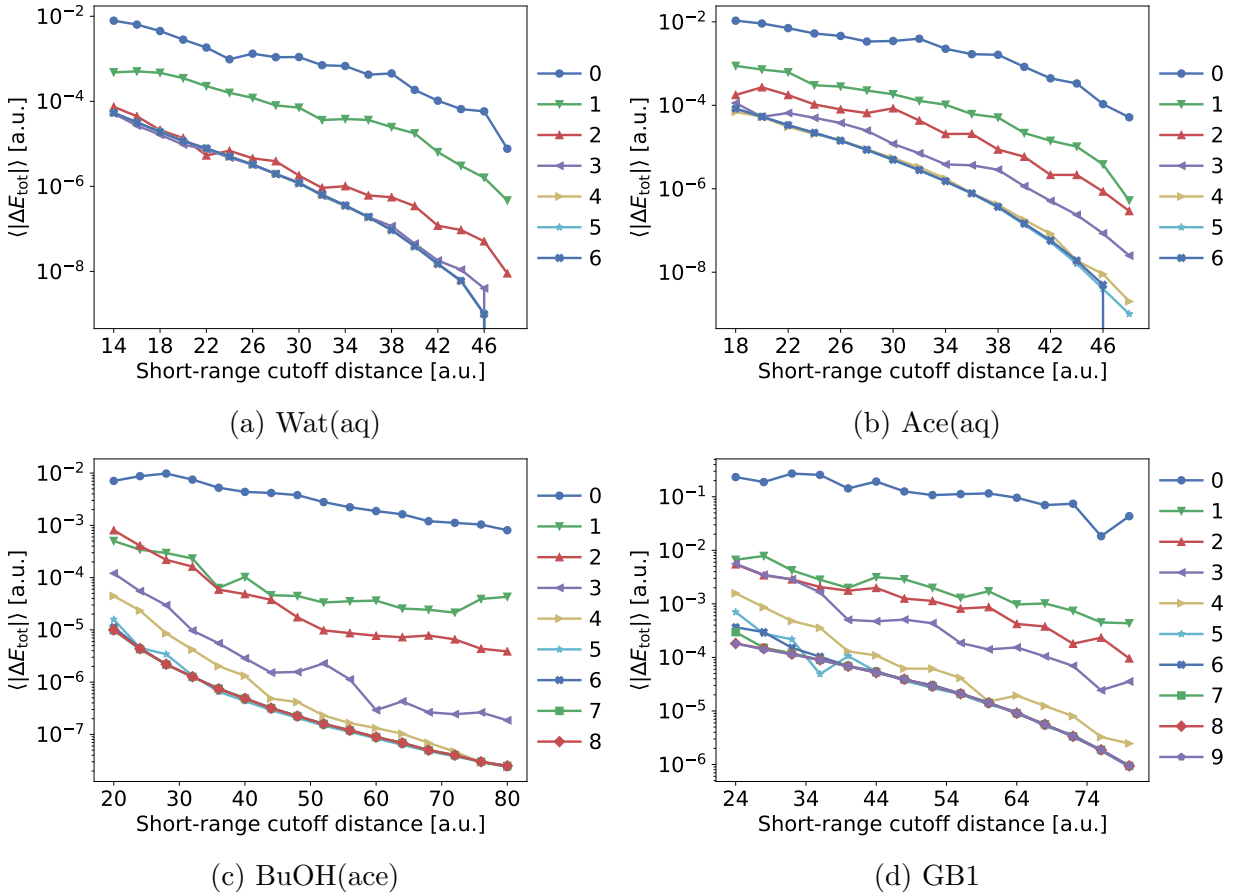


Figure 4: Fluctuation of the total energy per particle with respect to time for QM/MM-MD simulations in the NVE ensemble of the Wat(aq) (top panel), Ace(aq) (middle panel), and BuOH(ace) (bottom panel) systems within the BO or CP approach using short-range (SR) coupling for the entire system or including long-range (LR) coupling with short-range cutoffs of 28, 30, and 32 a.u. for Wat(aq), Ace(aq), and BuOH(ace), respectively. The change in energy per particle is given relative to the average energy ($\Delta E(t) = E(t) - \langle E \rangle$).

A new feature in our MiMiC-based QM/MM implementation compared to the current implementation in the CPMD program by Laio et al.[30] is the ability to use an open-ended multipole expansion for use in the long-range coupling. Using higher orders in the multipole expansion will in many cases produce a more accurate electrostatic potential thus allowing the long-range coupling to be used at shorter distances. Compared to the current implementation, this reduces the computational cost of the QM–MM interactions even further without compromising the accuracy. To investigate how much the short-range cutoff distances can be decreased while still retaining a high accuracy, we performed single-point calculations

on a series of snapshots using varying orders of the multipole expansion and short-range cutoff distances and compared key properties to calculations where the short-range coupling was used for the entire system. Specifically, we examined the total energy, forces, and, as a measure of how well the electron density is reproduced, the dipole moment of the QM subsystem. For the GB1 system, we investigate the Hirshfeld partial charge of the zinc ion instead of the dipole moment.



(a) Wat(aq)

(b) Ace(aq)

(c) BuOH(ace)

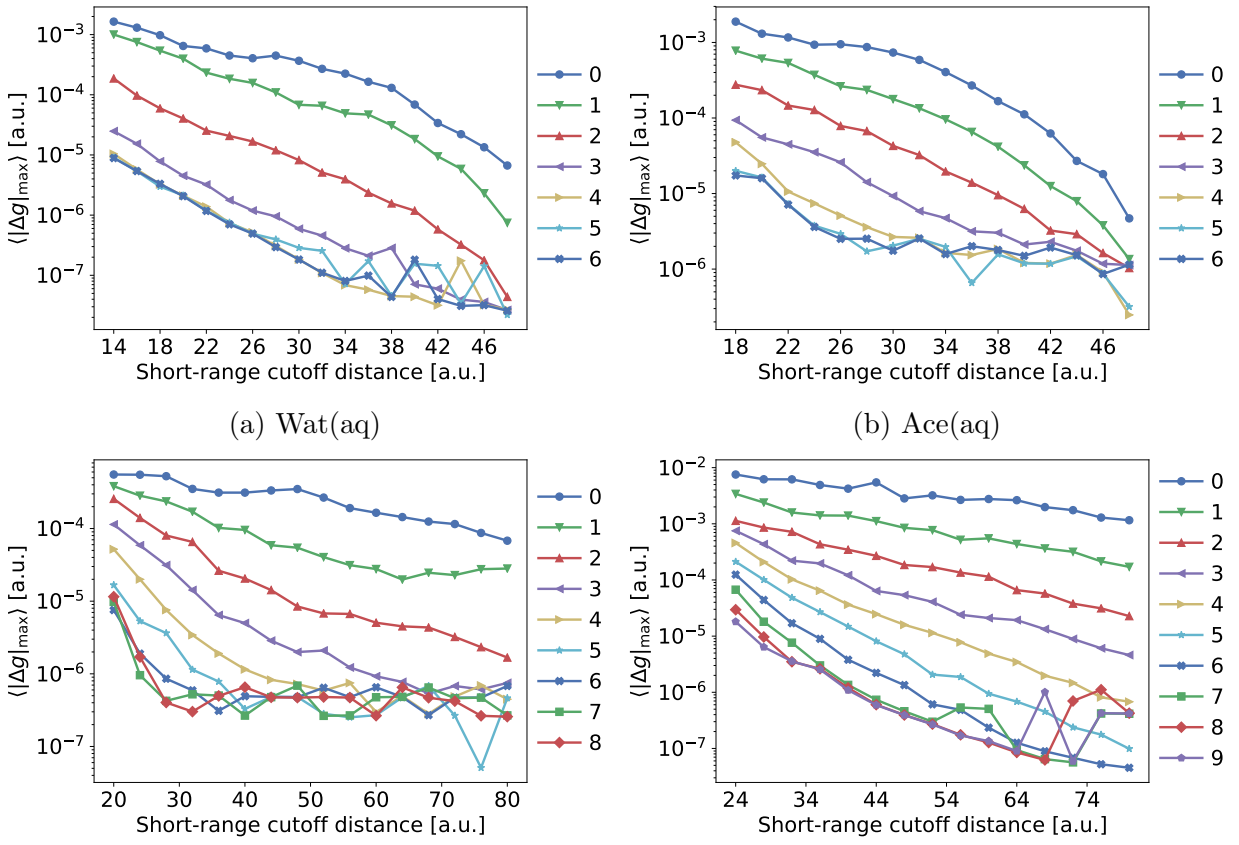
(d) GB1

Figure 5: Convergence of the total energy of (a) Wat(aq), (b) Ace(aq), (c) BuOH(ace), and (d) GB1 with respect to the short-range cutoff distance and the order of the multipole expansion used for the long-range coupling (which are indicated with different colors). The points show the mean absolute error based on single-point calculations on ten snapshots using the full short-range coupling as reference. The cutoff distance is based on the fragment-based partitioning scheme in the case of Wat(aq), Ace(aq), and BuOH(ace), whereas for GB1 it is based on the atom-based partitioning.

The results from the single-point calculations are shown in Figures 5–8. The lower limits

of the short-range cutoff distances were chosen so as to exclude distances that cause convergence issues in the SCF calculations. These issues were observed for some snapshots, mostly when using higher-order multipoles, which thus suggests that they are related to a divergence of the multipole expansion. It is possible that the divergence is exacerbated by electron spill-out or overpolarization of the electronic density. Electron spill-out can be avoided in most cases by using the short-range coupling for all MM atoms that are inside the QM cell. This is due to the modified interaction tensor (see eq 5) used in the short-range coupling which effectively prevents electron spill-out by ensuring that the potential stays finite as the QM electron – MM charge distance goes to zero. That being said, whether or not it is necessary to include *all* MM atoms that are inside the QM cell depends on the system. For the systems investigated here, the short-range cutoff distances have to be around 24, 30, and 32 a.u. using the fragment-based partitioning for the Wat(aq), Ace(aq), and BuOH(ace) systems, respectively, in order to go beyond the extent of the QM cell, while for the GB1 system, the short-range cutoff has to be roughly 30 a.u. using the atom-based partitioning. However, we do not observe any adverse effects by using shorter cutoffs apart from a slight increase in error.

Ideally, the long-range coupling should reproduce the short-range coupling but at a reduced computational cost. To achieve this, it is sufficient that the errors are equal to or lower than the numerical precision of the calculated properties which is determined by the SCF convergence threshold used in the single-point calculations. From Figure 5 it can be seen that the total energy converges only for the small solute–solvent systems. However, for these systems it converges at a cutoff distance where only very few solvent molecules are treated using the long-range coupling. The savings in computational cost are therefore negligible compared to using the short-range coupling for the entire system. The situation is different for the other properties as can be seen in Figures 6 and 7. The numerical precision of these properties is roughly of the order of 10^{-6} a.u. This point is first reached at cutoff distances of around 24 a.u. for the solute–solvent systems, while the GB1 system requires a



(a) Wat(aq)

(b) Ace(aq)

(c) BuOH(ace)

(d) GB1

Figure 6: Convergence of the forces in (a) Wat(aq), (b) Ace(aq), (c) BuOH(ace), and (d) GB1 with respect to the short-range cutoff distance and the order of the multipole expansion used for the long-range coupling (which are indicated with different colors). The points show the mean of the maximum absolute error based on single-point calculations on ten snapshots using the full short-range coupling as reference. The cutoff distance is based on the fragment-based partitioning scheme in the case of Wat(aq), Ace(aq), and BuOH(ace), whereas for GB1 it is based on the atom-based partitioning.

longer cutoff distance of about 40 a.u. to fully converge due to the much larger size of the QM subsystem in GB1.

The order at which the multipole expansion converges depends not only on the cutoff distance (i.e., the distance from the center of the expansion) but also on the composition of the QM subsystem (mainly the number of atoms) as well as the calculated property. For a given cutoff distance, increasing the multipole expansion above a certain order yields very little improvement, which suggests that the expansion has converged. To improve

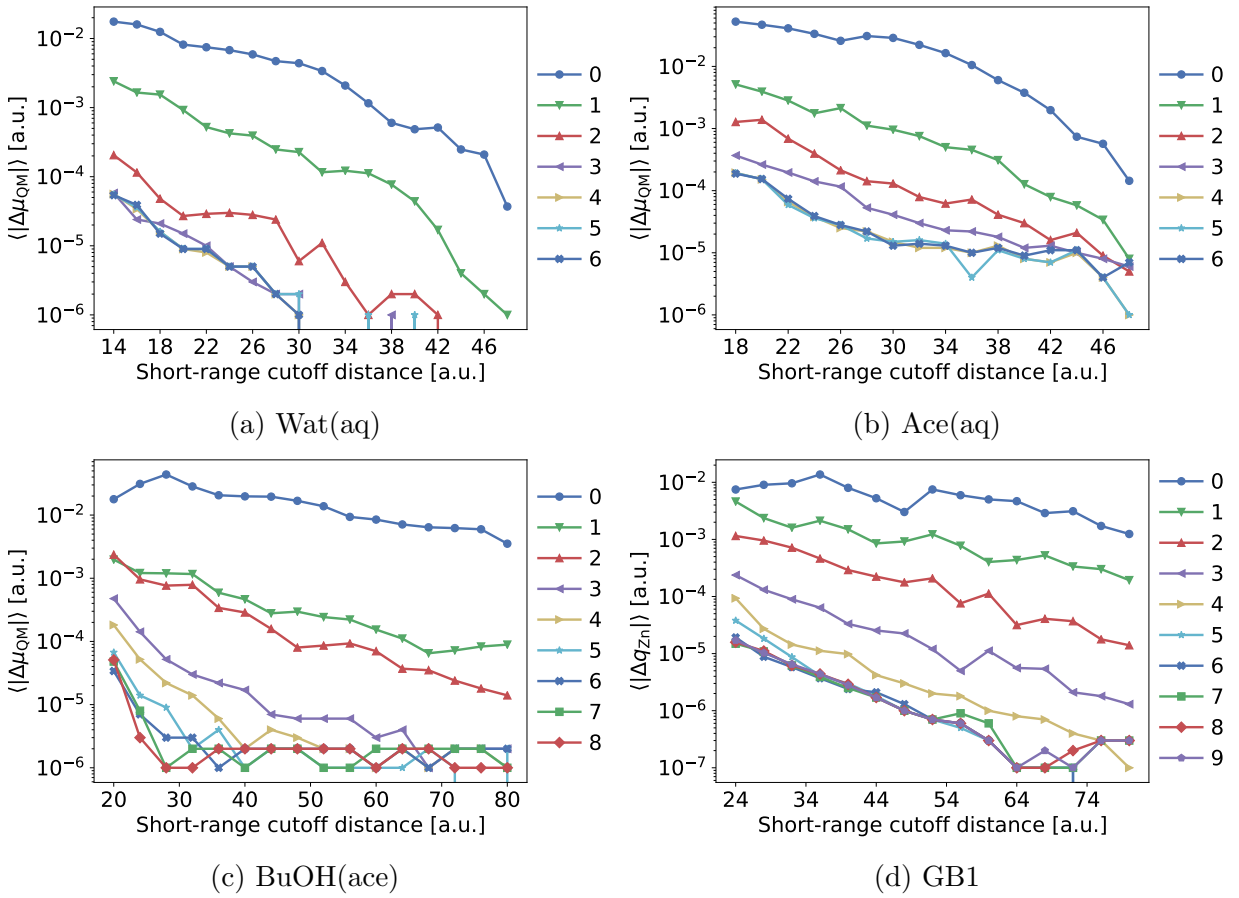


Figure 7: Convergence of the dipole moment strength of the QM (a) water molecule in the Wat(aq) system, (b) acetone molecule in the Ace(aq) system, (c) *n*-butanol molecules in the BuOH(ace) system, and (d) the Hirshfeld charge of the zinc ion in GB1 with respect to the short-range cutoff distance and the order of the multipole expansion used for the long-range coupling (which are indicated with different colors). The points show the mean absolute error based on single-point calculations on ten snapshots using the full short-range coupling as reference. The cutoff distance is based on the fragment-based partitioning scheme in the case of Wat(aq), Ace(aq), and BuOH(ace), whereas for GB1 it is based on the atom-based partitioning.

the accuracy further it is necessary to increase the cutoff distance until full convergence is reached both with respect to the order of the multipole expansion and the cutoff distance. To converge the forces at the shortest possible cutoff distance, which is at ~ 24 a.u. for the solute–solvent systems and ~ 40 a.u. for GB1, requires a fourth-order multipole expansion in the case of Wat(aq), fifth-order for Ace(aq), sixth-order for BuOH(ace), and seventh-order for GB1 (see Figure 6). The dipole moments (or zinc ion charge in the case of GB1) are also

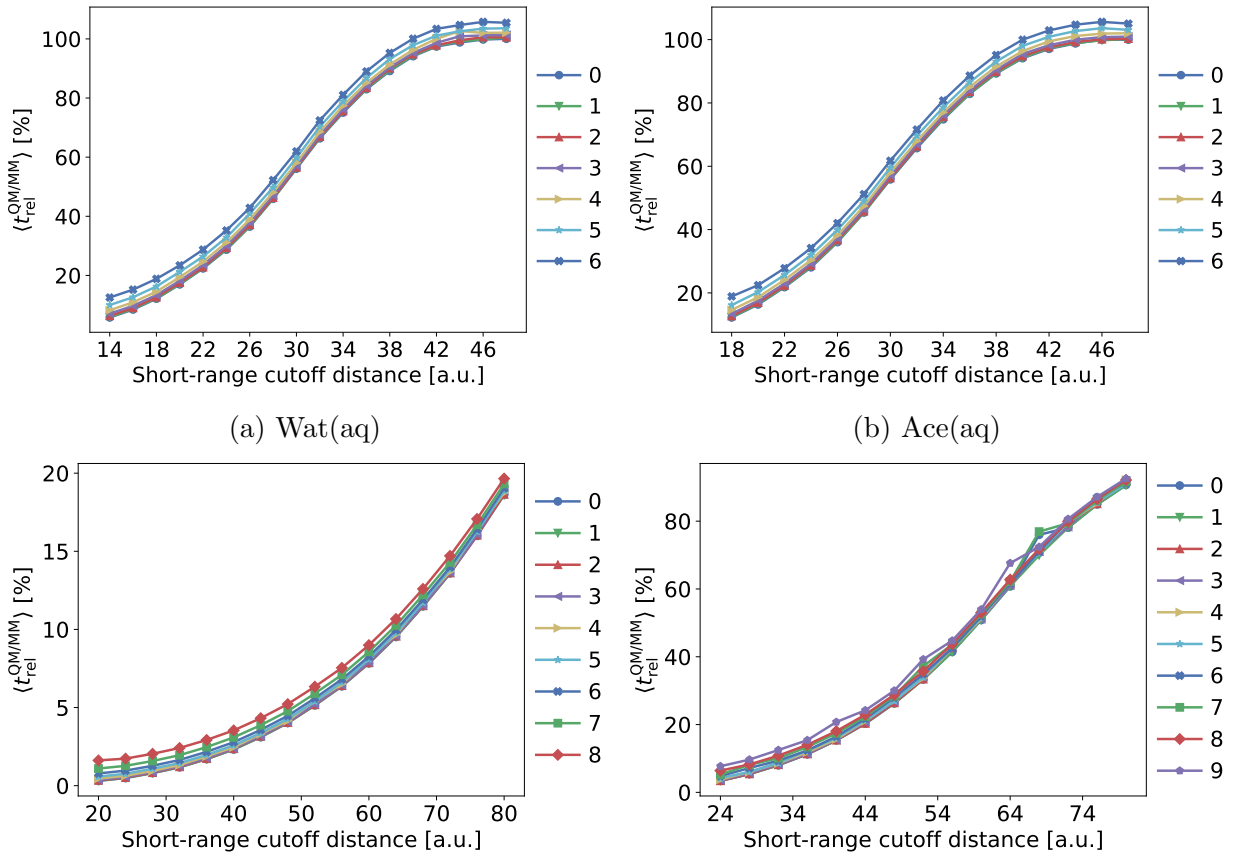
converged at around the same cutoff distances (see Figure 7). Using the cutoff distance and order of multipoles where the forces are fully converged yields an error in the total energy roughly of the order of 10^{-5} a.u. for Wat(aq) and BuOH(ace) and 10^{-4} a.u. for Ace(aq) and GB1. These errors are negligible compared to the thermal energy of the systems, and the dynamics of the two approaches can thus be considered equivalent as far as sampling is concerned.

Using the long-range coupling greatly improves the computational efficiency compared to using the short-range coupling for the entire system (see Figure 8). For the small solute–solvent and GB1 systems the computational cost is reduced by 70–80%. This reduction is further increased to $\sim 99\%$ for the large BuOH(ace) system.

Obtaining the same level of accuracy using a second-order multipole expansion, which is the limit in the implementation by Laio et al.,[30] would require cutoff distances that are much larger. In fact, for the small solute–solvent systems, the forces are only fully converged when almost all solvent water molecules are treated via the short-range coupling, while for the larger systems it requires larger cutoffs than those investigated here. The corresponding increase of the computational cost would by far exceed the cost associated with the use of a higher-order multipole expansion at shorter cutoff distances.

The optimal short-range cutoff distance and order of the multipole expansion varies depending on the system. Our results indicate that a relatively short cutoff can be applied for homogeneous systems (e.g., solute–solvent systems), while for heterogeneous systems (e.g., proteins) it may be necessary to use longer cutoffs. The number of atoms in the QM subsystem can be used as a guideline to decide at which order to truncate the multipole expansion, e.g. higher number of atoms generally require higher orders of multipoles. In general, it is advisable to investigate which cutoff distance and multipole order is suitable to achieve a desired accuracy.

Finally, to demonstrate the performance of the MiMiC-based QM/MM implementation, we ran a series of simulations of the GB1 system using an increasing number of compute

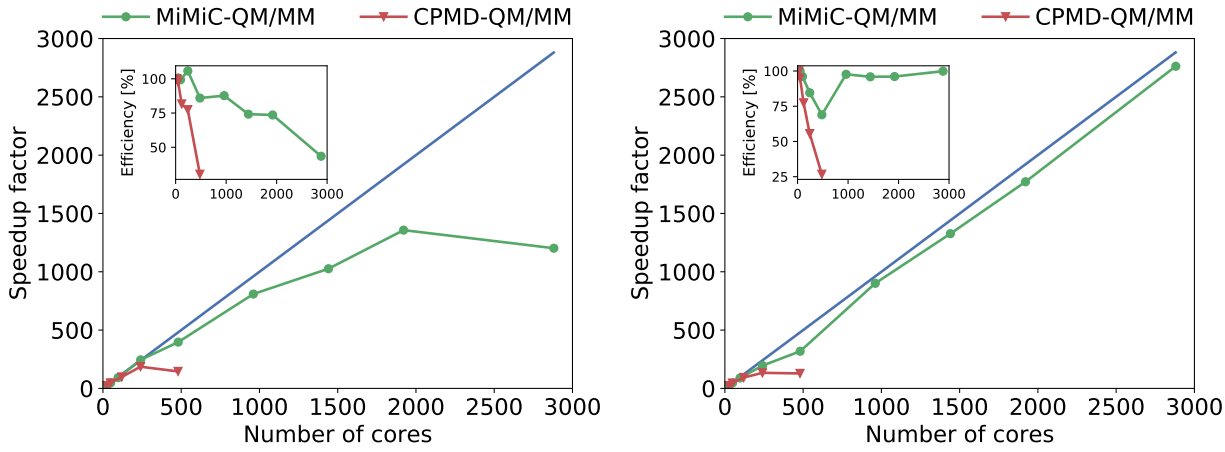


(a) Wat(aq) (b) Ace(aq)

(c) BuOH(ace) (d) GB1

Figure 8: Relative wall times for the calculation of the QM/MM contributions compared to the short-range coupling of the (a) Wat(aq), (b) Ace(aq), (c) BuOH(ace), and (d) GB1 systems with respect to the short-range cutoff distance and the order of the multipole expansion used for the long-range coupling (which are indicated with different colors). The points show the mean relative wall times based on single-point calculations on ten snapshots using the full short-range coupling as reference. The cutoff distance is based on the fragment-based partitioning scheme in case of Wat(aq), Ace(aq), and BuOH(ace) whereas for GB1 it is based on the atom-based partitioning.

cores. This benchmark provides preliminary insight into the performance of the MiMiC-based QM/MM. A thorough analysis will be presented in an upcoming study in order to fully demonstrate the improved efficiency of the present implementation. In Figure 9, we compare the parallel speedup of the MiMiC-based QM/MM and the current tight-coupling QM/MM approach in CPMD.[30] The system was set up so as to ensure a fair comparison between the two implementations. Thus, a second-order multipole expansion is used while all other settings correspond to those described in the Computational Details (section 4).



(a) Total (b) QM/MM

Figure 9: Parallel speedup of a time step of QM/MM-MD using CP dynamics. Speedups are given relative to one core using a single node (24 cores) as reference. Included are speedups of the (a) total time and (b) time spent computing QM–MM terms for the MiMiC-based QM/MM (green line/circles) and the current QM/MM implementation in CPMD (red line/triangles). The blue line indicates ideal speedup. The points show the mean speedup based on four time steps. The insets display the scaling efficiency. Calculations were run on JURECA (2 × 12-core Intel Xeon E5-2680 v3 CPUs per node).[69]

Due to differences in the partitioning schemes used in MiMiC and the current QM/MM in CPMD, we used a cutoff distance of 28 a.u. using the atom-based partitioning for the MiMiC-based QM/MM and 20 a.u. for the current QM/MM in CPMD, to ensure a similar partitioning of the MM atoms into short- and long-range domains. In Figure 9a, it can be seen that the MiMiC-based QM/MM-MD reaches a very good overall speedup with a parallel efficiency above 70% at 1920 cores when using a single node (24 cores) as the baseline. This is a substantial improvement compared to the current implementation in CPMD which stops scaling at 240 cores yielding a speedup factor of around 190. In Figure 9b, it is shown that the limiting factor in the current QM/MM implementation in CPMD is the computation of QM–MM interaction terms. In contrast, in the MiMiC-based QM/MM, the computation of interaction terms displays near linear scaling even up to $\sim$ 3000 cores, thus effectively eliminating this bottleneck. Here it is worth noting that the average time per time step of the MiMiC-based QM/MM-MD at the scaling limit is only 0.3 s. It is expected that scalability is further improved for larger QM subsystems. Roughly a third of the total time

per time step is spent computing QM–MM interactions terms at the scaling limit. This would be a much smaller fraction in a BO-based simulation where the QM part is more expensive due to the SCF cycles. Importantly, the communication overhead, which can be viewed as the cost of the loose coupling adopted in the present work, is only about 0.005 s. Since the communication overhead scales linearly with system size, it will be negligible even for large systems.

6 Summary and Outlook

We outlined a flexible and efficient framework for multiscale modeling in computational chemistry (MiMiC) which is based on a MPMD scheme using loose coupling between programs. Its high degree of flexibility allows coupling of potentially any set of molecular modeling programs with minimal interference in their source codes. Moreover, it enables efficient use of computational resources by exploiting the existing parallelization strategies used in the coupled programs. Fast communication between the coupled programs is achieved through a custom communication library that utilizes MPI intercommunicators.

The framework was demonstrated in a new electrostatic embedding QM/MM implementation combining the CPMD and GROMACS programs. Illustrative applications on simple solute–solvent systems as well as a more complex biologically relevant system were presented. The robustness of the implementation was verified through constant energy simulations that exhibited good conservation of energy. The QM/MM implementation features a generalized version of the electrostatic coupling scheme by Laio et al.[30] which approximates long-range QM–MM interactions through a multipole expansion of the electrostatic potential originating from the QM subsystem. We showed that this scheme can substantially reduce the computational cost associated with the electrostatic QM–MM interactions without losing accuracy compared to an exact treatment. At the same time, the parallel performance was substantially improved which will allow for much more efficient use of resources. Thus, an

overall speedup of roughly an order of magnitude was observed when compared to the current QM/MM implementation in CPMD showcasing that a loose coupling scheme does not have to be associated with significant communication overhead.

Extensions of the QM/MM methodology to polarized embedding including excited state treatments and combinations with MTS algorithms are currently ongoing. Thanks to its flexibility, the framework is also well suited for the development of models that go beyond basic QM/MM. For example, it allows the use of multiple subsystems, where each subsystem can be described at a different level of theory and resolution, as well as multilevel descriptions of the same parts of the system, which can be useful for subtractive schemes and MTS approaches.

Author Information

Author contributions

Conceptualization: J.M.H.O. (equal), S.M. (equal), and U.R. (equal); Methodology: J.M.H.O.; Software: J.M.H.O. (equal), V.B. (equal), S.M. (supporting), and M.P.B. (supporting); Investigation: J.M.H.O. (lead) and V.B. (supporting); Resources: J.M.H.O. (lead) and P.C. (supporting); Visualization: J.M.H.O. (lead) and V.B. (supporting); Writing – Original Draft: J.M.H.O. (lead) and V.B. (supporting); Writing – Review and Editing: J.M.H.O. (equal), V.B. (equal), S.M. (equal), E.I. (equal), M.P.B. (equal), P.C. (equal), and U.R. (equal); Supervision: J.M.H.O. (equal), P.C. (equal), and U.R. (equal); Funding Acquisition: J.M.H.O. (equal), P.C. (equal), and U.R. (equal).

Funding

J.M.H.O. acknowledges financial support from the Danish Council for Independent Research (DFF) through the Sapere Aude research career program (Grant ID: DFF–1325-00091 and DFF–1323-00744), the Carlsberg Foundation (Grant ID: CF15-0823), and the Research

Council of Norway through its Centres of Excellence scheme (Project ID: 262695). V.B. and P.C. acknowledge funding received within the European Union's Horizon 2020 program from project HPC-LEAP (MSCA-ITN-2014-EJD, Grant ID: 642069). E.I. and P.C. acknowledge funding received within the European Union's Horizon 2020 program from the BioExcel Center of Excellence (EINFRA-5-2015, Grant ID: 675728). P.C. also acknowledges funding by the Deutsche Forschungsgemeinschaft via FOR 2518 DynIon project P6. U.R. acknowledges funding from the Swiss National Science Foundation via the NCCR MUST and individual grants. Computational resources were provided by the DeIC National HPC Centre at the University of Southern Denmark, the Norwegian Supercomputing Program (NOTUR) through a grant of computer time (Grant ID: NN4654K), and Jülich Supercomputing Center.

Acknowledgement

The authors thank IBM Zürich Research Laboratory and especially Teodoro Laino, Valery Weber, and Alessandro Curioni for their help with the development of the interface in CPMD. The authors also thank Mark James Abraham, Berk Hess, and Erik Lindahl for valuable advice concerning the GROMACS interface. The authors are grateful to Guido Frisari for providing the pre-equillibrated structure of GB1.

Supporting Information Available

Description of the pre-equillibration procedure and convergence analysis with respect to the plane wave cutoff energy and the size of the quantum cell.

References

(1) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(2) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(3) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217–6263.

(4) Morzan, U. N.; de Armiño, D. J. A.; Foglia, N. O.; Ramírez, F.; Lebrero, M. C. G.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.

(5) Adhireksan, Z.; Davey, G. E.; Campomanes, P.; Groessl, M.; Clavel, C. M.; Yu, H.; Nazarov, A. A.; Yeo, C. H. F.; Ang, W. H.; Dröge, P.; Rothlisberger, U.; Dyson, P. J.; Davey, C. A. Ligand substitutions between ruthenium–cymene compounds can control protein versus DNA targeting and anticancer activity. *Nat. Commun.* **2014**, *5*.

(6) Campomanes, P.; Rothlisberger, U.; Alfonso-Prieto, M.; Rovira, C. The Molecular Mechanism of the Catalase-like Activity in Horseradish Peroxidase. *J. Am. Chem. Soc.* **2015**, *137*, 11170–11178.

(7) Genna, V.; Vidossich, P.; Ippoliti, E.; Carloni, P.; De Vivo, M. A Self-Activated Mechanism for Nucleic Acid Polymerization Catalyzed by DNA/RNA Polymerases. *J. Am. Chem. Soc.* **2016**, *138*, 14592–14598.

(8) Li, J.; Lyu, W.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.;

Sobott, F.; Grandori, R.; Carloni, P. Proton Dynamics in Protein Mass Spectrometry. *J. Phys. Chem. Lett.* **2017**, *8*, 1105–1112.

(9) Campomanes, P.; Neri, M.; Horta, B. A. C.; Röhrig, U. F.; Vanni, S.; Tavernelli, I.; Rothlisberger, U. Origin of the Spectral Shifts among the Early Intermediates of the Rhodopsin Photocycle. *J. Am. Chem. Soc.* **2014**, *136*, 3842–3851.

(10) Cupellini, L.; Caprasecca, S.; Guido, C. A.; Müh, F.; Renger, T.; Mennucci, B. Coupling to Charge Transfer States is the Key to Modulate the Optical Bands for Efficient Light Harvesting in Purple Bacteria. *J. Phys. Chem. Lett.* **2018**, *9*, 6892–6899.

(11) Loco, D.; Buda, F.; Lugtenburg, J.; Mennucci, B. The Dynamic Origin of Color Tuning in Proteins Revealed by a Carotenoid Pigment. *J. Phys. Chem. Lett.* **2018**, *9*, 2404–2410.

(12) Liem, S. Y.; Brown, D.; Clarke, J. H. R. Molecular dynamics simulations on distributed memory machines. *Comput. Phys. Commun.* **1991**, *67*, 261–267.

(13) Nelson, M. T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kalé, L. V.; Skeel, R. D.; Schulten, K. NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *Int. J. High Perform. Comput. Appl.* **1996**, *10*, 251–268.

(14) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(15) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **2005**, *122*, 084119.

(16) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for

Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(17) Bekas, C.; Curioni, A. Very large scale wavefunction orthogonalization in Density Functional Theory electronic structure calculations. *Comput. Phys. Commun.* **2010**, *181*, 1057 – 1068.

(18) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Grand, S. L.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(19) Phillips, J. C.; Sun, Y.; Jain, N.; Bohm, E. J.; Kale, L. V. Mapping to Irregular Torus Topologies and Other Techniques for Petascale Biomolecular Simulation. SC14: International Conference for High Performance Computing, Networking, Storage and Analysis. 2014.

(20) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. Electronic Absorption Spectra from MM and ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **2012**, *8*, 5092–5106.

(21) Götz, A. W.; Clark, M. A.; Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* **2013**, *35*, 95–108.

(22) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(23) Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. NAMD goes quantum: an integrative suite for hybrid simulations. *Nat. Methods* **2018**, *15*, 351–354.

(24) Torras, J.; Deumens, E.; Trickey, S. B. Software Integration in Multi-scale Simulations: the PUPIL System. *J. Computer-Aided Mater. Des.* **2006**, *13*, 201–212.

(25) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—a modular software package for QM/MM simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci* **2013**, *4*, 101–110.

(26) Lu, Y.; Farrow, M. R.; Fayon, P.; Logsdail, A. J.; Sokol, A. A.; Catlow, C. R. A.; Sherwood, P.; Keal, T. W. An Open-Source, Python-Based Redevelopment of the ChemShell Multiscale QM/MM Environment. *J. Chem. Theory Comput.* **2018**,

(27) CPMD, http://www.cpmd.org/, Copyright IBM Corp 1990-2018, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.

(28) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. *Lecture Notes in Computer Science*; Springer International Publishing, 2015; pp 3–27.

(29) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(30) Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.* **2002**, *116*, 6941–6947.

(31) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

(32) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Variational optimization of effective atom centered potentials for molecular properties. *J. Chem. Phys.* **2005**, *122*, 014113.

(33) Laio, A.; VandeVondele, J.; Rothlisberger, U. D-RESP: Dynamically Generated Electrostatic Potential Derived Charges from Quantum Mechanics/Molecular Mechanics Simulations. *J. Phys. Chem. B* **2002**, *106*, 7300–7307.

(34) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.

(35) Putrino, A.; Sebastiani, D.; Parrinello, M. Generalized variational density functional perturbation theory. *J. Chem. Phys.* **2000**, *113*, 7102–7109.

(36) Hutter, J. Excited state nuclear forces from the Tamm–Dancoff approximation to time-dependent density functional theory within the plane wave basis set framework. *J. Chem. Phys.* **2003**, *118*, 3928–3934.

(37) Tavernelli, I.; Röhrig, U. F.; Rothlisberger, U. Molecular dynamics in electronically excited states using time-dependent density functional theory. *Mol. Phys.* **2005**, *103*, 963–981.

(38) Tavernelli, I. Electronic density response of liquid water using time-dependent density functional theory. *Phys. Rev. B* **2006**, *73*, 094204.

(39) Frank, I.; Hutter, J.; Marx, D.; Parrinello, M. Molecular dynamics in low-spin excited states. *J. Chem. Phys.* **1998**, *108*, 4060–4069.

(40) Bircher, M. P.; Liberatore, E.; Browning, N. J.; Brickel, S.; Hofmann, C.; Patoz, A.; Unke, O. T.; Zimmermann, T.; Chergui, M.; Hamm, P.; Keller, U.; Meuwly, M.; Woerner, H.-J.; Vaníček, J.; Rothlisberger, U. Nonadiabatic effects in electronic and nuclear dynamics. *Struct. Dyn.* **2017**, *4*, 061510.

(41) Tapavicza, E.; Tavernelli, I.; Rothlisberger, U. Trajectory Surface Hopping within Lin-

ear Response Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **2007**, *98*, 023001.

(42) Tavernelli, I.; Curchod, B. F. E.; Rothlisberger, U. On nonadiabatic coupling vectors in time-dependent density functional theory. *J. Chem. Phys.* **2009**, *131*, 196101.

(43) Marx, D.; Parrinello, M. Ab initio path-integral molecular dynamics. *Z. Phys. B* **1994**, *95*, 143–144.

(44) Weber, V.; Bekas, C.; Laino, T.; Curioni, A.; Bertsch, A.; Futral, S. Shedding Light on Lithium/Air Batteries Using Millions of Threads on the BG/Q Supercomputer. 2014 IEEE 28th International Parallel and Distributed Processing Symposium. 2014.

(45) Bircher, M. P.; Rothlisberger, U. Plane-Wave Implementation and Performance of -la-Carte Coulomb-Attenuated Exchange-Correlation Functionals for Predicting Optical Excitation Energies in Some Notorious Cases. *J. Chem. Theory Comput.* **2018**, *14*, 3184–3195.

(46) Bircher, M. P.; López-Tarifa, P.; Rothlisberger, U. Shedding Light on the Basis Set Dependence of the Minnesota Functionals: Differences Between Plane Waves, Slater Functions, and Gaussians. *J. Chem. Theory Comput.* **2018**, *15*, 557–571.

(47) Bircher, M. P.; Rothlisberger, U. Exploiting Coordinate Scaling Relations To Accelerate Exact Exchange Calculations. *J. Phys. Chem. Lett.* **2018**, *9*, 3886–3890.

(48) Liberatore, E.; Meli, R.; Rothlisberger, U. A Versatile Multiple Time Step Scheme for Efficient ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 2834–2842.

(49) M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, GROMACS User Manual version 2019, www.gromacs.org (2019).

(50) Bozkurt, E. Reprogramming the B1 Domain of Streptococcal Protein G (GB1): A Joint Theoretical and Experimental Investigation. **2018**,

(51) Bozkurt, E.; Perez, M. A. S.; Hovius, R.; Browning, N. J.; Rothlisberger, U. Genetic Algorithm Based Design and Experimental Characterization of a Highly Thermostable Metalloprotein. *J. Am. Chem. Soc.* **2018**, *140*, 4517–4521.

(52) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(53) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6665–6670.

(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(55) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2011**, *8*, 61–74.

(56) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS molecule & liquid database. *Bioinformatics* **2012**, *28*, 752–753.

(57) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(58) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(59) Weinbach, Y.; Elber, R. Revisiting and parallelizing SHAKE. *J. Comput. Phys.* **2005**, *209*, 193–206.

(60) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(61) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(62) Troullier, N.; Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **1991**, *43*, 1993–2006.

(63) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(64) Martyna, G. J.; Tuckerman, M. E. A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. *J. Chem. Phys.* **1999**, *110*, 2810–2821.

(65) Pulay, P. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

(66) Pulay, P. Improved SCF convergence acceleration. *J. Comput. Chem.* **1982**, *3*, 556–560.

(67) Hutter, J.; Lüthi, H. P.; Parrinello, M. Electronic structure optimization in plane-wave-based density functional calculations by direct inversion in the iterative subspace. *Comput. Mater. Sci.* **1994**, *2*, 244–248.

(68) Arias, T. A.; Payne, M. C.; Joannopoulos, J. D. Ab initio molecular-dynamics techniques extended to large-length-scale systems. *Phys. Rev. B* **1992**, *45*, 1538–1549.

(69) JURECA Cluster Module, Jülich Supercomputing Centre (JSC), https://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JURECA/JURECA_node.html.

Graphical TOC Entry