New QM/MM Implementation of the DFTB3 Method in the Gromacs Package

Tomáš Kubař,*[a] Kai Welke,[b] and Gerrit Groenhof[c]

The approximate density-functional tight-binding theory method DFTB3 has been implemented in the quantum mechanics/molecular mechanics (QM/MM) framework of the Gromacs molecular simulation package. We show that the efficient smooth particle–mesh Ewald implementation of Gromacs extends to the calculation of QM/MM electrostatic interactions. Further, we make use of the various free-energy functionalities provided by Gromacs and the PLUMED plugin. We exploit the versatility and performance of the current framework in three typical applications of QM/MM methods to solve biophysical problems: (i) ultrafast proton transfer in malonaldehyde, (ii) conformation of the alanine dipeptide, and (iii) electron-induced repair of a DNA lesion. Also discussed is the further development of the framework, regarding mostly the options for parallelization. © 2015 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24029

Introduction

Most biochemical systems, such as enzymes, are too large to be described at any level of *ab initio* theory. At the same time, the available molecular mechanics (MM) force fields are not sufficiently flexible to model processes in which chemical bonds are broken or formed,[1,2] electrons are transferred[3] or excited electronic states are involved, such as in spectroscopy[4,5] or photodynamics.[6–8] To overcome the limitations of a full quantum chemical description on the one hand, and a full MM treatment on the other hand, methods have been developed that treat a small part of the system at the level of quantum chemistry (QM), while retaining the computationally cheaper MM force field for the larger part. While this QM/MM concept itself is simple and has been around for almost 40 years,[9] it inspires continuing development that is predominantly concerned with accurately integrating both "worlds" and finding more efficient ways of doing so. The importance of QM/MM methods, especially for biomolecular simulations, is highlighted best by the award of the 2013 Nobel prize in Chemistry to three of the pioneers of the field: Martin Karplus, Michael Levitt, and Arieh Warshel.

For studying dynamics beyond a few picoseconds, the QM method has to be fast enough to allow for extended sampling, while being able to treat the quantum-mechanical nature of the process under study to sufficient accuracy. The *density-functional tight-binding method* (DFTB), derived from the density-functional theory (DFT), is suitable for this task as shown in many applications.[10] As it avoids the laborious calculations of integrals, it is roughly three orders of magnitude faster than DFT and therefore extends the accessible timescale to nanoseconds.

After suitable QM method and MM force field have been chosen, the next choice concerns the coupling between them. Due to the long-range nature of the Coulomb interactions in particular, truncating these interactions in molecular dynamics (MD) simulations can lead to serious artefacts. Although the particle–mesh Ewald method (PME)[11] provides an accurate and highly efficient description of these long-range interactions, many QM/MM implementations truncate the Coulomb interactions beyond a given cutoff.

In this work, we address both accuracy and sampling efficiency of the QM/MM implementation in the Gromacs simulation package[12] by incorporating DFTB in combination with PME electrostatic coupling. At first, we will briefly summarize the involved methods, MM as well as DFTB. Then, we provide a detailed description of the new implementation of DFTB into Gromacs, with a special focus on the PME treatment of the QM/MM electrostatic interactions. To demonstrate the capabilities of the new implementation, we showcase our QM/MM implementation for the dynamics of several representative chemical systems and processes, namely liquid water, ultrafast proton transfer (PT) in malonaldehyde, the conformation of an alanine dipeptide, and bond breaking in a photo-damaged DNA oligonucleotide. For each of these systems, we compute free energies using various approaches, including free MD simulation, umbrella sampling (US), and well-tempered metadynamics, using the PLUMED plugin.[13]

Methods

Molecular mechanics force fields

Gromacs supports a wide variety of MM force fields: Gromos,[14] Amber,[15] Charmm,[16] and OPLS.[17] They share the

[a] *T. Kubař*
Institute of Physical Chemistry and Center for Functional Nanostructures, Karlsruhe Institute of Technology, Karlsruhe 76131, Germany
E-mail: tomas.kubar@kit.edu
[b] *K. Welke*
Department of Chemistry, Nagoya University, Nagoya 464-8602, Japan
[c] *G. Groenhof*
Nanoscience Center and Department of Chemistry, University of Jyväskylä, Jyväskylä 40014, Finland

same potential energy function, parameterizing the interatomic interactions as a sum of simple pairwise additive empirical potentials

$$E^{\mathrm{MM}}=\sum_r K_r(r-r^0)^2+\sum_\vartheta K_\vartheta(\vartheta-\vartheta^0)^2 +\sum_\varphi\sum_n\frac{1}{2}V_n[1+\cos(n\varphi-\gamma)] +\frac{1}{2}\sum_{A,B}\left(f_{\mathrm{LJ}}\cdot\left[\frac{C_{AB}}{r_{AB}^{12}}-\frac{D_{AB}}{r_{AB}^{6}}\right]+f_{\mathrm{QQ}}\cdot\frac{1}{4\pi\varepsilon_0}\cdot\frac{Q_AQ_B}{r_{AB}}\right) \quad (1)$$

where the first four sums run over all bonds r, angles ϑ and dihedral angles φ, respectively. The last two terms are the nonbonded interactions, including the van der Waals and electrostatic interactions. These interactions are excluded ($f=0$) for any pair of atoms that are connected via bonded or angle terms, while a scaling factor is applied for atoms that are connected via a dihedral term, with typical values of $f_{\mathrm{LJ}}=1/2$ and $f_{\mathrm{QQ}}=1/1.2$. The parameters of the force field (K, V, r^0, ϑ^0, γ, C, D, Q_A) are usually obtained by fitting to experimental or *ab initio* calculated references. After almost four decades of careful parameterization, today's force fields rival quantum chemistry in accuracy in many cases, at a fraction of the computational cost.

DFTB

As the theory behind the DFTB models,[18,19] its self-consistent charge variant (DFTB2 or SCC-DFTB)[20] and its third-order extension (DFTB3)[21] have been reviewed at great length recently, we restrict ourselves to a brief description; the interested reader may refer for details to the reviews.[10,22,23]

The DFTB3 model is based on a third-order expansion of the DFT exchange-correlation functional around a reference density, $\rho=\rho^0+\delta\rho\approx\sum_a\rho_a^0+\sum_a\Delta q_a\,F_a^{00}\,Y_a^{00}$. The reference density ρ^0 is a sum of densities of neutral atoms, and the charge fluctuations $\delta\rho$ are approximated as monopoles with normalized radial dependence F_a^{00} and normalized ($l=m=0$) spherical harmonics Y_a^{00}.[20] The total energy decomposes as

$$E^{\mathrm{DFTB3}}=\sum_{i,a,b}\sum_{\mu\in a}\sum_{\nu\in b}n_i\,c_{\mu i}\,c_{\nu i}\,H_{\mu\nu}^0+\frac{1}{2}\sum_{a,b}\gamma_{ab}\,\Delta q_a\,\Delta q_b +\frac{1}{3}\sum_{a,b}(\Delta q_a)^2\Delta q_b\,\Gamma_{ab}+E^{\mathrm{rep}}+E^{\mathrm{disp}} \quad (2)$$

where the Hamilton matrix elements $H_{\mu\nu}^0$ are subject to a two-center approximation. They are precalculated in DFT calculations of neutral atoms and stored together with the corresponding overlap matrix elements in a parameter file. The γ function in the second term of eq. (2) mediates the interactions of the atomic charge density fluctuations.[20] In the DFTB3 model,[21] γ is modified for all interactions involving a hydrogen atom, which improves the description of hydrogen-bonding interactions substantially.[24] The third-order term [third term in eq. (2)] contains the Γ-matrix, which accounts for the changes of chemical hardness with respect to the atomic charge.[21,24] The fourth term in eq. (2), the repulsive energy, is represented as a sum of pairwise repulsive potentials $E^{\mathrm{rep}}=\sum_{a<b}V_{ab}^{\mathrm{rep}}$ and depends only on the neutral reference density ρ^0. These repulsive potentials are fitted to the results of DFT calculations performed for a set of reference molecules and can thus be stored in the parameter files as well.[20] Attractive dispersion interaction, which is not accounted for in the DFTB formalism, is included by means of an empirical function E^{disp}, for which different adaptions have been proposed.[25–27] In this work, we implemented the popular D3 correction,[28] which was recently tested and applied in combination with DFTB3 as well.[29] It contains atom-pairwise contributions and is independent of the electronic structure.

$$E^{\mathrm{disp}}=-\frac{1}{2}\sum_{n=6,8}\sum_{a,b}s_n\,\frac{C_{ab}^n}{|\vec{r}_a-\vec{r}_b|^n+f(R_{ab}^0)^n} \quad (3)$$

Here, $C_{ab}^{6/8}$ are the leading-order dipole–dipole and dipole–quadrupole dispersion coefficients, s_n are scaling factors, and $f(R_{ab}^0)$ is the Becke–Johnson rational damping function.[30] The C^6 parameters are precalculated with time-dependent DFT.[28,29] In QM/MM calculations, this dispersion correction only affects the interactions between QM atoms. Van der Waals interactions between the QM and MM atoms are described using Lennard–Jones potentials.

Variation of the DFTB3 total energy with respect to the molecular orbital coefficients c_μ leads to a pseudoeigenvalue problem with matrix elements

$$H_{\mu\nu}=H_{\mu\nu}^0+\frac{1}{2}S_{\mu\nu}\sum_c^{\mathrm{QM}}\Delta q_c\left[\frac{1}{2}(\gamma_{ac}+\gamma_{bc})+\frac{1}{3}(\Delta q_a\,\Gamma_{ac}+\Delta q_b\,\Gamma_{bc})+\frac{1}{6}\Delta q_c(\Gamma_{ca}+\Gamma_{cb})\right] \quad (4)$$

This problem needs to be solved iteratively until the Mulliken charges are self-consistent, hence "self-consistent-charge DFTB" (SCC-DFTB). The necessary matrix elements $H_{\mu\nu}^0$ and $S_{\mu\nu}$ are provided in the available DFTB parameterizations as functions of interatomic distance, in the form of series of equidistant points. The values for a specific molecular geometry are obtained by interpolation and transformation using Slater–Koster rules.[31]

QM/MM algorithms

In a QM/MM calculation,[9,32,33] the system is partitioned into two regions: a QM region, typically consisting of a relatively small number of atoms relevant for the specific process being studied, and a MM region with all the remaining atoms.

The total Hamiltonian of the system

$$\hat{H}=\hat{H}^{\mathrm{QM}}+\hat{H}^{\mathrm{MM}}+\hat{H}^{\mathrm{QM/MM}} \quad (5)$$

is composed of the Hamiltonian for the QM region eq. (4), which is defined by the QM method of choice, the Hamiltonian for the MM region eq. (1), defined by the chosen force

field and a term describing the coupling of the two regions, $\hat{H}^{\mathrm{QM/MM}}$. The latter contains contributions from the bonded, van der Waals and electrostatic interactions across the QM/MM boundary.

$$\hat{H}^{\mathrm{QM/MM}} = \hat{H}^{\mathrm{QM/MM}}_{\mathrm{bonded}} + \hat{H}^{\mathrm{QM/MM}}_{\mathrm{vdw}} + \hat{H}^{\mathrm{QM/MM}}_{\mathrm{el}} \quad (6)$$

Bonded interactions. If the partition of the system into QM and MM regions intersects covalent bonds, the bonded interactions between QM and MM atoms are treated with a single link atom scheme.[32,33] On the vector of each bond between a QM and MM atom, a link atom is placed at a fixed distance from the QM atom, depending on its chemical identity. In the QM calculation, the link atom is considered as an ordinary hydrogen atom. The original bond between the QM and the MM atom is fixed at constant length by means of a constraint. As the position of the link atom is a function of the positions of the QM and MM boundary atoms, the link atom does not contribute independent degrees of freedom.

More sophisticated QM/MM boundary treatments do exist, such as the double link atom approach[34] and hybrid orbital schemes,[35,36] but it has been found that the treatment of long-range electrostatics is more important than boundary schemes;[37] this is especially true if the QM/MM boundary is made at some distance from the chemical active region and if the total charge is conserved.

Angle (and dihedral) terms eq. (1) are excluded from the MM calculation if two out of three (three out of four) atoms constituting that term are QM atoms. This avoids duplicating interactions covered in the QM calculation.[38] After the QM calculation, the forces acting on the (virtual) link atoms are relayed to the (real) QM and MM boundary atoms according to lever rule to conserve the total energy, force and torque.

To avoid overpolarization of the atoms at the QM boundary and conserve the overall charge, the charges around the QM/MM boundary have to be redistributed. To this end, the charges of the MM boundary atoms are set to zero and redistributed evenly among the other MM boundary atoms. This procedure follows the *divided frontier charge* scheme, which was found to be especially suitable for QM/MM simulations with DFTB.[37]

Van der Waals interactions. The van der Waals interactions between QM and MM atoms, $\hat{H}^{\mathrm{QM/MM}}_{\mathrm{vdW}}$ are calculated using a 12–6 Lennard–Jones potential with standard force field parameters for both QM and MM atoms. In principle, the effective atomic radii and polarizabilities of the QM atoms depend on the distribution of electrons, so that these parameters might differ for QM and MM atoms of the same type.[39]

Electrostatic interactions. The electrostatic interactions between QM and MM regions are calculated in DFTB3 as Coulomb interaction between the atomic charges obtained from the DFTB3 Mulliken population analysis, Δq_a and the electrostatic potential induced by all MM charges, Φ_a. Thus, an additional term is included in the DFTB3 Hamilton matrix elements eq. (4).

$$H^{\mathrm{QM/MM}}_{\mu\nu} = S_{\mu\nu} \cdot \frac{1}{2}(\Phi_a \, \Delta q_a + \Phi_b \, \Delta q_b) \quad (7)$$

$$\Phi_a = \sum_{A \in \mathrm{MM}} \frac{Q_A}{|\vec{r}_A - \vec{r}_a|} \quad (8)$$

In the simplest case, all electrostatic interactions between MM and QM atoms are considered. However, for a small number of QM atoms computed with a fast QM method, embedded in a large number of MM charges, such a double sum is computationally expensive to evaluate. In a calculation with periodic boundary conditions (PBC), in which also the interaction of the QM region with its own periodic copies is included, the electrostatic interactions are better handled by the PME method.[11,40]

As the potential induced by all MM atoms on the QM atoms [Φ_a, eq. (8)] does not depend on the QM charge distribution itself, it can be computed beforehand. Then, the contribution of the periodic images of QM atoms needs to be considered separately in the self-consistent QM procedure. This has been pursued in other DFTB/PME implementations—in CHARMM[41–43] and AMBER[44,45]—where the full Ewald method was used to describe the interaction of the periodic images of the QM region, rather than PME. In this work, we follow an alternative, entirely PME-based route, in which the periodic images of the QM region are considered part of the MM region for the purpose of calculation of Φ_a. Still, Φ_a needs to be re-evaluated in every iteration of the self-consistent procedure. Note again that the MM charges remain constant, and this makes it possible to use an efficient approach for dealing with these interactions, which will be discussed below.

Approximate treatment of QM/MM electrostatics. Early QM/MM schemes (including the existing QM/MM interface in Gromacs) treated the QM/MM electrostatics with schemes based on simple cutoffs. For benchmark and showcase purposes, we also implemented several variants in this work, namely force-switch, force-shift, and the related reaction-field approach. The details are described in the Supporting Information.

Implementation details

The scheme is built on the available QM/MM interface in Gromacs.[46–48] Link atoms are handled using Gromacs' virtual interaction site feature (type 2). Details regarding virtual sites are explained in the Gromacs manual.[49] Use is made of the existing Gromacs routines and data structures as much as possible, and only slight modifications of the PME routines were made to meet the specific requirements of QM/MM electrostatics evaluation.

On each QM atom *a*, a QM/MM electrostatic potential φ_a is induced by all MM atoms, including their periodic images, and additionally the periodic images of all QM atoms. It can be written as a sum of several components,

$$\varphi_a = \varphi_a^{\mathrm{lr}} + \varphi_a^{\mathrm{sr}} + \varphi_a^{\mathrm{iex}} + \varphi_a^{\mathrm{self}} + \varphi_a^{\mathrm{sur}} \quad (9)$$

namely the long-range, short-range, intramolecular exclusion, self-interaction, and surface contributions, respectively. In the

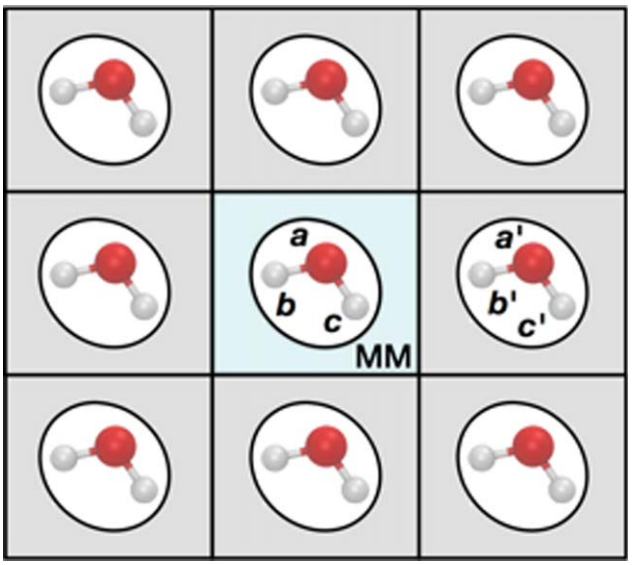

Figure 1. Schematic representation of a QM/MM setup with PBC. The QM/MM electrostatic interactions of atom a with MM atoms that are closer than neighborlist-cutoff distance are computed in the short-range part of the PME potential, φ^{sr}. The long-range part of the PME potential, φ^{lr} contains the interactions of atom a with the remaining MM atoms of the unit cell (blue box) as well as the interactions with all periodic images of all QM atoms $\{a', b', c'\}$ and the interactions with all periodic images of all MM atoms (grey boxes). Corrections have to be subtracted, to remove double counted interactions of QM atoms inside the unit cell (blue box) φ^{iex}, and the self-interaction of the QM atoms, φ^{self}. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Ewald framework, the long-range part of the potential φ_a^{lr} formally describes the interaction of Gaussian charge densities

$$G_i(\vec{r}) = \Omega_i \cdot \frac{\beta^3}{\pi^{\frac{3}{2}}} \exp\left(-\beta^2 |\vec{r} - \vec{r}_i|^2\right) \tag{10}$$

where Ω_i is the charge of atom i, so $\Omega_i = Q_i$ for MM atoms, and $\Omega_i = \Delta q_i$ for QM atoms. The parameter β controls the spatial extension of the densities. The long-range potential takes the form

$$\varphi_a^{\mathrm{lr}} = \frac{4\pi}{V} \sum_i^{\mathrm{QM} \cup \mathrm{MM}} \Omega_i \sum_{\vec{k} \neq \vec{0}} \frac{1}{|\vec{k}|^2} \cdot \exp\left(-\frac{|\vec{k}|^2}{4\beta^2}\right) \cdot \exp\left(-\mathrm{i} \cdot \vec{k} \cdot (\vec{r}_i - \vec{r}_a)\right) \tag{11}$$

The former summation runs over all of the atoms i in the system, the latter spans the reciprocal space by means of the vector $\vec{k}$, and V is the volume of the simulation box. The expression is evaluated with the smooth PME routines implemented in Gromacs. φ_a^{lr} is evaluated using all charges, including the charges of the QM atoms $\{a, b, c\}$ and their periodic images $\{a', b', c'\}$, see Figure 1.

In a typical situation, the number of QM charges Δq_a is much smaller than the number of MM partial charges Q_A. Furthermore, the MM charges Q_A are constant throughout the calculation. Exploiting these facts, we propose a more efficient way to obtain φ_a^{lr} in an SCC calculation, called "separated PME" in the following. In the first cycle, φ_a^{lr} is evaluated using all charges, $\varphi_a^{\mathrm{lr},1}$. In the following cycles, it is beneficial to only calculate the differential potential $\Delta\varphi_a^{\mathrm{lr}}$, which arises from the differential QM charge (with respect to the QM charge obtained in the first step of the cycle): $\Delta q_a = \Delta q_a^1 + \Delta\Delta q_a$.

$$\varphi_a^{\mathrm{lr}} = \varphi_a^{\mathrm{lr},1}(\Delta q_a, Q_A) + \Delta\varphi^{\mathrm{lr}}(\Delta\Delta q_a) \tag{12}$$

As most of the charges are MM charges and therefore can be discarded in all steps of the SCC but the first one, the computational cost of the subsequent grid calculations is decreased significantly.

The short-range component φ_a^{sr} is evaluated in the real space and includes interactions with MM atoms A that are in the neighbor list of QM atom a

$$\varphi_a^{\mathrm{sr}} = \sum_A^{\mathrm{nblist}} \frac{Q_A}{|\vec{r}_A - \vec{r}_a|} \cdot \mathrm{erfc}(\beta \cdot |\vec{r}_A - \vec{r}_a|) \tag{13}$$

As this potential does not depend on the QM charges Δq_a, it can be evaluated prior to the SCC. To reduce the computational effort, the short-range neighbor lists are built in regular intervals, here every 10 fs. The neighbor list of a QM atom contains all of the MM atoms that are closer than a certain distance, which is chosen such that φ_a^{sr} may be neglected at that distance. This is affected by the choice of β: A shorter cutoff is sufficient for the short-range component when β is large, while a smaller number of $\vec{k}$-points is sufficient in the long-range component when β is small.

The long-range potential φ_a^{lr} also contains interactions of QM atom a with the other QM atoms $\{b, c\}$ inside the unit cell (blue box in the center of Fig. 1), which are already described by the QM method. Therefore, these interactions, φ_a^{iex} have to be subtracted from φ_a^{lr}; this is sometimes called "exclusions":

$$\varphi_a^{\mathrm{iex}} = -\sum_{b \neq a}^{\mathrm{QM}} \frac{\Delta q_b}{|\vec{r}_b - \vec{r}_a|} \cdot \mathrm{erf}(\beta \cdot |\vec{r}_b - \vec{r}_a|) \tag{14}$$

Likewise, the interaction of QM atom a with itself has to be subtracted as well

$$\varphi_a^{\mathrm{self}} = -\frac{2}{\sqrt{\pi}} \beta \Delta q_a \tag{15}$$

Optionally, a surface term (or dipole correction) may be added to the PME potential for unit cells with zero overall charge and nonvanishing dipole moment,

$$\varphi_a^{\mathrm{sur}} = \frac{4\pi}{3\varepsilon_s} \frac{\vec{r}_a \cdot \vec{\mu}_{\mathrm{tot}}}{V} \tag{16}$$

where $\vec{\mu}_{\mathrm{tot}}$ is the total dipole moment of the simulation system. This yields the energy of the periodic system surrounded by a material with dielectric constant of ε_s (so, $\varepsilon_s = 1$ for vacuum). If the surface term is not included, then the situation corresponds to the system being surrounded by a conductor ($\varepsilon_s = \infty$), a.k.a. the "tin-foil" boundary conditions, which are mostly applied in biomolecular simulations.

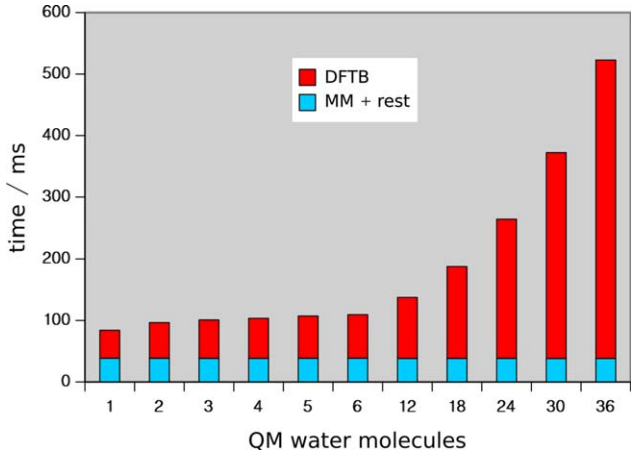

Figure 2. Computational cost of 1 time step. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Forces

The expressions for the gradients of DFTB3 energies derived in Ref. [21] were implemented. Analytic expressions were implemented for all but the derivatives of $H^0_{\mu\nu}$ and $S_{\mu\nu}$, which are computed numerically through finite differencing instead.

In addition, we also implemented an alternative approach based on Catmull–Rom spline interpolation of $H^0_{\mu\nu}$ and $S_{\mu\nu}$ in combination with analytical derivatives of these matrix elements. Both the numeric values of the forces and the conservation of energy are little affected by these modifications; a detailed report is presented in the Supporting Information.

For the long-range QM/MM contribution to the forces, two additional long-range PME calculations are necessary: one containing all charges for the forces on the QM atoms, and another containing only the charges of the QM atoms to obtain the forces on the MM atoms. The overhead associated with this is, however, negligible as will be shown in the Applications. The force due to φ_a^{sr} is obtained in the same double loop over the QM/MM neighbor lists as the contribution to the QM forces; the forces due to φ_a^{iex} and φ_a^{sur} are computed in a double and single loop over QM atoms, respectively. The gradients of $\varphi_a^{\mathrm{self}}$ vanish.

Applications

Simulation setup

All QM/MM simulations reported below were performed using the leap-frog integrator with a time step of 0.5 fs. The electrostatics were treated with PME, while the Lennard–Jones interactions were cutoff at 10 Å; the neighbor lists were created every 20 MD steps (10 fs). A constant temperature of 300 K and a constant pressure of 1 bar was maintained by means of the Nosé–Hoover thermostat with a time constant of 0.5 ps, and the Parrinello–Rahman barostat with a time constant of 0.5 ps, respectively. The calculations of both thermostat and barostat were performed in every step of MD (options nsttcouple = 1 and nstpcouple = 1).

Timing

The computational efficiency of the current QM/MM framework was assessed on a system of pure water, with various numbers of water molecules included in the QM region. The system consisted of a periodic box of $5 \times 5 \times 5$ nm^3 filled with 4142 water molecules, of which 1, 2, 3, 4, 5, 6, 12, 18, 24, 30, or 36 molecules are described with QM, and the remainder is modelled as TIP3P (MM) molecules. There are no covalent bonds cutting the QM–MM boundary. As there are no significant requirements for main memory or disk space (neither for the force field, nor for the DFTB calculation), the computational time is recorded and discussed as the measure of performance. All of the simulations are performed on a single core of an Intel Core i7-970 CPU (3.2 GHz).

The computational cost to perform one step of an MD simulation of above described water systems is presented in Figure 2. As expected, the DFTB calculation takes most of the computational time, however, the QM:MM timing ratio only reaches an order of magnitude for the largest QM region. Also rather surprising is the weak dependence on the number of QM water molecules.

It is instructive to see which components of the DFTB calculation itself are the most demanding. First, we will analyze the iterative SCC procedure within DFTB. From the point of view of computational efficiency, its main components are the PME calculation of electric potential on the QM atoms and the diagonalization of the charge-dependent Hamiltonian matrix. All of the other calculations involved (the set-up of the Hamiltonian, the identification of occupied orbitals, the calculation of Mulliken atomic charges, and the Broyden mixing of these charges) will be referred to as "rest" in the following; see Figure 3 for the data. Apparently, PME was the bottle neck for the QM size of up to 12 water molecules, while the diagonalization with its cubic computational complexity dominated starting at 24 water molecules. The PME calculation took the same amount of computational time for all of the systems in this set because of the identical box size and nearly identical number of MM atoms in the simulations.

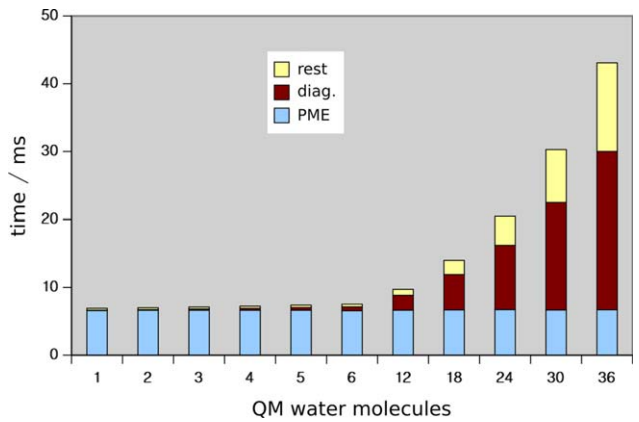

Figure 3. Computational cost of one SCC cycle in QM/MM simulations of the water systems. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

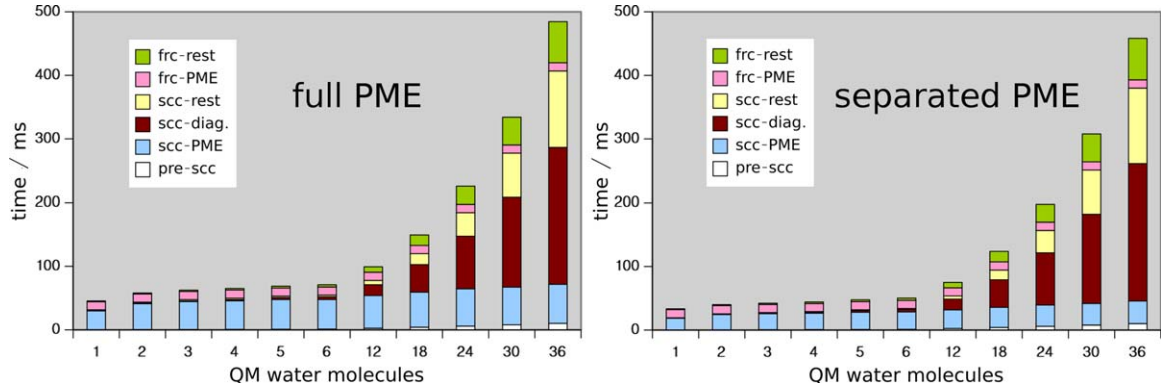

Figure 4. Computational cost of QM/MM SCC-DFTB calculation—full PME (left), separated PME (right). "prescc"—before SCC cycle (charge-independent calculations), "scc"—calculations within the SCC cycle (PME, diagonalization of the Hamiltonian and the rest), "frc"—post-SCC calculation of forces (PME and the rest).

Also, the newly proposed "separated PME" approach was tested. In the pure water systems considered in this work, full PME calculations took 6.7 ms each, and this was reduced to 3.4 ms when passing to the separated PME. This acceleration is caused most likely by the spreading of charges onto the PME grid taking a much shorter time due to the much smaller number of atoms considered. Note that the actual solution of Poisson's equation may not be accelerated at all.

The overall computational cost of the QM/MM calculations has been compared and decomposed, with both full PME and separated PME schemes, see Figure 4. The PME calculations dominated the total QM/MM computational cost in the simulations with small QM zones with up to 12 water molecules, and the acceleration due to separated PME was most notable (25% or more) for these smaller QM zones. For the largest QM zone considered, with 36 water molecules, the separated-PME simulations ran 5% faster than full PME. Note again that separated PME is no approximation and delivers results that are identical with full PME.

With the increasing size of the QM zone, the share of computational cost due to the calculation of forces decreased to ca. 20%, involving two additional PME runs. Thus, it seems that the meaningful target of future optimization is rather another component of the SCC procedure, the diagonalization of the charge-dependent Hamiltonian matrix. A possible option to accelerate and parallelize the diagonalization is the Jacobi rotation technique. Another strategy may be passing to an extended Lagrangian scheme replacing the standard Born–Oppenheimer ansatz, which was already done with an earlier DFTB model.[50] Nothing would have to change in our implementation to make use of that because the whole infrastructure supports parallel DFTB execution as soon as it is ready.

Malonaldehyde–ultrafast proton transfer

Featuring an ultrafast PT reaction, the enolic form of malonaldehyde molecule (3-hydroxy-2-propenal, Fig. 5) has long been a popular test system for quantum chemical methods aiming at the calculation of reaction rates and free energies,[51,52] as reviewed in Ref. 53. More involved studied include a recent application of extended sampling simulation techniques[54] and even quantum dynamics simulations in the excited state.[55] It should be noted that the PT possesses a non-negligible quantum character, making a tunneling mechanism of transfer possible.[56,57] To include tunneling, a QM treatment of the nuclei would be required, as demonstrated previously.[58]* As we use classical dynamics, tunneling cannot be described. Still, it will be demonstrated how the current implementation evaluates efficiently the potential energy of the system and its gradients, which may be used to run the dynamics of the molecule with an available classical or quantum engine.

The PT reaction in malonaldehyde was previously studied with a QM/MM scheme including the older second-order model SCC-DFTB and the corresponding parameterization.[59] Symmetric double-well potential was reproduced, and the barrier height was 4.2 and 3.3 kcal/mol in the solvated system and in the gas phase, respectively, in this previous work.

As for the choice of the reaction coordinate, a geometrical coordinate like the difference of O_1—H and O_2—H bond lengths is often used. Instead, it may be of advantage to use a nongeometrical coordinate based on the electronic structure, and this will be done here: the difference of self-consistent Mulliken charges of the oxygen atoms O_1 and O_2 will be considered (ΔQ). We note that despite its easy applicability in a free simulation, the use of such a coordinate would be difficult in a biased sampling simulation like US or metadynamics because the derivatives of Mulliken charges with respect to atom positions would have to be evaluated.

The difference of charges ΔQ is recorded along a free MD simulation of 10 ns, and its probability distribution function $P(\Delta Q)$ is obtained. This is then readily converted to the free energy, $\Delta G(\Delta Q) = -k_B T \cdot \log P(\Delta Q)$, where k_B is the Boltzmann constant and T is the temperature. The resulting free energy curves are presented in Figure 6 (left); the QM/MM electrostatic interactions were treated with PME. Additionally, a free energy dependence is obtained from a simulation of malonaldehyde in the gas phase, for comparison (red line).

Both in solution and in the gas phase, the free energy profile converges to a symmetric double-well shape within a

*In addition, it is also possible to run path-integral, or ring-polymer MD, but these approaches are beyond the scope of the present application.

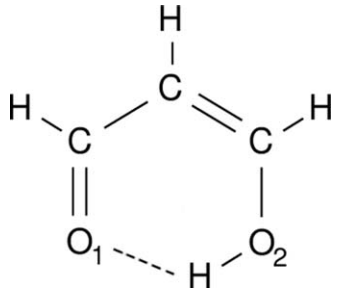

Figure 5. The internally hydrogen-bonded form of malonaldehyde. After a completed PT event, the H atom is bonded to O_1 and there is a hydrogen bond between H and O_2; the single and double carbon–carbon bonds exchange characters as well.

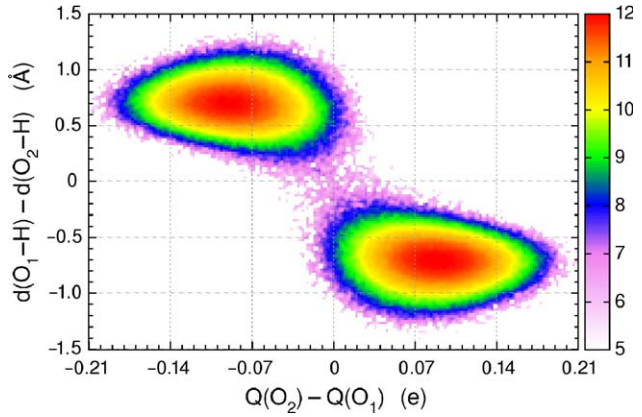

Figure 7. Correlation diagram of the reaction coordinates from the simulation of malonaldehyde in the gas phase; 2D histogram of the difference of charges of O atoms and the difference of O–H distances. Coded by color is the probability density (logarithmic scale, arb. units). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

simulation period of 10 ns. The barrier heights are 2.2 and 1.6 kcal/mol for the solvated system and the gas phase, respectively. When the course of PT is described with respect to the geometrical reaction coordinate, see Figure 6 (right), a different reaction pathway is considered, leading to different results: 4.0 and 3.2 kcal/mol, respectively. The latter values agree well with the previous DFTB work, despite the different DFTB model and parameterization used.[59] The nonlinear relationship of the two reaction coordinates is also illustrated by their correlation plot in Figure 7. In other words, some transfer of electron density between O_1 and O_2 is possible even when the proton is not being transferred physically. The charge difference thus appears to be a rather inferior reaction coordinate.

A popular way to enhance the sampling along geometrical reaction coordinates and to obtain free energies is the US.[60,61] To demonstrate the usefulness of the current QM/MM implementation, the free energy associated with PT in malonaldehyde was evaluated by means of US simulations as well. As the required complex reaction coordinate, the difference of two interatomic distances, is not available in the Gromacs package itself, the PLUMED plugin (version 2.1) was used for this purpose.[62] The reaction coordinate was scanned in steps of 0.05 Å, with a sampling of 25 ps in every window preceded by 5 ps of equilibration; the total simulation effort was thus 1.1 ns. Free energy curves obtained with the WHAM method[63,64] are presented in Figure 6 (right) together with those from the above presented free simulations of 10 ns. The US result is qualitatively the same as that from the free simulations; the barriers are slightly lower than in the free simulations, and the difference is 0.5 kcal/mol in the solvated system (QM/MM) and 0.25 kcal/mol in the gas phase. This suggests that the barrier region may have been undersampled in the free simulations.

Conformation of dialanine

Dialanine ($CH_3CO\text{-}NHCH(CH_3)CO\text{-}NHCH_3$) is one of the smallest molecules featuring a peptide-like pair of (φ, ψ) angles, and as such it has been subject of computational chemistry studies for over two decades.[65–69] It has been also used to assess the performance of MM force fields as well as of QM/MM schemes, even involving older DFTB models, to describe the secondary structure of peptides.[44,70,71] It was suggested that such a capped amino acid captures the main features of backbone conformation in proteins.[72,73] The effect of water on the conformation of dialanine was shown in a nuclear magnetic resonance study,[74] and the aforementioned previous work often

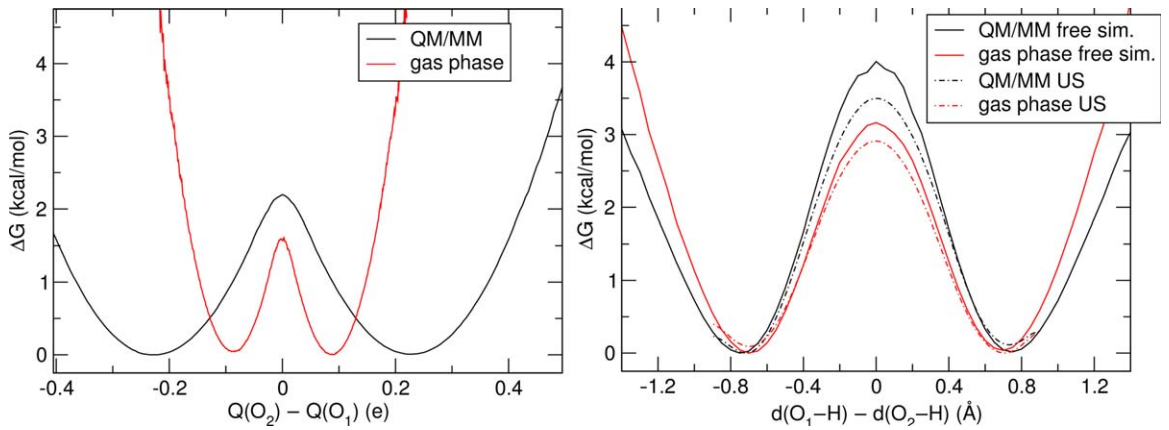

Figure 6. Free energy of PT in malonaldehyde (kcal/mol) with different choices of the reaction coordinate. Left—difference of DFTB Mulliken charges of the oxygen atoms; right—difference of distances of the hydrogen atom from the oxygen atoms ("US"—result from US simulations).

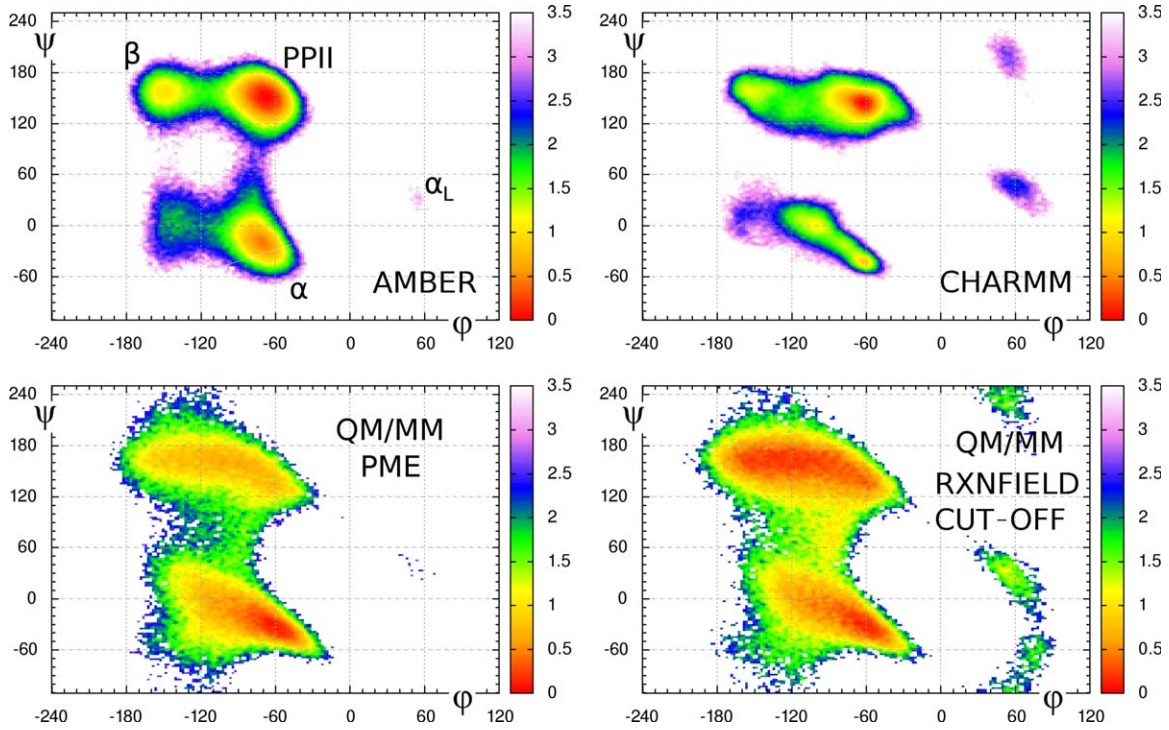

Figure 8. Conformation of dialanine as obtained with the different computational methods, expressed in terms of Ramachandran plots (φ and ψ in degrees). Coded by color is the free energy (kcal/mol), with $\Delta G = 0$ for the respective global minimum. The labeling of the different secondary structure types is shown in the Amber panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

concentrated on the conformational difference between dialanine in the gas phase and in aqueous solution.

Experimental studies reported that the polyproline II conformation (PPII) dominates in alanine oligopeptides[75,76] as well as various other small model peptides.[77] Other relevant conformations are the right-handed helix (α) and the extended conformation (β), but not the left-handed helix (α_L). An early study compared the structure of dialanine yielded by various force fields and by DFTB-based QM/MM.[70] Electrostatic QM/MM interaction were switched to zero from 8 to 12 Å. Major populations of the α and PPII–β states are observed with force fields as well as QM/MM. Notably, force fields yield much higher barriers between the two basins than the QM/MM method does.

To test the validity of our QM/MM implementation, we performed free MD simulations of dialanine in water (1168 TIP3P molecules) of 40 ns, using DFTB3/3OB with the D3 correction. Also performed were purely classical MD simulations with two commonly used force fields, Amber parm14SB[15,78] and charmm36,[16] in every case with TIP3P water.[79] Note that the treatment of nonbonded interactions differed in the MM simulations, following the recommended simulation protocols for the respective force fields.†

The topology of the molecule was prepared with AmberTools,[80] and converted to the Gromacs format with Amb-conv.[81] Prior to production simulations, an equilibration procedure was performed with the parm14SB force field; the time step was 1 fs. After drawing the velocities of atoms from the Maxwell–Boltzmann distribution at 10 K, the first phase of equilibration consisted of heating to 300 K during 10 ps, and was followed by further 10 ps of simulation at a constant temperature of 300 K maintained with the Berendsen thermostat. Finally, a simulation of 100 ps was performed with the Berendsen thermostat at 300 K and the Berendsen barostat at 1 bar. The conformation of the molecule was characterized in terms of the (φ, ψ) pairs along each of the MD trajectories (Ramachandran plot). Two-dimensional histograms were built from these data with a bin width of 2° in every dimension, and this probability distribution function P was converted to the free energy as $\Delta G(\varphi, \psi) = -k_B T \cdot \log P(\varphi, \psi)$. The resulting free energy maps are presented in Figure 8.

In all simulations, there are two main basins of minimal free energy, corresponding to the PPII helix together with the extended conformation (β), and to the right-handed helix (α). The force-field simulations (Amber as well as CHARMM) feature two distinct minima for PPII and β, while the QM/MM simulations show one large basin extending over PPII and β. Both Amber and CHARMM predict PPII as the most favored conformation, which agrees with the experimental observations; the QM/MM simulations provide similarly deep minima in the regions PPII+β and α, and the QM/MM with dipole corrected PME ($\varepsilon_s = 1$) shows nearly identical depths. A striking difference in the Ramachandran plot is apparent in the region between the two main basins: The barrier for passing between them is ca. 1.5 kcal/mol with QM/MM, while it is ca. 2.5 kcal/mol with Amber

†A plain cutoff at 10 Å was used for van der Waals interactions with parm14SB, while a cutoff switched to zero between 10 and 12 Å was applied with CHARMM36. This also resulted in lower computational efficiency of the CHARMM36 simulation.

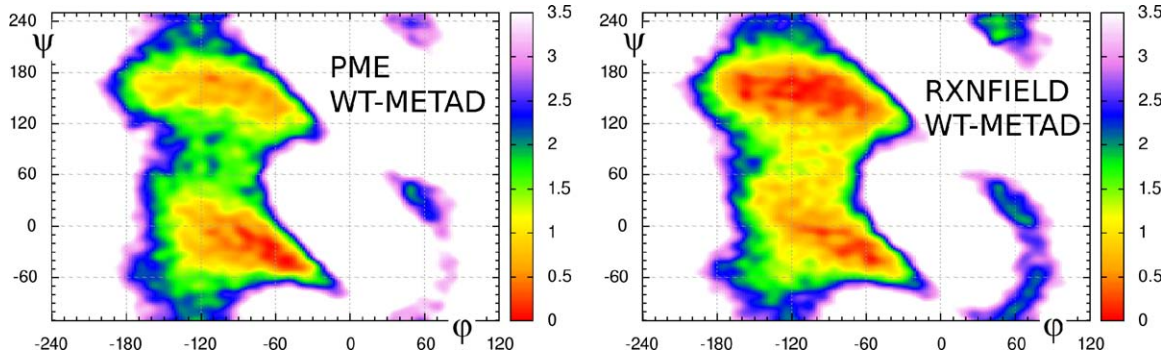

Figure 9. Free energy of dialanine as obtained from well-tempered metadynamics simulations with the different computational methods, expressed in terms of Ramachandran plots (φ and ψ in degrees). Coded by color is the free energy (kcal/mol), with $\Delta G = 0$ for the respective global minimum. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and at least 3.5 kcal/mol with CHARMM. A similar observation was made in the previous Amber work,[44] and it was noted that this may be due to the harmonic character of the force fields.

Several cutoff based treatments of QM/MM electrostatics were also tested, refer to the Supporting Information for the data. Apparently, the conformational dynamics is not very dependent on the long-range electrostatics for this uncharged molecule.

We also performed well-tempered metadynamics simulations[82–84] with PLUMED. Every 1 ps, a 2D Gaussian was added on the $\varphi-\psi$ subspace, with initial height of 0.2 kcal/mol and spatial extension of 0.35 rad in both dimensions; the bias factor was set to 4. The resulting free energy maps based on the sum of potential energies added along the simulations are shown in Figure 9. The agreement between the results from well-tempered metadynamics and free simulations is very good concerning the dominant conformations (α and PPII-β), while the nondominant conformations like the left-handed helix α_L are not converged in the free simulations of 10 ns. The convergence of the well-tempered metadynamics simulations is demonstrated in the Supporting Information.

Thymine dimer repair in DNA

The cyclobutane pyrimidine dimer (CPD) is a potentially mutagenic photoproduct of UV irradiation of DNA,[85] see Figure 10. Catalytic repair of this damage is essential, and one such repair mechanism involves the enzyme photolyase, which is believed to transfer an electron into the CPD. The resulting radical anion then undergoes a rapid dissociation.[86,87] Here, we performed QM/MM simulations at the DFTB3/3OB/D3 level of CPD embedded in a double stranded DNA oligomer, to verify whether our implementation can reproduce the findings in earlier QM/MM studies.[88] The DFTB3/3OB description of the radical anionic CPD in vacuo is compared with full-DFT data in the Supporting Information.

The computation in this work involved a CPD embedded in a double-helical DNA oligonucleotide of the sequence T_8:A_8. The initial structure was obtained by introducing distance restraints between the fourth and fifth thymine nucleobases of the first strand in an idealized B-DNA conformation; thus, the resulting nucleobase sequence was TTT(CPD)TTT:AAAAAAAA. The simulation setup altogether is an analogy to that in Ref. [88]. First, a force-field equilibration was performed. This consisted of an energy minimization, followed by a 20 ps MD simulation at a constant temperature of 300 K, and concluded by a 200 ps simulation at a constant temperature of 300 K and a constant pressure of 1 bar maintained with the Berendsen thermostat and Berendsen barostat, respectively.

First, the new QM/MM scheme was used to perform free simulations of the oligonucleotide containing the CPD lesion. The QM region consisted of the CPD itself while the remainder of the DNA molecule was modelled with the Amber parm99BSC0 force field.[89] Obviously, unlike the other applications in this work, the QM–MM boundary cut covalent bonds, namely the N-glycosidic bonds between the thymine nucleobases and the deoxyribose units. Hydrogen link atoms connected to the N1 atoms of the thymines were introduced and the cut bonds were constrained to their force-field equilibrium values.

The CPD moiety was stable over the entire length of 100 ps in such a free QM/MM simulation. To sample the anticipated process of disintegration of CPD slightly better, 100 independent QM/MM simulations with negatively charged CPD (radical anion) were performed, starting from structures collected along the previous simulation with neutral CPD. In a very simplified manner, this setup would simulate an electron uptake from an external electron donor. The length of the bonds between the two thymine moieties were measured along

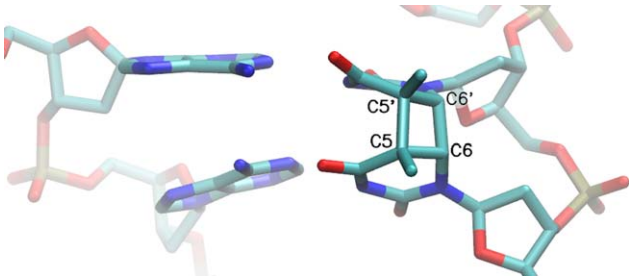

Figure 10. CPD embedded in a double-helical DNA oligonucleotide. Atoms are colored by element (C—cyan, N—blue, O—red, P—yellow), and hydrogens are not shown for clarity. The relevant carbon atoms are labeled consistently with the text. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

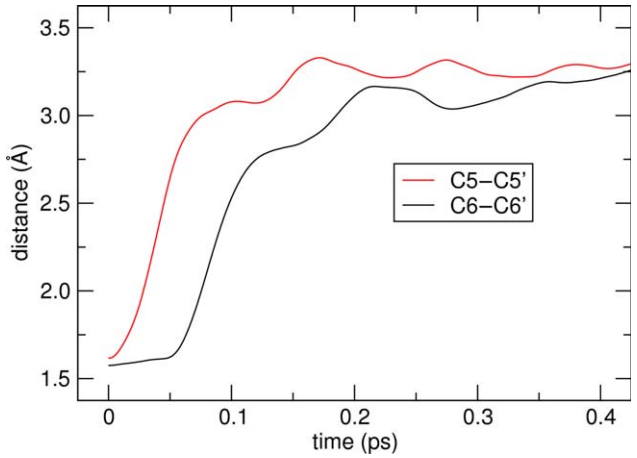

Figure 11. Length of the bonds C5—C5′ and C6—C6′ in the course of (non-equilibrium) QM/MM simulations performed with negatively charged CPD, which started from structures taken from a simulation with neutral CPD. The data represent an average over 100 of such simulations with different starting structures. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

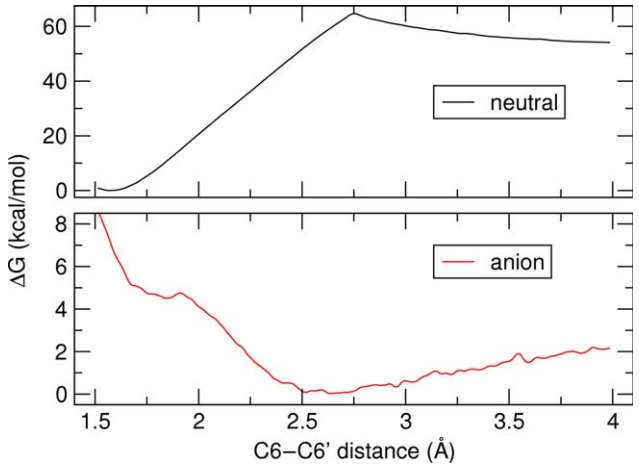

Figure 12. Free energy variation of a solvated double-helical DNA oligonucleotide as a function of the distance of C6 atoms in the centrally located thymine dimer. Data were obtained as potentials of mean force from series of US QM/MM simulations performed with DFTB3/3OB+D3. The QM system considered was the thymine dimer moiety, neutral as well as anion radical. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

these nonequilibrium simulations, and averaging over the 100 simulation was performed. The results are presented in Figure 11 while the data from all individual simulations may be referred to in the Supporting Information.

At the DFTB/Amber level of theory, it is the C5—C5′ bond that breaks immediately after the electron uptake, so that the two atoms are 3 Å apart already 100 fs after the electron uptake. The C6—C6′ bond, conversely, follows a slightly different fate: its length is hardly changing for some 50 fs before it eventually breaks as well. The entire process takes no more than 200 fs to complete.

To obtain the free energy profile of the transition from the CPD to two separate thymine nucleobases, US simulations were performed with the Gromacs "pull code." Following Ref. [88], the C6—C6′ distance was considered as the reaction coordinate because this bond broke more slowly than C5—C5′ in the free simulations. The reaction coordinate spanned the interval from 1.5 to 4.0 Å, which was divided into 51 simulations (windows) separated by 0.05 Å. In each window, a simulation of 10 ps was performed, from which the initial period of 1 ps was discarded (as equilibration); the accumulated simulation time thus amounted to twice 0.5 ns. The resulting free energy curves are presented in Figure 12.

The CPD is stable in the neutral state, and the results of the QM/MM simulation predict a deep minimum of free energy for the dimer structure accordingly. The histograms obtained from the individual biased simulations show sufficient overlap (data not shown), and the free energy curve is smooth, except for a kink at 2.75 Å. An increased sampling would be required to remove the possible uncertainty in the free energy in that region of the reaction coordinate. However, such endeavor goes beyond the scope of the current work.

The free energy curve obtained from current simulations of the radical anion of CPD is rather intriguing. The bound state seems to correspond to a very shallow minimum, separated from the dissociation to two unbonded molecules by a barrier lower than thermal energy (note that 1 k_BT=0.6 kcal/mol). This result supports the assumption of the metastable, extremely short-lived state in which the C6—C6′ bond is intact and the C5—C5′ has already broken. Still, the presence of a barrier means that this state could be (in principle) trapped and observed at very low temperatures.

Our observations agree qualitatively with the results of the previous QM/MM simulations,[88] which involved full DFT calculations at the level BLYP/TZV2P with a dispersion correction. In that work, a total of seven simulations of CPD radical anion over 2 ps each were performed, due to the large computational cost, and were complemented by metadynamics simulations. Much the same as in our DFTB QM/MM simulations, the C5—C5′ bond broke spontaneously while the C6—C6′ bond was metastable. The difference is that the barrier for C6—C6′ breaking is lower in our present work, and thus the breaking occurs on a shorter time scale. Still, the qualitative agreement of our results with those obtained using a full DFT-based scheme is encouraging for the new QM/MM implementation as well as DFTB3 itself.

Outlook

We have presented a new implementation of DFTB3 in the QM/MM framework of Gromacs. We have shown that the efficiency of the PME implementation for calculating the electrostatic interaction extends to the QM/MM part as well. In the presentation of three studies of typical biophysical problems, we have demonstrated the compatibility of the current implementation with the extensive sampling and free energy capabilities of both Gromacs itself and of the PLUMED plugin.

Further optimizations focus on the two time-dominating steps of the QM/MM simulations. The PME calculation of the QM/MM electrostatics is parallelizable with Gromacs efficiently,

and it is accelerated by the newly proposed separated PME treatment. The remaining bottle neck is the diagonalization of the Hamiltonian, and further development is going to deal with this issue.

The modified Gromacs program 5.0 as presented in this work can be obtained from our website at http://cbp.cfn.kit.edu/joomla/index.php/downloads along with the instructions for the installation and usage.

Acknowledgments

The authors are grateful for fruitful discussions to Carsten Kutzner, Honza Řezáč, Marcus Elstner, Max Kubillus, and Qiang Cui. G.G. acknowledges support from the Academy of Finland.

Keywords: quantum mechanics/molecular mechanics · density-functional tight-binding · free energy simulation · extended sampling · molecular dynamics

How to cite this article: T. Kubař, K. Welke, G. Groenhof. *J. Comput. Chem.* **2015**, *36*, 1978–1989. DOI: 10.1002/jcc.24029

Additional Supporting Information may be found in the online version of this article.

[1] M. W. van der Kamp, A. J. Mulholland, *Biochemistry* **2013**, *52*, 2708.
[2] S. C. L. Kamerlin, A. Warshel, *Proteins* **2010**, *78*, 1339.
[3] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.
[4] M. Hoffmann, M. Wanko, P. Strodel, P. H. König, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid, M. Elstner, *J. Am. Chem. Soc.* **2006**, *128*, 10808.
[5] I. Navizet, Y.-J. Liu, N. Ferré, H.-Y. Xiao, W.-H. Fang, R. Lindh, *J. Am. Chem. Soc.* **2010**, *132*, 706.
[6] G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. de Visser, H. J. C. Berendsen, M. Olivucci, M. A. Robb, *J. Am. Chem. Soc.* **2004**, *126*, 4228.
[7] A. M. Virshup, B. G. Levine, T. J. Martínez, *Theor. Chem. Acc.* **2014**, *133*, 1506.
[8] I. Schapiro, S. Ruhman, *Biochim. Biophys. Acta, Bioenergetics* **2014**, *1837*, 589.
[9] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227.
[10] M. Gaus, Q. Cui, M. Elstner, *WIREs Comput. Mol. Sci.* **2014**, *4*, 49.
[11] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.
[12] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435.
[13] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, M. Parrinello, *Comput. Phys. Commun.* **2009**, *180*, 1961.
[14] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, W. F. van Gunsteren, *J. Phys. Chem. A* **1999**, *103*, 3596.
[15] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712.
[16] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *J. Chem. Theory Comput.* **2012**, *8*, 3257.
[17] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657.
[18] G. Seifert, D. Porezag, T. Frauenheim, *Int. J. Quantum Chem.* **1996**, *58*, 185.
[19] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, R. Kaschner, *Phys. Rev. B* **1995**, *51*, 12947.
[20] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, *Phys. Rev. B* **1998**, *58*, 7260.
[21] M. Gaus, Q. Cui, M. Elstner, *J. Chem. Theory Comput.* **2011**, *7*, 931.
[22] P. Koskinen, V. Mäkinen, *Comput. Mater. Sci.* **2009**, *47*, 237.
[23] G. Seifert, J.-O. Joswig, *WIREs Comput. Mol. Sci.* **2012**, *2*, 456.
[24] Y. Yang, H. Yu, D. M. York, Q. Cui, M. Elstner, *J. Phys. Chem. A* **2007**, *111*, 10861.
[25] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, *J. Chem. Phys.* **2001**, *114*, 5149.
[26] S. M. Cybulski, T. M. Bledson, R. R. Toczyłowski, *J. Chem. Phys.* **2002**, *116*, 11039.
[27] L. Zhechkov, T. Heine, S. Patchkovskii, G. Seifert, H. A. Duarte, *J. Chem. Theory Comput.* **2005**, *1*, 841.
[28] S. Grimme, J. Anthony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[29] J. G. Brandenburg, S. Grimme, *J. Phys. Chem. Lett.* **2014**, *5*, 1785.
[30] E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2006**, *124*, 174104.
[31] J. C. Slater, G. F. Koster, *Phys. Rev.* **1954**, *94*, 1498.
[32] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1986**, *7*, 718.
[33] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700.
[34] D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chetfield, M. Hodošček, B. R. Brooks, *J. Chem. Phys.* **2002**, *117*, 10534.
[35] J. Gao, P. Amara, C. Alhambra, M. J. Field, *J. Phys. Chem. A* **1998**, *102*, 4714.
[36] Y. Wang, J. Gao, *J. Phys. Chem. B* **2015**, *119*, 1213.
[37] P. H. König, M. Hoffmann, T. Frauenheim, Q. Cui, *J. Phys. Chem. B* **2005**, *109*, 9082.
[38] K. P. Eurenius, D. C. Chatfield, B. R. Brooks, M. Hodošček, *Int. J. Quantum Chem.* **1996**, *60*, 1189.
[39] T. J. Giese, D. M. York, *J. Chem. Phys.* **2007**, *127*, 194101.
[40] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577.
[41] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus, *J. Phys. Chem. B* **2001**, *105*, 569.
[42] D. Riccardi, P. Schaefer, Q. Cui, *J. Phys. Chem. B* **2005**, *109*, 17715.
[43] K. Nam, J. Gao, D. M. York, *J. Chem. Theory Comput.* **2005**, *1*, 2.
[44] G. Seabra, R. C. Walker, M. Elstner, D. A. Case, A. E. Roitberg, *J. Phys. Chem. A* **2007**, *111*, 5655.
[45] R. C. Walker, M. F. Crowley, D. A. Case, *J. Comput. Chem.* **2008**, *29*, 1019.
[46] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.
[47] G. Groenhof, M. Boggio-Pasqua, L. V. Schäfer, M. A. Robb, In Advances in Quantum Chemistry, Vol. *59*; J. R. Sabian, E. Brändas, Eds.; Academic Press, Waltham, MA, **2010**; pp. 181–212.
[48] G. Groenhof, In Methods in Molecular Biology, Vol. *924*; L. Monticelli, E. Salonen, Eds.; Humana Press, New York, NY, **2013**; pp. 43–66.
[49] GROMACS, user manual, Version 5.0., Available at: http://www.gromacs.org, **2014**. Last accessed on 27 July 2015.
[50] G. Zheng, A. M. N. Niklasson, M. Karplus, *J. Chem. Phys.* **2011**, *135*, 044122.
[51] G. Karlström, H. Wennerström, B. Jönsson, S. Forsén, J. Almlöf, B. Roos, *J. Am. Chem. Soc.* **1975**, *97*, 4188.
[52] A. D. Isaacson, K. Morokuma, *J. Am. Chem. Soc.* **1975**, *97*, 4453.
[53] F. Fillaux, B. Nicolaï, *Chem. Phys. Lett.* **2005**, *415*, 357.
[54] Y. Mori, Y. Okamoto, *Phys. Rev. E* **2013**, *87*, 023301.
[55] J. D. Coe, T. J. Martínez, *J. Phys. Chem. A* **2005**, *110*, 618.
[56] M. Tuckerman, D. Marx, *Phys. Rev. Lett.* **2001**, *86*, 4946.
[57] A. Yamada, H. Kojima, S. Okazaki, *J. Chem. Phys.* **2014**, *141*, 084509.
[58] K. F. Wong, J. L. Sonnenberg, F. Paesani, T. Yamamoto, J. Vaníček, W. Zhang, H. B. Schlegel, D. A. Case, T. E. Cheatham, W. H. Miller, G. A. Voth, *J. Chem. Theory Comput.* **2010**, *6*, 2566.
[59] Ł. Walewski, P. Bała, M. Elstner, T. Frauenheim, B. Lesyng, *Chem. Phys. Lett.* **2004**, *397*, 451.
[60] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **1977**, *23*, 187.
[61] B. Roux, *Comput. Phys. Commun.* **1995**, *91*, 275.
[62] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604.
[63] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 1011.
[64] A. Grossfield, WHAM: The weighted histogram analysis method, Version 2.0.9., Available at: http://membrane.urmc.rochester.edu/content/wham, **2013**. Last accessed on 27 July 2015.
[65] P. J. Rossky, M. Karplus, *J. Am. Chem. Soc.* **1979**, *101*, 1913.
[66] I. K. Roterman, M. H. Lambert, K. D. Gibson, H. A. Scheraga, *J. Biomol. Struct. Dyn.* **1989**, *7*, 421.
[67] D. J. Tobias, C. L. Brooks, *J. Phys. Chem.* **1992**, *96*, 3864.

[68] M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren, R. A. Friesner, *J. Am. Chem. Soc*. **1997**, *119*, 5908.
[69] W.-G. Han, K. J. Jalkanen, M. Elstner, S. Suhai, *J. Phys. Chem. B* **1998**, *102*, 2587.
[70] H. Hu, M. Elstner, J. Hermans, *Proteins* **2003**, *50*, 451.
[71] G. M. de Seabra, R. C. Walker, A. E. Roitberg, *J. Phys. Chem. A* **2009**, *113*, 11938.
[72] Z. Shi, K. Chen, Z. Liu, A. Ng, W. C. Bracken, N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17964.
[73] F. Avbelj, S. G. Grdadolnik, J. Grdadolnik, R. L. Baldwin, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1272.
[74] C.-D. Poon, E. T. Samulski, C. F. Weise, J. C. Weisshaar, *J. Am. Chem. Soc*. **2000**, *122*, 5642.
[75] J. Graf, P. H. Nguyen, G. Stock, H. Schwalbe, *J. Am. Chem. Soc*. **2007**, *129*, 1179.
[76] D. Verbaro, I. Gosh, W. Nau, R. Schweitzer-Stenner, *Biophys. J*. **2011**, *100*, 518a.
[77] Z. Shi, R. W. Woody, N. R. Kallenbach, In Advances in Protein Chemistry, Vol. *62*; G. D. Rose, Ed.; Academic Press, Waltham, MA, **2002**; pp. 163–240.
[78] L. Wickstrom, A. Okur, C. Simmerling, *Biophys. J*. **2014**, *97*, 853.
[79] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys*. **1983**, *79*, 926.
[80] AmberTools 14, Available at: http://ambermd.org, **2014**, Last accessed on 28 April 2015.
[81] F. Ryjáček, Ambconv, Available at: http://www.gromacs.org/Downloads/User_contributions/Other_software, **2002**. Last accessed on 27 July 2015.
[82] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562.
[83] A. Laio, F. L. Gervasio, *Rep. Prog. Phys*. **2008**, *71*, 126601.
[84] A. Barducci, M. Bonomi, M. Parrinello, *WIREs Comput. Mol. Sci*. **2011**, *1*, 826.
[85] R. B. Setlow, *Science* **1966**, *153*, 379.
[86] Z. Liu, C. Tan, X. Guo, Y.-T. Kao, J. Li, L. Wang, A. Sancar, D. Zhong, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14831.
[87] V. Thiagarajan, M. Byrdin, A. P. M. Eker, P. Müller, K. Brettel, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9402.
[88] F. Masson, T. Laino, I. Tavernelli, U. Rothlisberger, J. Hutter, *J. Am. Chem. Soc*. **2008**, *130*, 3443.
[89] A. Pérez, I. Marchán, D. Svozil, J. Šponer, T. E. Cheatham, III, C. A. Laughton, M. Orozco, *Biophys. J*. **2007**, *92*, 3817.

Received: 11 June 2015
Accepted: 9 July 2015
Published online on 4 August 2015