JCTC Journal of Chemical Theory and Computation

Article

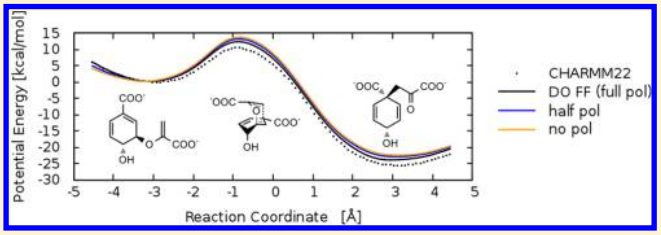
pubs.acs.org/JCTC

# Toward QM/MM Simulation of Enzymatic Reactions with the Drude Oscillator Polarizable Force Field

Eliot Boulanger and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

**ABSTRACT:** The polarization of the environment can influence the results from hybrid quantum mechanical/molecular mechanical (QM/MM) simulations of enzymatic reactions. In this article, we address several technical aspects in the development of polarizable QM/MM embedding using the Drude Oscillator (DO) force field. We propose a stable and converging update of the DO polarization state for geometry optimizations and a suitable treatment of the QM/MM-DO boundary when the QM and MM regions are separated by cutting through a covalent bond. We assess the performance of our approach by computing binding energies and geometries of three selected complexes relevant to biomolecular modeling, namely the water trimer, the N-methylacetamide dimer, and the cationic bis(benzene)sodium sandwich complex. Using a recently published MM-DO force field for proteins, we evaluate the effect of MM polarization on the QM/MM energy profiles of the enzymatic reactions catalyzed by chorismate mutase and by p-hydroxybenzoate hydroxylase. We find that inclusion of MM polarization affects the computed barriers by about 10%.




## 1. INTRODUCTION

Hybrid quantum mechanical/molecular mechanical (QM/MM) approaches have been established as a reliable tool for computing molecular properties and reaction mechanisms in the condensed phase.[1−4] A prime application example is provided by enzymatic reactions, for which it is difficult to represent the heterogeneous biological environment in an implicit manner.[5−7] QM/MM studies provide qualitative insight into such processes, as well as quantitative predictions that can be confronted with experimentation. In some cases, however, the standard QM/MM approach can fail to give the right answer, which may be due to several reasons. One of them is the neglect of polarization in the standard MM force fields, especially when the investigated reaction involves charged or very polar species.[2]

There are several ways to include polarization at the MM level.[8−10] The most prominent ones make use of induced dipoles,[11−15] fluctuating charges,[16−20] and Drude oscillators (DO).[21−27] The latter approach is adopted in this study and is also called the charge-on-spring[28−32] or shell model.[33−35] In the DO model,[21] a mobile charge, called a Drude particle (DP), is linked to a polarizable atom by a spring. A fixed charge of the same magnitude but opposite sign is added at the position of the atom, thus forming a dipole with the DP. Polarization arises from the electrostatic interactions of the DO with the rest of the system.

We included this model into the QM/MM framework some time ago using the GROMOS charge-on-spring force field.[36] The CHARMM-DO model was interfaced later in a separate development.[37] More recently, we proposed the extension of QM/MM to a fully polarizable three-layer treatment to better account for long-range electrostatics and to increase the computational efficiency.[38] Standard MM-DO polarizable force field parameters for proteins have become available only very recently,[39] and thus QM/MM-DO studies have mostly been carried out up to now for small molecules or ions solvated in water or other solvents.[36,40,41]

There are some special polarizable force fields that have been used in QM/MM studies of enzymatic reactions. For instance, in their pioneering work, Warshel and Levitt proposed a point-dipole approach to include polarization at the MM level.[42] Illingworth et al. employed an induced-charge model to include MM polarization in QM/MM computations on hydrogen-bonded model systems and found effects of approximately 10%.[43] In a later study on chorismate mutase, their induced-charge model for MM polarization gave a significant stabilization of all stationary points in the chorismate-to-prephenate conversion (15−17% of the total QM/MM interaction energies from electrostatics and polarization), but the computed barrier was not affected because of equal MM polarization contributions.[44] In the absence of well-parametrized and generally accepted standard polarizable force fields for proteins, there have not been, to our knowledge, any systematic studies on enzymatic reactions with MM polarization. In this article, we report pilot QM/CHARMM-DO applications to biologically relevant macromolecular systems, namely the enzymatic reactions catalyzed by chorismate mutase and by p-hydroxybenzoate hydroxylase.

QM/MM calculations on enzymatic reactions normally employ first-principles QM methods with QM regions encompassing typically 50 to 150 atoms. They often use geometry optimization techniques to explore the potential surface (PES),[1] even though free energy calculations can also

Received: December 19, 2013
Published: March 13, 2014

   dx.doi.org/10.1021/ct401095k | J. Chem. Theory Comput. 2014, 10, 1795−1809

be performed in an approximate manner using various sampling techniques along the reaction path.[45] Here, we focus on geometry optimizations at the QM/MM-DO level. In this case, the polarization of every DO has to be fully updated at every step.[36] In principle, this can be done by solving the corresponding system of equations, but in practice this is not efficient, and an iterative approach is usually preferred. Unfortunately, for DO-type polarizable force fields, which employ the Thole model[46] and include 1−2 and 1−3 bonded interactions in the polarization computation, we find that the iterative approach oscillates and does not converge for systems with a large number of bonded polarizable atoms. This will be the first issue covered in this manuscript.

The QM/MM combination of QM and MM subsystems has been thoroughly discussed in terms of the required embedding and boundary treatments.[1] The former define the QM/MM interactions at various levels of approximation (mechanical, electronic, and polarized embedding), while special protocols for the latter are needed, especially when a covalent bond is cut between the two subsystems. In this study, we focus on improvements for boundary and embedding treatments in the QM/MM-DO case. After briefly reviewing the methodological aspects of including DOs in a QM/MM scheme, we propose a method to update DPs for large systems such as enzymes. We then develop a special boundary treatment using butanol as a standard test case. Thereafter, we assess the QM/MM-DO interactions for three small but typical test cases: the water trimer, the N-methylacetamide dimer, and the cationic bis(benzene)sodium sandwich complex. Finally, we check the sensitivity of the QM/MM results with regard to the polarization of the enzymatic environment treated at the MM-DO level.

## 2. THEORY

**2.1. Polarizable Force Fields.** Atomic dipole polarizabilities can be introduced into force fields in several ways.[9,10,47] In most cases, the electrostatic part of the potential function is extended by including induced dipole/static multipole and induced dipole/induced dipole terms, as well as self-energy terms that account for the energy needed to create the dipoles. The static multipoles are usually monopoles, but expansions up to quadrupoles have been considered.[13,14] For force fields with localized polarizable centers, the induced dipoles ($\mu_i$) are obtained using classical electrostatics:

$$\mu_i = \alpha_i E(x_i) \tag{1}$$

where $E$ is the electric field at the position ($x_i$) of the polarizable atom $i$ and $\alpha_i$ is its polarizability, which is a parameter of the force field. The electric field is comprised of two parts, the static field due to the other permanent multipoles in the system ($E^0$) and the field due to all other induced dipoles.

$$E(x_i) = E^0(x_i) - \sum_{j \neq i} T_{ij}\mu_j \tag{2}$$

where $T_{ij}$ is the interaction tensor element between $\mu_i$ and $\mu_j$ that takes into account the interdependence of the induced dipole moments.

For additive force fields, the interactions between bonded atoms (1−2) and between next-nearest neighbor atoms (1−3) are commonly neglected in the computation of the electrostatic part of the potential. On the contrary, for polarizable force fields, it has been shown that such short-range interactions need to be included to obtain the proper polarization state.[46,48] However, their direct inclusion would lead to overpolarization due to the close distance between the polarizable centers, and it is thus necessary to use a damping function. The most popular choice is the Thole function containing additional parameters that are adjusted during the parametrization.[46,48] In this formalism, the product $T_{ij}\mu_j$ in eq 2 is multiplied by a prefactor $\gamma_{ij}$, i.e., the damping function for the interactions between nearby atoms.

The induced dipole moments of $N$ polarizable atoms in a given configuration of the system can be obtained by solving the following linear system of equations.

$$(\boldsymbol{\alpha}^{-1}\boldsymbol{\gamma}\mathbf{T})\boldsymbol{\mu} = \mathbf{E}^0 \tag{3}$$

where $\mathbf{T}$ is a $3N \times 3N$ tensor containing the elements $T_{ij}$, $\boldsymbol{\alpha}^{-1}$ is a diagonal matrix containing the inverse of the atomic polarizability tensors, and $\boldsymbol{\gamma}$ represents the interatomic Thole damping functions; $\boldsymbol{\mu}$ and $\mathbf{E}^0$ are $3 \times N$ matrices containing the Cartesian components of the induced dipoles and of the static electric field at each atom, respectively. The exact solution of eq 3 by matrix algebra is often not practical for large systems with thousands of atoms ($N$), and an iterative self-consistent (SC) approach is therefore normally preferred.[49]

When using Thole-type models, we find that the straightforward SC approach often oscillates and does not converge for large systems. A common alternative is to use the Successive Over-Relaxation (SOR) method, in which the induced dipole moment $\mu_i^B$ of any polarizable center $i$ is taken at each step as

$$\mu_i^B = m_A\mu_i^A + m_B\mu_i^{B'} \tag{4}$$

where $m_A + m_B = 1$, $\mu_i^A$ is the dipole moment obtained at the previous step of the iterative cycle, and $\mu_i^{B'}$ is the predicted dipole moment for the current step using the standard SC procedure.[49] This update procedure helps to achieve convergence of the SC method. The required number of cycles strongly depends on the choice of the $m$ coefficients. Their optimum values can be different for different force fields and different systems. Occasionally this approach still fails to converge, and then eq 3 has to be solved by matrix algebra.

Xie et al. proposed a coupled method for converging the induced dipoles of a polarizable force field.[50] In their iterative scheme, intermolecular interactions are taken into account by the standard SC procedure, while intramolecular interactions are handled by matrix inversion. This scheme is directly applicable to solvents, but there is also a variant for polymers such as proteins, in which each monomer is treated separately but with its closest neighbors included in the matrix inversion procedure. We revisit these aspects in section 4.1 when developing a method suitable for geometry optimization at the QM/MM-DO level.

**2.2. Drude Oscillators in a QM/MM Framework.** The Drude oscillators provide a polarizable force field model, in which the induced dipoles are represented by two point charges of the same magnitude but opposite sign close in space and linked by a spring.[21,51] One of them is maintained at the position of the polarizable atom while the other, the Drude particle (DP), is free to move in the external electric field. In geometry optimizations, the DPs are allowed to adjust, and the ideal polarization state is computed at every step. Since a point-charge approximation is used to represent the induced dipole,

only monopole interactions appear in the electrostatic part of the potential function, which takes the following form:

$$E_{\mathrm{DO}}^{\mathrm{elec}} = \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} + \sum_i \sum_{j'} \left( \frac{q_i q_{j'}}{r_{ij'}} - \frac{q_i q_{j'}}{r_{ij}} \right) + \sum_{i'} \sum_{j'>i'} \left( \frac{q_{i'} q_{j'}}{r_{i'j'}} - \frac{q_{i'} q_{j'}}{r_{ij}} \right) - \sum_{i'} \sum_{j'} \left( \frac{q_{i'} q_{j'}}{r_{ij'}} \right) + \frac{1}{2} \sum_{i'} k_{d,i'} d_{i'}^2 \tag{5}$$

where $q_i$ is the permanent charge at atom $i$, $q_{i'}$ is the charge associated with the Drude oscillator, $r$ is the distance between the two considered charges, $d$ is the DO charge separation, and $k_{d,i'}$ is the force constant of the DO spring, which is a parameter of the force field. The iterative update of the various positions proceeds as follows: (1) compute the electric field at every DO position; (2) based on the resulting induced dipole moment, update the DP position; (3) check convergence with respect to criteria based on energy, electric field, and/or position, and if not converged, restart at the first step.

DO force fields employ the Thole model and include the 1−2 and 1−3 interactions for induced dipole/induced dipole interactions.[27,52] The screening function is based on an exponential charge distribution and a damping function, which for monopole interactions is given by

$$\gamma_{ij} = 1 - \left( \frac{u_{ij}}{2} + 1 \right) e^{-u_{ij}} \tag{6}$$

where $\mu_{ij} = r_{ij} t_{ij}$, $r_{ij}$ is the inter-DP-distance, and $t_{ij}$ the Thole parameter:

$$t_{ij} = \frac{t_i + t_j}{\sqrt[6]{\alpha_i \alpha_j}} \tag{7}$$

with $t_i$ and $t_j$ being force field parameters of the respective polarizable centers.

The electrostatic potential function of the CHARMM-DO force field is extended not only by a point-charge representation of the atomic dipole polarizability but also by additional point charges that represent lone pairs.[27] These latter charges are rigidly linked to the heteroatoms of the system and allow a better description of the fixed charge distribution. Their positions are determined at every step of a geometry optimization from the associated set of internal coordinates, and their gradient components are distributed among the neighbor atoms such that the total force and the total torque are conserved. Another technical advantage is that the lone pair positions can be used to define local internal coordinate systems centered at heteroatom positions, which allows the use of anisotropic polarization for DOs.[53]

In a QM/MM framework, the mutual polarizations of the QM and MM regions have to be taken into account.[36] The point-charge representation of the DO model allows for a straightforward combination with most of the QM methods, since it can be treated in the same way as the point charges of the additive force fields in the electronic embedding scheme.[49] Several approaches have been proposed to find the ideal polarization state of both the QM and MM parts at every step.[37] We use the dual-SC approach, i.e., we run a full QM calculation and update the DP position as in the standard iterative procedure, but include the QM field contribution in the static electric field at each DP position. This procedure is iterated until the QM energy and DP position both converge. This scheme is expensive but leads to accurate results. It can be easily interfaced with any QM code and can be used for geometry optimization.

Another issue to consider in QM/MM simulations is the boundary treatment in cases when the QM/MM partitioning cuts through a covalent bond. This aspect has not yet been addressed for QM/MM-DO approaches. We will discuss this point in section 4.2 and propose a treatment for this situation.

## 3. COMPUTATIONAL DETAILS

All computations were run using the ChemShell package interfaced to several programs.[54−56] The MNDO program[57] was used for semiempirical QM calculations, while all other QM results were obtained with TURBOMOLE6.3.[58] The additive parts of the MM potential were computed using DL-POLY.[59] All DO-related computations were implemented separately in the hybrid module of ChemShell.[54] If not stated otherwise, the resolution-of-identity (RI) approximation[60] was applied in the MP2 calculations,[61] as is often done in QM/MM simulations at this level.[62]

## 4. RESULTS

**4.1. Converging the Drude Oscillators.** During geometry optimization, the SC procedure may fail to converge the DO positions as soon as a system with bonded polarizable atoms becomes large enough to justify the use of a force field. The SOR approach[49] is a good alternative but requires more iterations than the standard SC procedure, and in some cases it may also fail to converge to the proper polarization state. As the speed of convergence is a major practical issue in QM/MM computations, we investigate other options in the following.

We first consider the noniterative approach of solving eq 3 analytically by matrix algebra. Although computationally demanding, this can be useful in difficult cases, and it should also lead to fewer cycles in QM/MM dual-SCF procedures. Equation 3 can formally be solved by matrix inversion, which becomes very expensive for large systems. Since the considered matrix is symmetric and positive definite, it is much better to solve the corresponding system of linear equations, using Cholesky factorization followed by forward and backward substitutions. For typical system sizes in QM/MM studies of enzymes, this leads to 30−40 fold improvements in efficiency when both procedures are properly implemented. In the following, any reference to an exact or analytic solution by matrix algebra implies the use of the latter Cholesky-based approach.

The analytic procedure does not give exactly the same results as the SC procedure, in which the electric field is computed at the position of the DP for the update of DO; this position changes in each iteration. Note that even if one computes the electric field at the atomic position, as in the GROMOS charge-on-spring force field,[9] the problem remains due to the point-charge approximation so that eq 3 is not solved exactly. As an alternative, we propose an iterative approach in which at each step the electric field is evaluated at the new DP position obtained from the previous step. To assess the method, we used the 30 Å water sphere from our previous work on the solvated glycine test system (903 water molecules without the glycine solute).[38,63] The water molecules were described by the SWM4-NDP model.[64] This system has the advantage of

being large without having any 1−2/1−3 interactions, and thus the iterative approach converges without problems. We also considered computing the permanent electric field (not the one due to other dipoles) at the atomic positions. In terms of polarization energy, the results deviate by less than 0.01% for both techniques. The mean absolute deviation of the gradient components is $2.64 \times 10^{-5}$ au with a maximum value of $1.25 \times 10^{-4}$ au If the permanent electric field is also evaluated at the atomic positions, values of $5.68 \times 10^{-5}$ and $2.78 \times 10^{-4}$ au are obtained, thus roughly doubling the deviations. In both cases, these values are clearly below the commonly adopted criteria for geometry optimization ($3.0 \times 10^{-4}$ and $4.5 \times 10^{-4}$ au). This confirms that the point-charge approximation to the induced dipole is accurate enough to simulate point dipoles, which in turn suggests that we can use the iterative Cholesky factorization (ICF) approach as a reference for large systems that do not converge with SC techniques. It can also be used together with the SOR method in the case of nonconvergence. For both types of electric field computation, the ICF procedure took 6 steps for a total of 3.5 h on one 2.9 MHz Xeon processor. We also tried to evaluate the electric field at positions different from the DP, but any deviation from this position decreased the accuracy of the method.

Xie et al. proposed a hybrid approach to tackle systems with bonded polarizable centers.[50] They used the SC technique for intermolecular interactions and solved eq 3 by matrix algebra for intramolecular interactions. In the case of polymers, they suggested to determine intramolecular interactions by matrix algebra including the two neighbors of each monomer in the matrix to be inverted. For a system containing a solute surrounded by small solvent molecules, they discussed the gain in efficiency obtained, thanks to the good convergence of their method, but they did not evaluate its accuracy. Here, we take a bottom-up approach using a more general related model and compare its results with the ICF scheme.

We consider a system of polarizable DO centers which may be bonded to each other. We define blocks as any subsets of these centers. The partitioning into blocks need not be based on chemical intuition, and it is not necessary that they correspond to molecules or residues. To obtain the polarization of each block, the computation is run for a superblock containing also the neighboring blocks (based on connectivity or distance). The polarizable atoms not included in a superblock form the outer shell. The electric field ($E$) at any DP position ($x_i$) in a given block can be split in an additive fashion:

$$E(x_i) = E^{PC}(x_i) + E^{SB}(x_i) + E^{OS}(x_i) \quad (8)$$

where $E^{PC}$ is the electric field due to permanent charges (atoms and lone pairs), $E^{SB}(x_i)$ is the electric field due to all other DOs in the superblock, and $E^{OS}(x_i)$ is generated by the DOs in the outer shell. To obtain the ideal electric field at each DP position and to get the polarization state of the system, we use an iterative hybrid approach. We first compute the electric field due to every element of the system at every DP position using monopole interactions such as in the SC approach ($E_{MP}(x_i)$). From the precomputed electric field, we remove all DO electric field contributions from within the superblock using

$$E_{MP}(x_i) - E^{SB}_{MP}(x_i) = E^{PC}_{MP}(x_i) + E^{OS}_{MP}(x_i) = E^{0}_{CF}(x_i) \quad (9)$$

and thus obtain the permanent electric field for the superblock for each DP, which now includes outer-shell DO contributions. We solve the linear system of equations for the superblock with Cholesky factorization using this $E^0_{CF}(x_i)$ as permanent electric field in eq 3. In this way, we get the induced dipole moment for each block separately. If necessary, anisotropy can be included *a posteriori* with an update of the DP positions. The calculation is iterated until we achieve convergence of the DP positions. Note that for QM/MM computations, the field due to QM region should be included into $E^0_{CF}(x_i)$ and kept constant in a given iteration.

In this approach, we need to define the partitioning into blocks and the rules for generating the superblocks. Our only constraint is that all Thole-type interactions (1−2, 1−3) involving a given block should be taken into account within its superblock. To specify the proper combination law, we have chosen to define each polarizable center as a block and to generate the superblock either through connectivity or distance from this atom. As a test system, we have chosen the chorismate mutase enzyme as described below in section 4.4. We have removed the substrate and tested our approach at the full MM level using a preliminary version of the CHARMM-DO protein force field, noting that the conclusions are directly portable to the QM/MM level. To check the dependence of the results on the combination law (connectivity vs distance), we have computed the polarization state of the enzyme in the gas phase and in water. Figure 1 shows the maximum absolute deviation of the gradient components for these two setups (blue/green, with/without water) between the two combination laws (connectivity: plain lines; distance, dashed lines). The units on the $x$ axis are Ångstroms in the case of distance and refer to the connectivity order up to which atoms were included into the superblock (1−5 means that 1−2, 1−3, 1−4, and 1−5 are included). Considering that our standard convergence criterion in geometry optimizations is $4.5 \times 10^{-4}$ au for the gradient components, it is clear that the connectivity-based approach does not converge properly with increasing superblock size. The distance-based selection seems much better in this regard, as it converges in both cases for distances of more than 6 or 7 Å from the polarizable centers. This is confirmed by comparing the results from the computations in the gas phase and in water: while the distance-based computations converge to the same values, there is a significant gap in the connectivity-based results (implying that distant water molecules play an important role for the polarization state). In terms of efficiency, the ICF scheme took 12 h for the full solvated system and 2 h for the gas-phase system. Solving the linear system of equation using matrix algebra took 385 h for the full system and gave precisely the same result. Convergence of the full system with the distance-based selection of superblocks is reached after 12, 25, and 75 min for cutoffs of 7, 8, and 9 Å, respectively. For comparison, we note that the SOR approach does not converge with a damping factor of 0.1 and takes 15 (20) min with a damping factor of 0.2 (0.4). This similarity in efficiency comes from the fact that the SOR method requires a larger number of steps, each of which is less costly.

To compare with the work of Xie et al.,[50] we considered the case in which the blocks are chosen as amino acid residues or water molecules. For the definition of superblocks, we used their method of taking the bonded blocks as well as the more accurate distance-based approach suggested here, selecting any residue having at least one polarizable center within a predetermined cutoff distance from any polarizable atom of the block. For the protein in the gas phase, the bonded approach gave a mean absolute deviation of $2.6 \times 10^{-5}$ au with

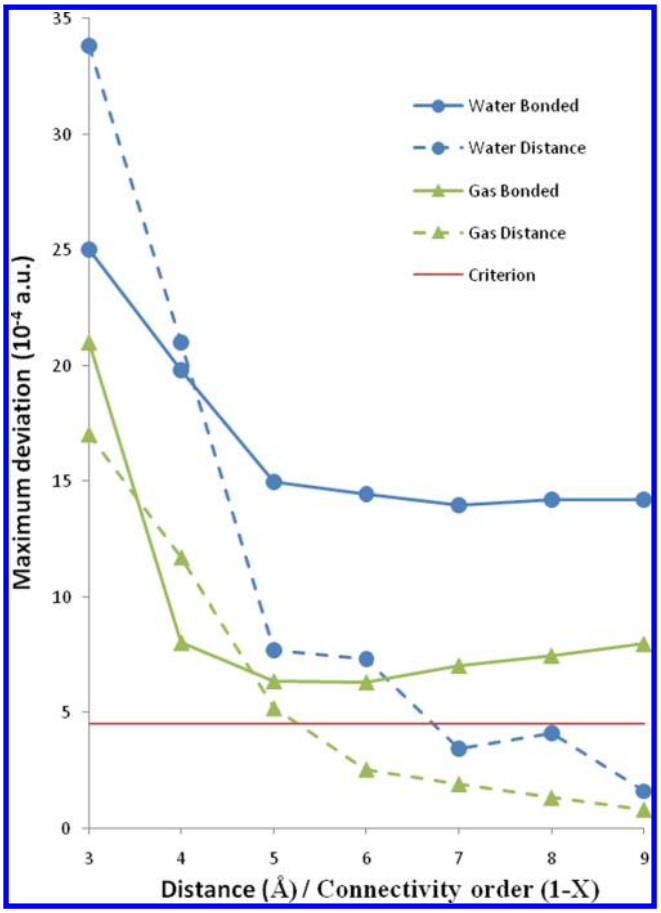



**Figure 1.** Maximum absolute deviation of the gradient components when comparing the hybrid CF-SC to the ICF approach (see text) for chorismate mutase, without substrate, using a preliminary version of the CHARMM-DO polarizable force field (provided to us by A. D. MacKerell in 2012). Gas phase (green triangles) and solvated enzyme (blue circles) were considered with different selection criteria for the hybrid model. Plain (dashed) lines refer to a connectivity-based (distance-based) superblock selection, with corresponding units of the $x$ axis (maximum connectivity order $1 - X$/distance in Å; see text). We target maximum absolute deviations below the red line, which represents the standard convergence criteria for geometry optimization.

a maximum value of $6.84 \times 10^{-4}$ au, while the distance-based approach with a cutoff of only 5 Å gave values of $1.07 \times 10^{-5}$ and $2.4 \times 10^{-4}$ au, respectively, using the same computation time (86 and 87 s). As in the case of the atom-based block definition, the bonded approach failed at producing sufficiently low gradients while the distance-based approach gave acceptable results (lower than the commonly used convergence criteria). The results were less satisfactory for the solvated enzyme, with a maximum deviation in the computed gradient components of more than $10 \times 10^{-5}$ au for the distance-based approach (cutoff: 7 Å) and more than $15 \times 10^{-5}$ au for the bonded approach.

According to our results, the atom-based block definition with a carefully chosen distance criterion should give the best results for this kind of hybrid approach. In studies of enzymes, it can be used together with the SOR technique if there are convergence problems. It could also be attractive for larger systems in general, since it can reduce the computation time significantly. Finally, it may also be useful for other polarizable force fields that do not employ the DO point-charge approximation.

**4.2. QM/MM Boundary Treatment for the Drude Oscillator Model.** In the development of the QM/MM method, special care has been taken to develop boundary treatments that allow cutting a covalent bond at the frontier between QM and MM regions.[65,66] Several approaches have been proposed, which normally work reasonably well if the frontier is chosen properly (e.g., cutting at an unpolar single bond that is as far away as possible from the electronically relevant part of the QM region).[1]

At a given atom, the DO model may involve the following electrostatic entities: the atomic point charge, the lone pairs, and the DP as well as its counter charge.[51] To define a boundary treatment, we have chosen not to take the lone pairs into account, assuming that no bond to a heteroatom will be cut. For the atomic point charges, we apply the commonly adopted charge shift scheme without any further modification.[67] In this scheme, the charge on the MM atom in the frontier bond (M1) is distributed to the other MM atom(s) that is (are) bonded to it (M2); a point dipole is added at these atoms (M2) to compensate for the charge shift, and the valence of the frontier QM atom is satisfied by adding a QM hydrogen atom (link atom). As this charge shift scheme performs well with additive force fields, we did not see any need to modify it. Therefore, we only have to develop a model for treating the DOs at the boundary.

We have investigated five different models. The first one (model 0) neglects the DO on the M1 atom, without any other modification. In model 1, the polarizability of M1 is transferred to the polarizable centers in the M2 position (without using this polarizability when computing the Thole screening function). As these M2 atoms become extremely polarizable in model 1, it may be more appropriate to compute the Thole function with the full M2 polarizability (including the contributed shifted from M1), which leads to model 2. The full transfer including the $t_i$ parameters gives rise to model 3. Finally, as the M2 atoms are rather close to the virtually bonded QM region, we considered another model, model 4, in which the polarizable M2 atoms interact with the QM region according to the Thole model (without any transfer of parameters).

In the development of boundary treatments, the proton affinity (PA) and deprotonation enthalpy (DE) of n-propanol or n-butanol are commonly used as test systems.[65,68] In these molecules, different C−C bonds can be cut, 2 in the case of n-propanol and 3 for n-butanol. Cutting the C−C bond closest to oxygen (cut1) is considered as an extreme case, while the other options (cut2 and cut3) are more representative of typical QM/MM applications. In previous work, boundary treatments that were successful for these test systems have also performed well in other QM/MM applications.[65,66] Therefore, we only consider n-butanol in the following. In analogy to a previous study on propanol,[65] we used the semiempirical AM1 method[69] to evaluate DE and PA for frozen geometries of butanol, butanolate anion, and butanolium cation. Again following the literature,[65] we also included a sodium cation ($Na^+$) to simulate an "extreme" environment. In our tests, this cation was put either in the QM or the MM region. Instead of choosing a set of a few predetermined positions for the cation, we generated for each test 100 positions that were randomly picked at distances of at least 3 Å away from any atom of the molecule and within 9 Å of the geometric center of the molecule (the

statistical results remain essentially unchanged when running tests with 150 positions). Note that when the sodium ion is included in the QM region, the computations crashed occasionally when the MM part of n-butanol was situated between the QM part of n-butanol and the QM cation; in these cases (10−15%), extra computations were run to obtain 100 sampling points.

Table 1 compiles the deviations from the full QM results of DE (upper part) and PA (lower part) as well as their standard deviation (in parentheses) for the different boundary treatments (models 0−4), cation treatments (QM vs MM), and positions of the QM/MM boundary (cut1−cut3). The results are consistent with previous experience from additive force fields. The deviations are largest for cut1, and they decrease as the distance of the QM/MM boundary from the OH group increases (smallest for cut3). Regardless of other options, the best boundary treatment is always provided by model 0. Transferring polarizability to the M2 atom in model 1 systematically increases the deviation from the full QM reference results. Applying Thole damping at M2 slightly improves the results in model 2 (but not much), while transferring the complete Thole parameter set in model 3 makes things even worse. Applying Thole damping to the M2/QM interaction in model 4 is also detrimental, especially when the cation is part of the QM region.

**Table 1. Average Deviation from the QM Results and Associated Standard Deviation (in Parentheses) for Deprotonation Enthalpy (Upper Part) and Proton Affinity (Lower Part) of *n*-Butanol in the Presence of a Sodium Cation at Different Positions (See Text) Computed with Different QM/MM Boundary Treatments (Models 0−4, See Text)[a]**

| $Na^+$ | model | cut1 | | cut2 | | cut 3 | |
|---|---|---|---|---|---|---|---|
| QM | 0 | 6.91 | (0.53) | 3.72 | (0.20) | 2.20 | (0.12) |
| | 1 | 7.24 | (0.57) | 4.00 | (0.28) | 2.21 | (0.16) |
| | 2 | 7.21 | (0.45) | 4.02 | (0.24) | 2.20 | (0.10) |
| | 3 | 7.48 | (1.25) | 4.01 | (0.28) | 2.22 | (0.21) |
| | 4 | 8.45 | (1.29) | 6.20 | (1.04) | 1.38 | (2.17) |
| MM | 0 | 9.74 | (1.57) | 4.59 | (0.37) | 2.75 | (0.31) |
| | 1 | 10.47 | (1.30) | 4.49 | (0.37) | 2.78 | (0.33) |
| | 2 | 10.33 | (1.37) | 4.51 | (0.35) | 2.77 | (0.32) |
| | 3 | 11.25 | (1.21) | 4.50 | (0.35) | 2.79 | (0.34) |
| | 4 | 9.79 | (1.51) | 4.60 | (0.37) | 2.76 | (0.31) |
| QM | 0 | 2.98 | (0.48) | 3.10 | (0.20) | 1.49 | (0.09) |
| | 1 | 8.79 | (0.46) | 4.99 | (0.19) | 1.51 | (0.11) |
| | 2 | 7.62 | (0.28) | 4.98 | (0.20) | 1.50 | (0.09) |
| | 3 | 14.33 | (4.75) | 4.99 | (0.19) | 1.51 | (0.11) |
| | 4 | 16.39 | (0.78) | 11.00 | (0.64) | 1.55 | (0.15) |
| MM | 0 | 3.04 | (0.46) | 3.20 | (0.26) | 1.65 | (0.26) |
| | 1 | 7.79 | (1.36) | 4.73 | (0.50) | 1.67 | (0.27) |
| | 2 | 6.85 | (0.98) | 4.76 | (0.47) | 1.66 | (0.27) |
| | 3 | 12.92 | (4.70) | 4.73 | (0.48) | 1.68 | (0.28) |
| | 4 | 3.08 | (0.47) | 3.20 | (0.26) | 1.65 | (0.26) |

[a]Values are given in kcal/mol, and the full QM computation is taken as a reference. AM1 was used for the QM part and the Drude Oscillator force field for the MM region. The cuts are defined by the number of bonds from the hydroxyl group (cut1 being the nearest, and cut3 the farthest with only one methyl group in the MM region). The sodium cation ($Na^+$) can be part of the MM or QM region.

The cut1 is obviously not advisable, consistent with the rules of thumb known for additive force fields. As expected, the cut3 gives the most realistic results: the deviations from the QM reference results are smallest for both PA and DE, and the choice of boundary treatment has only little influence (disregarding model 4 with QM $Na^+$). Overall, the present results suggest that the inclusion of Drude oscillators in the force field has little effect on the boundary treatment, since the quality of the results is similar to what is found for additive force fields with standard boundary treatments. The simplest being the best, we recommend the use of model 0, i.e., simply to remove any DO at the M1 position in QM/MM-DO calculations.

**4.3. Biologically Relevant Test Systems.** To evaluate QM/MM-DO compatibility, we have chosen a set of simple, biologically relevant systems. The purpose of these tests is not to obtain any insight into the properties of the studied molecules but to systematically investigate the ability of CHARMM-DO to give more accurate results than the CHARMM22 additive force field in a QM/MM framework.[70] Usually, MM molecules are not involved or located in the direct vicinity of the chemical reaction being studied. We therefore assume that reproducing the binding energies and geometries of complexes will give valid insight into QM/MM-DO compatibility.

*4.3.1. Water Trimer.* Apart from its biological relevance, water is one of the cornerstones of force field parametrization.[22,28,64] Indeed, it is generally the first molecule to be parametrized and is then included in the parametrization of all others. Checking water is thus the logical first step when it comes to evaluating QM/MM-DO compatibility. The SWM4-NDP[64] water model (called SWM4 in this study and used in the CHARMM-DO force field) has been the subject of a few QM/MM studies. The water dimer was investigated in the tests by Lu and Zhang using BLYP/6-31G(d,p) as a QM component.[37] QM/MM studies on the solvation of QM ions have employed SWM4 as a polarizable solvent.[40,41]

Our test system is the cyclic water trimer.[71] This simple complex has the advantage that each of the three water molecules is equivalent, donating one hydrogen bond and accepting another one (see Figure 2). It has been investigated in several theoretical studies.[72−74] It has been included in the preliminary parametrization of SWM4 using a positive DP,[22] but not in the final one with a negative DP.[64] Yu et al. have shown that the SWM4 model properly reproduces both the binding energy (−14.35 vs −14.92 kcal/mol) and the geometry of the complex (RMSD 0.07 Å) using MP2/CBS as a benchmark.[75]

Two reasonable QM/MM separations are possible, with either one QM and two MM water molecules or *vice versa*. We have investigated both cases using the SWM4[64] and TIP3P[76] water models to compare with the standard additive force field (TIP3P). To differentiate between the possible combinations, we adopt a three-letter notation to define which model has been used for each individual water molecule; Q for QM, S for SWM4, and T for TIP3P (e.g., QSS means one QM and two SWM4 water molecules). For the QM part, we used standard DFT methods that are commonly employed for QM/MM simulations of enzymes (BP86,[77,78] BLYP,[79] B3LYP,[80] PBE,[81] and PBE0[82]). As the water trimer is a noncovalent complex, we included the Grimme D2 dispersion correction for BP86, BLYP, B3LYP, and PBE.[83] We also considered *ab initio* methods, Hartree−Fock (HF) and MP2;[60,61] the latter is

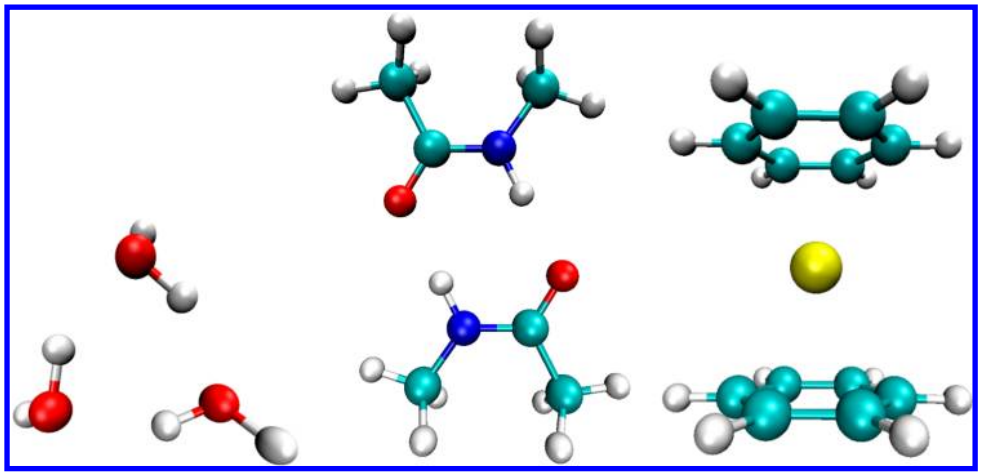


**Figure 2.** Complexes used in this study to evaluate QM/MM-DO compatibility. From left to right, the water trimer in its most stable cyclic form, the *cis*-NMA dimer, and the cationic bis(benzene)sodium sandwich complex.

known to give very good results for this complex. The SVP,[84] TZVP, and TZVPP[85] basis sets were used in all of these calculations. Further tests were performed using the semiempirical QM methods MNDO,[86,87] AM1,[69] PM3,[88] OM1,[89] OM2,[90] and OM3.[91,92] Since both the TIP3P and SWM4 models have been parametrized with rigid geometries, we constrained the internal geometry of the water molecules during the optimization but also considered the case when they are flexible (for consistency with the QM approach). We discuss the results of the geometry optimizations in terms of the binding energy of the complex and its structure. The binding energy was calculated from the energies of the optimized complex and the optimized water molecule. The geometry of the complex was assessed by two criteria, namely the O−O distances and the angles between the O−O−O plane and the hydrogen atoms not involved in hydrogen bonding. The latter angles are reported in absolute value, without distinguishing between the up or down orientation of the corresponding O−H bonds (note that the up−up−up and down−down−down orientations were never encountered). We do not focus here on the ability of the methods to reproduce experimental or high-level theoretical data but rather on the compatibility of the QM and MM potentials in a QM/MM framework to reproduce the QQQ results obtained with the same QM method.

Binding energies are listed in Table 2. Two key tendencies are observed. The first one is that the QM/MM prediction of the binding energy is improved upon basis set extension. This improvement is systematic for SWM4 when used in combination with DFT, DFT-D2, or MP2. For TIP3P, the binding energy is underestimated with MP2 and DFT, except when applying the D2 correction term. The second general trend concerns the changes in the QM/MM binding energies when replacing TIP3P by SWM4 in the MM part. For the first-principles QM methods, switching from QQT to QQS increases the binding energy by typically 2.05−2.46 kcal/mol (and even by 3.27 kcal/mol for B3LYP-D2/TZVPP). When going from QTT to QSS, this increase is even larger (3.31−3.44 kcal/mol). These changes in the binding energy are not due to the inclusion of polarization in SWM4 (always stabilizing in the complex and zero for an isolated water molecule) but rather to the parametrization of the MM model and the SWM4 treatment of the oxygen lone pair. For the semiempirical QM methods, the changes are in the same direction but less pronounced (see Table 2).

We now briefly address the performance of specific QM/MM combinations with regard to the reproduction of the pure QM reference energies. For DFT QM components, the functionals with Becke exchange tend to perform better than the parameter-free PBE approaches when combined with SWM4, while there is no such clear trend for TIP3P. When applying the D2 dispersion correction, the results for SWM4 deteriorate, and overall the DFT-D2/TIP3P combinations seem to perform better than DFT-D2/SWM4 (without clear distinction between different types of functionals). Concerning *ab initio* methods, HF/MM calculations give rather large deviations from the HF reference energies (regardless of the chosen MM model). The MP2/MM results are satisfactory for QQS but not for QSS, QQT, and QTT. There are no obvious specific patterns when using semiempirical QM methods; here, PM3 and OM3 seem to perform best.

Upon removing the geometry constraints on the MM water molecules during the optimizations (see the results given in parentheses in Table 2), the binding energies are generally increased slightly (as expected). The changes are typically on the order of 0.5 (0.3) kcal/mol for QQS (QQT), between 0 and 1.0 kcal/mol for QSS, and around 0.6 kcal/mol for QTT. These changes do not generally lead to a better reproduction of the QQQ reference results, and for the sake of consistency, it seems preferable to retain the constraints on the MM water geometries in QM/MM geometry optimizations (i.e., to use frozen MM water geometries). We note that this convention is usually not adopted during standard QM/MM minimizations with additive force fields.

With regard to the reproduction of the QQQ reference binding energies, the QM/SWM4 combination is clearly superior to QM/TIP3P for the QQX system when using the BP86, BLYP, and B3LYP functionals with the TZVPP basis set or *ab initio* methods with TZVP or TZVPP. This also holds for the QXX test systems in the case of DFT but not for *ab initio* methods. Overall, the best performance among all tested QM/MM variants is found when combining the SWM4 water model with the following QM components: DFT/TZVPP with DFT = BP86, BLYP, or B3LYP; MP2/TZVP; and MP2/TZVPP. The semiempirical PM3 and OM3 methods also give acceptable results. Generally, QM/SWM4 performs slightly better than QM/TIP3P (compared with the QQQ reference

Table 2. Binding Energies (kcal/mol) of the Water Trimer with Different QM Methods (See Text)[a]

| Hamiltonian | basis | QQQ | QQS | | QQT | | QSS | | QTT | |
|---|---|---|---|---|---|---|---|---|---|---|
| BP86 | SVP | −28.45 | −19.02 | (−19.57) | −21.47 | (−21.79) | −14.33 | (−15.21) | −17.7 | (−18.22) |
| | TZVP | −18.64 | −17.3 | (−17.84) | −19.49 | (−19.82) | −15.78 | (−16.75) | −19.2 | (−19.77) |
| | TZVPP | −16.63 | −16.16 | (−16.7) | −18.28 | (−18.59) | −15.24 | (−16.36) | −18.61 | (−19.16) |
| BLYP | SVP | −29.24 | −19.45 | (−19.98) | −21.83 | (−22.15) | −14.09 | (−14.21) | −17.44 | (−17.95) |
| | TZVP | −18.04 | −17.24 | (−17.78) | −19.39 | (−19.71) | −15.67 | (−16.62) | −19.1 | (−19.65) |
| | TZVPP | −16.14 | −16.14 | (−16.66) | −18.23 | (−18.54) | −15.16 | (−16.11) | −18.53 | (−19.07) |
| B3LYP | SVP | −27.41 | −19.04 | (−19.57) | −21.47 | (−21.79) | −14.4 | (−15.26) | −17.74 | (−18.27) |
| | TZVP | −18.5 | −17.36 | (−17.87) | −19.58 | (−19.91) | −15.75 | (−15.93) | −19.18 | (−19.73) |
| | TZVPP | −16.43 | −16.19 | (−16.7) | −18.35 | (−18.65) | −15.21 | (−15.29) | −18.6 | (−19.14) |
| PBE | SVP | −31.37 | −20.23 | (−20.77) | −22.63 | (−22.95) | −14.28 | (−15.13) | −17.63 | (−18.15) |
| | TZVP | −20.92 | −18.21 | (−18.75) | −20.4 | (−20.73) | −15.72 | (−16.68) | −19.14 | (−19.7) |
| | TZVPP | −18.99 | −17.11 | (−17.65) | −19.23 | (−19.54) | −15.19 | (−15.15) | −18.56 | (−19.1) |
| PBE0 | SVP | −27.52 | −19.15 | (−19.71) | −21.61 | (−21.94) | −14.57 | (−15.39) | −17.94 | (−18.48) |
| | TZVP | −19.72 | −17.73 | (−18.27) | −19.99 | (−20.32) | −15.77 | (−15.88) | −19.2 | (−19.76) |
| | TZVPP | −17.63 | −16.56 | (−17.05) | −18.75 | (−19.06) | −15.23 | (−16.17) | −18.62 | (−19.16) |
| BPE86-D2 | SVP | −31.97 | −20.19 | (−20.76) | −22.64 | (−22.97) | −14.33 | (−15.22) | −17.67 | (−18.22) |
| | TZVP | −21.84 | −18.37 | (−18.94) | −20.57 | (−20.91) | −15.77 | (−16.75) | −19.18 | (−19.77) |
| | TZVPP | −19.89 | −17.25 | (−17.8) | −19.38 | (−19.69) | −15.24 | (−16.2) | −18.59 | (−19.16) |
| BLYP-D2 | SVP | −33.31 | −20.82 | (−21.35) | −23.18 | (−23.49) | −14.1 | (−14.95) | −17.41 | (−17.95) |
| | TZVP | −21.57 | −18.44 | (−19) | −20.6 | (−20.92) | −15.67 | (−15.8) | −19.07 | (−19.65) |
| | TZVPP | −19.71 | −17.35 | (−17.89) | −19.45 | (−19.76) | −15.16 | (−15.27) | −18.5 | (−19.07) |
| B3LYP-D2 | SVP | −30.87 | −20.18 | (−20.72) | −22.61 | (−22.94) | −14.4 | (−15.22) | −17.72 | (−18.27) |
| | TZVP | −21.55 | −18.4 | (−18.95) | −20.63 | (−20.95) | −15.76 | (−16.71) | −19.15 | (−19.73) |
| | TZVPP | −19.52 | −16.14 | (−16.65) | −19.41 | (−19.72) | −15.21 | (−15.31) | −18.57 | (−19.14) |
| PBE-D2 | SVP | −33.89 | −21.07 | (−21.62) | −23.47 | (−23.8) | −14.28 | (−15.15) | −17.61 | (−18.16) |
| | TZVP | −23.19 | −18.98 | (−19.55) | −21.18 | (−21.5) | −15.73 | (−15.91) | −19.12 | (−19.7) |
| | TZVPP | −21.31 | −17.88 | (−18.41) | −20.02 | (−20.32) | −15.2 | (−15.15) | −18.54 | (−19.11) |
| HF | SVP | −17.83 | −16.65 | (−17.19) | −18.9 | (−19.24) | −15.02 | (−15.9) | −18.39 | (−18.95) |
| | TZVP | −14.04 | −16.02 | (−16.55) | −18.15 | (−18.48) | −15.84 | (−16) | −19.25 | (−19.82) |
| | TZVPP | −12.07 | −14.89 | (−15.38) | −16.98 | (−17.29) | −15.34 | (−16.27) | −18.73 | (−19.28) |
| MP2 | SVP | −24.29 | −18.38 | (−18.92) | −20.8 | (−21.13) | −14.76 | (−15.55) | −18.15 | (−18.7) |
| | TZVP | −18.2 | −17.25 | (−17.78) | −19.43 | (−19.76) | −15.83 | (−16.7) | −19.27 | (−19.84) |
| | TZVPP | −17.41 | −16.54 | (−17.05) | −18.68 | (−18.99) | −15.33 | (−15.45) | −18.73 | (−19.28) |
| MNDO | / | −1.42 | −7.75 | (−7.85) | −9.24 | (−9.35) | −11.18 | (−11.71) | −13.47 | (−13.82) |
| AM1 | / | −15.37 | −12.07 | (−12.22) | −13.55 | (−13.68) | −11.44 | (−11.9) | −13.91 | (−14.28) |
| PM3 | / | −10.07 | −9.22 | (−9.3) | −10.83 | (−10.89) | −10.68 | (−11.09) | −13.13 | (−13.45) |
| OM1 | / | −10.64 | −12.63 | (−12.96) | −14.47 | (−14.71) | −13.37 | (−13.91) | −16.18 | (−16.65) |
| OM2 | / | −14.19 | −12.7 | (−12.99) | −14.58 | (−14.81) | −12.95 | (−13.62) | −15.77 | (−16.22) |
| OM3 | / | −14.66 | −13.38 | (−13.73) | −15.59 | (−15.88) | −13.78 | (−14.46) | −16.75 | (−17.25) |

[a]They were computed at the full QM level (QQQ) or using a hybrid QM/MM approach with the TIP3P water force field (T) or the SWM4-NDP polarizable force field (S). Two kinds of QM regions were considered, one with two water molecules (QQT, QQS) and the other with one water molecule (QTT, QSS). Binding energies were obtained from geometry optimizations, in which the internal structure of each MM water molecule was constrained (values in parentheses from calculations without such constraints).

data). Finally, we note that, thanks to the careful parametrization, the QM/SWM4 binding energies are closer to the experimental values than the full QM or the QM/TIP3P results, regardless of the chosen QM method.

The optimized O−O distances are generally between 2.65 and 2.85 Å, except for the *ab initio* HF and the semiempirical methods. A significant difference is observed when it comes to symmetry. For pure QM calculations (QQQ), all three O−O distances are identical (within the precision of the optimization). This also holds for pure MM calculations, which yield equivalent O−O distances (TIP3P 2.75 Å, SWM4 around 2.80 Å). However, this is no longer true at the QM/MM level, where we generally find some spread between the different O−O distances. In the case of one MM water molecule (QQS and QQT), both MM models show a similar performance. When there are two MM water molecules (QTT and QSS), the spread is much smaller, and the computed O−O distances are close to their values from the full MM calculations. Interestingly, the values obtained do not depend much on the chosen QM method, and the QM/SWM4 results for QSS are thus generally close to experimental results (2.8 Å).[71] Contrary to the binding energies, basis set extension does not affect the O−O distances much.

The difference between the TIP3P and SWM4 water models becomes more pronounced when considering the angles between the O−O−O plane and the hydrogen atoms not involved in hydrogen bonding. In full MM optimizations, SWM4 gives realistic geometries, while TIP3P produces a planar trimer. As TIP4P is also known to give proper geometries,[71] this difference is probably due to the explicit treatment of the lone pairs in SWM4 (rather than polarization effects). Similar trends are observed at the QM/MM level. The

Table 3. Binding Energies (kcal/mol) and Hydrogen Bond Distances (Å) between the Two Monomers in the *cis*-NMA Dimer[a]

| | | full QM | | | QM/MM-DO | | | QM/MM | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hamiltonian | basis | energy | dist1 | dist2 | energy | dist1 | dist2 | energy | dist1 | dist2 |
| BP86 | SVP | −18.19 | 1.76 | 1.76 | −17.64 | 1.72 | 1.30 | −15.35 | 1.70 | 1.61 |
| | TZVP | −14.41 | 1.82 | 1.82 | −19.64 | 1.62 | 1.19 | −16.58 | 1.72 | 1.60 |
| | TZVPP | −14.11 | 1.80 | 1.80 | −19.37 | 1.63 | 1.27 | −16.40 | 1.73 | 1.60 |
| BLYP | SVP | −17.75 | 1.81 | 1.81 | −17.22 | 1.76 | 1.53 | −15.06 | 1.71 | 1.61 |
| | TZVP | −13.40 | 1.87 | 1.87 | −19.38 | 1.63 | 1.20 | −16.39 | 1.72 | 1.60 |
| | TZVPP | −13.00 | 1.86 | 1.86 | −19.30 | 1.64 | 1.25 | −16.24 | 1.73 | 1.60 |
| B3LYP | SVP | −17.69 | 1.84 | 1.84 | −17.60 | 1.75 | 1.56 | −15.43 | 1.70 | 1.62 |
| | TZVP | −14.06 | 1.86 | 1.86 | −19.67 | 1.60 | 1.19 | −16.62 | 1.72 | 1.61 |
| | TZVPP | −13.60 | 1.86 | 1.86 | −19.30 | 1.67 | 1.40 | −16.42 | 1.72 | 1.61 |
| PBE | SVP | −19.84 | 1.76 | 1.76 | −17.69 | 1.76 | 1.39 | −15.30 | 1.71 | 1.61 |
| | TZVP | −15.75 | 1.83 | 1.83 | −19.54 | 1.62 | 1.19 | −16.51 | 1.73 | 1.60 |
| | TZVPP | −15.46 | 1.80 | 1.80 | −19.29 | 1.65 | 1.32 | −16.32 | 1.73 | 1.60 |
| PBE0 | SVP | −18.66 | 1.79 | 1.79 | −17.83 | 1.72 | 1.39 | −11.83 | 1.85 | 1.86 |
| | TZVP | −15.36 | 1.83 | 1.83 | −19.62 | 1.68 | 1.42 | −16.68 | 1.71 | 1.61 |
| | TZVPP | −14.96 | 1.82 | 1.82 | −19.43 | 1.66 | 1.35 | −16.50 | 1.72 | 1.61 |
| BP86-D2 | SVP | −22.02 | 1.72 | 1.72 | −17.94 | 1.72 | 1.41 | −15.41 | 1.71 | 1.61 |
| | TZVP | −18.09 | 1.77 | 1.77 | −19.78 | 1.63 | 1.18 | −16.61 | 1.72 | 1.60 |
| | TZVPP | −17.88 | 1.76 | 1.76 | −19.66 | 1.61 | 1.20 | −16.45 | 1.73 | 1.59 |
| BLYP-D2 | SVP | −21.83 | 1.77 | 1.77 | −17.95 | 1.73 | 1.36 | −15.09 | 1.71 | 1.62 |
| | TZVP | −17.36 | 1.82 | 1.82 | −19.61 | 1.68 | 1.40 | −16.48 | 1.73 | 1.60 |
| | TZVPP | −17.02 | 1.82 | 1.82 | −20.50 | 1.69 | 1.38 | −16.28 | 1.73 | 1.60 |
| B3LYP-D2 | SVP | −21.32 | 1.79 | 1.79 | −17.71 | 1.75 | 1.54 | −15.47 | 1.70 | 1.62 |
| | TZVP | −17.53 | 1.82 | 1.82 | −20.27 | 1.68 | 1.37 | −16.66 | 1.72 | 1.61 |
| | TZVPP | −17.22 | 1.82 | 1.82 | −19.48 | 1.61 | 1.19 | −16.47 | 1.72 | 1.61 |
| PBE-D2 | SVP | −22.58 | 1.73 | 1.73 | −18.08 | 1.71 | 1.38 | −15.32 | 1.71 | 1.61 |
| | TZVP | −18.39 | 1.78 | 1.78 | −20.36 | 1.71 | 1.38 | −16.55 | 1.73 | 1.60 |
| | TZVPP | −18.20 | 1.77 | 1.77 | −19.41 | 1.62 | 1.22 | −16.36 | 1.73 | 1.59 |
| HF | SVP | −13.24 | 1.99 | 1.99 | −18.45 | 1.73 | 1.60 | −16.32 | 1.68 | 1.64 |
| | TZVP | −11.30 | 2.01 | 2.01 | −19.93 | 1.65 | 1.30 | −16.94 | 1.69 | 1.64 |
| | TZVPP | −7.08 | 1.86 | 1.86 | −19.63 | 1.67 | 1.41 | −16.72 | 1.70 | 1.63 |
| MP2 | SVP | −18.17 | 1.84 | 1.84 | −17.06 | 1.86 | 1.86 | −15.46 | 1.71 | 1.62 |
| | TZVP | −15.46 | 1.85 | 1.85 | −17.08 | 1.64 | 1.37 | −16.30 | 1.72 | 1.61 |
| | TZVPP | −16.38 | 1.82 | 1.82 | −19.28 | 1.62 | 1.18 | −16.26 | 1.72 | 1.61 |
| MNDO | | −1.17 | 3.42 | 3.42 | −10.73 | 1.84 | 1.90 | −10.43 | 1.78 | 1.82 |
| AM1 | | −7.99 | 2.07 | 2.07 | −11.82 | 1.85 | 1.87 | −11.40 | 1.77 | 1.76 |
| PM3 | | −6.58 | 1.80 | 1.80 | −10.35 | 1.87 | 1.87 | −10.79 | 1.84 | 1.78 |
| OM1 | | −6.87 | 2.08 | 2.08 | −13.65 | 1.79 | 1.80 | −13.15 | 1.71 | 1.71 |
| OM2 | | −13.75 | 1.64 | 1.64 | −14.31 | 1.79 | 1.79 | −13.73 | 1.70 | 1.70 |
| OM3 | | −12.93 | 1.50 | 1.50 | −14.61 | 1.78 | 1.74 | −13.95 | 1.69 | 1.66 |

[a]Different QM methods are used to describe the QM monomer. Results are given for full QM and for QM/MM computations using either the polarizable Drude Oscillator force field or the CHARMM additive force field to represent the MM monomer. Distances are taken between the hydrogen atom and the oxygen acceptor atom.

QM/TIP3P optimizations always give planar trimer structures, while the QM/SWM4 calculations (in combination with any first-principle QM method) yield realistic out-of-plane angles that are typically 1−3° too large (compared with the QQQ reference data).

In the hybrid QM/MM computations, the SWM4 water model performs better than TIP3 overall, because it can properly reproduce the geometries, thanks to the explicit lone-pair treatment. Being partly parametrized with respect to high-level *ab initio* data, it tends to give accurate binding energies in a QM/MM framework (compared with the full QM data). The best QM/SWM4 results are obtained when employing the QM method and basis set used for its parametrization, namely MP2 with a large basis set. The SWM4 model is also compatible with DFT methods, especially when a large basis set is used. For fast QM/SMW4 computations, the semiempirical PM3 and OMX methods appear to be efficient alternatives for this system.

*4.3.2. NMA Dimer.* N-methylacetamide (NMA) often serves as a prototype test system when parametrizing a new force field for proteins, as it provides the smallest possible representation of the peptide bond.[93] It has also been used in benchmark studies that target biologically relevant data.[94,95] Here, we investigate the NMA dimer, in which both monomers are in their *cis* conformation. As shown in Figure 2, two equivalent hydrogen bonds are formed between the two monomers. The NMA dimer is particularly relevant for our purposes as this type of hydrogen bond is often encountered in QM/MM studies of enzymes.

In QM/MM work, one normally avoids QM/MM boundaries that cut through a hydrogen bond directly involving the substrate or other reactive species.[1] Here, we deliberately

perform a demanding test on the NMA dimer where this convention is violated, treating one NMA as a QM molecule and the other one as an MM molecule (thus cutting through both hydrogen bonds). Since the dimer is symmetric, the two possible assignments are equivalent. We again compare QM/MM results obtained with a polarizable DO force field and the additive CHARMM force field (QM/MM-DO vs QM/CHARMM). Parameters for the additive force field were taken from the distributed CGenFF set,[96] and DO parameters were a refined version of those developed by Harder et al.[97] We focus on the binding energy and the hydrogen bond lengths. The results are listed in Table 3. Full MM computations lead to a binding energy of −14.88 kcal/mol and a hydrogen bond distance of 1.75 Å for the DO model, compared with −11.90 kcal/mol and 1.74 Å for the CHARMM22 force field.

Concerning the binding energies, the full QM reference results are reproduced very well at the QM/MM-DO level when using DFT/SVP as the QM method (contrary to what has been found for the water trimer). Upon basis set extension from SVP to TZVP, the dimer is destabilized for any QM method (both in pure QM and hybrid QM/MM calculations, regardless of the chosen force field). Including dispersion corrections generally improves the QM/MM results, but not the QM/MM-DO results. The polarizable force field performs better than the additive one whenever SVP is used as the basis. Among the semiempirical QM methods, both force fields give good results in combination with OM2, which is known to perform well for these kinds of systems.

Table 3 also lists the hydrogen bonding distances in the NMA dimer, which are identical by symmetry in the pure QM and MM calculations. When using a hybrid QM/MM model, the symmetry is broken, and the difference (splitting) between the two computed hydrogen bond distances is an excellent criterion to assess the compatibility of the QM and MM descriptions. Compared to the full QM reference results, the QM/CHARMM calculations produce acceptable geometries. The splitting is 0.09 Å on average (maximum: 0.14 Å for BP86-D2/TZVPP and PBE-D2/TZVPP), and the deviation from the QM reference distances amounts to 0.16 Å on average (maximum: 0.37 Å for OM1, disregarding the pure QM results from MNDO which fails to give hydrogen bonds). The QM/MM-DO calculations generally perform less well: the splitting is 0.28 Å on average (maximum: 0.45 Å for BP86/TZVP), and the average deviation from the QM reference distances is 0.29 Å (maximum: 0.53 Å for HF/TZVP). The best QM/MM-DO distances are generally obtained with the SVP basis, which may at least partially explain the good results for the binding energy obtained with this basis set. Comparing the performance of different QM methods in the QM/MM-DO calculations, all DFT functionals fail to reproduce the pure QM(DFT) geometries (best match: splitting of 0.19 Å and deviation of 0.185 Å for B3LYP/SVP), and the inclusion of empirical dispersion corrections does not improve the results at all. Among the *ab initio* QM methods, MP2/SVP gives the best QM/MM-DO results with zero splitting and a deviation of only 0.02 Å, while the others do not perform well. On the other hand, semiempirical QM methods lead to a very small splitting at both the QM/MM-DO and QM/CHARMM levels, but they are less accurate at reproducing the corresponding QM reference distances. In the QM/MM-DO framework, PM3 and OM2 give no splitting and are closer to the reference distances than the corresponding full QM results.

To summarize, QM/MM-DO calculations with MP2/SVP as a QM component best reproduce both the energetic and geometric QM reference results, and they also yield excellent agreement with the experimental values. It should be noted in this context that MP2 with a basis of SVP-type quality has been used as the QM method for parametrizing the geometry of such compounds in the MM-DO force fields.[98] For other first-principles QM methods, QM/MM-DO tends to perform less well than QM/CHARMM, especially when it comes to the geometry of the NMA dimer. Among the semiempirical QM methods, OM2 seems to be the best choice.

*4.3.3. π-Cation Interactions.* The preceding comparisons indicate that the QM/MM-DO results tend toward those obtained by the DO model alone when the size of the MM region is expanded. We verify those aspects with another type of interaction relevant for enzyme catalysis: the π-cation interactions. We have chosen the cationic bis(benzene)sodium sandwich complex as a representative test system (Figure 2). This complex has no heteroatoms with lone pairs so that we can directly assess the effect of MM polarization.

At the MM-DO level, several benzene−sodium complexes were investigated by Orabi and Lamoureux.[99] Their results show a clear improvement over the CHAMM22/27 results in comparisons with experimental and high-level *ab initio* data.[100] This was achieved by a specific parametrization for the DO model, which is not used here as we would like to assess the difference between the reference data and the DO results. We use the initially proposed DO parameters, designed for condensed phase simulations.[101]

As in the case of the water trimer, we investigated several definitions of the QM and MM regions. We again use a three-letter notation: the first and third letters refer to the benzene moieties and the middle one to the sodium cation. In this scheme, Q stands for QM and D for DO. We tested every possible combination except QDQ, which showed some instability.

For such systems, MP2 is known to give fairly accurate results at reasonable computation cost.[100] It was employed to obtain the geometries of the benzene rings in the DO parametrization.[101] Therefore, we used it here in combination with the SVP, TZVP, TZVPP, and QZVP basis sets. The counterpoise correction for the basis set superposition error was computed for the QQQ and QQD systems. In the QQD case, this correction was evaluated for the QQ system without taking into account the benzene ring represented at the MM-DO level.

As previously, we first consider the binding energy of the complex. Figure 3 shows the MP2-based results for all basis sets and QM/MM partitionings, along with the experimental and the MM-DO value. The pure MP2 results (QQQ) are in good agreement with experimental results, especially for the larger basis sets. The MM-DO approach (DDD) overestimates the binding energy by around 15 kcal/mol when using the original DO parameters that had been calibrated for proper interaction with water molecules (solvation energy).[101] The QM/MM-DO binding energies (QQD, QDD, DQD) apparently interpolate between the QQQ and DDD values, approaching the MM-DO result when including two fragments in the MM region.

When treating only the sodium cation at the QM level (DQD), the computed binding energy is essentially identical to the MM-DO value, regardless of the chosen basis set, indicating that the MM-DO parameters for $Na^+$ are consistent with its QM description. Improvements in the parametrization should

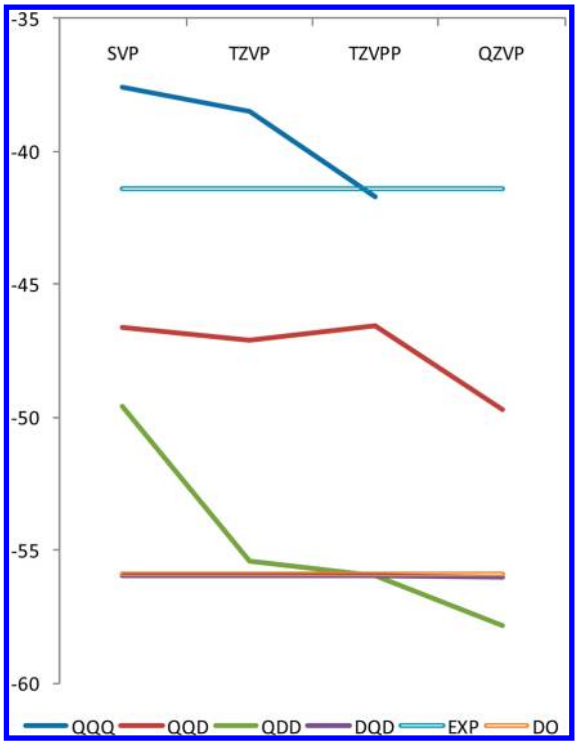



**Figure 3.** RI-MP2-based results for the energy of the cationic bis(benzene)sodium sandwich complex relative to the isolated fragments (in kcal/mol) obtained with four different basis sets. Shown are the full QM results (QQQ, blue), the QM/MM-DO results for different partitionings (QQD, red; QDD, green; DQD, purple; see text), the experimental value (EXP), and the pure MM-DO value (DO). Energies for complexes with more than one QM fragment were corrected for basis set superposition error (BSSE). For technical reasons, the BSSE could not be determined for the QQQ/QZVP combination.

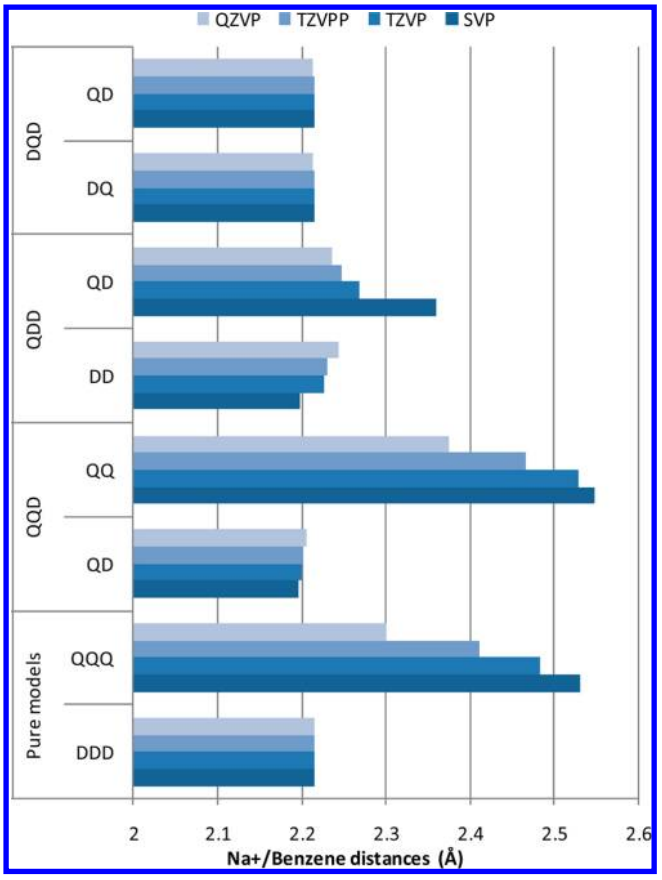



**Figure 4.** RI-MP2-based distances between the sodium cation and the center of the benzene rings in the cationic bis(benzene)sodium complex for four different basis sets (SVP, TZVP, TZVPP, and QZVP; see color code). Shown are results from the pure QM treatment (QQQ), from the pure MM-DO treatment (DDD), and from QM/MM-DO calculations (QQD, QDD, and DQD; the labels QQ, QD, and DQ specify which distance is plotted). See text for further details.

thus focus on the benzene part.[99] We note in this context that the QM/MM-DO nonbonded interactions are calculated from the same Lennard-Jones potential that is used in the DO model. As known from other QM/MM-DO studies,[99] the readjustment of these parameters for the QM atoms may improve the results through a more realistic description of the $Na^+$/benzene interaction. With proper reparameterization of the force field, QM/MM-DO calculations should give binding energies of an accuracy similar to the QQQ approach.

Figure 4 shows the optimized distances between the sodium cation and the center of the benzene rings for all currently investigated approaches. In the case of the pure models (QQQ and DDD), the complex is symmetric, and hence only one distance value is given. The symmetry is lost for the hybrid QM/MM models (QQD, QDD, DQD) for which both distances are shown with the corresponding assignments. In the pure QM model (QQQ), basis set extension from SVP to QZVP shortens the distance from 2.5 to 2.3 Å. The QQ distance in QQD shows similar behavior. In the pure MM model (DDD), the optimized distance is 2.2 Å. Similar values are obtained in the QM/MM models for distances involving a benzene ring described at the MM level, regardless of the chosen basis set (QQD, QDD, DQD; see Figure 4). The angle between the three moieties (benzene center−$Na^+$−benzene center) is always found to be around 180°.

**4.4. Influence of Force Field Polarization on Enzymatic Reactions.** In this section, we investigate the effect of DO polarization on enzymatic reactions. Our goal is to assess the influence of DO polarization on single-point QM/MM energies along previously determined enzymatic reaction pathways. In this initial study, we refrain from geometry optimizations and from free energy calculations (sampling), and hence also from comparison with experimental data, because we consider it most important to first gauge the basic effect of DO polarization on the QM/MM energetics.

As the main test enzyme, we have chosen chorismate mutase from *Bacillus subtilis*, which catalyzes the pericyclic Claisen rearrangement of chorismate into prephenate through a six-membered transition state (see Figure 5). This enzyme is well characterized experimentally[102] and has been extensively studied by QM/MM computations.[1,103] The rather small substrate (24 atoms) is a natural choice for the QM region; it is bound through noncovalent interactions so that there is no need for cutting bonds at the QM/MM boundary.[104]

For the MM region, we used the recently published CHARMM-DO parameters for proteins[39] and the SWM4 parameters for water.[64] No further DO parameters are required, since there are no other species in the MM region. For the QM-MM nonbonded interactions, we applied the CHARMM22 Lennard-Jones parameters of the substrate as in previous studies. These may not be the optimum choice, but we expect them to be realistic enough for a qualitative


dx.doi.org/10.1021/ct401095k | J. Chem. Theory Comput. 2014, 10, 1795−1809

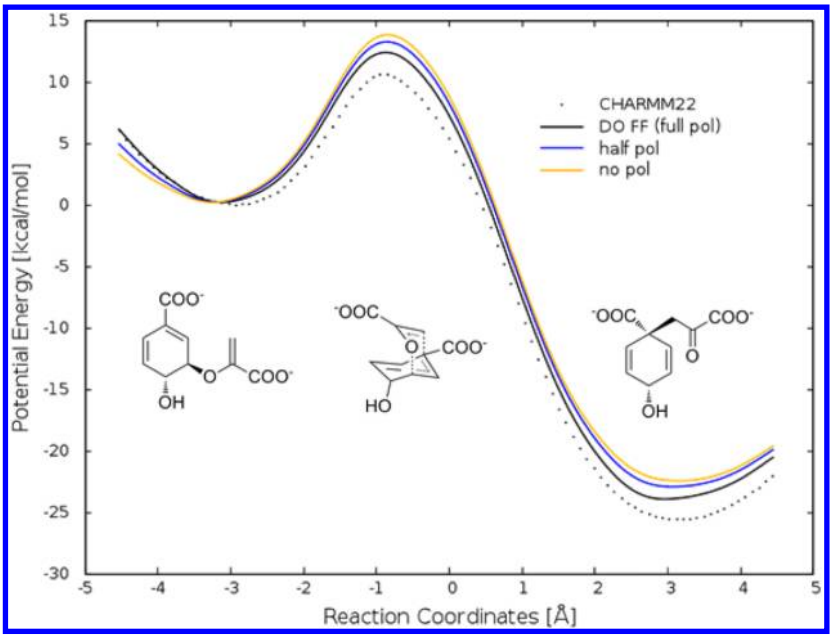



**Figure 5.** Potential energy profile of the Claisen rearrangement in chorismate mutase from QM(RI-MP2/SVP)/MM optimizations using the fixed-charge CHARM22 and the polarizable CHARMM-DO force fields. Also shown are QM/MM results obtained with the CHARMM-DO force field with the DO contributions switched off (no pol) or scaled down by a factor of 0.5 (half pol). The reaction coordinate is the difference between the distances for the breaking and forming bonds. The insets show the QM regions for the reactant (left), transition state (middle), and product (right). See text for further details.

assessment of the effect of MM polarization on the reaction profile.

In previous work of our group on chorismate mutase,[62] we had taken five snapshots from a classical molecular dynamics run of the reactant system and had used them as starting points for QM(MP2/SVP)/CHARMM22 geometry optimizations to determine reaction paths for the chorismate-prephenate rearrangement. We have now performed single-point calculations at the these previously optimized geometries using the QM(MP2/SVP)/CHARMM-DO approach. These calculations were done for all five pathways at all available points (intervals of 0.05 Å). The choice of QM method is supported by the fact that the MP2/SVP level of theory was the one that gave the best QM/MM-DO results for the NMA dimer (see section 4.2).

We adopted the following procedure for evaluating the effects of MM polarization. At any given geometry, we first replaced the fixed-charged CHARMM22 force field for the MM region by the nonpolarizable part of the CHARMM-DO force field (without the DO contributions but including the lone pair terms at the heteroatoms) and evaluated the corresponding QM/MM single-point energies (labeled "no pol"). We then included the DO contributions and reevaluated the single-point energies with full MM polarization using the SOR approach with a damping factor of 0.2 to obtain the DP positions (results labeled "full pol"). To check for consistency, we also considered the case in which the DO contributions are scaled down by applying a factor of 0.5, with appropriate scaling of the DO charges to preserve the force constant of the DO spring (results labeled "half pol"). We also tried to double the DO contributions, but this invariably led to nonconvergence of the iterative SOR procedure ("polarization catastrophe"). Figure 5 gives an example of the results obtained for one of the snapshots. Since the computed single-point energies show some minor irregular fluctuations ("noise" mostly arising from the replacement of the nonpolarizable terms), we used spline interpolation to produce smooth curves in Figure 5; this does not affect the following qualitative assessment of the effects of MM polarization.

Table 4 lists the changes in the computed barrier height and reaction energy for the five snapshots considered when switching from the "no pol" to the "full pol" model (in parentheses: from "no pol" to "half pol"). Evidently, the inclusion of full MM polarization consistently diminishes the barrier height, and it also tends to make the reaction more exothermic. The "full pol" results are close to those obtained with the additive CHARMM22 force field, which are known to be in good agreement with experimental results. Inclusion of half of the MM polarization normally leads to changes in the same direction, which are however less pronounced than might have been expected (typically ca. 40% except for $\Delta\Delta E$ of snapshot 4, see Table 4), indicating the nonlinear nature of polarization effects. Quantitatively, the effects of MM polarization on the computed barriers and reaction energies in chorismate mutase are rather small: for example, the calculated

**Table 4. Effect of MM Polarization on the Barrier ($\Delta\Delta E^{\ddagger}$) and the Reaction Energy ($\Delta\Delta E$) of the Claisen Rearrangement in Chorismate Mutase in Five Independent Snapshots (in kcal/mol)[a]**

| snapshot | $\Delta\Delta E^{\ddagger}$ | | $\Delta\Delta E$ | |
|---|---|---|---|---|
| 1 | −1.3 | (−0.29) | −1.21 | (−0.41) |
| 2 | −1.25 | (−0.51) | 0.23 | (0.09) |
| 3 | −1.42 | (−0.62) | −2.01 | (−0.73) |
| 4 | −0.52 | (−0.13) | −0.46 | (0.02) |
| 5 | −1.15 | (−0.51) | −2.99 | (−1.16) |

[a]The values correspond to the differences between "full pol" and "no pol" results (in parentheses: between "half pol" and "no pol"). See text for details.

barriers are lowered by 0.5 to 1.4 kcal/mol. The computed barrier heights for the five snapshots are around 10 kcal/mol; the contributions from MM polarization thus amount to about 5−15%, which is consistent with previous studies.[43]

The Claisen rearrangement catalyzed by chorismate mutase involves relatively little charge transfer,[105] and one may thus suspect that the impact of MM polarization could be more pronounced in enzymatic reactions that exhibit more pronounced charge redistribution. To check for this possibility, we have investigated the enzyme p-hydroxybenzoate hydrolase (PHBH) in a completely analogous manner. PHBH catalyzes the transformation of p-hydroxybenzoate into 3,4-dihydroxybenzoate, by formally moving an "$OH^+$" moiety from the cofactor to the substrate, with a concomitant charge transfer of one electron in the opposite direction. We selected the four snapshots from our previous QM/MM work on PHBH[105] and performed single-point QM/MM computations with and without MM polarization using the same QM method (B3LYP/6-31G*), the same QM region, and the same geometries as before.[105] Like in the case of chorismate mutase, the effect of MM polarization was assessed by switching it on and off. In these single-point QM/MM energy evaluations, the PHBH protein was represented by the CHARMM-DO force field, whereas the ribityl side chain was described by the standard CHARMM force field (due to the lack of CHARMM-DO parameters). This inconsistency is not expected to be severe, because the main impact of MM polarization should come from the polarizable PHBH residues surrounding the reactive center of the system (and not from the rather distant ribityl side chain at the opposite side of the cofactor). The single-point results for PHBH are collected in Table 5.

**Table 5. Effect of MM Polarization on the Barrier ($\Delta\Delta E^{\ddagger}$) and the Reaction Energy ($\Delta\Delta E$) of the Electrophilic Substitution Reaction in p-Hydroxybenzoate Hydroxylase in Four Independent Snapshots (in kcal/mol)[a]**

| snapshot | $\Delta\Delta E^{\ddagger}$ | | $\Delta\Delta E$ | |
|---|---|---|---|---|
| 1 | 2.1 | (1.7) | 1.7 | (1.2) |
| 3 | 0.8 | (0.6) | 2.8 | (1.9) |
| 4 | 0.8 | (0.6) | 0.5 | (0.3) |
| 5 | 2.2 | (1.3) | 2.7 | (1.7) |

[a]The values correspond to the differences between "full pol" and "no pol" results (in parentheses: between "half pol" and "no pol"). See text for details. Snapshot 2 was discarded already in the original work for technical reasons.[105]

In analogy to chorismate mutase (see Table 4), the effects of MM polarization in PHBH are rather small on an absolute scale both for the barrier (0.8 to 2.2 kcal/mol) and for the reaction energy (0.5 to 2.7 kcal/mol). The QM/CHARMM-DO single-point values range from 7.6 to 11.0 kcal/mol for the barrier and from −26.1 to −31.0 kcal/mol for the reaction energy so that the contributions from MM polarization typically amount to 10−20% and 5−10%, respectively. It seems noteworthy that inclusion of MM polarization leads to a slight decrease of the barrier in chorismate mutase, but to a slight increase for PHBH. Incorporating MM polarization may thus shift barriers in different enzymes into different directions. More importantly, however, our two initial tests indicate that these shifts tend to be rather small regardless of whether the reaction involves little charge redistribution (chorismate mutase) or strong charge transfer (PHBH).

## 5. CONCLUSIONS

In this article, we have addressed several issues connected with the use of the polarizable DO force fields in QM/MM simulations of enzymatic reactions.

First, we investigated the convergence of the DO scheme in QM/MM geometry optimizations. We suggested and assessed two approaches that can be used in addition or instead of the one previously proposed.[49] The first one consists of iteratively solving the system of equations of polarization using Cholesky factorization instead of matrix inversion (for the sake of efficiency). The second one is a hybrid approach in which short-distance interactions are treated by Cholesky factorization and the remaining ones through an iterative self-consistent approach.

We further studied possible QM/MM boundary treatments involving MM atoms that carry Drude oscillators. By a series of tests for n-butanol in the presence of a sodium cation at different positions, we showed that the simplest possible model, namely the removal of the DO located at the MM frontier position (M1), gave the most satisfactory results, with an accuracy similar to what is normally achieved in standard QM/MM boundary treatments for additive force fields.

The systematic tests on the water trimer and the NMA dimer indicate that the DO model performs best in a QM/MM framework when employing the QM method and basis set used in the underlying MM parametrization. The match with experimental results is not perfect but benefits from the calibration of the MM parameters against experimental data. We thus propose to preferentially use this QM method in QM/MM calculations. This is consistent with the approach of Rowley and Roux who computed the solvation structure of sodium and potassium ions in water using MP2/def2-TZVP QM for the ions and neighboring water molecules in combination with the SWM4 model as a reference system.[40] A similar observation was made by Illingworth et al. when considering the water dimer with different polarizable force fields.[43] They concluded that some force fields were more compatible with certain basis sets than with others. We generalize this further by suggesting that best results will be obtained with the method used during MM parametrization. These conclusions are based on a few representative test systems, and further validation will be needed to support them.

In the selected test systems, the QM/MM-DO results for the binding energies and geometries tend to converge to the pure MM-DO values when describing an increasing part of the system at the MM-DO level. This will often lead to rather accurate results since the MM-DO force field generally gives results close to experimental ones (more so than the CHARMM22/27 additive force field).

In the chorismate mutase case, we find that switching on MM polarization in QM/MM single-point calculations affects the computed barrier height for the enzymatic Claisen rearrangement of chorismate to prephenate, but only to a rather minor extent: the barrier is lowered by 0.5−1.4 kcal/mol (i.e., by 5−15%) in the five snapshots studied. In PHBH, MM polarization leads to a slight increase in the computed barriers by 0.8−2.2 kcal/mol. We expect that inclusion of MM polarization may generally cause small changes on the order of 5−20% in the computed energies, but this will need to be confirmed by further studies that should also include geometry optimizations and free energy calculations.

## AUTHOR INFORMATION


### Corresponding Author

*E-mail: thiel@mpi-muelheim.mpg.de.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Alex MacKerell and Pedro Lopes for helpful discussions and for providing the CHARMM-DO parameters for proteins prior to publication.

## REFERENCES


(1) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.
(2) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
(3) Shurki, A.; Warshel, A. In *Advances in Protein Chemistry*; Valerie, D., Ed.; Academic Press: San Diego, CA, 2003; Vol. *66*, p 249.
(4) Senn, H.; Thiel, W. In *Atomistic Approaches in Modern Biology*; Reiher, M., Ed.; Springer: Berlin, 2007; Vol. *268*, p 173.
(5) Mennucci, B. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6583.
(6) Senn, H. M.; Thiel, W. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182.
(7) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425.
(8) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *Theor. Chem. Acc.* **2009**, *124*, 11.
(9) Yu, H.; van Gunsteren, W. F. *Comput. Phys. Com.* **2005**, *172*, 69.
(10) Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034.
(11) Kaminski, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787.
(12) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.
(13) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.
(14) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549.
(15) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2013**, *9*, 4046.
(16) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1504.
(17) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
(18) Rick, S. W.; Stuart, S. J.; Bader, J. S.; Berne, B. *J. Mol. Liq.* **1995**, *65*, 31.
(19) Stuart, S. J.; Berne, B. *J. Phys. Chem.* **1996**, *100*, 11934.
(20) Patel, S.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1.
(21) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025.
(22) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185.
(23) Vorobyov, I.; Anisimov, V. M.; Greene, S.; Venable, R. M.; Moser, A.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2007**, *3*, 1120.
(24) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D., Jr.; Schulten, K.; Roux, B. *J. Phys. Chem. Lett.* **2011**, *2*, 87.
(25) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 774.
(26) Lopes, P. E. M.; Zhu, X.; Lau, A.; Roux, B.; MacKerell, A. D. *Biophys. J.* **2011**, *100*, 612.
(27) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2005**, *1*, 153.
(28) Yu, H. B.; Hansson, T.; van Gunsteren, W. *J. Chem. Phys.* **2003**, 118.
(29) Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69.
(30) Straatsma, T. P.; McCammon, J. A. *Mol. Simul.* **1990**, *5*, 181.
(31) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2007**, *3*, 1499.
(32) Geerke, D. P.; van Gunsteren, W. F. *J. Phys. Chem. B* **2007**, *111*, 6425.
(33) Nüsslein, V.; Schröder, U. *Phys. Status Solidi B* **1967**, *21*, 309.
(34) Schröder, U. *Solid State Commun.* **1993**, *88*, 1049.
(35) de Leeuw, N. H.; Parker, S. C. *Phys. Rev. B* **1998**, *58*, 13901.
(36) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2007**, *3*, 1499.
(37) Lu, Z.; Zhang, Y. *J. Chem. Theory Comput.* **2008**, *4*, 1237.
(38) Boulanger, E.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 4527.
(39) Lopes, P. E.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2013**, *9*, 5430.
(40) Rowley, C. N.; Roux, B. t. *J. Chem. Theory Comput.* **2012**, *8*, 3526.
(41) Riahi, S.; Roux, B.; Rowley, C. N. *Can. J. Chem.* **2013**, *91*, 552.
(42) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(43) Illingworth, C. J. R.; Gooding, S. R.; Winn, P. J.; Jones, G. A.; Ferenczy, G. G.; Reynolds, C. A. *J. Phys. Chem. A* **2006**, *110*, 6487.
(44) Illingworth, C. J. R.; Parkes, K. E.; Snell, C. R.; Marti, S.; Moliner, V.; Reynolds, C. A. *Mol. Phys.* **2008**, *106*, 1511.
(45) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452.
(46) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.
(47) Antila, H.; Salonen, E. In *Biomolecular Simulations*; Monticelli, L., Salonen, E., Eds.; Humana Press: New York, 2013; Vol. *924*, p 215.
(48) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399.
(49) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
(50) Xie, W.; Pu, J.; Gao, J. *J. Phys. Chem. A* **2009**, *113*, 2109.
(51) Zhu, X.; Lopes, P. E. M.; MacKerell, A. D. *WIREs Comput. Mol. Sci.* **2012**, *2*, 167.
(52) Baker, C. M.; MacKerell, A. D. *J. Mol. Model.* **2010**, *16*, 567.
(53) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1587.
(54) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1.
(55) ChemShell. www.chemshell.org (accessed Dec 17, 2013).
(56) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. *WIREs Comput. Mol. Sci.* **2014**, *4*, 101−110.
(57) Thiel, W. *MNDO program*; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2004.
(58) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(59) Forester, T. R.; Smith, W. *DL-POLY program*; Daresbury Laboratory: Daresbury, Warrington, England, 1996.
(60) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.
(61) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(62) Polyak, I.; Benighaus, T.; Boulanger, E.; Thiel, W. *J. Chem. Phys.* **2013**, *139*, 064105.
(63) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2009**, *5*, 3114.
(64) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. *Chem. Phys. Lett.* **2006**, *418*, 245.
(65) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
(66) König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.
(67) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. *J. Phys. Chem. B* **1999**, *103*, 6133.
(68) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.
(69) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(70) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe,

M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(71) Keutsch, F. N.; Cruzan, J. D.; Saykally, R. J. *Chem. Rev.* **2003**, *103*, 2533.
(72) Xantheas, S. S.; Dunning, J. T. H. *J. Chem. Phys.* **1993**, *98*, 8037.
(73) Fowler, J. E.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **1995**, *117*, 446.
(74) Wales, D. J. *J. Am. Chem. Soc.* **1993**, *115*, 11180.
(75) Yu, W.; Lopes, P. E. M.; Roux, B.; MacKerell, J. A. D. *J. Chem. Phys.* **2013**, *138*, 034508.
(76) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(77) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(78) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
(79) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(80) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(81) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(82) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.
(83) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.
(84) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.
(85) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.
(86) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.
(87) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907.
(88) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
(89) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775.
(90) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495.
(91) Scholten, M. Ph.D. Thesis, Universität Düsseldorf: Düsseldorf, Germany, 2003.
(92) Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751.
(93) Ponder, J. W.; Case, D. A. In *Advances in Protein Chemistry*; Valerie, D., Ed.; Academic Press: San Diego, CA, 2003; Vol. *66*, p 27.
(94) Möhle, K.; Hofmann, H.-J.; Thiel, W. *J. Comput. Chem.* **2001**, *22*, 509.
(95) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182.
(96) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671.
(97) Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D., Jr.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3509.
(98) Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3509.
(99) Orabi, E. A.; Lamoureux, G. *J. Chem. Theory Comput.* **2011**, *8*, 182.
(100) Lamoureux, G.; Orabi, E. A. *Mol. Simul.* **2012**, *38*, 704.
(101) Lopes, P. E. M.; Lamoureux, G.; Roux, B.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2007**, *111*, 2873.
(102) Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691.
(103) Claeyssens, F.; Ranaghan, K. E.; Lawan, N.; Macrae, S. J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. *Org. Biomol. Chem.* **2011**, *9*, 1578.
(104) Senn, H. M.; Kästner, J.; Breidung, J.; Thiel, W. *Can. J. Chem.* **2009**, *87*, 1322.
(105) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 238−249.