JCTC Journal of Chemical Theory and Computation

Long-Range Electrostatic Effects in QM/MM Studies of Enzymatic Reactions: Application of the Solvated Macromolecule Boundary Potential

Tobias Benighaus and Walter Thiel*

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

Received September 23, 2010

Abstract: Long-range electrostatic interactions are important in simulations of enzymatic reactions. They can be divided into the effects due to bulk solvent and those due to the electrostatic potential of the outer macromolecule. We study and quantify the importance of these two effects for two test systems by application of the solvated macromolecule boundary potential (SMBP) [*J. Chem. Theory Comput.* **2009**, *5*, 3114−3128]. We validate the accuracy of the SMBP for these test systems and present a transferable protocol for determination of optimal SMBP parameters as well as recommended default values for these parameters. Two enzymatic reactions with different characteristics are studied: the intramolecular Claisen rearrangement in chorismate mutase that is associated with little charge transfer and the hydroxylation reaction in *p*-hydroxybenzoate hydroxylase that corresponds to a formal "OH^+" transfer and thus involves significant charge transfer. It is found that the effects of the electrostatic potential of the outer macromolecule and of bulk solvent are only important in the latter case, where their neglect causes deviations in the computed barriers on the order of 1−2 kcal/mol, respectively. Even larger deviations on the order of several kilocalories per mole are observed for the reaction energies in *p*-hydroxybenzoate hydroxylase if the electrostatic potential of the outer macromolecule is neglected.

1. Introduction

Experience in classical simulations of biomolecular systems indicates that a reliable description of long-range electrostatic interactions is of crucial importance.[1−6] Without an accurate description of these effects, simulations can yield results that are even qualitatively wrong. Examples range from the stability of an α helix[7,8] over the thermodynamics of peptide folding processes[9] to properties of lipid bilayers.[10,11]

Efficient approaches to describing electrostatic interactions accurately have been developed for classical force field simulation methods. If all solvent molecules are modeled explicitly, the Ewald summation method in combination with periodic boundary conditions (PBC) is the most established choice.[12−14] Unfortunately, Ewald summation suffers from high computational costs since the enzyme has to be solvated in a water box of adequate size[15−17] to avoid artifacts from the artificially imposed periodicity.[1,18] Alternative approaches include fast multipole methods[19−22] and the use of stochastic boundary conditions.[23−26]

The treatment of long-range electrostatic interactions is far less developed in the context of hybrid quantum mechanical/molecular mechanical (QM/MM) approaches, which are now routinely applied to study enzymatic reactions that necessitate a QM description of bond breaking and forming processes.[6,27,28] The technical difficulties introduced by the QM atoms are the main reason for this situation. Recently, the Ewald summation method has been adapted for QM/MM approaches using semiempirical[29−31] and density functional theory (DFT)[32−34] QM methods.

The region in the enzyme that is described by an accurate QM method in hybrid QM/MM approaches is necessarily small and encompasses at most a few hundred atoms. Therefore, QM/MM methods are best suited for treating

* To whom correspondence should be addressed. E-mail: thiel@mpi-muelheim.mpg.de.

10.1021/ct1005455
Published on Web 12/03/2010

localized electronic processes. In these cases, boundary potentials are an attractive approach to handling long-range electrostatic interactions.[23−26,35−42] In the boundary potential ansatz, the full system is subdivided into an inner region, which comprises the active site and its neighboring residues, and an outer region, which contains the rest of the enzyme and the outer solvent molecules. While all atoms in the inner region are modeled atomistically, the effect of the outer region atoms on the inner region is mimicked by the boundary potential. Assuming an ideal boundary potential, statistical properties derived from simulations of the inner region in interaction with the boundary potential are the same as those from simulations of the full system. Although an exact and rigorous boundary potential can be constructed by integrating over all outer region degrees of freedom,[42] approximations are indispensable to constructing an efficient implementation.

The generalized solvent boundary potential (GSBP)[43] has been successful in classical force field simulations.[43−48] It was recently adapted for QM/MM approaches using semiempirical QM methods.[49,50] In the GSBP, the outer region solvent molecules are represented by a polarizable dielectric continuum (PDC) and the outer macromolecule region by fixed point charges. The electrostatic potential of the outer region is separated into a static potential, which is induced by the outer region charges being shielded by the PDC, and a dynamic reaction field potential, which is induced by the interaction of the inner region charge distribution with the PDC. The great advantage of the GSBP is its closed-form expression for the dielectric response, which is also valid for irregularly shaped dielectric boundaries. However, construction of this expression is connected with a significant computational overhead.

Therefore, we recently introduced the solvated macromolecule boundary potential (SMBP),[51] which is based on approximations similar to those in the GSBP but has been designed to be efficient in geometry optimizations. Moreover, the SMBP can be used in combination with every QM/MM Hamiltonian.

The effect of long-range electrostatics can be divided into two contributions: interactions with bulk solvent and interactions with the electrostatic potential of the outer macromolecule region (EPOM). By construction, the SMBP allows us to study both effects independently by an appropriate choice of the dielectric constant of the PDC. Here, we apply the SMBP to investigate and quantify the significance of these two effects in QM/MM studies of enzymatic reactions. Two enzymatic reactions were selected as test cases that have been the focus of much theoretical and experimental research: the Claisen rearrangement of chorismate to prephenate in chorismate mutase (CM) and the hydroxylation reaction in *p*-hydroxybenzoate hydroxylase (PHBH). Before applying the SMBP, its accuracy for these two systems is evaluated, optimal values for its inherent parameters are determined, and a transferable protocol to validate the SMBP for other enzymatic systems is presented.

2. Methods

In this section, we briefly review the theoretical foundation of the SMBP and the GSBP.

2.1. Generalized Solvent Boundary Potential. The GSBP introduces two main approximations with the objective of providing an accurate and efficient approximation of the electrostatic potential, which is induced by the outer macromolecule and solvent region. The outer solvent molecules are replaced by a PDC, and the outer macromolecule region is represented by fixed point charges. Then, the GSBP consists of two terms: the interaction of the inner region charges (q_A) with the electrostatic potential of the outer macromolecule being shielded by the PDC (ϕ_s^o) and the interaction with the self-induced reaction field ϕ_{rf}^i.

$$\Delta W_{elec}^{GSBP} = \sum_{A\in \text{inner}} q_A \phi_s^o(\mathbf{r}_A) + \sum_{A\in \text{inner}} q_A \phi_{rf}^i(\mathbf{r}_A) \qquad (1)$$

With the outer macromolecule region being fixed, ϕ_s^o is constant and needs to be computed only once at the beginning of a molecular dynamics (MD) simulation. Direct computation of the second term, however, is prohibitively expensive in MD runs since the reaction field potential has to be updated for each configuration, i.e., in each step.

To circumvent repeated solution of the Poisson−Boltzmann (PB) equation to update ϕ_{rf}^i, a Green's function has been introduced to express the inner reaction field potential for a given charge density of the inner region (ρ_i).[43]

$$\phi_{rf}^i(\mathbf{r}) = \int d\mathbf{r}'\, \rho_i(\mathbf{r}')\, G_{rf}(\mathbf{r}, \mathbf{r}') \qquad (2)$$

The reaction field Green's function and the inner region charge density are projected onto the same basis set $\{b_n\}$. The solvation energy of the inner region can then be calculated from the reaction field matrix, M_{rf}, and the generalized multipole moments of the inner region charge density (Q_n).

$$\Delta W_{elec}^{GSBP} = \sum_{A\in \text{inner}} q_A \phi_s^o(\mathbf{r}_A) + \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n \qquad (3)$$

In this way, the GSBP offers an analytical expression for the electrostatic potential and avoids solution of the PB equation in each MD step. However, it is important to point out that application of the GSBP is connected with a significant overhead since computation of the reaction field matrix requires solving the PB equation a few hundred times.[43,50]

2.2. Solvated Macromolecule Boundary Potential. Although free energy computations are becoming more popular in the QM/MM field,[28,52−55] the majority of QM/MM studies still rely on (constrained) geometry optimizations to compute potential energy differences and molecular structures of stationary points. Hence, we developed the solvated macromolecule boundary potential (SMBP) that targets QM/MM geometry optimizations with ab initio or DFT QM methods. Moreover, conceptual proximity to the GSBP was a requirement to allow application of the GSBP in free energy perturbation studies on the basis of QM/MM/SMBP potential energy profiles.[51]

The SMBP relies on the same basic approximations as the GSBP: The outer macromolecule region is represented by fixed point charges and the outer solvent molecules by the PDC. However, the SMBP differs in the computation of the inner self-induced reaction field potential, which is updated in each geometry optimization step by solving the PB equation. This is more efficient than the initial construction of the reaction field matrix by repeated solution of the PB equation (typically about 800 times).[50]

In addition to efficiency in geometry optimizations, the SMBP was designed to be applicable with all QM/MM Hamiltonians. To achieve this, the interaction with the static outer region potential and the reaction field potential is expressed as the interaction of the QM and MM charge densities with the effective potential that they experience, respectively.

$$\Delta W_{\mathrm{elec}}^{\mathrm{SMBP}} = \int \mathrm{d}\mathbf{r}\, \rho_{\mathrm{QM}}(\mathbf{r})\, \phi_{\mathrm{tot}}^{\mathrm{QM}}(\mathbf{r}) + \int \mathrm{d}\mathbf{r}\, \rho_{\mathrm{MM}}\, \phi_{\mathrm{tot}}^{\mathrm{MM}}(\mathbf{r}) \tag{4}$$

with

$$\phi_{\mathrm{tot}}^{\mathrm{QM}}(\mathbf{r}) = \phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}) + \phi_{\mathrm{rf}}^{\mathrm{MM}}(\mathbf{r}) + \frac{1}{2}\phi_{\mathrm{rf}}^{\mathrm{QM}}(\mathbf{r}) \tag{5}$$

$$\phi_{\mathrm{tot}}^{\mathrm{MM}}(\mathbf{r}) = \phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}) + \frac{1}{2}\phi_{\mathrm{rf}}^{\mathrm{MM}}(\mathbf{r}) \tag{6}$$

Since the QM and MM reaction field potentials, $\phi_{\mathrm{rf}}^{\mathrm{QM}}(\mathbf{r})$ and $\phi_{\mathrm{rf}}^{\mathrm{MM}}(\mathbf{r})$, depend on the configuration of the inner region, they have to be updated in each step. Computation of the QM reaction field potential is further exacerbated by the mutual dependence of the QM charge density and the QM reaction field potential. To solve this problem, a self-consistent reaction field (SCRF) procedure is applied: the wave function and the effective QM potential are updated alternately, with the QM charge density being fixed during computation of $\phi_{\mathrm{tot}}^{\mathrm{QM}}$ and vice versa. This approach enables a flexible interface of SMBP and QM codes: ESP charges are used to represent the QM density in the PB equation, and a small set of virtual surface charges is used to represent the SMBP during optimization of the wave function. Hence, the SMBP can be used in combination with virtually any QM code that is suitable for QM/MM calculations.[51]

The QM/MM free energy perturbation (QM/MM-FEP) approach introduced by Yang et al.[56] allows the computation of QM/MM free energy profiles using ab initio or DFT QM methods at moderate computational costs. QM/MM-FEP is based on three assumptions: (1) Sampling over the QM and MM degrees of freedom can be separated in a product ansatz, (2) finite-temperature effects in the QM region can be estimated using the harmonic approximation, and (3) the electrostatic interactions of the MM region with the QM atoms can be approximated as interactions of MM point charges with fixed ESP point charges on the QM atoms. The QM/MM-FEP calculation provides a free energy correction to the potential energy profile along a predefined reaction coordinate.

The SMBP employs the same approximations as the GSBP. Therefore, the SMBP and GSBP can be used complementarily. If one uses a QM/MM/SMBP Hamiltonian to compute the potential energy profile, the efficient GSBP can be applied to sample over the MM degrees of freedom in the subsequent FEP step. The resulting QM/MM/GSBP-FEP method reduces the computational costs of the FEP step by typically 1 order of magnitude in the cases studied.[51]

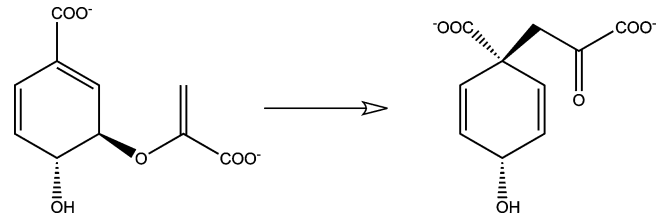

Figure 1. Intramolecular Claisen rearrangement catalyzed by chorismate mutase.

3. Computational Details

GSBP and SMBP have been implemented in a developmental version of the modular program package ChemShell.[57,58] In the present study, the QM energies and gradients were evaluated with the MNDO[59] and Turbomole 5.7.1 programs.[60] The DL_POLY[61] code was employed to run the CHARMM22 force field in all calculations of the MM part.[62] Hydrogen link atoms were applied in combination with the charge-shift scheme[57] to saturate the QM region. The HDLCOpt optimizer was used to optimize stationary points in hybrid delocalized internal coordinates.[63] The PB equation was solved with the ChemShell PB module, which applies the optimal successive over-relaxation method in combination with Gauss-Seidel relaxation to compute the electrostatic potential.[64,65] A maximum absolute change in every grid point of 2×10^{-5} au was adopted as the convergence criterion, and third-order B splines were used to interpolate between the grid points.[66] The dielectric boundary was defined as the superposition of the van der Waals radii from the CHARMM22 force field. MD simulations were performed unter NVT conditions with a time step of 1 fs at a temperature of 300 K controlled by a Nosé−Hoover chain thermostat.[67−70] Free water molecules were kept rigid with SHAKE constraints,[71] and the mass of deuterium was assigned to all hydrogen atoms. The QM reaction field potential was considered converged when the root-mean-squared deviation was below 2×10^{-5} au. At the beginning of the SCRF procedure, a neutral charge was assigned to all QM atoms.

4. Results

4.1. Chorismate Mutase: Validation and Parameter Determination. CM catalyzes the intramolecular Claisen rearrangement from chorismate to prephenate (Figure 1). A recent review provides a summary of theoretical work on the reaction mechanism and the origin of catalysis in CM.[28] The Claisen rearrangement is a pericyclic reaction without significant charge transfer, and one would thus expect only minor effects of the EPOM and bulk solvent on this enzymatic reaction.

The setup for CM was based on a system that has been used in previous work.[52] Initial coordinates were taken from

the crystallographic structure of *Bacillus subtilis* CM (PDB code 2CHT) with a bound transition state analog (TSA). Only the first of four trimers in the asymmetric unit was retained, and the TSA between chains A and B was transformed into a chorismate molecule. The other TSAs were removed. The system was solvated with a 30 Å water sphere and then subjected to a 200 ps QM/MM MD simulation using self-consistent charge-density functional tight binding (SCC-DFTB)[72] as a QM method. Our setup started from the resulting structure corresponding to snapshot 1 in the previous study.[52] To generate 10 initial configurations for the QM/MM calculations that are independent of the previous study, snapshots were taken every 10 ps after 100 ps of extra equilibration. The QM region consisted only of the chorismate molecule, which was modeled by the semiempirical AM1 Hamiltonian.[73] All protein atoms were assigned to the MM region. For each configuration, the inner region was centered on the initial position of the C_1 atom (following IUPAC nomenclature) with an extended dielectric cavity radius of 21.0 Å. All atoms within 19.0 Å of the center were assigned to the inner region and modeled explicitly. Due to the inaccuracies of the SMBP and the GSBP at the boundary,[50,51] the inner region was further subdivided into an active inner region and a frozen inner region. All atoms within 17.0 Å of the center belong to the active region so that it is surrounded by an "insulation" region of 2 Å. A spherical restraint with a radius of 17.0 Å and a force constant of 50 kcal/(mol Å^2) was applied to the oxygen atoms of all active water molecules in MD calculations with the QM/MM/GSBP method. The reaction is described by means of a reaction coordinate (RC) defined as the difference of the lengths of the breaking C−O and the forming C−C bonds. Potential energy profiles of the reaction are computed by constraining the RC to values from −2.5 to +2.5 Å in steps of 0.1 Å.

The electrostatic potentials that constitute the SMBP (see eqs 5 and 6) are obtained as grid-based solutions of the PB equation. A focusing approach is applied to allow usage of fine grids for the inner region.[74] The PB equation is first solved for a coarse outer grid that covers the full system and then for a fine inner grid that focuses on the inner region. The boundary values of the inner grid are set by interpolation from the outer grid. Therefore, the mesh sizes of the grids are the main parameters that determine the accuracy as well as the efficiency of the SMBP. They have to be chosen carefully. Hence, one of the objectives of this study is the development of a transferable protocol to estimate adequate mesh sizes based only on fast single-point energy and gradient calculations.

In a vacuum environment, that is, with a dielectric constant of 1 anywhere in space, the electrostatic potential of the SMBP and of exact Coulombic electrostatics has to be identical. This allows simple evaluation of the accuracy of the SMBP by direct comparison to standard QM/MM results. Table 1 provides the mean absolute (MAD) and maximum absolute deviations (MAX) of the gradient components for mesh size combinations of 0.15, 0.25, 0.4, 0.6, and 0.8 Å for the inner grid and 0.8, 1.25, 1.5, 2.0, and 2.5 Å for the outer grid. Here, all atoms within 18 Å of the center were considered to account for fluctuations of the active atoms that may occur later in the simulations. MAD and MAX results show no significant dependence on the mesh size of the outer grid but a strong dependence on the mesh size of the inner grid. For all mesh size combinations, the MAD values are between 10^{-5} and 5×10^{-5} au. The default convergence criterion for QM/MM geometry optimizations is a maximum absolute gradient component of less than 4.5×10^{-4} au.[63] From this perspective, the MAD results can be deemed accurate. However, they do not allow determination of the optimal mesh size. The MAX deviations are more helpful for this task. For inner mesh sizes of ≤0.25 Å, MAX deviations are below the convergence criterion so that this mesh size excels as a safe choice at acceptable computational costs. Since neither the accuracy nor the computational demands depend strongly on the mesh size of the outer grid, a relatively fine outer grid with a mesh size of 1.25 Å was selected in combination with an inner grid spacing of 0.25 Å for all calculations (unless noted otherwise).

Table 1. Mean Absolute (MAD) and Maximum Absolute (MAX) Deviations [10^{-4} au] of the Electrostatic Forces Computed with the SMBP for the Chorismate Mutase System and Averaged over 10 Configurations (Relative to QM/MM Results with Full Coulombic Electrostatics)[a]

	inner grid size [Å]				
outer grid size [Å]	0.15	0.25	0.40	0.60	0.80
	MAD—atoms within 18 Å				
0.80	0.11	0.12	0.19	0.31	0.45
1.25	0.16	0.16	0.23	0.34	0.48
1.50	0.17	0.17	0.23	0.35	0.48
2.00	0.15	0.16	0.22	0.34	0.47
2.50	0.18	0.18	0.24	0.36	0.49
	MAX—atoms within 18 Å				
0.80	2.19	3.67	8.46	16.29	20.99
1.25	3.15	4.01	8.58	16.40	21.12
1.50	3.37	4.04	8.59	16.44	21.07
2.00	2.94	4.01	8.50	16.37	21.09
2.50	3.26	4.22	8.58	16.37	21.17

[a] Different mesh size combinations were used.

Detailed examination shows that the accuracy of the SMBP strongly depends on the radial position of the atoms. This point is illustrated in Figure 2, which shows the MAD and MAX values for different inner grid mesh sizes as a function of the radial position. In the center of the inner region, the electrostatic potential varies only slowly and is described accurately by all tested mesh sizes. At the boundary, however, the electrostatic potential becomes more complex and the deviation increases significantly for mesh sizes >0.4 Å. Therefore, fine inner mesh sizes appear to be necessary. Since the chemical process occurs in the center of the inner region, however, it is possible that moderate deviations at the boundary are tolerable. To test this hypothesis, reaction energies and activation energies were computed for grid size combinations ranging from 0.25/1.25 Å to 1.8/3.5 Å. The results are documented in Table S1 of the Supporting Information (SI). In comparison to full Coulombic electrostatics, the MAD and MAX deviations of the potential energy differences are below 0.3 and 0.8 kcal/mol, respectively, if the inner mesh size is ≤0.8 Å. Such an inner grid mesh size

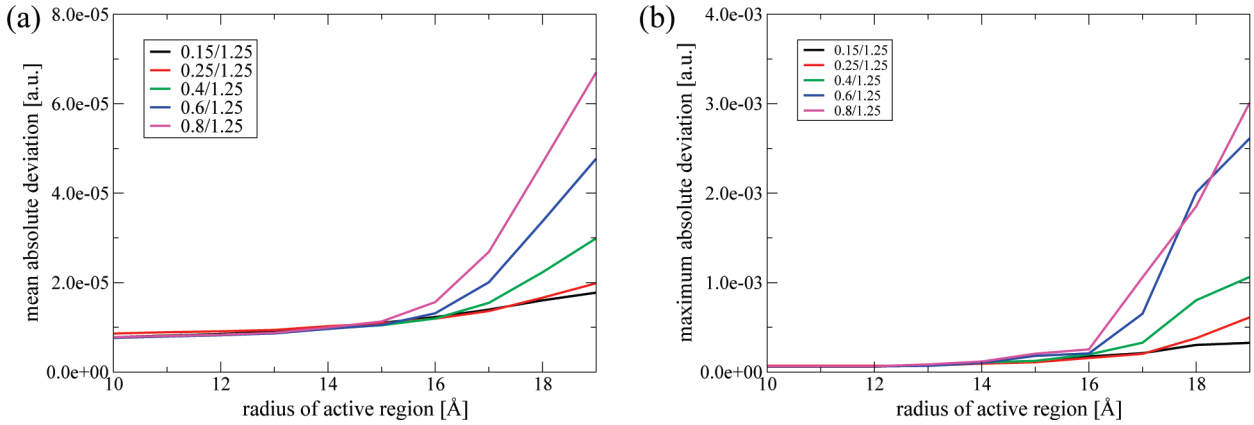

Figure 2. Mean absolute deviations (a) and maximum absolute deviations (b) of the electrostatic forces at all atoms inside the active region relative to the exact QM/MM values for the chorismate mutase (CM) test system. Results are shown for different mesh sizes of the inner grid and are plotted as a function of the radius of the active region. An outer grid size of 1.25 Å is used. All calculations were performed on configuration 1 of the CM system. The radius of the inner region was 19 Å (see text).

corresponds to a MAX deviation of the gradient components in the range of 2×10^{-3} au (see Table 1). We conclude that the SMBP will provide results of high accuracy if the mesh size is chosen such that the MAX values do not exceed 4.5×10^{-4} au anywhere in the inner region. The SMBP will still give results of adequate accuracy with coarser mesh sizes if the MAX values do not exceed 4.5×10^{-4} au for atoms more than 3 Å away from the boundary and if the MAX values do not exceed 2×10^{-3} au for atoms more than 1 Å away from the boundary.

In the GSBP, another important parameter has to be determined: the size of the basis set which models the inner region charge distribution. The chosen orthonormal basis functions are based on spherical harmonics so that the size of the basis set is determined by the order of the highest multipole moment that is included. Table 2 gives the MAD and MAX deviations for basis sets of increasing size determined by maximum multipole moments from order $L = 1$ to $L = 19$. Moreover, it provides the fraction of inaccurate gradient components whose deviation is larger than the standard convergence criterion (see above). SMBP values serve as the reference, since the SMBP represents the basis set limit of the GSBP. In previous applications of the GSBP, multipole moments up to order $L = 19$ were usually included. The MAX results confirm this choice and show that with $L = 19$, MAX deviations on the order of 4.5×10^{-4} au are observed, which is sufficiently accurate. The fractions of inaccurate gradient components indicate that the reaction field potential converges in this system at $L = 17$. The residual error can be attributed to technical differences of SMBP and GSBP. It seems interesting to check whether looser criteria can also yield results of adequate accuracy. Table S2 in the SI shows that the basis-set-dependent error converges for $L = 10$ if only those atoms are considered that are at least 3 Å away from the boundary. The impression that $L = 10$ provides sufficient accuracy is further supported by the fact that the MAX deviation for all atoms at least 1 Å away from the boundary is 1.3×10^{-3} au (Table 2) and, therefore, below the criterion of 2×10^{-3} au that was used to determine the mesh sizes. Hence, $L = 19$ emerges as the safe and $L = 10$ as the efficient choice. In section 4.2, both values are tested, and it will be shown that they describe bulk solvent effects in CM with similar accuracy.

Table 2. Mean Absolute (MAD) and Maximum Absolute (MAX) Deviations [10^{-4} au] and the Fraction of Inaccurate Gradient Components (x_{grad}) [%] of the Electrostatic Forces of the Chorismate Mutase System Computed with the GSBP with Different Basis Set Sizes (Relative to SMBP Results, See Text)[c]

L[a]	MAD	MAX	x_{grad}[b]
1	4.00	69.77	26.47
2	3.19	60.99	21.03
3	2.39	50.28	15.60
4	1.74	39.47	11.05
5	1.22	30.00	6.76
6	0.95	27.21	4.67
7	0.76	22.59	3.16
8	0.61	18.56	1.97
9	0.50	14.48	1.13
10	0.42	13.31	0.70
11	0.36	10.97	0.40
12	0.32	8.75	0.25
13	0.29	7.34	0.12
14	0.27	5.95	0.07
15	0.25	5.26	0.05
16	0.24	5.03	0.04
17	0.23	4.72	0.03
18	0.22	4.55	0.03
19	0.22	4.53	0.03

[a] highest order multipole moment. [b] fraction of gradient components with a deviation >4.5 × 10^{-4} a.u. [%]. [c] All values are averaged over 10 configurations and computed with a mesh size combination of 0.25/1.25 Å.

4.2. Chorismate Mutase: Results. We now apply the SMBP to study the effects of the EPOM and bulk solvent on the enzymatic reaction in CM. Table 3 gives the reaction energies and activation energies that were computed with different methods to describe these effects. In each case, the energy differences were averaged over 10 configurations. The standard deviation of any given mean value defines a confidence interval corresponding to a confidence limit of 68%. Statistically significant are differences between mean values that are larger than the confidence interval of the reference value.

Standard QM/MM calculations with full Coulombic electrostatics (Coulomb in Table 3) give reaction and

Table 3. Reaction Energies (ΔE) and Activation Energies ($\Delta E^{\ddagger}$) of the Claisen Rearrangement in Chorismate Mutase [kcal/mol] Computed with Different Treatments of Long-Range Electrostatics

	Coulomb		SMBP(vac)[a]		NOR[b]		SMBP(solv)[c]		SMBP(solv,app)[d]	
configuration	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
1	−20.0	33.0	−19.3	33.1	−19.9	33.4	−20.1	32.6	−19.1	32.8
2	−17.3	34.8	−17.2	34.8	−11.5	41.6	−18.3	33.1	−17.2	34.3
3	−17.5	40.1	−17.1	40.4	−18.9	37.2	−16.8	39.5	−16.5	40.2
4	−17.7	38.2	−17.7	38.3	−14.9	36.5	−17.3	37.5	−17.3	37.8
5	−21.9	32.4	−21.8	32.5	−21.8	33.1	−21.9	32.4	−21.9	32.7
6	−18.0	36.2	−17.8	36.3	−17.1	36.6	−17.7	35.4	−17.7	35.4
7	−19.6	35.3	−19.5	35.4	−19.3	35.3	−19.9	34.7	−19.6	35.1
8	−19.3	35.8	−19.3	35.7	−19.2	37.2	−19.4	34.9	−19.0	36.0
9	−16.3	34.4	−16.1	34.5	−17.4	34.7	−17.9	33.5	−17.2	34.0
10	−20.7	30.6	−20.7	30.6	−20.6	31.3	−21.0	30.4	−20.0	31.5
mean value	−18.8	35.1	−18.6	35.2	−18.1	35.7	−19.0	34.4	−18.5	35.0
standard deviation of data[e]	1.8	2.8	1.8	2.8	3.0	2.9	1.7	2.6	1.7	2.6
standard deviation of mean[f]	0.6	0.9	0.6	0.9	0.9	0.9	0.5	0.8	0.5	0.8
MAD[g]			0.2	0.1	1.3	1.5			0.5	0.6

[a] $\varepsilon = 1$. [b] Neglect of outer region. [c] $\varepsilon = 80$. [d] $\varepsilon = 80$, $\phi_{rf}^{QM} = 0$. [e] Standard deviation of individual energy values. [f] Standard deviation of the mean value (68% confidence limit). [g] Mean absolute deviation relative to full Coulombic electrostatics. For SMBP(solv,app), SMBP(solv) values are used as a reference.

activation energies of −18.8 and +35.1 kcal/mol, respectively. In the standard QM/MM approach using electronic embedding, the EPOM is computed without approximation, but the effect of the bulk solvent is neglected. Compared to the experimental value of 12.7 kcal/mol,[75] the computed value for the activation energy is strongly overestimated because of the use of the AM1 Hamiltonian in the QM region.[28] Despite this shortcoming, we expect the AM1 Hamiltonian to be adequate for studying the relative effects caused by long-range electrostatics, which should be captured at this level in a realistic manner. For two of the 10 configurations, the conclusions drawn from AM1 results were confirmed by more accurate DFT calculations using the B3LYP functional[76−78] in combination with a 6-31G* basis (see below).

When the SMBP is applied under vacuum conditions (SMBP(vac)), the effect of the EPOM is approximated and bulk solvent effects are neglected. The SMBP describes the EPOM accurately and reproduces the standard QM/MM results with very small deviations. The mean values for the reaction and activation energy differ by only 0.2 and 0.1 kcal/mol, respectively. Also, the results for the individual configurations are very similar, as indicated by MAD values of similar size.

Going one step further, one could completely neglect the electrostatic influence from the outer solvent and macromolecule region (neglect of outer region, NOR). This simplest approximation should lead to significant deviations from the standard QM/MM results, if the energetics are sensitive to long-range electrostatics. With the NOR approximation, the mean reaction and activation energies increase by 0.7 and 0.6 kcal/mol, respectively (Table 3). These differences are at the boundary of the confidence interval of the standard QM/MM results and thus cannot be considered significant. However, the MAD values of 1.3 and 1.5 kcal/mol for reaction and activation energies show that the results deviate significantly for the individual configurations. With less fortuitous error cancellation, significant discrepancies of more than 1 kcal/mol are therefore possible when applying the NOR approximation.

Table 4. Reaction Energies (ΔE) and Activation Energies ($\Delta E^{\ddagger}$) in Chorismate Mutase [kcal/mol] Computed at the QM(B3LYP/6- 31G*)/MM Level of Theory with Different Treatments of Long-Range Electrostatics (See Table 3 for Notation)

	Coulomb		SMBP(vac)		NOR[a]		SMBP(solv)	
configuration	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
3	−14.0	15.9	−14.8	15.5	−14.3	14.8	−13.5	15.7
8	−16.2	13.0	−16.2	13.0	−17.1	12.8	−16.3	12.6

[a] Neglect of outer region.

The effect of bulk solvent on this reaction was studied using the SMBP with a dielectric constant of 80 for the outer solvent region (SMBP(solv)). The inclusion of bulk solvent lowers the reaction energy only marginally by 0.2 kcal/mol. A stronger change is observed for the activation energy, which is reduced by 0.7 kcal/mol to a value of 34.4 kcal/mol. However, both values lie still within the statistical error bars of the standard QM/MM results so that the effect of bulk solvent has to be deemed insignificant for this reaction.

For configurations 3 and 8, the effects of the EPOM and bulk solvent were checked by DFT/MM calculations using B3LYP/6-31G* for the QM region. The results are summarized in Table 4. Focusing on the computed barriers, we first note that the standard DFT/MM treatment with Coulombic electrostatics yields values (13.0−15.9 kcal/mol) close to experimental results (see above). These barriers are lowered slightly when applying the SMBP under vacuum conditions (by 0.0−0.4 kcal/mol) and more so when using the NOR approximation (by 0.2−1.1 kcal/mol). The barrier lowering due to bulk solvent is again small (0.2−0.4 kcal/mol). These DFT results are fully consistent with the AM1/MM results (Table 3).

Since the QM region is usually located far away from the dielectric boundary, one can assume that the QM contribution to the reaction field potential is small and can be neglected. This is the SMBP(solv,app) approximation which neglects the ϕ_{rf}^{QM} term in eq 5 and thus avoids the SCRF procedure, thereby accelerating the SMBP calculations. This approach yields reaction and activation energies of −18.5 and +35.0

Table 5. Reaction Free Energies (ΔA) and Activation Free Energies ($\Delta A^{\ddagger}$) in Chorismate Mutase [kcal/mol] at $T = 300$ K Computed with Different Treatments of Long-Range Electrostatics

	GSBP(vac)		GSBP(solv)[a]		GSBP(solv,fast)[b]		NOR[c]	
configuration	ΔA	$\Delta A^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$
1	−19.0	31.1	−17.6	31.1	−17.0	31.6	−16.9	31.7
2	−17.1	32.6	−16.2	33.1	−15.7	33.1	−14.5	36.8
3	−16.0	37.8	−14.6	36.2	−15.6	36.6	−16.3	33.9
4	−18.0	35.1	−15.2	35.7	−15.9	34.6	−15.8	33.7
5	−18.3	33.5	−19.2	32.2	−18.3	32.7	−18.9	32.2
6	−17.0	34.8	−17.0	32.4	−16.6	33.4	−15.8	34.6
7	−18.6	33.2	−18.1	32.5	−18.5	32.1	−19.1	31.7
8	−18.1	33.2	−16.6	33.4	−17.3	32.9	−16.2	33.1
9	−17.5	32.2	−16.6	32.3	−17.3	31.7	−16.2	31.9
10	−19.7	30.4	−18.4	30.2	−18.0	31.1	−17.8	31.9
mean value	−17.9	33.4	−16.9	32.9	−17.0	33.0	−16.8	33.1
standard deviation of data[d]	1.1	2.1	1.4	1.8	1.1	1.6	1.4	1.6
standard deviation of mean[e]	0.3	0.7	0.5	0.6	0.3	0.5	0.5	0.5

[a] Grid size: 0.25/1.25 Å, 400 basis functions ($L = 19$). [b] Grid size: 0.6/2.0 Å, 121 basis functions ($L = 10$). [c] Neglect of outer region. [d] Standard deviation of individual energy values. [e] Standard deviation of the mean value (68% confidence limit).

kcal/mol, respectively. These values are within the statistical error bars of the standard QM/MM and the SMBP(solv) results. Hence, this approximation is valid for this reaction.

Finally, we checked whether the effect of the EPOM and bulk solvent is more pronounced for the dynamical behavior of this system, that is, when computing free energy differences. We applied the QM/MM-FEP method to compute the free energies of reaction and activation at $T = 300$ K (Table 5). Along the RC, the reaction was split up into discrete windows. For each of these, ESP charges for the QM atoms were derived by fitting to the electrostatic potential at the 200 MM atoms closest to the QM region. For each window, the active region was equilibrated for 10 ps with the QM region held fixed and the QM-MM electrostatic interactions modeled classically using the ESP charges. Subsequently, energy differences were sampled for 10 ps, and the data were subjected to statistical tests for lack of correlation and trend.[79] If necessary, data were discarded to obtain a series of values without a trend. Although the effective sampling length was determined by the statistical tests, it was ensured that at least 5 ps of data were retained.

In the QM/MM-FEP calculations, the GSBP was applied to sample the MM phase space efficiently.[51] Again, the accuracy of the QM/MM/GSBP-FEP ansatz was validated against standard QM/MM-FEP results using vacuum conditions. However, since QM/MM-FEP calculations are computationally intense, free energy profiles without GSBP were obtained only for two configurations. The results are documented in Table S3 of the SI and show that QM/MM/GSBP reproduces QM/MM results in FEP calculations well, with deviations of 0.0−0.2 kcal/mol in the reaction free energies (ΔA) and 0.3−0.5 kcal/mol in the activation free energies ($\Delta A^{\ddagger}$). We now consider the changes in the calculated mean values when going from potential energies (Table 3) to free energies (Table 5). Using the GSBP in a vacuum, the reaction energy increases by +0.7 to −17.9 kcal/mol while the activation energy decreases by 1.8 to 33.4 kcal/mol. Four terms contribute to the difference between free and potential energy: the zero point vibrational energy (ZPE, ΔE_{ZPE}), the thermal contribution to the internal energy (ΔU_{th}), the QM entropy contribution ($-T\Delta S_{QM}$), and the entropy of the QM−MM interactions ($-T\Delta S_{QM-MM}$). In the QM/MM-FEP approach, we assume that $-T\Delta S_{QM-MM}$ is the difference of free and potential QM−MM interaction energies and neglect other contributions. Table 6 shows that the total effect on the barrier is dominated by the ZPE. The entropic QM contribution from the harmonic approximation (0.8 kcal/mol) cancels the entropic QM−MM contribution from the sampling (−0.8 kcal/mol) so that entropy does not contribute significantly to finite-temperature effects on the activation energy. This result contradicts the experimental observation of an entropic contribution of $(-T\Delta S)_{exp} = 2.7$ kcal/mol.[75] The negative change of entropy in the transition state has been ascribed to the loss of conformational flexibility due to the partial formation of two covalent bonds.[52] Since the degrees of freedom of the QM region are held fixed during QM/MM-FEP sampling, this phenomenon cannot be captured in the free energy difference of QM−MM interactions. The entropic QM contribution based on the harmonic approximation shows the correct trend but underestimates the magnitude. Therefore, one may tentatively assume that a significant contribution to the negative change in entropy in the activation energy of CM comes from the degrees of freedom that involve coupled motions of QM and MM atoms which are not sampled in the QM/MM-FEP ansatz. The deviation of QM/MM/GSBP-FEP results from experimental results would then arise from the QM/MM-FEP ansatz itself, not from the approximations in the boundary potentials. This view is supported by the results of previous semiempirical QM/MM umbrella sampling simulations (QM = SCC-DFTB) that do not restrict the flexibility of the QM region and reproduce the entropic

Table 6. Contributions [kcal/mol] to the Differences between Free and Potential Energies in Chorismate Mutase Using the QM/MM/GSBP-FEP Method (See Text)[a]

	reaction	activation
ΔE_{ZPE}	0.3	−1.5
ΔU_{th}	−0.1	−0.3
$-T\Delta S_{QM}$	0.3	0.8
$-T\Delta S_{QM-MM}$	0.2	−0.8
total	0.7	−1.8

[a] All values are averaged over 10 configurations.

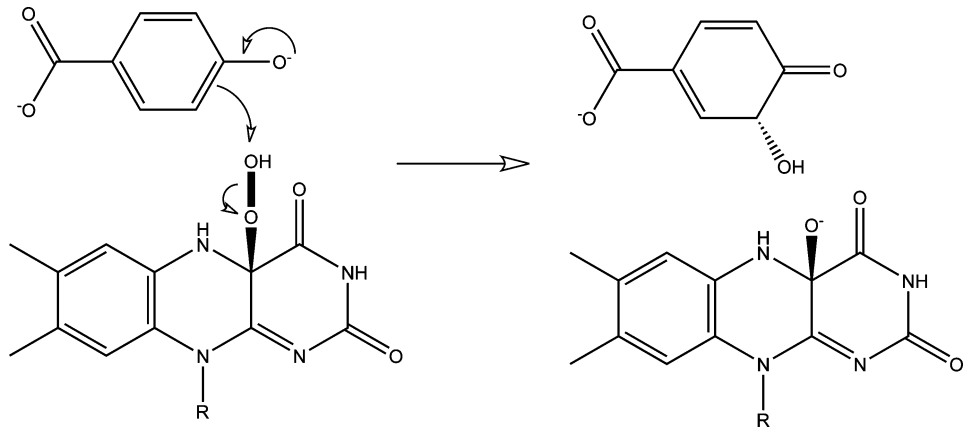

Figure 3. Hydroxylation reaction catalyzed by *p*-hydroxybenzoate hydroxylase. R denotes the ribityl side chain of the hydroperoxo flavine-adenine cofactor.

contribution to the barrier with good accuracy.[52] It is further substantiated by the observation that QM/MM-FEP predicts a negative entropic QM−MM contribution ($-T\Delta S_{QM-MM}$) to the activation free energy independent of the boundary potential and the QM/MM level of theory (Table S4, SI).

The effects of bulk solvent on the free energy differences were studied using a dielectric constant of 80 for the outer solvent region (GSBP(solv)). The activation free energy decreases upon such inclusion of bulk solvent by 0.5 kcal/mol and is thus still within the confidence interval of the vacuum result. The reaction energy, however, increases significantly by +1.0 to −16.9 kcal/mol. These results are reproduced quite accurately if coarser mesh sizes of 0.6/2.0 Å and a smaller basis set ($L = 10$) are used (GSBP-(solv,fast)). However, simply neglecting the electrostatic effect of the outer region (NOR) yields results of very similar accuracy. This can be explained by the shielding effect of the outer solvent, which is mimicked most simply by neglecting the outer region charges. The quantitative accuracy is probably fortuitous given that the results for the potential energy differences deviate significantly from the SMBP(solv) results (Table 3).

On the basis of these results, we may draw the following conclusions for CM: The SMBP and the GSBP reproduce long-range electrostatic interactions with high accuracy when using fine grids with mesh sizes of 0.25/1.25 Å as well as coarser grids with mesh sizes of 0.6/2.0 Å. It should be emphasized, however, that the influence of the EPOM on this reaction is limited. This can be attributed to the "neutral" character of the intramolecular Claisen rearrangement in CM. Neglecting the electrostatic influence of the outer region does not have much effect on the mean potential energy differences although the results for the individual configurations differ significantly. Similarly, bulk solvent effects do not affect the potential energy differences. A statistically significant influence of bulk solvent on the order of 1 kcal/mol was only observed for the reaction free energy.

4.3. *p*-Hydroxybenzoate Hydroxylase. The hydroxylation reaction in the catalytic cycle of PHBH has been the focus of many theoretical investigations. It has become a prototypical test system to benchmark theoretical methods for computational enzymology. A summary of theoretical work on PHBH is provided in a recent review.[28] In this reaction, a formal OH^+ unit is transferred from the flavin-adenine hydroperoxide cofactor (FADHOOH) to the *p*-hydroxybenzoate substrate (pOHB; see Figure 3). Since the reaction is associated with a considerable charge transfer, it is a potentially good test case for reactions with stronger long-range electrostatic interactions.

The PHBH setup was based on a system that has been used in previous studies.[79−81] It was generated by solvating the enzyme containing the FADHOOH cofactor and the pOHB substrate in a 90 Å solvent box and equilibrating it with gradually decreasing harmonic restraints on the non-water atoms. In a following MD run, harmonic restraints were acting only on the cofactor and substrate, and in the resulting structure, which served as starting point for our setup, all water molecules outside 11 Å from any protein atom were discarded.

Due to a change of force field from GROMOS (previously) to CHARMM (this study), the system was re-equilibrated for 500 ps with constraints on the cofactor, substrate, and all water molecules with oxygen atoms outside 2.9 Å from any protein atom. Five configurations were selected from this MD run after 420, 440, 460, 480, and 500 ps, which were used as starting structures to compute potential energy profiles of the hydroxylation reaction. The QM region comprised pOHB and the isoalloxazine part of FADHOOH up to the first methylene unit of the ribityl side chain, which was saturated with a hydrogen link atom. The B3LYP functional[76−78] in combination with a 6-31G* basis set was used to model the QM atoms. When applying the SMBP, the inner region was centered on the initial position of the distal oxygen atom of the FADHOOH cofactor with an extended dielectric cavity radius of 22.5 Å. All charge groups with any atom within 18.5 Å of the center were assigned to the inner region and were modeled explicitly. The remaining set of atoms constituted the outer macromolecule region and was represented by the SMBP. All charge groups with any atom within 16 Å of the center were assigned to the active region and were allowed to move during optimization. The hydroxylation reaction was described by means of a RC, which was chosen to be the difference in the bond lengths of the breaking bond between the proximal (O_p) and distal (O_d) oxygen atoms of FADHOOH and the forming bond

Table 7. Reaction Energies (ΔE) and Activation Energies ($\Delta E^{\ddagger}$) of the Hydroxylation Reaction in *p*-Hydroxybenzoate Hydroxylase [kcal/mol] Computed with Different Treatments of Long-Range Electrostatics

	Coulomb		SMBP(vac)[a]		NOR[b]		SMBP(solv)[c]		SMBP(solv,app)[d]	
configuration	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
1	−27.0	9.5	−26.9	9.4	−28.9	11.4	−28.5	8.0	−29.7	7.9
3	−21.2	11.4	−20.9	11.4	−35.5	4.6	−21.8	10.7	−30.6	7.9
4	−28.1	7.9	−27.9	7.9	−32.1	6.6	−29.8	5.9	−31.3	6.1
5	−25.3	10.1	−25.2	10.2	−29.5	7.9	−25.2	9.0	−28.9	8.1
mean value	−25.4	9.7	−25.2	9.7	−31.5	7.6	−26.3	8.4	−30.1	7.5
standard deviation of data[e]	3.0	1.5	3.1	1.5	3.0	2.9	3.6	2.0	1.0	1.0
standard deviation of mean[f]	1.5	0.7	1.5	0.7	1.5	1.4	1.8	1.0	0.5	0.5
MAD[g]	–	–	0.2	0.0	6.1	3.0	–	–	3.8	1.0

[a] $\varepsilon = 1$. [b] Neglect of outer region. [c] $\varepsilon = 80$. [d] $\varepsilon = 80$, $\phi_{rf}^{QM} = 0$. [e] Standard deviation of individual energy values. [f] Standard deviation of the mean value (68% confidence limit). [g] Mean absolute deviation relative to full Coulombic electrostatics. For SMBP(solv,app), SMBP(solv) values are used as a reference.

between the distal oxygen atom and the meta carbon atom of pOHB (C_m).

$$\xi = d(O_d - O_p) - d(O_d - C_m) \tag{7}$$

The potential energy profile was scanned for RC values between −1.7 Å and +1.7 Å in steps of 0.1 Å. We applied the same statistical test as in the CM system to validate the boundary potentials for PHBH and determined the optimal mesh and basis set sizes. For the sake of brevity and clarity, the data are relegated to the SI (Tables S4−S6, Figure S1), and only the results are briefly summarized here. Analysis of the deviations of the gradient components (Table S5) indicates that deviations in PHBH are very similar to those in CM. The relationship of accuracy and radial position is also similar: deviations are very small in the center of the inner region and increase significantly only for coarse mesh sizes and only at the boundary separating the inner and outer regions (Figure S1). Therefore, a mesh size combination of 0.25/1.25 Å excels again as a safe choice that reproduces full Coulombic electrostatics very accurately in all parts of the inner region. However, coarser mesh sizes also reproduce the electrostatic potential accurately everywhere except in close proximity to the boundary. As in CM, potential energy differences are less sensitive to the mesh sizes: computations with mesh size combinations ranging from 0.25/1.25 Å to 1.8/3.5 Å yield accurate results with MAX deviations that do not exceed 0.3 kcal/mol for all mesh sizes (Table S7). In GSBP calculations, basis sets of similar size are adequate for CM and PHBH (Table S6).

On the basis of these results, we conclude that SMBP and GSBP behave very similarly with respect to their accuracy and its dependence on the parameters for these two enzymes, despite the different nature of the two reactions and the different QM methods used (AM1 and B3LYP, respectively). This substantiates our expectation that not only the validation protocol but also the resulting parameters are transferable to other enzymatic systems. We suggest a mesh size combination of 0.25/1.25 Å with multipole moments up to order $L = 19$ as a safe choice for accurate calculations. As default options, we recommend a mesh size combination of 0.6/1.25 Å with multipole moments up to order $L = 10$ for efficient application of SMBP and GSBP. Since even coarser mesh sizes reproduce energy differences with only marginal deviations, we are confident that these values yield accurate results at reduced computational costs for general enzymatic systems.

A mesh size combination of 0.25/1.25 Å was applied to compute the potential energy differences of the hydroxylation step in PHBH with the SMBP in solution and in a vacuum, with full Coulombic electrostatics, and with the simple NOR approach. We found it unavoidable to discard configuration 2, since we failed to compute continuous energy profiles for this configuration despite many attempts. The results in Table 7 show that the SMBP reproduces standard QM/MM results with high accuracy. The mean values of the reaction and activation energy deviate by only 0.2 and 0.0 kcal/mol. MAD values of similar magnitude show that the results for the individual configurations also agree well. In contrast to CM, the EPOM has a significant effect on the energetics of the hydroxylation reaction in PHBH. If the electrostatic influence of the outer region is simply neglected (NOR), the mean value of the reaction energy changes by more than 6 kcal/mol from −25.4 to −31.5 kcal/mol. The activation energy is reduced by more than 2 kcal/mol from 9.7 to 7.6 kcal/mol. MAD values of 6.1 and 3.0 kcal/mol for reaction and activation energies show that the deviations are even larger for the individual configurations.

A more detailed analysis reveals that two different effects are responsible for the observed differences in the reaction energies. In configurations 1 and 3, the hydrogen bonding networks connecting the QM and MM region are different after geometry optimization with full Coulombic electrostatics and with the NOR approximation. By contrast, in configurations 4 and 5, the hydrogen bonding networks between the QM and MM region are very similar for both electrostatic treatments. In these two configurations, the differences in the reaction energies are dominated by the differences in the electrostatic QM−MM interaction energies, as shown in Table S8. Here, the electrostatic QM−MM interaction energy includes direct QM−MM interactions as well as the polarization effect of the MM point charges. These results suggest that with the NOR approximation the MM atoms are more flexible in adapting to the electrostatic potential of the QM region since they do not feel the EPOM. This effect becomes more pronounced when the QM region becomes more polar. During the hydroxylation reaction in PHBH, the less polar peroxide group separates into more

polar alcohol and alcoholate groups, which form hydrogen bonds with the neighboring MM residues. Therefore, the difference in the QM−MM interaction energies of the product and reactant is greater with the NOR approximation compared to that with full Coulombic electrostatics. The NOR approximation favors the more polar product state and thus shifts the reaction energies to more negative values. In conclusion, we face two possible consequences of neglecting the EPOM: either geometry optimizations may lead to structures that already differ in the hydrogen bonding network or changes in the QM−MM interaction energies may bias reaction energetics significantly, even when the hydrogen bonding network is similar. These results underline the fact that the EPOM can have a significant influence on potential energy differences. They cast doubt on the quantitative accuracy of QM/MM and pure QM studies that neglect the EPOM.

In PHBH, we also observe a significant solvent effect on the reaction. When bulk solvent is modeled with the SMBP, the activation energy decreases by more than 1 kcal/mol from 9.7 to 8.4 kcal/mol. This agrees qualitatively with chemical reasoning: the bulk solvent stabilizes the charged OH^+ species and lowers the energy of the transition state. Given the distinct charge transfer in this reaction, the effect of bulk solvent that is observed in PHBH should be in the upper range of what can be expected in enzymatic reactions. Even stronger effects seem possible if the inner region is chosen to be rather small and the charge transfer thus occurs closer to the dielectric boundary.

The significant solvent effect renders this reaction an interesting test for the SMBP(solv,app) method, which neglects the QM contribution to the reaction field potential. A satisfying agreement with SMBP(solv) results can only be observed for configuration 1. For the other configurations, deviations reach up to 10 kcal/mol. The mean value for the reaction energy is −30.1 kcal/mol and therefore far outside the confidence interval of the SMBP(solv) mean value. In the case of PHBH, the QM contribution to the reaction field potential thus has a significant influence on the energetics of the hydroxylation reaction. Hence, the SMBP(solv,app) method is not a valid approximation in the PHBH system.

5. Conclusion

In this study, we evaluated the electrostatic effect of the outer macromolecule region and of bulk solvent on two enzymatic systems. We applied the newly developed SMBP, which allows us to model bulk solvent as a PDC and to distinguish the effects of bulk solvent and the EPOM. The SMBP introduces approximations to describe electrostatic interactions with the outer macromolecule region more efficiently. Therefore, the accuracy of the SMBP was evaluated for both enzymatic test systems, and a protocol for validation and determination of adequate values for its inherent parameters was presented. This protocol was applied to generate a set of optimal parameters that is transferable to general enzymatic systems. The SMBP was found to describe the EPOM with high accuracy. Typically, deviations of mean values on the order of 0.1 to 0.2 kcal/mol are observed for both enzymes. Deviations for individual configurations may be slightly higher but rarely exceed 0.3 kcal/mol.

Two enzymatic reactions with rather different characteristics were used to study the effect of the EPOM and bulk solvent. The Claisen rearrangement in CM is a pericyclic reaction without much charge transfer. For this kind of reaction, the electrostatic influence of the outer macromolecule region on the reaction energetics is not significant when considering the mean values of all 10 configurations. However, deviations on the order of 1.5 kcal/mol are observed for individual configurations so that the neglect of long-range electrostatics can be detrimental in the absence of adequate sampling. Bulk solvent effects on the reaction energetics in CM are found to be small.

The hydroxylation reaction in PHBH, in contrast, is associated with a stronger charge transfer since the reaction formally corresponds to an OH^+ transfer. As a consequence, the EPOM has a strong influence on the reaction energetics, and its neglect causes errors of several kilocalories per mole due to a systematic overstabilization of the more polar product state (arising from the higher flexibility of the MM residues without the EPOM). Moreover, bulk solvent stabilizes the transition state and reduces the reaction barrier by about 1 kcal/mol in PHBH.

Depending on the charge transfer characteristics of the chemical process, the EPOM and bulk solvent can thus have a significant effect on the energetics of enzymatic reactions. Among these two contributions, the EPOM is clearly more important. The SMBP offers a convenient way to evaluate both contributions accurately and efficiently in QM/MM calculations.

Acknowledgment. This work was supported by the Max Planck Initiative on Multiscale Materials Modeling and the Volkswagenstiftung. T.B. gratefully acknowledges a Kekulé scholarship from the Fonds der Chemischen Industrie. We thank J. Breidung and S. Thiel for CM and PHBH input data and A. Koslowski for his program to perform spline fits of energy profiles.

Supporting Information Available: Energy differences of the reactions in CM and PHBH with different mesh size combinations; free energy differences of the reaction in CM using full Coulombic electrostatics for two configurations; evaluation of the accuracy of the electrostatic forces for different basis set sizes using the GSBP for CM and PHBH; evaluation of the accuracy of the electrostatic forces with different mesh sizes for PHBH; contributions to the reaction energies of the hydroxylation reaction in PHBH; plot of MAD and MAX values of electrostatic forces as a function of the radial position for PHBH. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Struct.* **1999**, *28*, 155–179.

(2) Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211–217.

(3) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509–521.

(4) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186–195.

(5) Monticelli, L.; Simãţes, C.; Belvisi, L.; Colombo, G. *J. Phys.: Condens. Matter* **2006**, *18*, S329–S345.

(6) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

(7) Schreiber, H.; Steinhauser, O. *Biochemistry* **1992**, *31*, 5856–5860.

(8) Schreiber, H.; Steinhauser, O. *Chem. Phys.* **1992**, *168*, 75–89.

(9) Baumketner, A.; Shea, J.-E. *J. Phys. Chem. B* **2005**, *109*, 21322–21328.

(10) Patra, M.; Karttunen, M.; Hyvönen, M.; Falck, E.; Vattulainen, I. *J. Phys. Chem. B* **2004**, *108*, 4485–4494.

(11) Patra, M.; Karttunen, M.; Hyvönen, M.; Falck, E.; Lindqvist, P.; Vattulainen, I. *Biophys. J.* **2003**, *84*, 3636–3645.

(12) Ewald, P. *Ann. Phys.* **1921**, *369*, 253–287.

(13) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(14) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(15) Hünenberger, P. H.; McCammon, J. A. *Biophys. Chem.* **1999**, *78*, 69–88.

(16) Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.

(17) Kuwajima, S.; Warshel, A. *J. Chem. Phys.* **1988**, *89*, 3751–3759.

(18) Weber, W.; Hünenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.

(19) Esselink, K. *Comput. Phys. Commun.* **1995**, *87*, 375–395.

(20) Appel, A. W. *SIAM J. Sci. Stat. Comput.* **1985**, *6*, 85–103.

(21) Schmidt, K. E.; Lee, M. A. *J. Stat. Phys.* **1991**, *63*, 1223–1235.

(22) Ding, H. Q.; Karasawa, N.; Goddard, W. A., III. *J. Chem. Phys.* **1992**, *97*, 4309–4315.

(23) Berkowitz, M.; McCammon, J. A. *Chem. Phys. Lett.* **1982**, *90*, 215–217.

(24) Brooks, C. L., III; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312–6325.

(25) Brunger, A.; Brooks, C. L., III; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495–500.

(26) Essex, J. W.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 951–972.

(27) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173–290.

(28) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.

(29) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2–13.

(30) Gao, J.; Alhambra, C. *J. Chem. Phys.* **1997**, *107*, 1212–1217.

(31) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019–1031.

(32) Yarne, D. A.; Tuckerman, M. E.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 3531–3539.

(33) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176–1184.

(34) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2006**, *2*, 1370–1378.

(35) Friedman, H. L. *Mol. Phys.* **1975**, *29*, 1533–1543.

(36) Wang, L.; Hermans, J. *J. Phys. Chem.* **1995**, *99*, 12001–12007.

(37) Brunger, A.; Brooks, C. L., III; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 8458–8462.

(38) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100–3107.

(39) Alper, H.; Levy, R. M. *J. Chem. Phys.* **1993**, *99*, 9847–9852.

(40) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124–129.

(41) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

(42) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050–9063.

(43) Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924–2937.

(44) Banavali, N. K.; Im, W.; Roux, B. *J. Chem. Phys.* **2002**, *117*, 7381–7388.

(45) König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548–563.

(46) Ma, L.; Cui, Q. *J. Am. Chem. Soc.* **2007**, *129*, 10261–10268.

(47) Zhu, X.; Jethiray, A.; Cui, Q. *J. Chem. Theory Comput.* **2007**, *3*, 1538–1549.

(48) Riccardi, D.; König, P.; Prat-Resina, X.; Yu, H.; Elstner, M.; Frauenheim, T.; Cui, Q. *J. Am. Chem. Soc.* **2006**, *128*, 16302–16311.

(49) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905/1−14.

(50) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2008**, *4*, 1600–1609.

(51) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2009**, *5*, 3114–3128.

(52) Senn, H. M.; Kästner, J.; Breidung, J.; Thiel, W. *Can. J. Chem.* **2009**, *87*, 1322–1337.

(53) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *56*, 467–505.

(54) Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 2925–2929.

(55) Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.

(56) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483–3492.

(57) Sherwood, P.; et al. *THEOCHEM* **2003**, *632*, 1–28.

(58) ChemShell. http://www.chemshell.org (accessed August 14, 2010).

(59) Thiel, W. *MNDO2004*; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2004.

(60) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(61) Smith, W.; Forester, T. *J. Mol. Graph.* **1996**, *14*, 136–141.

(62) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(63) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.

(64) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterlig, W. T. Partial Differential Equations. In *Numerical Recipes in C*; Cambridge University Press: Cambridge, England, 1988; pp 673−680.

(65) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435–445.

(66) Im, W.; Seefeld, S.; Roux, B. *Biophys. J.* **2000**, *79*, 788–801.

(67) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(68) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.

(69) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(70) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(71) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(72) Elstner, M.; Porezag, D.; Jungnickel, G.; Elstner, J.; Haugk, M.; Frauenheim, T.; Suhai, T.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(73) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(74) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327–335.

(75) Kast, P.; Aisf-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691–2694.

(76) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(77) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(78) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(79) Senn, H. M.; Thiel, S.; Thiel, W. *J. Chem. Theory Comput.* **2005**, *1*, 494–505.

(80) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Chem.* **2006**, *45*, 6856–6859.

(81) Mata, R. A.; Werner, H.-J.; Thiel, S.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 025104/1−8.

CT1005455