JCTC Journal of Chemical Theory and Computation

QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction

Johannes Kästner, Hans Martin Senn, Stephan Thiel, Nikolaj Otte, and Walter Thiel*

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

Received October 14, 2005

Abstract: We used the free-energy perturbation (FEP) method in quantum mechanics/molecular mechanics (QM/MM) calculations to compute the free-energy profile of the hydroxylation reaction in the enzyme *p*-hydroxybenzoate hydroxylase (PHBH). *k* statistics were employed to analyze the FEP sampling including estimation of the sampling error. Various approximations of the free-energy perturbation method were tested. We find that it is adequate not only to freeze the density of the QM part during the dynamics at frozen QM geometry but also to approximate this density by electrostatic-potential-fitted point charges. It is advisable to include all atoms of a QM/MM link in the perturbation. The results of QM/MM-FEP for PHBH are in good agreement with those of thermodynamic integration and umbrella sampling.

I. Introduction

The free energy is the measure for the driving force of a chemical reaction. It can be calculated by a variety of methods. As bonds are broken and formed in chemical reactions, quantum mechanical methods are required for the calculation of free-energy differences. Standard electronic-structure methods provide the internal energy ΔU at zero temperature, $T = 0$ K. The Helmholtz free energy, $\Delta A = \Delta U - T\Delta S$, at a finite temperature also includes the entropy change ΔS. While an approximation for ΔS may be obtained from the harmonic frequencies of the system, it can be more accurately computed by sampling along a reaction coordinate.

In principle, the free-energy change along a reaction coordinate ξ may be calculated directly from the distribution function of ξ obtained from a molecular-dynamics (MD) simulation. Such a calculation may be accelerated by umbrella sampling (US).[1,2] This method applies a restraint (bias) to the reaction coordinate. In the limit of an infinitely strong bias, that is, a constraint, the method becomes equivalent[3] to thermodynamic integration (TDI).[4−7] The free-energy change may then be determined by integration of the mean force on this constraint. Both methods, thermodynamic integration and umbrella sampling, are based on an exhaustive sampling of the phase space. For large molecules, this is currently still impractical when using computationally demanding ab initio or density functional methods because of the prohibitive computational effort. This holds true even if such methods serve as QM components in QM/MM approaches where the reactive center is described by quantum mechanics (QM) and the environment by molecular mechanics (MM).

The sampling problem has been addressed by number of different approaches.[8−17] Here, we focus on a QM/MM-FEP treatment[8] that applies the free-energy perturbation (FEP) method[18] to QM/MM simulations (see ref 8 for a comparison with previously available approaches[9,11−14]). In QM/MM-FEP, only the computationally less demanding MM part is sampled, while the demanding QM part is kept frozen. Free-energy perturbation used in this manner includes the following approximations: (1) The entropy change within the QM part is not sampled but estimated from the harmonic approximation. (2) Commonly,[8,16,19−22] the density of the fixed QM system is not only frozen but approximated by electrostatic potential (ESP) charges when calculating its interaction with the MM part.

We have tested these approximations on a biological system, the enzyme *p*-hydroxybenzoate hydroxylase[23] (PHBH; EC 1.14.13.2). While natively catalyzing the transformation

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

10.1021/ct050252w CCC: $33.50
Published on Web 02/15/2006

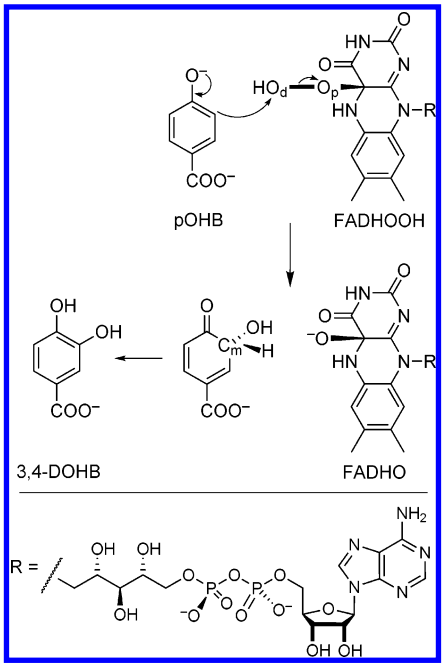

Figure 1. Schematic view of the rate-determining OH-transfer reaction catalyzed by the enzyme PHBH.

of *p*-hydroxybenzoate (pOHB) to 3,4-dihydroxybenzoate (3,4-DOHB), PHBH has also been proposed as a biocatalyst for the hydroxylation of halogenated pOHB derivatives.[24] During the catalytic cycle, the flavine cofactor (FAD: flavine−adenine dinucleotide) is reduced to $FADH_2$ by NADPH (nicotinamide−adenine dinucleotide phosphate, reduced form). It then reacts with molecular oxygen to form the flavin hydroperoxide (FADHOOH), shown in Figure 1. In what is believed to be the rate-determining step,[25−27] FADHOOH hydroxylates the substrate pOHB, yielding FADHO and a hydroxycyclohexadienone that tautomerizes rapidly to the aromatic 3,4-DOHB. FADHO is finally protonated to FADHOH, loses water, and regenerates the oxidized form FAD.

The rate-determining step is shown in Figure 1. The substrate, in its dianionic form,[25,26] is hydroxylated in an aromatic electrophilic substitution reaction. From temperature-dependent measurements of the overall rate, the activation energy was estimated as 49 kJ mol^{-1} at pH 8.0.[28] AM1 has been shown to overestimate the reaction barrier[29,30] but to yield fairly accurate structures, except for the underestimated peroxide O−O bond length.[29,31]

While the QM/MM-FEP method itself is not new,[8] the purpose of this work is to validate its approximations by calculating their influence on the resulting free-energy difference. We carefully estimate the effects of different treatments of the electrostatic QM/MM interaction. Problems occurring in the perturbation of link atoms and their solutions are also discussed. Moreover, we point out how to use *k* statistics to analyze the FEP results. The QM/MM-FEP method is designed to be applicable to QM/MM setups with demanding QM methods. However, to test the method, we employed the fast semiempirical AM1 Hamiltonian.[32] The methodological issues raised by QM/MM-FEP are expected to be similar for AM1 and higher-level QM methods, and the choice of AM1 allows us to investigate these issues efficiently and to assess various QM/MM-FEP approximations by comparisons against full QM calculations. Thus, the primary aim of this study is to test QM/MM-FEP on a real-world example, rather than to accurately reproduce the experimental activation barrier.

II. Methods

II.A. QM/MM Free-Energy Perturbation. In its original formulation,[18] free-energy perturbation is defined via an unperturbed Hamiltonian and a perturbation term ΔE_{pert}. Sampling ΔE_{pert} makes it possible to calculate the free-energy difference of the perturbation by exponential averaging: $\Delta A = -1/\beta \ln\langle\exp(-\beta\Delta E_{pert})\rangle$. $\langle x\rangle$ denotes a canonical average, $\beta = (k_B T)^{-1}$, and k_B is the Boltzmann constant. The term "perturbation" is somewhat misleading since the theory is exact and does not correspond to a perturbation ansatz in the usual sense. Applied to QM/MM simulations,[8] the unperturbed Hamiltonian corresponds to the QM/MM energy expression of a system where the QM atoms are fixed. The perturbation corresponds to a geometry step of the QM atoms. The phase space sampled is restricted to the degrees of freedom of the MM atoms. This allows one to calculate a free-energy profile in QM/MM simulations using demanding QM methods.

We briefly restate the formalism of QM/MM-FEP to point out the approximations used. The total energy of a QM/MM calculation may be written as

$$E_{total} = E_{qm}(\mathbf{r}_{qm}) + E_{qm/mm}(\mathbf{r}_{qm},\mathbf{r}_{mm}) + E_{mm}(\mathbf{r}_{mm}) \quad (1)$$

with E_{qm} depending on the coordinates of the QM atoms $\mathbf{r}_{qm}$, and E_{mm} depending on the coordinates of the MM atoms $\mathbf{r}_{mm}$. The term $E_{qm/mm}(\mathbf{r}_{qm},\mathbf{r}_{mm})$ includes all energy contributions coupling the QM and the MM parts:

$$E_{qm/mm} = E_{vdW} + E_Q + E_{FF} \quad (2)$$

that is, the van der Waals interaction E_{vdW}, the electrostatic interaction E_Q, and the force field terms E_{FF} of the junctions. The latter come from covalent bonds between QM and MM atoms. Within the electrostatic embedding scheme, the MM point charges polarize the QM part, and the electrostatic interaction between the QM and MM parts, E_Q, is therefore included in the energy provided by the QM code. Under the convention of eq 2, E_Q thus has to be calculated separately for obtaining $E_{qm/mm}$. From the QM electronic energy in the point-charge field, $\langle\Psi|\mathscr{H}|\Psi\rangle$, one obtains $E_{qm} = \langle\Psi|\mathscr{H}|\Psi\rangle - E_Q$.

We will divide our discussion of QM/MM-FEP into three steps: (1) calculation of an energy profile of the reaction using constrained optimizations, (2) calculation of the energy of the perturbation, ΔE_{pert}, and (3) sampling of ΔE_{pert}. While this separation is conceptually sensible, steps 2 and 3 are coupled in practical simulations.

II.A.1. Optimization. A reaction coordinate $\xi(\mathbf{r}_{qm})$ depending only on QM positions is defined. The reaction is split into discrete windows, each characterized by a value ξ_i. Constraining the reaction coordinate to some ξ_i, all other QM and MM degrees of freedom are optimized for each window *i*. This results in a set of minimum-energy geom-

etries and a profile of the internal energy of the reaction at zero temperature.

II.A.2. Perturbation. The optimized structures serve as unperturbed and perturbed structures in turn: When structure i is perturbed with structure $i + 1$, the energy of perturbation is given by

$$\Delta E_{\text{pert}}^{i\to i+1} = \underbrace{E_{\text{qm/mm}}(\mathbf{r}_{\text{qm}}^{i+1}, \mathbf{r}_{\text{mm}}^{i})}_{\text{perturbed}} - \underbrace{E_{\text{qm/mm}}(\mathbf{r}_{\text{qm}}^{i}, \mathbf{r}_{\text{mm}}^{i})}_{\text{unperturbed}} \quad (3)$$

Thus, the unperturbed energy is calculated with all atoms at their positions of the optimized window i. The perturbed energy is calculated with the MM positions of window i and the QM positions of window $i + 1$. Equation 3 defines the "forward perturbation". In the "backward perturbation", window $i + 1$ is perturbed with window i:

$$\Delta E_{\text{pert}}^{i+1\to i} = E_{\text{qm/mm}}(\mathbf{r}_{\text{qm}}^{i}, \mathbf{r}_{\text{mm}}^{i+1}) - E_{\text{qm/mm}}(\mathbf{r}_{\text{qm}}^{i+1}, \mathbf{r}_{\text{mm}}^{i+1}) \quad (4)$$

II.A.3. Sampling. The free-energy change between window i and window $i + 1$ is given by

$$\Delta A^{i\to i+1} \approx \Delta E_{\text{qm}}^{i\to i+1} + \Delta A_{\text{qm/mm}}^{i\to i+1} \quad (5)$$

$\Delta E_{\text{qm}}^{i\to i+1}$ is the difference of the QM energies between the windows i and $i + 1$. $\Delta A_{\text{qm/mm}}^{i\to i+1}$ incorporates the change in the free energy due to the QM/MM interactions as well as the MM part. It is obtained from sampling

$$\Delta A_{\text{qm/mm}}^{i\to i+1} = -\frac{1}{\beta} \ln\langle \exp(-\beta \Delta E_{\text{pert}}^{i\to i+1}) \rangle_{\text{mm},i} \quad (6)$$

The energy difference is sampled at window i, meaning that the MM atoms move according to the forces from the QM part of window i. The average is only taken over the MM coordinates, since the QM coordinates are always frozen in the MD simulations. Forward ($i \to i + 1$) and backward ($i + 1 \to i$) perturbation converge to the same energy difference, with opposite sign. As $\Delta E_{\text{qm/mm}}^{i\to i+1} + \Delta E_{\text{mm}}^{i\to i+1}$ is known from the reaction profile, the knowledge of $\Delta A_{\text{qm/mm}}^{i\to i+1}$ allows an estimate of the corresponding entropic contributions.

Values for $\Delta A_{\text{qm/mm}}^{i\to i+1}$ of a typical simulation, including the error bar defined by eq 17, are shown in Figure 2. Summation of the results of eq 5 provides $\Delta A(\xi)$ on a grid provided by the ξ_i values of the different windows. Minima and maxima are determined by interpolating three consecutive values of $\Delta A(\xi)$ with a second-order polynomial.

II.B. Entropic Effects of the QM Part. In eq 5, the entropy and finite-temperature effects in the QM part have been neglected. These can be taken into account by calculating the harmonic frequencies of the QM part and applying standard methods from statistical thermodynamics to evaluate the difference $\Delta A_{\text{qm}} - \Delta E_{\text{qm}}$ for the stationary points of interest (minima, transition states).[8,33]

II.C. Electrostatic Interaction in QM/MM-FEP. During the MD sampling, the structure of the QM part is kept frozen at the optimized geometry of either window i or $i + 1$ for the forward and backward perturbation, respectively. Instead of calculating E_{Q} from a density obtained from full self-consistent field (SCF) iterations in each MD step, one may introduce the approximation to freeze the density ρ, that is, to neglect changes in the polarization of the density caused by the varying MM coordinates during the MD run for a given window. Calculating E_{Q} from a fixed density requires evaluating one-electron integrals only, but no SCF iterations, and is thus computationally less expensive.

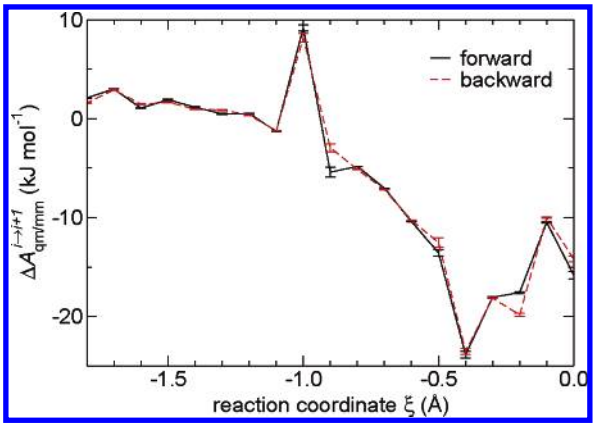

Figure 2. $\Delta A_{\text{qm/mm}}^{i\to i+1}$ obtained from forward perturbation and backward perturbation including the error bar. Differences between forward and backward perturbations are mainly caused by incomplete sampling. The spike at $\xi = -1.0$ Å is caused by the perturbation between two manifolds, see section III.A.

In a further approximation, charges which reproduce the electrostatic potential (ESP charges) are commonly used instead of the full density to calculate E_{Q}.[8,16,19−22] As the point charges should reproduce the energy and forces generated by the full density, ESP charges are well-suited. This allows one to completely avoid QM calculations in the sampling runs. When an accurate—but slow—QM method is used, the vast majority of the calculation time is, therefore, spent on obtaining the optimized structures.

We tested both approximations, fixed ρ and ESP charges, by comparison to the full SCF density using the fast AM1 method. In section III.C, we show that both approximations are well-justified, at least for the system under investigation.

II.D. Link Atoms and Their Perturbation. In our QM/MM setup,[34] covalent bonds between the QM and the MM parts, so-called junction bonds, are capped by link atoms. These are only treated by the QM code and are invisible to the MM code. The link atom is placed on the line connecting the QM atom and the MM atom of the junction bond at a constant distance from the former. Forces on the link atom are remapped to the QM and MM junction atoms. The stretching of the junction bond is described at the MM level. As the link atom is intrinsically of QM nature, its position is constrained in the MD sampling. We also freeze the MM atom of the junction.

To calculate $E_{\text{qm/mm}}(\mathbf{r}_{\text{qm}}^{i+1}, \mathbf{r}_{\text{mm}}^{i})$ in eq 3, the QM geometry is taken from window $i + 1$, while the MM geometry is the one from window i. The question of what to do with the link atoms arises. There are four possibilities to treat the position of the junction atoms, illustrated in Figure 3: (1) Only the QM atom is moved to its position in window $i + 1$; the link and the MM atom remain at their positions in window i. (2) The link atom is placed on the line connecting

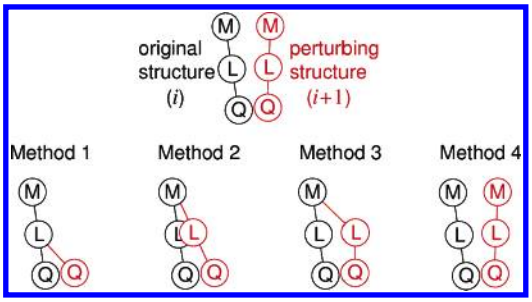

Figure 3. Four methods of perturbing link atoms. Q, L, and M refer to the QM atom, the link atom, and the MM atom of the junction, respectively.

the QM atom of window $i + 1$ and the MM atom of window i. The MM atom remains at its position in window i. (3) QM and link atoms are moved to their positions in window $i + 1$; the MM atom remains at i. (4) All three atoms are moved to their positions in window $i + 1$.

The drawback of methods 1 and 2 is that the link atom is moved from the position where its density has been calculated. In methods 1 and 3, the position of the link atom is inconsistent as it does not lie anymore on the line connecting the QM and MM junction atoms. Method 4 thus emerges as the most consistent and promising choice, although it involves the largest perturbation as all three atoms are moved.

II.E. System Setup. The enzyme PHBH was treated in analogy to our previous work.[29,35] The initial geometry was based on a crystal structure (see PDB file 1IUW).[36] The enzyme, consisting of 394 amino acids, 219 crystallographic water molecules, the FADHOOH cofactor, and fully deprotonated *p*-hydroxybenzoate, was solvated in a cubic water box. After a series of structure optimizations and MD runs, a production run was performed under periodic boundary conditions in the canonical (*NVT*) ensemble at $T = 300$ K with restraints acting on the cofactor and the substrate. A snapshot after 40 ps was used as the starting point for this study. All water molecules further than 11 Å from any protein atom were discarded. The water molecules in the outer solvation shell between 2.9 and 11 Å were kept rigid, and all atoms inside were left free. This resulted in 6245 protein atoms, 102 atoms of the cofactor and the substrate, and 2445 water atoms, all of which were free to move. In total, there are 22 772 atoms (8792 free and 13 980 fixed in the outer solvation shell).

The QM/MM setup was chosen as follows. The QM region consisted of 49 atoms: the substrate *p*-hydroxybenzoate (in its dianionic form) and the isoalloxazine part of the cofactor FADHOOH. There is one covalent bond between the QM and the MM parts, which was saturated by a H link atom. The QM part was described with the semiempirical AM1 Hamiltonian[32] and the protein environment with the GROMOS force field. We used the QM/MM approach as implemented in the ChemShell software package.[34] ChemShell provided the optimizer, the MD driver, and the interfaces to the MNDO99[37] and GROMOS96[38] codes. The ability to perform QM/MM-FEP calculations was implemented into ChemShell. The electrostatic interaction between the QM and the MM atoms was treated by including all MM point charges in the QM Hamiltonian. The charge-shift scheme[34] was applied at the junction. The MM-MM electrostatic interactions were evaluated explicitly for all atom pairs with a distance of up to 14 Å and approximated by a generalized Poisson−Boltzmann reaction field[39] with $\epsilon_r = 54.0$ beyond. The SCF convergence criterion was 10^{-8} eV.

The MD snapshot (see above) of the reactant was the starting point for structure optimizations. The whole system, that is, all atoms except the frozen outer solvation shell, was optimized in hybrid delocalized internal coordinates[40,41] (HDLC) to a convergence criterion for the maximum gradient component of $0.45 \times 10^{-3}\ E_h\ a_0^{-1}$, using a limited-memory quasi-Newton algorithm[42] (L-BFGS; BFGS: Broyden−Fletcher−Goldfarb−Shanno). This led to the reactant state. For the following transition-state search and optimizations, we defined an active region of about 3300 atoms, composed of the QM part and all residues with at least one atom within 15 Å of the substrate.

For the transition-state search, we chose a reaction core of nine atoms directly involved in the OH transfer. The OH group was manually displaced toward the expected transition state. The microiterative transition-state search in HDLCs proceeded as follows: L-BFGS steps were performed for all atoms of the active region except the reaction core until the environment was converged to within a maximum gradient component of $0.45 \times 10^{-3}\ E_h\ a_0^{-1}$. One partitioned rational function optimizer[43] (P-RFO) step was then performed for the reaction core, using an explicit Hessian. Continuing with the L-BFGS optimization, this process was iterated until the largest gradient component in the core was less than $1.35 \times 10^{-3}\ E_h\ a_0^{-1}$.

The difference of two bond lengths was defined as the reaction coordinate:

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

See Figure 1 for atom labeling. Note that the sign of ξ is opposite in ref 35. The transition state is at $\xi(\mathrm{TS}) = -0.41$ Å. A reaction profile in intervals of 0.1 Å was calculated to define the windows for the FEP simulations. The starting point was $\xi = -0.4$ Å, which was generated from the TS geometry by moving O_d by 0.01 Å along the reaction coordinate. In each of the windows, the reaction coordinate was constrained, and all other degrees of freedom in the active region were optimized. The profile ranged from $\xi = -1.8$ to $+1.8$ Å. At these optimized structures, ESP charges were calculated by fitting the potential at the positions of the 200 MM atoms nearest to the QM atoms.

The MD simulations were started from these optimized structures. The QM part, the first MM atom of the junction, and the outer solvation shell were frozen. All other 8742 atoms were equilibrated for 30 ps for the window at $\xi = -0.4$ Å. During the heating phase, the first 10 ps of this equilibration, a Berendsen thermostat[44] was used. In all other MD simulations, a canonical (*NVT*) ensemble at $T = 300$ K was generated by a Nosé−Hoover chain thermostat[45−48] with a chain length of 4 and a characteristic period of 20 fs, corresponding to a thermostat wavenumber of 375 cm^{-1}. Newton's equations of motion were integrated with a reversible noniterative leapfrog-type integrator[49] with a time

step of 1 fs. To ensure energy conservation at this time step, all hydrogen atoms were assigned the mass of deuterium and the free water molecules were kept internally rigid using SHAKE constraints.[50] To prepare the subsequent window, the QM part of the equilibrated window with $\xi = -0.4$ Å was replaced by the QM part with $\xi = -0.3$ Å. This system was again equilibrated for 10 ps. In this manner, all windows were equilibrated consecutively.

The FEP production runs were performed for the forward and the backward perturbation for 10 ps in each window, unless noted otherwise. Equilibration of the system with respect to ΔE_{pert} was tested as described elsewhere[51] by testing for the lack of a trend in the coarse-grained average and its variance, for normality, and for a lack of correlation. Whenever the tests for trend showed that ΔE_{pert} or its variance were not stationary over the whole range of the production run, MD steps from the beginning of the simulation were dropped until stationarity was reached. This was done separately for the forward and the backward perturbation data of each window. The range with stationary ΔE_{pert} was then used for the analysis.

To compare the results of FEP with other free-energy methods, we performed TDI and US simulations. Methodology and detailed results of TDI simulations have been reported previously.[35] The reaction coordinate was constrained in intervals of 0.1 Å. The force of constraint was sampled and integrated along ξ to compute the free energy. In the simulations for umbrella sampling, the constraint was replaced by a harmonic restraint of the form $w_i = K/2(\xi - \xi_i)^2$ with $K = 0.18\ E_h\ a_0^{-2}$. The values of ξ_i were chosen to be the same as in the TDI simulations. The structures resulting from the TDI sampling were used as starting structures for the US simulations. After a re-equilibration of 2 ps, which is necessary because the constraints were replaced by restraints, the system was sampled until the mean and the variance of ξ were trend-free according to the Mann−Kendall test[51,52] over at least 8 ps. The data of these 8 ps were then used for the analysis. We used the weighted histogram analysis method (WHAM)[53,54] as well as umbrella integration[3] to combine the different windows of the umbrella sampling simulations.

To test the approximation of the full QM density by ESP charges, we used not only AM1 but also density functional theory (DFT).[55,56] These calculations were done with the TURBOMOLE[57−61] code (version 5.7.1) interfaced to ChemShell, using the BP86 functional.[62−66] The DFT optimization was started from the AM1 geometry and carried out with the TZVP[67] basis set. At the optimized geometry, the self-consistent electron density was computed with the aug-cc-pVTZ[68,69] basis, which includes polarization and diffuse functions. ESP charges were fitted to this density to compare the forces on the MM atoms obtained from the density and the ESP charges.

II.F. Statistical Analysis of FEP Results. In free-energy perturbation, exponential averages of the form $\langle \exp(-\beta \Delta E_{\text{pert}}) \rangle$ have to be evaluated.[70] These are dominated by small values of ΔE_{pert}, which are poorly sampled. Thus, the result may strongly depend on the random occurrence of low values of ΔE_{pert} in the trajectory.

It is more efficient to use an expansion of $\langle \exp(-\beta \Delta E_{\text{pert}}) \rangle$ rather than the direct exponential average. $\langle \exp(-\beta \Delta E_{\text{pert}}) \rangle$ can be expressed as a cumulant expansion[18,71−74]

$$\ln\langle e^{-\beta \Delta E_{\text{pert}}} \rangle = \sum_{i=1}^{\infty} \frac{(-\beta)^i}{i!} \kappa_i \tag{8}$$

with the κ_i being the cumulants. They depend on the first and higher moments of the distribution of ΔE_{pert}.

In the case of an MD simulation, the distribution itself is not available; thus, the cumulants cannot be calculated directly. Only the sample values drawn from the distribution are available. The ith k statistic k_i is the unique symmetric unbiased estimator of the cumulant κ_i.[71,74] The first four k statistics are given[71,74] in terms of the sample size N, the sample mean $\langle \Delta E_{\text{pert}} \rangle$, and the sample central moments m_i, with $m_i = 1/N \sum_{j=1}^{N} (\Delta E_{\text{pert},j} - \langle \Delta E_{\text{pert}} \rangle)^i$:

$$k_1 = \langle \Delta E_{\text{pert}} \rangle \tag{9}$$

$$k_2 = \frac{N}{N-1} m_2 \tag{10}$$

$$k_3 = \frac{N^2}{(N-1)(N-2)} m_3 \tag{11}$$

$$k_4 = \frac{N^2[(N+1)m_4 - 3(N-1){m_2}^2]}{(N-1)(N-2)(N-3)} \tag{12}$$

For a normally distributed sample, all cumulants κ_i for $i \geq 3$ vanish. In general, the distribution of ΔE_{pert} taken from an MD simulation in a given window is very close to a normal distribution. (A special case where this is not true will be discussed in section III.C.) We, therefore, truncate the cumulant expansion after the second term. Only the mean and the variance of ΔE_{pert} are then required to calculate the estimate of $\langle \exp(-\beta \Delta E_{\text{pert}}) \rangle$. These, however, depend much less on the equilibration of the system than the direct average of $\exp(-\beta \Delta E_{\text{pert}})$ or the higher cumulants, which are more strongly influenced by rarely occurring small values of ΔE_{pert}.

To calculate an error bar for ΔA, we use the estimators of the variances of the k statistics:[71,74]

$$\widehat{\text{var}}(k_1) = \frac{k_2}{N} \tag{13}$$

$$\widehat{\text{var}}(k_2) = \frac{2N{k_2}^2 + (N-1)k_4}{N(N+1)} \tag{14}$$

In the special case of a normal parent distribution, the estimator of $\text{var}(k_3)$ is

$$\widehat{\text{var}}(k_3) = \frac{6N(N-1){k_2}^3}{(N-2)(N+1)(N+3)} \tag{15}$$

Thus, we use $-1/\beta \ln\langle \exp(-\beta \Delta E_{\text{pert}}) \rangle = \overline{\Delta A} \pm 2s$ as the confidence interval, approximating the Student t fractile at a confidence level of 95% by 2 for large N and estimating s^2 from error propagation.

$$\overline{\Delta A} = \langle \Delta E_{\text{pert}} \rangle - \frac{\beta}{2} k_2 \tag{16}$$

$$s^2 = \widehat{\text{var}}(k_1) + \frac{\beta^2}{4}\widehat{\text{var}}(k_2) + \frac{\beta^4}{6^2}\widehat{\text{var}}(k_3) \tag{17}$$

Note that this error measure only accounts for the statistical fluctuations of the MD run. It includes neither errors caused by incomplete sampling nor errors caused by the method itself, such as the choice of the QM region or the intrinsic accuracy of the QM or the MM method.

III. Results and Discussion

III.A. Structural Issues. In the reaction, the transfer of an OH group from the hydroperoxy group to *p*-hydroxybenzoate takes place. The stationary points that emerged from the structure optimizations are shown in Figure 4. In the transition state, the hydrogen atom of the OH group is stabilized by a hydrogen bond to the backbone amide oxygen of Pro293. During the stepwise structural optimizations along the reaction coordinate, this hydrogen bond remained present. In the case of the product state, a local minimum was found around $\xi = 1.0$ Å, see Figure 5. The hydrogen bond broke at the optimization for $\xi = 1.7$ Å with a distinct energy lowering. Stepwise backward optimization resulted in a second minimum around $\xi = 1.5$ Å, which is 8.5 kJ mol^{-1} lower than the first local minimum. In this case, a different hydrogen bond, OH···O_p, was formed. Further backward optimization increased the energy above the curve obtained with the OH···O(Pro293) hydrogen bond. The system thus shows a hysteresis. Note that all values given here refer to energies and reaction coordinates on the reaction profile and were obtained from calculations with constrained ξ values in intervals of 0.1 Å. An analogous behavior was found for the reactant state. In this case, H of the OOH group does not participate in any hydrogen bond in the more stable minimum.

The occurrence of such a hysteresis shows that the chosen reaction coordinate does not account for all structural changes during the reaction. The hydrogen bonds change but do not contribute to the reaction coordinate. Introducing the term "manifold" for the set of geometries with a given hydrogen-bond pattern, the transition state belongs to a different manifold [with an OH···O(Pro293) hydrogen bond] than the product (with OH···O_p) and the reactant state (no such hydrogen bond involving OH). In the FEP simulations, a change between these manifolds has to be accomplished.

In the case of the reactant state, the structural changes between the two manifolds are small, and FEP calculations between them remain possible. The change from the manifold with OH···O(Pro293) to the one of the reactant state occurs in the perturbation between $\xi_i = -0.9$ Å and $\xi_{i+1} = -1.0$ Å. This results in a spike in $\Delta A^{i\to i+1}_{\text{qm/mm}}$ at $\xi = -1.0$ Å, see Figure 2, which however does not lead to a noticeable discontinuity in $A(\xi)$, shown in Figure 6, due to compensating changes in $\Delta E^{i\to i+1}_{\text{qm/mm}}$, see eq 5.

Near the product state, the structural changes between the two manifolds are too large to be overcome in a single perturbation simulation, causing a high statistical error, as obtained from eq 17. Therefore, we used 10 intermediate windows with structures obtained from linear interpolation between the windows with $\xi_i = 0.6$ Å and $\xi_{i+1} = 0.7$ Å. With this choice, the perturbation runs converged.

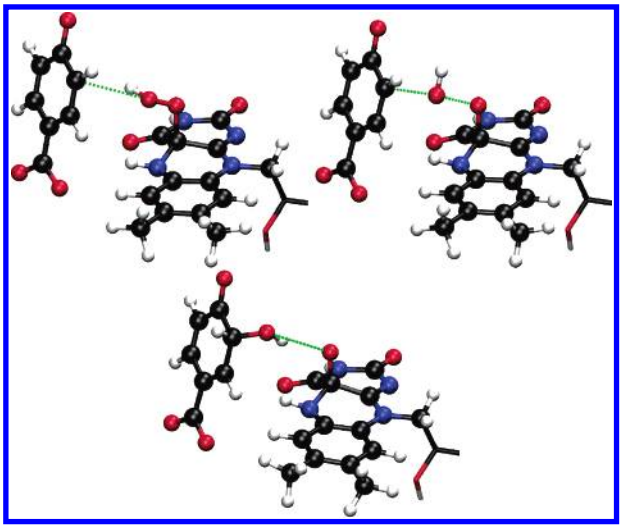

Figure 4. Reactant state, transition state, and product state of the OH-transfer reaction of PHBH. The substrate and the truncated cofactor are shown. The reaction coordinate is indicated by a dotted green line. The atoms included in the QM part are drawn as a ball-and-stick model.

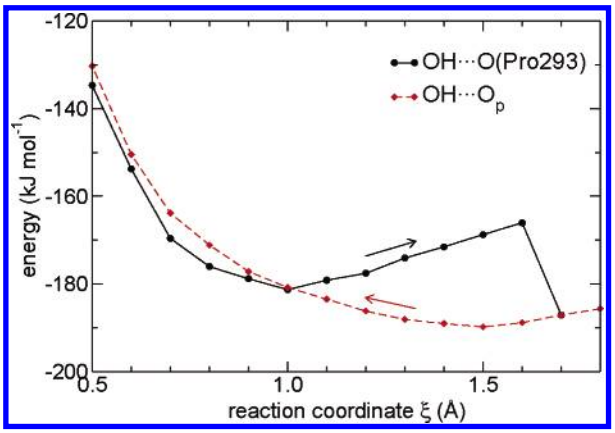

Figure 5. The two manifolds near the product state (see text). The one with a hydrogen bond OH···O(Pro293) leads to the transition state. The arrows indicate the direction in which the energy profile was calculated.

The reaction path between the reactant state and the transition state will be used in section III.C to discuss the approximations used in QM/MM-FEP. First, however, we compare different methods of free-energy sampling.

III.B. Comparison of Methods. We calculated the energy profile of the complete reaction with several methods. In the FEP calculations, the link atoms were perturbed with method 4, and the sampling was done with the full, but frozen, QM density.

Figure 6 shows the comparison between the energy profiles obtained by optimization, FEP, TDI, and US. The curves have been shifted in energy to match best in the reactant and the transition state. The stationary points obtained by each method are marked, and the corresponding numerical results are given in Tables 1−3. With the use of the harmonic

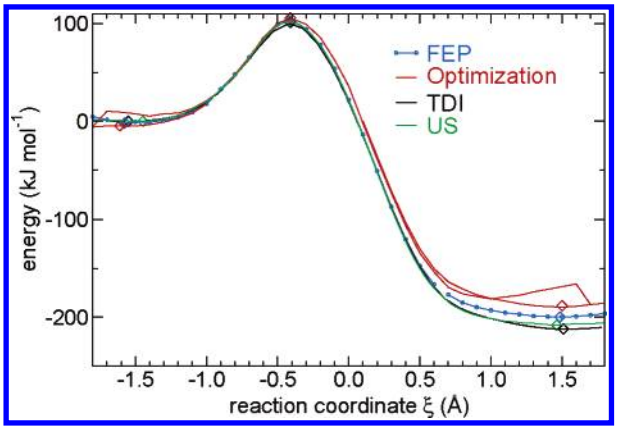

Figure 6. Energy profiles with different methods. Blue, FEP; red, optimization; black, TDI; and green, US. The two optimization curves represent two local minima near the product and the reactant (with different H bonds, see text). The gap in the FEP graph at $\xi = 0.6-0.7$ Å corresponds to the change between these two manifolds in the FEP calculations.

Table 1. Free Energies of Activation ($\Delta^{\ddagger}A$) and Reaction ($\Delta_r A$) in kJ mol^{-1} Obtained with Different Methods[a]

method	$\Delta^{\ddagger}A$	$\Delta_r A$
optimization	112.3	−184.3
FEP	108.2 ± 1.0	−198.6 ± 1.3
TDI[35]	101 ± 2	−212 ± 2
US	101.5	−208.1

[a] Thermal and entropic corrections for the QM region are included in the values given for FEP and optimization (harmonic approximation).

Table 2. Contributions to the Free-Energy Changes ΔA in kJ mol^{-1}[a]

	forward barrier	reaction energy
ΔA	108.2	−198.6
ΔE_{qm}	80.2	−350.3
$\Delta A_{qm/mm}$	26.1	152.4
$\Delta A_{qm} - \Delta E_{qm}$	1.9	−0.7
ΔE_{mm}	−10.8	−12.7
$\Delta A_{qm/mm} - \Delta E_{qm/mm} - \Delta E_{mm}$	−4.1	−14.3
ΔE_{qm}(ZPE)	5.8	1.1

[a] $\Delta A_{qm/mm}$ was calculated by FEP; ΔE_{qm}, $\Delta E_{qm/mm}$, and ΔE_{mm} were calculated by optimization; and $\Delta A_{qm} - \Delta E_{qm}$ as well as ΔE_{qm}(ZPE) were calculated from the harmonic frequencies of the QM part at the stationary points.

Table 3. Reaction Coordinate at the Reactant State, the Transition State, and the Product State (Å) Obtained with Different Methods

method	ξ(RS)	ξ(TS)	ξ(PS)
optimization	−1.61	−0.41	1.50
FEP	−1.58	−0.42	1.49
TDI[35]	−1.55	−0.41	1.51
US	−1.45	−0.42	1.46

approximation, the thermal and entropic contributions of the QM region have been included in the values for optimization and FEP in Table 1 (but not in Figure 6).

The TDI and US approaches are expected to yield the same free-energy changes because they both sample the entire system. The differences between TDI and US results are indeed very small (Table 1) and are most probably caused by incomplete sampling. US was analyzed by umbrella integration.[3] WHAM analysis of the umbrella sampling data leads to a range of 100.1−102.3 kJ mol^{-1} for the activation barrier and −205.7 to −209.4 kJ mol^{-1} for the reaction energy, depending on the number of bins used for the analysis.

In the FEP approach, the thermal and entropic contributions to the free energy are evaluated in harmonic approximation for the QM region ($\Delta A_{qm} - \Delta E_{qm}$) and are determined by sampling the environment ($\Delta A_{qm/mm} - \Delta E_{qm/mm} - E_{mm}$). It can be seen from Table 1 that the FEP results are close to the TDI and US reference values, which supports the validity of the FEP approximations for the QM and MM regions (see above). It is obvious from Table 1 that FEP accounts for some but not all of the differences between optimization on one hand and TDI or US on the other hand.

Table 2 lists the individual contributions to the FEP free-energy changes ΔA, which are the sums of the three following energies (lines 2−4). The contribution ΔE_{qm} from the QM energy is dominant, and the term $\Delta A_{qm/mm}$ is also substantial. The thermal and entropic contributions from the QM region ($\Delta A_{qm} - \Delta E_{qm}$) are small compared to those from the environment ($\Delta A_{qm/mm} - \Delta E_{qm/mm} - \Delta E_{mm}$), emphasizing the importance of including the latter (as done in FEP). These QM/MM thermal and entropic contributions lower the barrier by 4.1 kJ mol^{-1} and make the reaction more exergonic by 14.3 kJ mol^{-1} (see Table 2), which may be related to the changes in the hydrogen bond network during the reaction.[35] Finally, it should be pointed out that zero-point vibrational corrections have not been applied to the results given (Table 1). Such corrections are not available for TDI or US but can be deduced for FEP at least for the QM region from the computed QM Hessian: the corresponding ΔE_{qm}(ZPE) values are fairly small in the present case (Table 2).

The values of the reaction coordinate are listed in Table 3. The curvature of the energy surface around the transition state is rather high; thus, ξ(TS) is nearly independent of the method. In the reactant and product states, the energy surface is flat because the reaction coordinate contains a distance which does not correspond to a chemical bond. This causes larger variations in the reaction coordinate for the minima.

III.C. Test of Approximations. *III.C.1. Frozen Density.* The use of free-energy perturbation as implemented here requires freezing the QM geometry during the perturbation sampling. We assume that it is adequate to neglect polarization of the QM density in response to the moving MM atoms in the environment. To assess this assumption, we also sampled the system with full SCF iterations. With the frozen density, we obtain a free-energy barrier $\Delta^{\ddagger}A = 106.3 \pm 0.99$ kJ mol^{-1}, while we obtain $\Delta^{\ddagger}A = 103.8 \pm 0.99$ kJ mol^{-1} for full SCF iterations, see Table 4. The difference is not negligible but would seem to be tolerable in practice, especially in view of the computational savings: SCF iterations are avoided and only one-electron integrals have to be evaluated.

Table 4. Reaction Coordinate at the Reactant State and the Transition State (Å) and Forward Free-Energy Barrier (kJ mol^{-1}) for Different Approximations Used in FEP

density	link method	ξ(RS)	ξ(TS)	$\Delta^{\ddagger}A$
frozen	4	−1.578	−0.414	106.3 ± 0.99
full SCF	4	−1.501	−0.423	103.8 ± 0.99
ESP	4	−1.564	−0.431	105.7 ± 1.18
frozen	1	−1.580	−0.416	106.5 ± 0.82
frozen	2	−1.577	−0.415	105.3 ± 0.82
frozen	3	−1.576	−0.415	105.0 ± 0.82
frozen	4	−1.578	−0.414	106.3 ± 0.99

III.C.2. Density Replaced by Point Charges. For a frozen density, the computational demands can be further reduced if $E_Q(\mathbf{r}_{qm},\mathbf{r}_{mm})$ is calculated from point charges approximating the density and not from the full density itself. This reduces the sampling to a pure MM simulation from the computational point of view and, thus, significantly reduces the computation effort. Re-equilibration of the system previously sampled with the full density was necessary after switching to point charges, as this changes the gradient of the MM atoms near the QM region. We re-equilibrated each window for 20 ps and also used production runs of 20 ps. The energies obtained with ESP charges are very close to those obtained with the full density. We obtain $\Delta^{\ddagger}A = 106.3 \pm 0.99$ kJ mol^{-1} with the full density and 105.7 ± 1.18 kJ mol^{-1} when calculating all electrostatic interactions from ESP charges, see Table 4.

In an alternative approach, one might consider calculating ΔE_{pert} from the full frozen density and the forces for the dynamics from ESP charges. Testing this approach is sensible as the evaluation of the energy is much less demanding than the evaluation of the gradient. However, numerical problems occur that can be rationalized by consideration of the theoretical basis of FEP: The exponential average of ΔE_{pert} only yields the free-energy difference when sampled over the canonical ensemble.[70] If the dynamics, that is, the averaging, is performed with ESP charges while ΔE_{pert} is still calculated from the full frozen density, the ensemble does not match with ΔE_{pert}. Thus, it is not advisable to use different expressions of E_Q for the dynamics and the perturbation.

The ESP charges are generally able to reproduce the multipoles of the QM density very well. Thus, forces on distant MM point charges are essentially correct. If, however, the point charges penetrate into the QM density, they are partially shielded, and their gradients (forces) differ from those obtained from the ESP charges. In Figure 7, the differences between the absolute values of the electric fields caused by the ESP charges, $|\vec{E}_{ESP}|$, and the density, $|\vec{E}_\rho|$, at the position of the MM atoms of the window with $\xi_i = -0.4$ Å are shown. The electrostatic force on the MM atom i with charge Q_i is obtained as $\vec{F}_i = Q_i\vec{E}$.

As a more spatially extended density is expected to lead to a more pronounced shielding, we calculated the effect not only for AM1 but also for BP86/aug-cc-pVTZ, see Figure 7. The values are obtained from geometries with $\xi_i = -0.4$ Å optimized with AM1 and BP86/TZVP, respectively. It can be seen from Figure 7 that the extended basis set, which includes diffuse functions, leads only to slightly larger errors in the electrostatic forces than the minimum-basis set of AM1. Thus, we expect that ESP charges are also a good approximation for extended basis sets.

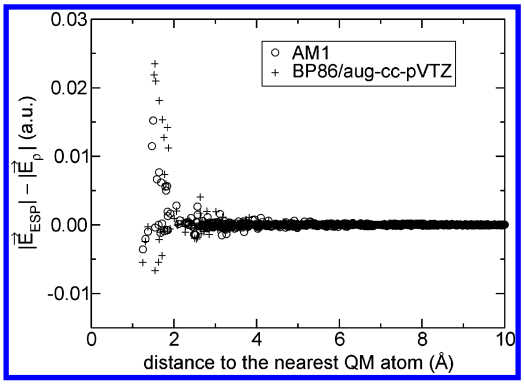

Figure 7. Differences between the absolute values of the electric fields caused by the ESP charges, $|\vec{E}_{ESP}|$, and the density, $|\vec{E}_\rho|$. The difference is evaluated at the positions of the MM atoms.

QM/MM-FEP with full SCF iterations considers all energy contributions that enter the QTCP-U approach of Rod and Ryde.[22] Our finding that ESP charges are a good approximation is in agreement with their work.

III.C.3. Link Atoms. The method used to treat the link atoms does not influence the results appreciably, as seen from Table 4 (last four lines). There is only one junction between the QM part and the MM part in our system; thus, PHBH may not represent a severe test for the link atom treatment. However, its influence on the free energy is so small that we do not expect a significant effect even if the system contains more link atoms. The simulations have been done with the full frozen density. The differences between the simulations are caused only by the link atom treatment as we sampled along the same trajectories. This was achieved by starting from the same configuration and velocity distribution. The link atom treatment only affects the perturbation, but not the dynamics. A comparison of the geometry after the simulation verified that the same trajectories had been sampled. Because of the small effect of the link atom treatment, see Table 4, formal aspects (section II.D) recommend method 4.

IV. Conclusion

We used the example of PHBH to show that QM/MM-FEP is a reliable and efficient method to calculate reaction free energies and free-energy barriers. We tested several methodological approximations on the system PHBH. Freezing the density during the FEP sampling caused an error of 2.5 kJ mol^{-1} in the computed barrier. This error was even smaller, 1.9 kJ mol^{-1}, when the frozen density was approximated by point charges. Different choices of the link-atom treatment altered the barrier by up to 1.5 kJ mol^{-1}. These numbers should be judged in the light of a typical sampling error of 1.0 kJ mol^{-1} in our simulations. We find that it is advisable to perturb all three atoms of a QM/MM junction (method 4) and that it is adequate to approximate the QM density by ESP charges in the FEP sampling.

When ESP charges are used, the computer time required for the geometry optimizations exceeds the time required for the MD simulations when using a demanding QM method. Thus, QM/MM-FEP is affordable at any level of QM theory where one can afford the geometry optimizations. The choice of the reaction coordinates is only limited by the optimizer. For the HDLC optimizer used in this study, it is straightforward to implement any linear combination of internal coordinates as a constraint. Intrinsically, constraints are unnecessary in the MD simulation, with the exception of freezing the Cartesian coordinates of the QM part and the first MM atom of each junction. This junction atom should then be included in the calculation of the Hessian for the harmonic approximation of the QM entropy.

Acknowledgment. This work was supported by the Volkswagenstiftung, Grant I/80454, and the Deutsche Forschungsgemeinschaft, Grant SFB 663/C4.

References

(1) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578.

(2) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.

(3) Kästner, J.; Thiel, W. *J. Chem. Phys.* **2005**, *123*, 144104.

(4) den Otter, W. K.; Briels, W. J. *J. Chem. Phys.* **1998**, *109*, 4139.

(5) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(6) Schlitter, J.; Klähn, M. *J. Chem. Phys.* **2003**, *118*, 2057.

(7) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737.

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

(9) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293.

(10) Strajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333.

(11) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049.

(12) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154.

(13) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.

(14) Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448.

(15) Kuhn, B.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 2586.

(16) Donini, O.; Darden, T.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 12270.

(17) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72.

(18) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(19) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384.

(20) Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2000**, *122*, 6560.

(21) Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* **2005**, *94*, 138302.

(22) Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* **2005**, *1*, 1240.

(23) Entsch, B.; van Berkel, W. J. *FASEB J.* **1995**, *9*, 476.

(24) Jadan, A. P.; Moonen, M. J. H.; Boeren, S.; Golovleva, L. A.; Rietjens, I. M. C. M.; van Berkel, W. J. H. *Adv. Synth. Catal.* **2004**, *346*, 367.

(25) Entsch, B.; Ballou, D. P.; Massey, V. *J. Biol. Chem.* **1976**, *251*, 2550.

(26) Entsch, B.; Palfey, B. A.; Ballou, D. P.; Massey, V. *J. Biol. Chem.* **1991**, *266*, 17341.

(27) Husain, M.; Entsch, B.; Ballou, D. P.; Massey, V.; Chapman, P. J. *J. Biol. Chem.* **1980**, *255*, 4189.

(28) van Berkel, W. J. H.; Müller, F. *Eur. J. Biochem.* **1989**, *179*, 307.

(29) Billeter, S. R.; Hanser, C. F. W.; Mordasini, T. Z.; Scholten, M.; Thiel, W.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2001**, *3*, 688.

(30) Ridder, L.; Mulholland, A. J.; Vervoort, J.; Rietjens, I. M. C. M. *J. Am. Chem. Soc.* **1998**, *120*, 7641.

(31) Ridder, L.; Mulholland, A. J.; Rietjens, I. M. C. M.; Vervoort, J. *J. Mol. Graphics Modell.* **1999**, *17*, 163.

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

(33) Atkins, P. W.; de Paula, J. *Physical Chemistry*, 7th ed.; Oxford University Press: Oxford, U. K., 2002; Chapter 19.

(34) 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. *THEOCHEM* **2003**, *632*, 1.

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

(36) Gatti, D. L.; Entsch, B.; Ballou, D. P.; Ludwig, M. L. *Biochemistry* **1996**, *35*, 567.

(37) Thiel, W. *MNDO99*, version 6.1; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2004.

(38) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf and BIOMOS b.v.: Zürich, Switzerland; Groningen, The Netherlands, 1996.

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

(40) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192.

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

(42) Liu, D. C.; Nocedal, J. *Math. Prog. B* **1989**, *45*, 503.

(43) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *J. Phys. Chem.* **1985**, *89*, 52.

(44) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(45) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

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

(47) Nosé, S. *Mol. Phys.* **1984**, *52*, 255.

(48) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511.

(49) Jang, S.; Voth, G. A. *J. Chem. Phys.* **1997**, *107*, 9514.

(50) Ryckaert, J.-P.; Ciccottiand, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(51) Schiferl, S. K.; Wallace, D. C. *J. Chem. Phys.* **1985**, *83*, 5203.

(52) Mann, H. B. *Econometrica* **1945**, *13*, 245.

(53) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011.

(54) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40.

(55) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

(56) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

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

(58) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652.

(59) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119.

(60) Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *118*, 9136.

(61) Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346.

(62) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(63) Dirac, P. A. M. *Proc. R. Soc. London, Ser. A* **1929**, *123*, 714.

(64) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.

(65) Slater, J. C. *Phys. Rev.* **1951**, *81*, 385.

(66) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(67) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

(68) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(69) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(70) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.

(71) Kenney, J. F.; Keeping, E. S. *Mathematics of Statistics*; Van Nostrand: Princeton, NJ, 1951; Vol. 2.

(72) Marcinkiewicz, J. *Math. Z.* **1939**, *44*, 612.

(73) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.* **2003**, *119*, 3559.

(74) Weisstein, E. W. k-Statistic; MathWorld-A Wolfram Web Resource. http://mathworld.wolfram.com/k-Statistic.html (accessed Aug 26, 2005).

CT050252W