

Enzymatic Hydroxylation in *p*-Hydroxybenzoate Hydroxylase: A Case Study for QM/MM Molecular Dynamics

Hans Martin Senn,* Stephan Thiel, and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany

Received December 21, 2004

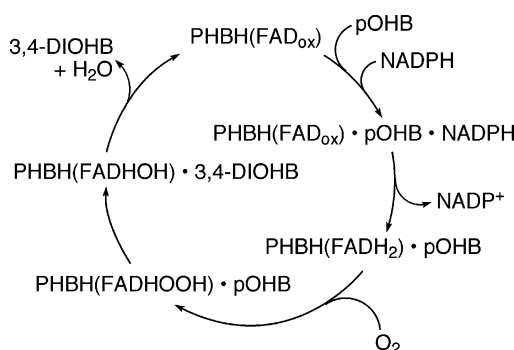
Abstract: We investigate the OH transfer step of the hydroxylation reaction of *p*-hydroxybenzoate in the enzyme *p*-hydroxybenzoate hydroxylase (PHBH) using QM/MM molecular dynamics methods. The QM region (49 atoms) is treated at the AM1 level, while the MM part (ca. 23 000 atoms) is described by the GROMOS force field. Performing pointwise thermodynamic integration from 10 starting structures, we have obtained an average value of the free-energy barrier for this reaction of 101 kJ mol⁻¹. The simulations provide insight into the dynamics of the hydrogen bonding network in the active site along the course of the reaction. In addition, we describe statistical techniques to analyze molecular dynamics data that assess the convergence of averages and yield an error measure. We discuss the effect of different error sources on the free energy.

I. Introduction

The enzyme *p*-hydroxybenzoate hydroxylase¹ (PHBH, EC 1.14.13.2) catalyzes the transformation of *p*-hydroxybenzoate (pOHB) to 3,4-dihydroxybenzoate (3,4-DOHB, also known as protocatechuate). It is active, for example, in soil bacteria where it plays a crucial role in the oxidative degradation of aromatic compounds, such as lignin, 3,4-DOHB being the substrate for subsequent catechol ring-cleavage reactions. PHBH has recently also been proposed as a biocatalyst for the hydroxylation of fluorinated and chlorinated pOHB derivatives.² The enzyme contains flavin-adenine dinucleotide (FAD) as a cofactor. The oxygen of the transferred hydroxy group stemming from O₂, PHBH is a flavoprotein classified as a monooxygenase. The second oxygen atom is not utilized for product formation but is ultimately reduced to H₂O.

The catalytic cycle may be divided into a reductive and an oxidative branch, as assessed by the oxidation state of the cofactor (Scheme 1). In the reductive half-reaction, PHBH with oxidized FAD binds the substrate pOHB as well as the two-electron reductant NADPH (reduced nicotinamide-adenine dinucleotide phosphate). The flavin cofactor is reduced by NADPH to FADH₂, and NADP⁺ is released. In

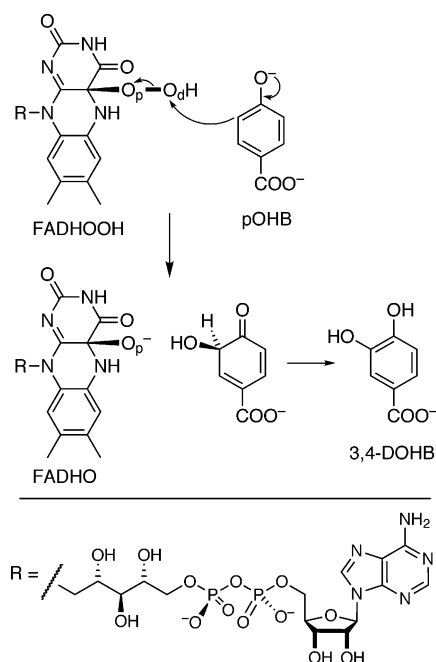
Scheme 1. Schematic Catalytic Cycle for pOHB Hydroxylation by PHBH



the subsequent oxidative half-reaction, FADH₂ reacts with molecular oxygen to form a flavin hydroperoxide (FAD-HOOH) that acts as the active hydroxylation agent. FAD-HOOH hydroxylates pOHB in the meta position, yielding FADHO (FADHOH after protonation) and a hydroxycyclohexadienone intermediate that tautomerizes rapidly into the aromatic 3,4-DOHB (Scheme 2). To close the cycle, FAD-HOH loses water, regenerating the oxidized form of FAD, and 3,4-DOHB is liberated.

PHBH has been the subject of numerous experimental investigations over the last almost three decades and has become a paradigm for flavoprotein monooxygenases; we

* Corresponding authors e-mail: thiel@mpi-muelheim.mpg.de (W.T.) and senn@mpi-muelheim.mpg.de (H.M.S.).

Scheme 2. Hydroxylation Step

refer here to the review by Entsch and van Berkel.¹ As our focus in the present study lies exclusively on the hydroxylation step, we summarize here pertinent experimental results from structural, mechanistic, kinetic, and mutation studies. (1) The hydroxylation reaction follows the aromatic electrophilic substitution mechanism, that is, FADHOOH acts formally as an “OH⁺” donor.³ (2) The substrate reacts in its dianionic (phenolate) form,^{4,5} which concurs with (1), the oxido substituent being strongly activating and ortho directing in S_EAr reactions. The pK_a of the phenolic proton in enzyme-bound pOHB is lowered relative to free pOHB.^{5,6} (3) A number of specific interactions with active-site residues stabilize the pOHB dianion and contribute to its binding and activation.^{5–8} Both the dianionic substrate and the oxido flavin resulting from OH transfer are stabilized by a positive electrostatic potential in the active site.⁹ (4) The different stages of the reaction are coupled with (at least) three distinct conformational states of the flavin, which on one hand makes the active site accessible for substrate binding and interaction of the oxidized FAD with NADPH and on the other hand shields the active hydroperoxy-FAD from reacting with solvent water once it is formed.^{10–12} (5) The relative rates of the different reaction steps, and thus the rate-determining step, appear to depend on the external conditions. Under very mildly acidic conditions (pH 6.5) at low temperature (275–279 K), the rate constants for hydroxylation and loss of H₂O from FADHOH are around 47 s^{–1} and 14 s^{–1}, respectively,^{4,5,13} that is, hydroxylation is not rate-limiting. The overall turnover rate of the enzyme under these conditions⁵ is around 6 s^{–1}. For more basic conditions, the overall turnover rate was determined as 7.5 s^{–1} (pH 8.6, 277 K),⁹ 44 s^{–1} (pH 7.9, 298 K),⁵ and 55 s^{–1} (pH 8.0, 298 K).¹⁴ Enzyme activity is highest in the pH range 7.5 to 8.5.¹⁵ From temperature-dependent measurements of the overall rate, the activation energy was obtained as 49 kJ mol^{–1} at pH 8.0.¹⁵

Also from the theoretical side the hydroxylation step of the PHBH catalytic cycle has found considerable attention

and has become a kind of prototype system for computational studies on enzyme reactions. An early study found a good correlation of experimental turnover rates of fluoro-substituted pOHB derivatives with the energies of the highest occupied molecular orbital (HOMO) calculated for the free substrates in the gas phase at the semiempirical AM1 level.¹⁶ This observation supported the notion that the hydroxylation is rate-determining under the conditions of the measurements¹⁴ (pH 8.0, 298 K). It also provided evidence for the dianionic state of the active substrate as well as for the S_EAr mechanism. A series of contributions by Ridder et al.^{17–20} realized a considerable methodical step forward by using a combined molecular-mechanical/quantum-mechanical (QM/MM) approach. Applying AM1 to the substrate and the isoalloxazine part of FADHOOH and the CHARMM force field to the remaining part of the cofactor as well as for the whole protein including crystal water, they calculated reaction profiles along a predefined reaction coordinate describing the attack of the hydroperoxy-“OH⁺” on the meta position of the substrate, optimizing at each point along the reaction coordinate the atoms within a given distance of the active region. The activation barriers thus obtained correlated again well with the experimental turnover rates for fluoro-substituted pOHB derivatives¹⁷ and with experimental rate constants of the hydroxylation step³ for a series of modified flavins and substrates,¹⁹ corroborating the S_EAr mechanism and the dianionic state of the substrate. These conclusions were strengthened by analyzing in more detail¹⁸ the structural and electronic changes and interactions with specific residues along the reaction path. The overall effect of the surrounding protein on the energetics of the reaction was found to be relatively small and the dominant contributions to stem from the QM part. The role of the enzyme in this case thus appears to consist mainly in activating the substrate by deprotonation, preparing the “OH⁺” donor FADHOOH, and keeping the two reactants in a position favorable for reacting. The latest contribution²⁰ improved upon the computational level, employing HF/6-31G(d) instead of AM1 for the QM part in the reaction-profile calculations. In addition, single-point calculations were done on the isolated QM part in the gas phase with B3LYP/6-311+G(d, p) and LMP2/6-31+G(d). The results largely confirmed the conclusions obtained before with regard to the mechanism and the role of the protein.

The barrier calculated for the hydroxylation of pOHB is 73 kJ mol^{–1} with AM1/CHARMM.¹⁷ HF/CHARMM provides a considerably higher value (about 125 kJ mol^{–1}), while the single-point calculations on HF/CHARMM geometries yield 51 and 47 kJ mol^{–1} with B3LYP and LMP2, respectively.²⁰ They provide a value of about –230 kJ mol^{–1} for the reaction energy. The comparison of different methods against gas-phase calculations up to MP2/6-31++G(d, p)//HF/6-31++G(d, p) on model reactions for the PHBH hydroxylation showed that HF drastically overestimates the barrier height, whereas B3LYP underestimates it; AM1 also overestimates, by about the same amount as B3LYP is too low.²¹ Structurally, AM1 is fairly accurate, except for the length of the peroxide O–O bond, which is underestimated.^{18,21} Together with the finding that AM1/CHARMM reproduces also the H-bonding patterns obtained with HF/

CHARMM,²⁰ AM1 can be considered sufficiently accurate and reliable for the reaction under study, at significantly lower computational cost than DFT or correlated ab initio methods.

A slightly different methodological approach was taken by Billeter et al.,²¹ who combined AM1 with the GROMOS force field to optimize the geometry of reactant and product states and the transition state for the entire enzymatic system. In addition, they ran molecular dynamics (MD) simulations at the same level of theory, performing thermodynamic integration (TI) along the $C_{\text{meta}}-\text{O}_d$ distance as the reaction coordinate. They obtained 89 kJ mol⁻¹ for the (static) potential energy barrier and 49 kJ mol⁻¹ for the free-energy barrier. The corresponding reaction energies were -286 kJ mol⁻¹ and -211 kJ mol⁻¹, respectively.

The purpose of the present contribution is 2-fold. First, we have investigated the hydroxylation step in PHBH with more extensive MD simulations and thermodynamic integration calculations. While retaining AM1/GROMOS, we have enlarged the model by including a shell of water molecules around the protein. To account better for changes in the protein environment, we have performed a series of simulations starting from several snapshots derived from a classical MD run. We have also employed a more accurate reaction coordinate. Second, we have used these calculations as a case study for clarifying a number of methodological and procedural issues that arise in the context of QM/MM MD simulations. We have focused on well-defined criteria for monitoring and establishing equilibration and stationary averages as well as on a careful identification and quantification of errors.

II. Methods

A. System Definition and Preparation. The setup of the PHBH system was based on the one described in ref 21. The system was prepared by means of fully classical MD simulations using GROMOS96.²² The GROMOS87²³ force field parameters (set 37C4 with SPC water) were used, augmented by aliphatic hydrogen atoms lacking van der Waals and electrostatic interactions (which are absorbed into the atom the H is bound to). This setup facilitates QM/MM coupling as it allows one to define and change the QM region without modifying the MM topology. The enzyme (394 amino acids), including 219 crystallographic water molecules, FADHOOH, and dianionic pOHb (7004 atoms in total), was placed in a cubic box ($a = 9$ nm) of water molecules. The whole system was subjected to a series of minimizations and MD runs (periodic boundary conditions, *NVT*, 300 K) with gradually decreasing harmonic positional restraints acting on the nonwater atoms. Finally, the restraints were released except for the cofactor and the substrate (102 atoms), which were kept restrained with $k = 1.046 \times 10^4$ kJ mol⁻¹ nm⁻². With this setup, an MD run of 200 ps was performed, from which snapshots were taken every 40 ps.

In each snapshot, a solvation shell was defined by the water molecules located between 0.29 and 1.1 nm from any protein atom. Water molecules further away were discarded; those closer to, or inside, the protein were left free. The water molecules in the shell were kept fixed, forming a rigid

Table 1. Composition of the Snapshots

snapshot (ps)	no. of atoms				total
	protein	FADHOOH + pOHb	free water	fixed water	
40	6245	102	2445	13 980	22 772
80	6245	102	2556	13 863	22 766
120	6245	102	2625	14 016	22 988
160	6245	102	2619	14 001	22 967
200	6245	102	2598	14 136	23 081

enclosing wall around the protein, preventing the free water from escaping into the vacuum and effectively providing a kind of boundary potential to the system. Details about the composition of the snapshots are summarized in Table 1.

Each snapshot was then used as starting structure for locating stationary points (minima and transition state) of the hydroxylation reaction at the AM1/GROMOS level by means of a protocol involving a linear-scaling microiterative optimization algorithm working in hybrid delocalized coordinates²⁴ as implemented in ChemShell.²⁵ The QM region included the substrate and the isoalloxazine part of FADHOOH (49 atoms); full details on these calculations will be reported elsewhere.²⁶ The stationary points thus obtained served as initial structures for the QM/MM MD simulations described here.

B. QM/MM MD Setup. All calculations reported were done at the AM1²⁷/GROMOS level within the modular program package ChemShell,^{25,28} which provided the QM/MM coupling engine and the MD driver. The energy and gradient evaluations for the QM and MM part were performed by the MNDO99²⁹ and GROMOS96²² codes, respectively, interfaced to ChemShell. The QM region contained the dianionic substrate and the isoalloxazine part of FADHOOH up to the first methylene unit of the ribityl side chain (see Scheme 2) and was terminated by a H link atom (49 QM atoms in total). The QM density was electrostatically embedded into the field of the fixed MM point charges by including them into the QM Hamiltonian. The charge-shift scheme²⁵ was applied at the QM/MM boundary; no electrostatic QM/MM cutoff was employed. The MM electrostatic interactions were evaluated in GROMOS96 for all atom pairs within a distance of 1.4 nm and approximated by a generalized Poisson–Boltzmann reaction field³⁰ beyond. The parameters for the reaction field were $\epsilon_1 = 1.0$ and $\epsilon_2 = 54.0$ (i.e., the permittivity of SPC water³¹) for the relative permittivities of the inner region and the continuum, respectively; the ionic strength was zero (i.e., inverse Debye screening length $\kappa = 0$).

The MD simulations were performed under *NVT* (canonical) conditions at $T = 300$ K. The temperature was controlled by a Berendsen thermostat³² (coupling time $\tau_c = 0.1$ ps) during the heating phase. For equilibration and production, a Nosé–Hoover chain (NHC) thermostat^{33–36} was implemented into the ChemShell dynamics module, together with a reversible noniterative leapfrog-type integrator³⁷ (thermostat chain length: 4; characteristic time of the thermostat coupling to the physical system $\tau_{\text{NHC}} = 0.02$ ps, corresponding to a thermostat wavenumber $\tilde{\nu}_{\text{NHC}} = 375$ cm⁻¹). All hydrogen

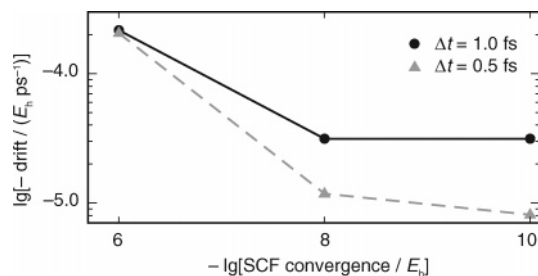


Figure 1. Drift of the total energy in AM1 Born–Oppenheimer NVE simulations as a function of the SCF convergence criterion and the time step.

atoms were assigned the mass of deuterium, $m_H = 2.014 \text{ u}$; the free water molecules were kept rigid using SHAKE constraints.^{38,39} The time step was $\Delta t = 1 \text{ fs}$.

Each starting structure (optimized reactant and transition states of snapshots 40, 80, 120, 160, and 200 ps and the product state of snapshot 40 ps) was equilibrated as follows: Starting with random velocities corresponding to a Boltzmann distribution at 300 K, the system was run for 10 ps with the Berendsen thermostat, followed by 25 ps of NHC dynamics. For simulations starting from a transition state (TS), the reaction coordinate was kept fixed at its value in the TS using a SHAKE constraint for a difference of distances. The equilibration process was assessed by monitoring (1) the temperatures of different subsystems (protein, free water, FADHOOH + pOHB) and (2) the value of the (unconstrained) reaction coordinate (RC) for simulations of a reactant or product state or the constraint force acting on the RC for simulations of a TS, respectively. For the system to be considered equilibrated, these quantities were required to be stationary according to the criteria and the procedure described below (section I.D). The average subsystem temperatures were in addition required to coincide and be equal to 300 K within the confidence interval. Typically, it took around 25 ps for the temperatures to become equilibrated. For most systems, 35 ps of total equilibration time were sufficient for all criteria to be met. If not, the simulation was extended until this was the case.

An important issue in QM/MM MD simulations (or, more generally, in any type of Born–Oppenheimer MD) is the balance between the time step, which controls the accuracy of the integration of the equations of motion, and the SCF convergence, which determines the accuracy of the QM forces. Using FADHOOH (88 atoms, $m_H = 2 \text{ u}$) as a test system, we ran Born–Oppenheimer NVE (microcanonical) simulations with AM1 to evaluate the influence of the time step and the SCF convergence criterion on the long-term stability of the simulations as measured by the drift in the total energy. As shown in Figure 1, either of the two parameters can be the limiting factor. If the SCF convergence is not strict enough, decreasing the time step does not improve stability. On the other hand, tightening the convergence criterion is only able to reduce the drift down to a certain level, at which the time step becomes limiting. Since the computation time is not significantly increased by a stricter SCF convergence criterion, whereas depending linearly on the time step, we chose as the optimum combination a time step of 1 fs and an SCF convergence of

$10^{-8} E_h$. Furthermore, we modified ChemShell to avoid any loss of accuracy in energy, gradient, and coordinate data when they are exchanged between the different modules (QM code, MM code, QM/MM coupling engine, MD driver), which proved to be crucial for the stability of long simulations.

C. Thermodynamic Integration. The free energy of activation (and in one case also the reaction free energy) was obtained from thermodynamic integration along the difference-of-distances reaction coordinate $RC = d(C_{\text{meta}} - O_d) - d(O_p - O_d)$, which captures both the formation of the bond between C_{meta} of pOHB and the distal oxygen of the hydroperoxy group and the cleavage of the peroxide bond, thus providing an accurate description of the full reactive process from reactants to products. Note that according to this definition, the reaction runs from positive values of the RC at the reactants to negative values at the products, that is, from “right to left” on a RC scale.

The formalism of thermodynamic integration⁴⁰ can be summarized by the relations

$$\Delta A_{a \rightarrow b} = \int_a^b d\xi \frac{\partial A}{\partial \xi} = \int_a^b d\xi \left\langle \frac{\partial V}{\partial \xi} \right\rangle_{\xi} \approx \int_a^b d\xi \langle F_c \rangle_{\xi} \quad (1)$$

where $A = A(\xi)$ is the Helmholtz free energy parametrized by ξ , V is the potential energy, and F_c is the force of constraint; $\langle \cdot \rangle$ designates averaging over the canonical ensemble. Identifying ξ with the reaction coordinate, the free-energy difference $\Delta A_{a \rightarrow b}$ between two states a and b , characterized by ξ_a and ξ_b , is evaluated as the potential of the mean force of constraint,^{41–43} that is, the average force acting on the constraint that keeps the reaction coordinate at a certain value. In the last equality in eq 1, we made the approximation that the metric-tensor correction^{41,44–47} is negligible (see Appendix A.1 for explicit expressions of the constraint force and the metric-tensor correction).

To evaluate the integral of the average constraint force over the reaction coordinate, it is discretized. One performs simulations at a series of fixed values of ξ , calculates $\langle F_c \rangle$ at each point, and numerically integrates the average constraint force over the reaction coordinate.⁴⁸

We used an integration step along the RC of 0.1 Å. At each point, the system was sampled for at least 5 ps, until the force of constraint was stationary according to the procedure described in section I.D below. After 1 ps, the system was stepped to the next point. The end point of the integration on the reactant side was determined from the average value of the RC in the reactant (typically 1.7 Å), obtained from the converged equilibration run. Correspondingly, the end point on the product side was at $RC = -1.6 \text{ Å}$. For the cases where we were interested only in the forward barrier, we integrated from the reactant up to $RC = 0.0 \text{ Å}$, the TS being typically at $RC \approx 0.4 \text{ Å}$. For all snapshots, we integrated forward (from the equilibrated reactant to the TS) and backward (from the equilibrated TS to the reactant), obtaining two independent values for the forward free energy of activation, $\Delta^\ddagger A_{\text{fw}}$. A typical force curve and its integral are shown in Figure 2. For the 40-ps snapshot, we also integrated further from the TS to the product, which yielded the reaction free energy for the forward reaction, $\Delta_r A_{\text{fw}}$.

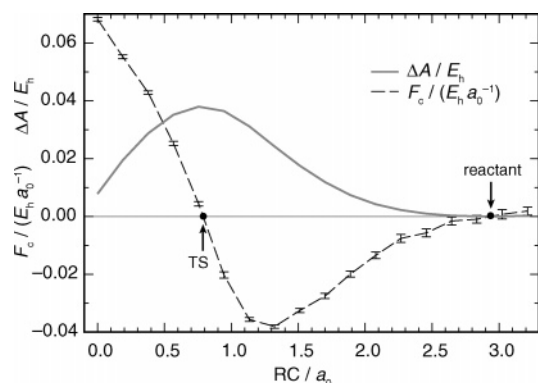


Figure 2. Average constraint force and its integral (i.e., the free-energy difference) along the RC. 40-ps snapshot, sampling from reactant to TS.

Details on the numerical integration and the error analysis are provided in Appendix A.1.

D. Analysis of MD Data. When calculating the average of any dynamical quantity from an MD simulation, e.g., for monitoring the equilibration process or obtaining a “measurement” such as the force of constraint, one needs a procedure to calculate the average and its variance from the MD data and criteria to assess if the average has become stationary. The former issue is somewhat complicated by the fact that MD data is inherently correlated, which means that the usual statistical formulas, e.g., to calculate the variance of the mean, are not applicable. Two main approaches to deal with this problem are conceivable. One can either (1) treat the correlations explicitly by extracting correlation information from the data and accounting for them in the calculation of the variance or (2) transform the data such as to remove the correlations and apply the standard formulas for statistically independent data. Both strategies have been used in the literature and their relative merits discussed.^{49–54} The weakness of the first approach is that obtaining reliable estimates for the correlation coefficients⁵¹ is not straightforward (see, e.g., the succinct treatment in ref 53) as different estimators involving different approximations may be chosen depending on the nature and the properties of the correlations present in the data. We have therefore chosen a decorrelation approach, which appears to be more robust. In particular, we have adopted the procedure suggested by Schiferl and Wallace,⁵⁰ which we summarize here.

The main idea of this treatment is to remove correlations by coarse-graining and then to subject the decorrelated data to statistical tests checking for stationarity and lack of correlation. The coarse-grained data set is generated simply by collecting successive original data points $\{X_i, i = 1, 2, \dots, N\}$ into n nonoverlapping segments or bins of width m . (N is adjusted to make m an integer divider by deleting extraneous points from the beginning of the series.) We denote the bin means and variances by \bar{x}_k and s_k^2 ($k = 1, 2, \dots, n$). The \bar{x}_k now constitute the coarse-grained data set, which is free of correlations, provided an appropriate choice of m , as discussed below.

The procedure can be outlined as follows:

(1) Plot the X_i and determine visually when the data series is apparently constant (no obvious drift or change in

bandwidth). Discard the data before that point. Also determine the oscillation period, e.g., by counting the number of maxima over a certain interval at the end of the series.

(2) Coarse-grain the data, using the oscillation period as a trial value for m .

(3) Apply statistical tests for (a) lack of trend in the \bar{x}_k (Mann–Kendall test for trend); (b) lack of trend in the s_k^2 (Mann–Kendall test for trend); (c) normal distribution of the \bar{x}_k (Shapiro–Wilk W test for normality); and (d) lack of correlation in the \bar{x}_k (von Neumann test for serial correlation). Details on these tests are provided in Appendix A.2.

(4) If either of the trend tests fails, drop more points from the start of the series. If they both succeed, but the normality and/or the correlation test fails, increase m .

(5) If so many points need to be eliminated or m increased so much that n falls below 24, extend the simulation to generate more data. (This minimum value for n secures the resolving power of the statistical tests.)

(6) If all tests pass, calculate mean, variance, and confidence interval for the series:

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n \bar{x}_k \quad (2)$$

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (\bar{x}_k - \bar{X})^2 \quad (3)$$

$$\langle X \rangle = \bar{X} \pm t_{n-1}^{(\alpha/2)} \frac{S}{\sqrt{n}} \quad (4)$$

$t_{n-1}^{(\alpha/2)}$ is the quantile of Student’s t distribution with $n - 1$ degrees of freedom at the chosen significance level α . We have worked with $\alpha = 0.025$ (i.e., a confidence level of 95%) throughout.

This procedure affords a well-defined and well-justified protocol to assess whether a given quantity has converged to a stationary value and at the same time yields this average value and a measure for its statistical error. The importance of applying well-defined statistical criteria has recently been highlighted also by Karplus and co-workers,⁵⁵ who elaborated on the idea of reverse cumulative averaging.⁵⁶

Regarding the sampling times required to achieve stationary averages, we note that after the relatively long initial equilibration phase (tens of picoseconds), the sampling time at each value of the RC could be kept rather short (a few picoseconds). We did not observe any systematic dependence of the sampling time on the RC, the snapshot, or the starting point. The complete data from the thermodynamic integration simulations including the number of steps accepted in the final average, bin width, and average constraint force are provided as Supporting Information.

Results for PHBH and Discussion

A. Energetics. The free energies of activation for the hydroxylation of pOHB in PHBH calculated for the different snapshots (40, 80, 120, 160, and 200 ps) and starting points (reactant R and transition state TS) are collected in Table 2. The values lie within a range of 95 to 108 kJ mol^{−1} with an average of 101 kJ mol^{−1}, which is thus our best estimate for

Table 2. Free Energies of Activation, Values of the RC at the End-Points of the Thermodynamic Integration, and Values of the Components of the RC at the TS for the Hydroxylation of pOH in PHBH

snapshot (ps) (start ^b)	$\Delta^\ddagger A_{fw}$ (kJ mol ⁻¹) ^c	RC (Å)		$\langle d \rangle_{TS}^a$ (Å)	
		R	TS	O _p -O _d	C _m -O _d
40 (R)	101 ± 2	1.56	0.42	1.66	2.06
40 (TS)	101 ± 2	1.55	0.41	1.63	2.05
80 (R)	108 ± 2	1.62	0.42	1.66	2.03
80 (TS)	100 ± 2	1.53	0.42	1.64	2.04
120 (R)	106 ± 2	1.48	0.42	1.65	2.05
120 (TS)	105 ± 2	1.54	0.41	1.64	2.04
160 (R)	98 ± 2	1.77	0.44	1.65	2.05
160 (TS)	95 ± 2	1.65	0.45	1.68	2.08
200 (R)	102 ± 2	1.51	0.42	1.65	2.05
200 (TS)	97 ± 2	1.45	0.42	1.65	2.05

^a Average distance at RC = 0.42 Å for 40 ps (TS), at RC = 0.4 Å otherwise. The error according to eq 4 is 0.01 Å, except for $\langle d(C_m-O_d) \rangle_{TS}$ in 80 ps (R), where it is 0.05 Å. ^b R: Integration from reactant to TS; TS: integration from TS to reactant. ^c Activation free energy of the forward reaction; error calculated according to eq 14.

$\Delta^\ddagger A_{fw}$. This value is significantly higher than the 49 kJ mol⁻¹ obtained previously²¹ also with AM1/GROMOS. However, those calculations differ from the ones reported here in several aspects, such as the absence of a solvation shell, a different reaction coordinate, and positional restraints on the atoms outside a region of about 7 Å around the active site, which precludes a direct comparison. With respect to corresponding experimental data, there is the value for the overall activation energy¹⁵ of 49 kJ mol⁻¹ (derived from an Arrhenius plot covering a temperature range of 21 K) that could serve as a reference, provided that hydroxylation is rate-limiting under the experimental conditions (pH 8.0).^{16,17} Accepting this assumption, one can also convert⁵⁷ overall rate constants obtained under the same conditions^{5,14} into a free energy of activation, affording about 64 kJ mol⁻¹. Our calculations thus appear to overestimate the barrier.

This discrepancy is partly accounted for by the known deficiency of AM1 to yield barriers that are too high. For instance, AM1 overestimates the barrier for the hydroxylation of pOH by H₂O₂ by 45% compared to MP2.²¹ Another factor that may contribute is the uncertainty about the protonation state of the enzyme. There are five histidine residues within 15 Å of the active site that were all taken as neutral but can carry an additional proton depending on pH and local pK_a. A mutation study that replaced a negatively charged glutamate some 10 Å away from pOH by a neutral glutamine (Glu49Gln⁹), thus enhancing the positive electrostatic field in the active site, found that this seemingly subtle change leads to a 4-fold increase of the hydroxylation rate. Protonating one or more of the histidines around the active site is hence likely to decrease the barrier. However, we have not investigated this possibility in the present work.

The barriers obtained for both starting points of the same snapshot agree within 5 kJ mol⁻¹ in all cases except the 80-ps snapshot, where the difference is 8 kJ mol⁻¹. This variability as well as the spread between different snapshots is a consequence of incomplete sampling. In the ideal case of exhaustive sampling, we would expect to obtain the same value irrespective of the history of the system. For real

Table 3. Convergence Behavior of the Activation Free Energy with Increasing Sampling Time for the 80-ps Snapshot

sampling length ^a	$\Delta^\ddagger A_{fw}$ (kJ mol ⁻¹)	
	R ^b	TS ^b
first 5 ps	108.0 ± 1.9	99.8 ± 2.1
first 10 ps	105.6 ± 1.6	100.8 ± 1.5
first 20 ps	105.6 ± 1.0	101.5 ± 1.1
first 30 ps	104.6 ± 1.0	102.2 ± 1.1
first 40 ps	103.8 ± 0.9	101.1 ± 1.1
all data	103.6 ± 0.8	100.6 ± 0.7

^a Amount of raw data considered. Following the analysis procedure, a variable number of steps is discarded from the beginning of the series at each integration point. ^b R: Integration from reactant to TS; TS: integration from TS to reactant.

sampling conditions, however, the system retains some memory of the previous treatment (that is, the classical MD simulations, the QM/MM optimizations, and the previous TI steps) that is not averaged out fully. The observed variability thus reflects differences in the environment that affect the height of the free-energy barrier.

To investigate the influence of sampling time, we chose the 80-ps snapshot, where the results from the two starting points differed most. We sampled for an additional 40 ps at each value of the RC, thus accumulating at least 45 ps per integration point. We evaluated the free energy by including successively more data into the analysis. The results of this convergence study are shown in Table 3. While the value for snapshot 80 ps (TS) remains essentially constant upon extending the sampling time, the activation free energy calculated from snapshot 80 ps (R) decreases by about 4 kJ mol⁻¹. This brings it close to the average of the other snapshots, and the difference between the results from the two starting points is reduced from 8 to 3 kJ mol⁻¹. We conclude that the observed variability in the results is indeed due to incomplete sampling and can be reduced by increasing the sampling time. Furthermore, the approach to use several starting points provides a relatively cost-efficient way to validate the results and identify “outliers”; bringing these to convergence can, however, require considerable additional effort.

From the full thermodynamic integration of snapshot 40 ps from the TS to both the reactant and the product state, we obtained the free energy of reaction as $\Delta_r A_{fw} = -212$ kJ mol⁻¹, which is in perfect in agreement with the result of the previous AM1/GROMOS study.²¹ Combining the reaction energy with the barrier for the forward reaction, we obtain for the activation free energy of the backward reaction $\Delta^\ddagger A_{bw} = 313$ kJ mol⁻¹. The value of the RC at the product was -1.51 Å.

B. Structural Features. With regard to structural features, we focus on interactions of pOH with active site residues and their changes along the RC. There is quite a detailed understanding from experimental studies of how the dianionic substrate is accommodated in the active site (see Figure 3). The phenolate oxygen is stabilized by a hydrogen bond to Tyr 201, from which a H-bond chain extends via Tyr 385 and several water molecules to His 72.⁵ The carboxy group forms a salt bridge with Arg 214⁶ and accepts hydrogen

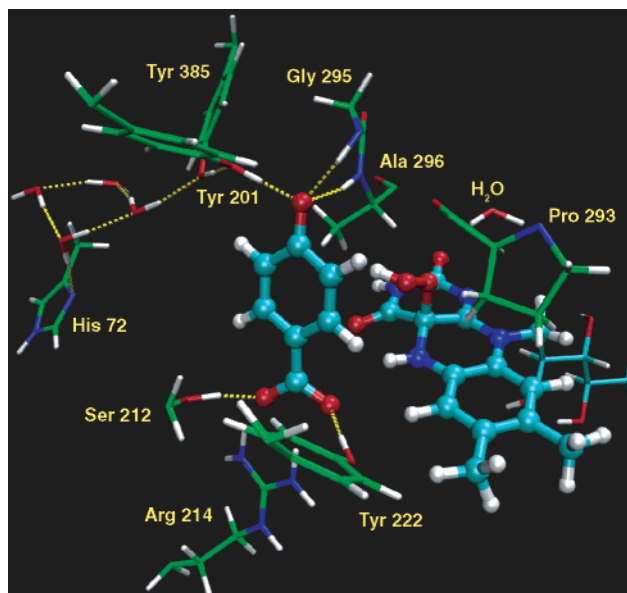


Figure 3. Active site of PHBH with FADHOOH, bound pOHB (both cyan), and selected residues (green). QM region in ball-and-sticks representation, MM part as sticks. Unless necessary, only the side chains of amino acids are shown.

bonds from Ser 212⁸ and Tyr 222.⁷ In addition, we find hydrogen bonds from the backbone amide NH of Gly 295

and Ala 296 to the pOHB phenolate. Special attention has been given to the role of Pro 293. Previous calculations^{18,20} found that the FADHOOH hydroperoxy group forms a hydrogen bond to the backbone amide carbonyl oxygen of Pro 293 as the reaction proceeds toward the transition state. This stabilization was identified as the only significant specific interaction of the protein with the transition state.²⁰ Moreover, as the peroxide O–O bond breaks, the negative charge developing on the proximal oxygen is stabilized by a hydrogen bond from a nearby water molecule.¹⁸

We have followed the hydrogen bonding pattern in the active site along the reaction coordinate. Figure 4 shows, as a representative example, pertinent distances for snapshot 40 ps (R). The features we discuss here, however, apply to all 10 simulations, if not mentioned otherwise. pOHB is kept in place by several permanent interactions that remain stable over the course of the reaction. The carboxylate is stabilized through hydrogen bonds from the hydroxy groups of Tyr 222 and Ser 212 as well as by the salt bridge to the guanidinium moiety of Arg 214 (Figure 4e). The phenolate oxygen accepts stable hydrogen bonds from the hydroxy group of Tyr 201 and the amide NH of Gly 295 (Figure 4c). The hydrogen bond from the relatively acidic phenol-OH is somewhat shorter than the one from the amide NH. Until before the transition state, there is also a hydrogen bond from

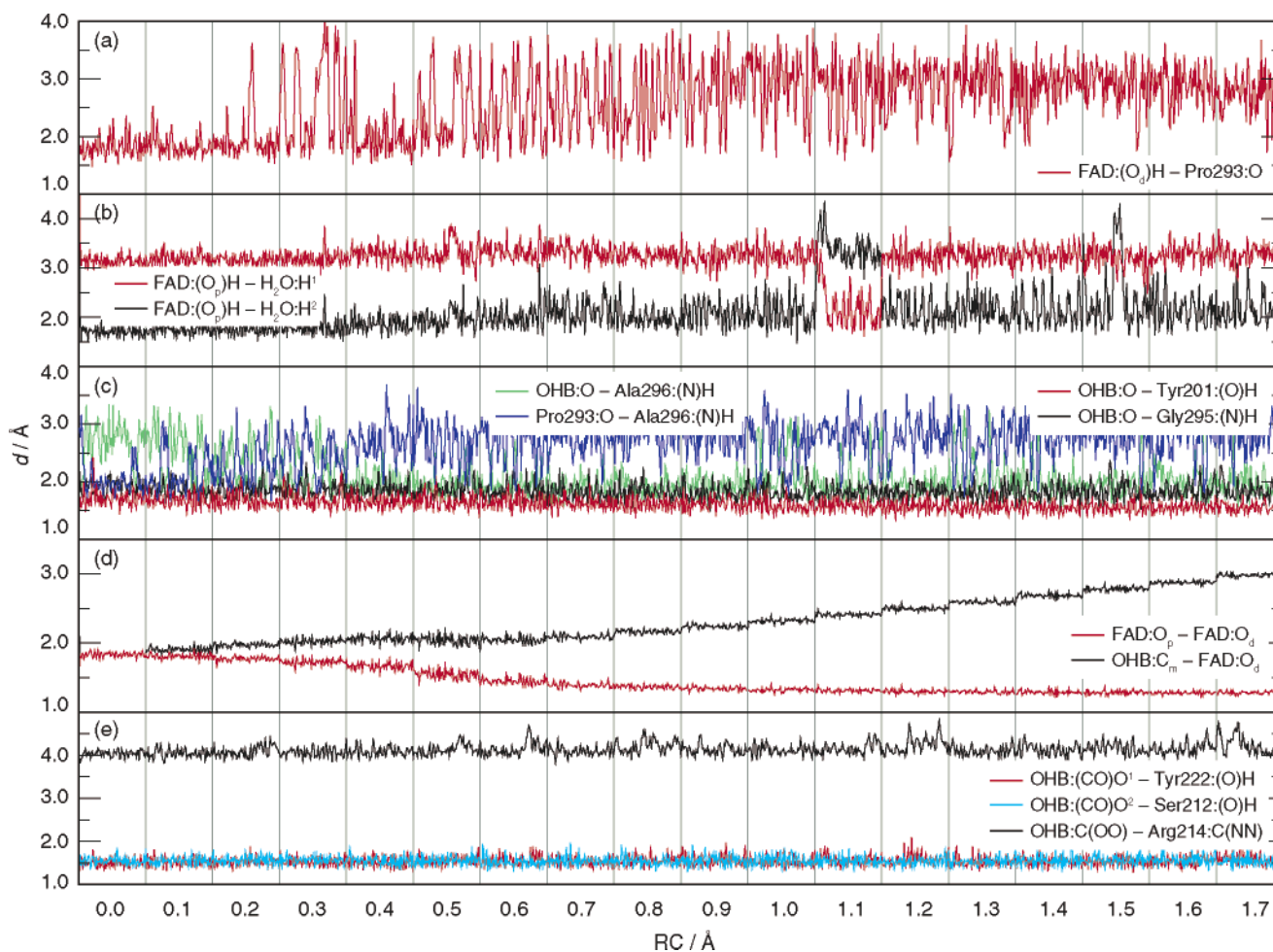


Figure 4. Selected distances along the RC for snapshot 40 ps (R). The reaction proceeds from right to left; the TS is at 0.4 Å. Each vertical panel shows the last 4 ps of sampling at the given value of the RC.

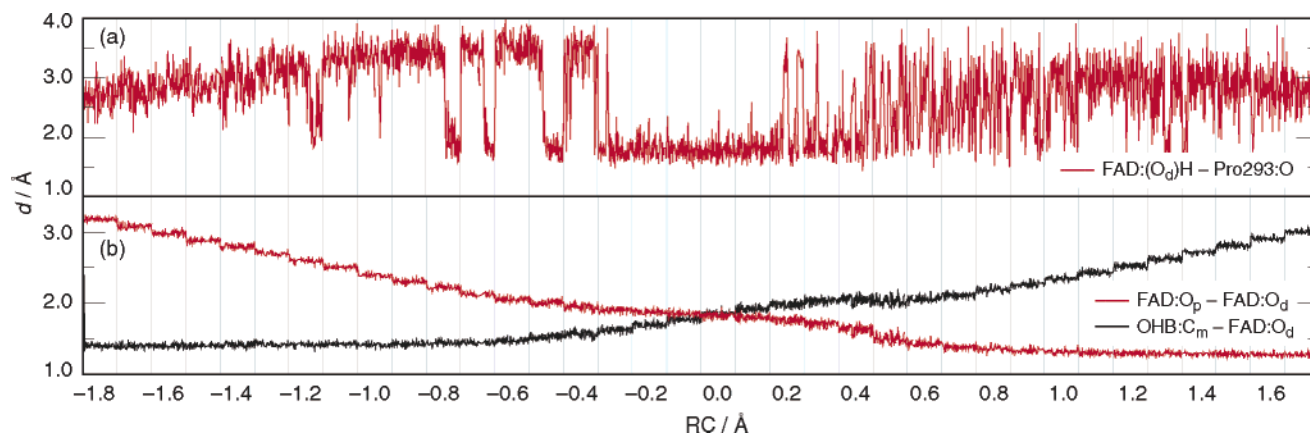


Figure 5. Selected distances along the RC for snapshot 40 ps (TS). The complete reaction from reactants (on the right) to products (on the left) is covered; the TS is at 0.4 Å. Each vertical panel shows the last 4 ps of sampling at the given value of the RC.

the amide NH of Ala 296 to the pOHB phenolate. This interaction is lost around the TS, when the phenolate is transformed into a carbonyl moiety, which is less basic and hence a weaker hydrogen-bond acceptor. Likewise, the remaining two hydrogen bonds from Tyr 201 and Gly 295 lengthen by about 0.2 Å as the system proceeds from reactants to the TS.

On the other hand, the oxido oxygen of FADHO resulting from breaking the O–O bond in FADHOH is a much better hydrogen-bond acceptor than the hydroperoxy group. This is reflected in the development of a strong hydrogen bond from a water molecule (Figure 4b). Initially, this water molecule is only weakly associated with the proximal hydroperoxy oxygen. The average distance is around 2.3 Å, with large fluctuations, and it occasionally rotates (and binds through the other hydrogen atom, e.g., at RC = 1.1 Å) or drifts away (RC = 1.5 Å). Toward the TS, the interaction becomes more stable, the distance fluctuations subside, and a strong, stable hydrogen bond is established from the TS onward.

The proposed capability of Pro 293 of “shepherding” the transferring OH moiety by a hydrogen bond to its amide oxygen is of particular interest. As seen in Figure 4a, the frequency of hydrogen bonding events from the transferred OH to the Pro 293 amide oxygen increases as the reaction proceeds toward the TS. However, only at the TS (RC = 0.4 Å) a stable hydrogen bond is established for the first time. It is broken and reformed several times after the TS, before becoming permanent toward RC = 0.0 Å. From the complete thermodynamic integration carried out for snapshot 40 ps (TS) (see Figure 5a), we find the reverse pattern as the system moves away from the TS toward the products, the hydrogen bond between OH and Pro 293 breaking more and more frequently and finally staying broken. Unlike the other hydrogen bonds discussed above, the OH–Pro 293 interaction varies widely among different snapshots. In some cases, a stable interaction exists at the TS; in other cases, the bond is only established later along the RC or not at all.

Another factor of interest in this context is the hydrogen bond from the Ala 296 amide NH to the amide oxygen of Pro 293 as it competes for Pro 293 against the transferred OH group. This interaction is complementary to the bond

from Ala 296 to the pOHB phenolate oxygen, that is, one of the bonds is present at any given time (see Figure 4c). In most simulations, the Ala 296–pOHB interaction is almost stable until the TS, with a few switching events to Ala 296–Pro 293, and is permanently replaced by Ala 296–Pro 293 somewhere after the TS. However, the exact location of the switch from Ala 296–pOHB to Ala 296–Pro 293 varies for different snapshots or does not happen at all.

To obtain a more quantitative measure for the presence or absence of these two hydrogen bonds, we determined for each point along the RC a bonding index (or bonding probability), that is, the fraction of the time the bond is present over the last 4 ps. The criterion for a bond is a distance of 2.5 Å or less. We find the following general trends for the bond indices along the RC: (i) The OH–Pro 293 bond index is around 0.1 from the reactants up until RC = 1.0 Å, when it starts rising. (ii) At the TS (RC = 0.4 Å), the OH–Pro 293 bond index scatters widely, between 0.44 and 0.93; at RC = 0.0 Å, the bond index is greater than 0.8 for most snapshots. (iii) The Ala 296–Pro 293 bond index remains around 0.1 up until the TS, where it increases sharply to 0.9 or larger at RC = 0.0 Å in all cases but one.

To elucidate specifically the effect on the transition state and the barrier, we averaged the bond indices over the TS and the two adjacent points (i.e., over RC = 0.5, 0.4, and 0.3 Å). The results of this analysis are shown in Table 4. The average bond index for OH–Pro 293 shows some correlation with the activation free energy, e.g., snapshots 160 ps (TS) and 200 ps (TS) have a low barrier and a high bond index, which would be expected from the argument that this hydrogen bond stabilizes the TS. However, it fails in other cases, where the bond index is low, but the barrier is not, e.g., 160 ps (R) or 80 ps (TS). The corresponding inverse correlation of the barrier height with the Ala 296–Pro 293 bond index is similarly inconclusive. The most consistent trend is found for the difference between the two bond indices, but there are exceptions here as well. For instance, 40 ps (TS) should have a particularly low activation free energy and 80 ps (TS) a high one, which is not the case. A remarkable detail in Table 4 is the observation that only in the two 80-ps snapshots both hydrogen bonds are equally

Table 4. Bond Indices for the Hydrogen Bonds to the Pro 293 Amide Oxygen from the FADHOOH OH and the Ala 296 Amide NH, Averaged over the TS Region

snapshot (ps) (start ^b)	average bond index ^a		difference
	O _d H–Pro293	Ala296–Pro293	
40 (R)	0.68	0.38	0.30
40 (TS)	0.72	0.20	0.52
80 (R)	0.49	0.55	–0.06
80 (TS)	0.48	0.54	–0.06
120 (R)	0.56	0.32	0.24
120 (TS)	0.68	0.42	0.25
160 (R)	0.51	0.12	0.38
160 (TS)	0.75	0.38	0.37
200 (R)	0.66	0.41	0.25
200 (TS)	0.78	0.49	0.29

^a The bond index was determined for each point of the RC as described in the text. Tabulated are the values averaged over RC = 0.3, 0.4, 0.5 Å. ^b R: Integration from reactant to TS; TS: integration from TS to reactant.

probable around the TS, while the bond to the transferred OH is otherwise clearly more abundant.

In conclusion, we can identify two distinct groups of hydrogen bonds in the active site. There is a set of permanent hydrogen bonds that stabilize the substrate and keep it in place. Another set of hydrogen bonds rearranges as the reaction proceeds, adapting to the changing charge distribution in the active site. In particular, we find a rearrangement of the hydrogen bonding network around the amide oxygen of Pro 293 in the TS region that is very likely to stabilize specifically the transferred OH group and thus the transition state. However, it has not been possible to single out this interaction as the determining factor controlling the barrier height.

Finally, we consider the course of the two components of the reaction coordinate, the hydroperoxy O_p–O_d and pOHB C_m–O_d distances (Figure 4d). Until RC = 1.0 Å, the O–O bond is almost unaffected, the RC acting to bring the substrate closer to the cofactor. The O–O bond then starts lengthening, slowly until RC = 0.7 Å, more rapidly between RC = 0.7 and 0.3 Å. During this latter period, the C_m–O_d distance remains almost constant. This is the phase where the O–O bond is broken. From the full simulation on snapshot 40 ps (TS) (Figure 5b) we find that from RC = 0.2 to –0.6 Å, the C–O distance smoothly shortens, that is, the C–O bond is forming, while the O–O distance increases, that is, 3,4-DOHB drifts away from the cofactor. Only at about RC = –0.7 Å, the C–O bond is fully formed and remains constant. From RC = 0.3 Å onward, we can also observe the change in hybridization of the attacked C_m atom from sp² to sp³, as reflected in the dihedral angle C_o–C_m–C_p–H_m.

IV. Conclusions

We have investigated the OH transfer step of the hydroxylation reaction of *p*-hydroxybenzoate in the enzyme *p*-hydroxybenzoate hydroxylase (PHBH). Using thermodynamic integration with AM1/GROMOS QM/MM molecular dynamics simulations, we have calculated the activation free energy of the reaction. We have used 10 different starting

structures, obtained from classical MD simulations and QM/MM optimizations of stationary points. Our model included a shell of solvating water molecules around the protein, and the full system except the outer water shell was left unconstrained. From these calculations, we have obtained a free-energy barrier of $\Delta^\ddagger A_{\text{fw}} = 101 \text{ kJ mol}^{-1}$ as our best estimate. As AM1 is known to overestimate reaction barriers, this value is likely to be too high. For the reaction free energy, we have obtained $\Delta_r A_{\text{fw}} = -212 \text{ kJ mol}^{-1}$. The simulations have also provided detailed insight into the changes in the hydrogen bonding network in the active site and the cleaving and forming of the covalent bonds along the course of the reaction. With respect to a stabilization of the transition state by hydrogen bonds, we have confirmed that Pro 293 interacts with the transferring OH in the transition state region. However, it has not been possible to pinpoint this interaction, or any other, as the controlling factor for the barrier height.

Moreover, we have described a procedure involving coarse-graining of the data and statistical tests to ensure that converged averages are obtained from MD simulations. The procedure also yields a measure for the statistical error and hence allowed us to assess the magnitude of different sources of error. For the simulation conditions used, the error in the final free-energy difference is dominated by the error due to incomplete sampling, as seen by the differences in the results for different starting points. The pure statistical error, on the other hand, is mainly caused by the variance of the average constraint force, whereas the error from the numerical integration procedure is negligible.

QM/MM calculations treating the QM region at a higher level of theory are currently underway in our laboratory, which will provide a better assessment of the performance of AM1.

Acknowledgment. We thank Dr. S. Billeter, IBM Zürich Research Laboratory, Rüschlikon, Switzerland, for helpful discussions and the implementation of the difference-of-distances constraint in ChemShell.

Appendix

A.1. Calculation of the Free Energy and Error Analysis. The difference-of-distances constraint involving the three atoms *A*, *B*, and *C* has the holonomic form

$$\sigma(\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C) = |\mathbf{R}_{BA}| - |\mathbf{R}_{CB}| - \xi = 0 \quad (5)$$

where $|\mathbf{R}_{ij}| = |\mathbf{R}_i - \mathbf{R}_j|$ and ξ is the value of the constraint. The iterative calculation of the constraint force with the SHAKE algorithm^{38,58} requires the derivatives with respect to the atomic positions involved

$$\begin{aligned} \nabla_A \sigma &= \hat{\mathbf{R}}_{BA} \\ \nabla_B \sigma &= -(\hat{\mathbf{R}}_{BA} + \hat{\mathbf{R}}_{CB}) \\ \nabla_C \sigma &= \hat{\mathbf{R}}_{CB} \end{aligned} \quad (6)$$

where $\hat{\mathbf{R}}$ designates unit vectors. The constraint force F_c is identified with the converged Lagrangean multiplier γ that guarantees that the constraint is exactly satisfied in the next

time step. Generalizing to L constraints, the positions are propagated as

$$\mathbf{R}_I(+) = \bar{\mathbf{R}}_I(+) - \frac{\Delta^2}{M_I} \sum_{k=1}^L \gamma_k \nabla_k \sigma_k|_0 \quad (7)$$

where $\bar{\mathbf{R}}_I(+)$ is the new position propagated *without* constraints, M_I is the mass of atom I , and Δ is the time step.

The complete expression for the free-energy difference including the metric-tensor correction reads⁴⁷

$$\Delta A_{a \rightarrow b} = \int_a^b d\xi \langle F_c \rangle_\xi - k_B T \ln \frac{\langle |\mathbf{Z}|^{-1/2} \rangle_{\xi=b}}{\langle |\mathbf{Z}|^{-1/2} \rangle_{\xi=a}} \quad (8)$$

where $|\mathbf{Z}|$ is the determinant of the matrix

$$Z_{mn} = \sum_J M_J^{-1} \nabla_J \xi_m \nabla_J \xi_n \quad (9)$$

(also referred to as the Fixman determinant⁵⁹), where the sum runs over all atoms involved in any constraint. For the case of only one constraint, $|\mathbf{Z}|$ simplifies to

$$z = \sum_J M_J^{-1} (\nabla_J \xi)^2 \quad (10)$$

Evaluating explicitly for the difference-of-distances constraint, eq 5, yields

$$\begin{aligned} z &= M_A^{-1} \hat{\mathbf{R}}_{BA}^2 + M_B^{-1} (-\hat{\mathbf{R}}_{BA} - \hat{\mathbf{R}}_{CB})^2 + M_C^{-1} \hat{\mathbf{R}}_{CB}^2 \\ &= M_A^{-1} + 2M_B^{-1} [1 + \cos(\hat{\mathbf{R}}_{BA}, \hat{\mathbf{R}}_{CB})] + M_C^{-1} \end{aligned} \quad (11)$$

Apart from constant mass terms that will cancel out in eq 8, the correction thus depends only on the angle between the two bonds involved in the constraint or, rather, on the change in this angle between the end points of the integration. In the present application, the angle $\angle(\mathbf{R}_{00}, \mathbf{R}_{C0})$ varies between about 140° and 170° for reactant and TS, respectively, which amounts to a correction of the order of -1 kJ mol⁻¹.

Provided the average constraint force at a series of points ξ_i along the RC, $\langle F_c \rangle_i$, polynomials of third degree were used to interpolate between successive points with Mathematica.⁶⁰ The resulting analytical representation was used to determine the value of the RC at the stationary points (where $\langle F_c \rangle = 0$) and to obtain the free-energy difference between these points by integration.

To quantify the error in the free energy, we consider two contributions: (1) The error of the numerical integration procedure and (2) the error due to the uncertainty in the $\langle F_c \rangle_i$. To assess the magnitude of these two error sources, we perform the analysis for integration according to Simpson's rule⁶¹

$$I = \int_a^b dx f(x) \approx \frac{h}{3} (f_1 + 4 \sum_{\substack{i=2 \\ i \text{ even}}}^{N-1} f_i + 2 \sum_{\substack{i=3 \\ i \text{ odd}}}^{N-2} f_i + f_N) \quad (12)$$

where $f_i = f(x_i)$ are the values of the function $f(x)$ at the (equispaced) grid points x_i , h is the grid spacing, and N is the total number of points; N must be odd. Simpson's rule

approximates the function by second-degree polynomials defined by three successive points. The error of Simpson's rule is given by

$$|e_I| \leq \frac{(b-a)h^4}{180} |f^{(4)}(x^*)| = \frac{(N-1)h^5}{180} |f^{(4)}(x^*)| \quad (13)$$

where $|f^{(4)}(x^*)|$ is an upper limit for the absolute value of the fourth derivative of $f(x)$ on $[a, b]$. For values as they have typically occurred in the present application (integration interval $2.28 a_0$, $h = 0.19 a_0$, $|f^{(4)}(x^*)| \approx 3.5 E_h a_0^{-5}$), we estimate $|e_I| \leq 0.15$ kJ mol⁻¹. ($|f^{(4)}(x^*)|$ was obtained numerically from typical force curves using Mathematica.)

To calculate the error contribution of the uncertainty in the f_i , given by their variances s_i^2 , to the value of the integral, we apply Gaussian error propagation to Simpson's rule, obtaining

$$s_I^2 = \sum_i \left(\frac{\partial I}{\partial f_i} \right)^2 s_i^2 = \left(\frac{h}{3} \right)^2 (s_1^2 + 16 \sum_{\substack{i=2 \\ i \text{ even}}}^{N-1} s_i^2 + 4 \sum_{\substack{i=3 \\ i \text{ odd}}}^{N-2} s_i^2 + s_N^2) \quad (14)$$

With $h = 0.19 a_0$, $N = 13$, and assuming a constant, typical value of $s_i = 10^{-3} E_h a_0^{-1}$ for the uncertainty of the constraint force, we estimate $s_I = 1.8$ kJ mol⁻¹.

We therefore conclude from this analysis that the error of the integral, that is, of the free-energy difference, is dominated by the error due to the uncertainty in the constraint force for the simulation conditions and parameters used in the present work. Accordingly, the reported error bars for ΔA have been calculated from eq 14.

A.2. Statistical Tests for Establishing Converged Averages. The first test to be applied to the coarse-grained data is the Mann–Kendall test for trend^{62,63} because it is a nonparametric (i.e., distribution-free) test, unlike alternative trend tests which assume a normal distribution, and because the absence of a trend is a prerequisite for the following tests. The Mann–Kendall test was implemented as in ref 50.

The Shapiro–Wilk W statistic⁶⁴ is a very powerful and universal measure for departure from normality, testing only for shape, but not for a specific mean and variance of the normal distribution. We have used a modern implementation of the Shapiro–Wilk test^{65–67} that is valid also for larger sample sizes. Code for the algorithm⁶⁵ and the required auxiliary routines^{68–70} was obtained from StatLib.⁷¹

To check for correlation in the trend-free, normal data, we have used the von Neumann test for serial correlation⁷² as described in ref 50.

Supporting Information Available: Full data from the thermodynamic integration simulations for each sampling run, including number of steps sampled, information on the averaging procedure, and average force of constraint. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Entsch, B.; van Berkel, W. J. H. *FASEB J.* **1995**, *9*, 476–483.

- (2) 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*, 367–375.
- (3) Ortiz-Maldonado, M.; Ballou, D. P.; Massey, V. *Biochemistry* **1999**, *38*, 8124–8137.
- (4) Entsch, B.; Ballou, D. P.; Massey, V. *J. Biol. Chem.* **1976**, *251*, 2550–2563.
- (5) Entsch, B.; Palfey, B. A.; Ballou, D. P.; Massey, V. *J. Biol. Chem.* **1991**, *266*, 17341–17349.
- (6) Clarkson, J.; Palfey, B. A.; Carey, P. R. *Biochemistry* **1997**, *36*, 12560–12566.
- (7) van der Bolt, F. J. T.; Vervoort, J.; van Berkel, W. J. H. *Eur. J. Biochem.* **1996**, *237*, 592–600.
- (8) Moran, G. R.; Entsch, B.; Palfey, B. A.; Ballou, D. P. *Biochemistry* **1999**, *38*, 6292–6299.
- (9) Ortiz-Maldonado, M.; Cole, L. J.; Dumas, S. M.; Entsch, B.; Ballou, D. P. *Biochemistry* **2004**, *43*, 1569–1579.
- (10) Palfey, B. A.; Moran, G. R.; Entsch, B.; Ballou, D. P.; Massey, V. *Biochemistry* **1999**, *38*, 1153–1158.
- (11) Palfey, B. A.; Basu, R.; Frederick, K. K.; Entsch, B.; Ballou, D. P. *Biochemistry* **2002**, *41*, 8438–8446.
- (12) Ortiz-Maldonado, M.; Entsch, B.; Ballou, D. P. *Biochemistry* **2003**, *42*, 11234–11242.
- (13) Entsch, B.; Ballou, D. P. *Biochim. Biophys. Acta* **1989**, *999*, 313–322.
- (14) Husain, M.; Entsch, B.; Ballou, D. P.; Massey, V.; Chapman, P. J. *J. Biol. Chem.* **1980**, *255*, 4189–4197.
- (15) van Berkel, W. J. H.; Müller, F. *Eur. J. Biochem.* **1989**, *179*, 307–314.
- (16) Vervoort, J.; Rietjens, I. M. C. M.; van Berkel, W. J. H.; Veeger, C. *Eur. J. Biochem.* **1992**, *206*, 479–484.
- (17) Ridder, L.; Mulholland, A. J.; Vervoort, J.; Rietjens, I. M. C. M. *J. Am. Chem. Soc.* **1998**, *120*, 7641–7642.
- (18) Ridder, L.; Mulholland, A. J.; Rietjens, I. M. C. M.; Vervoort, J. *J. Mol. Graphics Modell.* **1999**, *17*, 163–175.
- (19) Ridder, L.; Palfey, B. A.; Vervoort, J.; Rietjens, I. M. C. M. *FEBS Lett.* **2000**, *478*, 197–201.
- (20) Ridder, L.; Harvey, J. N.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Phys. Chem. B* **2003**, *107*, 2118–2126.
- (21) Billeter, S. R.; Hanser, C. F. W.; Mordasini, T. Z.; Scholten, M.; Thiel, W.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2001**, *3*, 688–695.
- (22) 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, and Groningen, The Netherlands, 1996.
- (23) van Gunsteren, W. F.; Berendsen, H. J. C. *GROMOS87 Program and Manual*; Laboratory of Chemical Physics, University of Groningen: Groningen, The Netherlands, 1987.
- (24) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.
- (25) 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. R.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1–28.
- (26) Thiel, S.; Senn, H. M.; Thiel, W. Manuscript in preparation.
- (27) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (28) *ChemShell*, V. 3.0a3, 2004.
- (29) Thiel, W. *MNDO99*, V. 6.1; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2003.
- (30) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5469.
- (31) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 3169–3174.
- (32) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (33) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (34) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
- (35) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (36) Martyna, G. J.; Klein, M. L.; Tuckerman, M. J. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (37) Jang, S.; Voth, G. A. *J. Chem. Phys.* **1997**, *107*, 9514–9526.
- (38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (39) Kutteh, R.; Straatsma, T. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1998; Vol. 12; pp 75–136.
- (40) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
- (41) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- (42) Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R. *Chem. Phys.* **1989**, *129*, 241–251.
- (43) Paci, E.; Ciccotti, G.; Ferrario, M.; Kapral, R. *Chem. Phys. Lett.* **1991**, *176*, 581–587.
- (44) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737–7744.
- (45) den Otter, W. K.; Briels, W. J. *J. Chem. Phys.* **1998**, *109*, 4139–4146.
- (46) Schlitter, J.; Klähn, M. *J. Chem. Phys.* **2003**, *118*, 2057–2060.
- (47) Schlitter, J.; Klähn, M. *Mol. Phys.* **2003**, *101*, 3439–3443.
- (48) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175–1188.
- (49) Jorgensen, W. L. *Chem. Phys. Lett.* **1982**, *92*, 405–410.
- (50) Schiferl, S. K.; Wallace, D. C. *J. Chem. Phys.* **1985**, *83*, 5203–5209.
- (51) Straatsma, T. P.; Berendsen, H. J. C.; Stam, A. J. *Mol. Phys.* **1986**, *57*, 89–95.
- (52) Bishop, M.; Frinks, S. *J. Chem. Phys.* **1987**, *87*, 3675–3676.
- (53) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461–466.

- (54) Morales, J. J.; Nuevo, M. J.; Rull, L. F. *J. Comput. Phys.* **1990**, 89, 432–438.
- (55) Yang, W.; Bitetti-Putzer, R.; Karplus, M. *J. Chem. Phys.* **2004**, 120, 2618–2628.
- (56) Pearlman, D. A. *J. Comput. Chem.* **1994**, 15, 105–123.
- (57) $\Delta^\ddagger G = RT \ln[k_B T / (h k_{\text{obs}})]$, where k_{obs} is the rate constant, T is the temperature, R is the universal gas constant, k_B is Boltzmann's constant, and h is Planck's constant.
- (58) Ryckaert, J.-P. *Mol. Phys.* **1985**, 55, 549–556.
- (59) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, 71, 3050–3053.
- (60) *Mathematica*, V. 5.0; Wolfram Research: Champaign, IL, 2003.
- (61) Weisstein, E. W. *Simpson's Rule*; 2004; <http://mathworld.wolfram.com/SimpsonsRule.html>.
- (62) Mann, H. B. *Econometrica* **1945**, 13, 245–259.
- (63) Kendall, M. G.; Gibbons, J. D. *Rank correlation methods*; Arnold: London, 1990.
- (64) Shapiro, S. S.; Wilk, M. B. *Biometrika* **1965**, 52, 591–611.
- (65) Royston, P. *Appl. Statist.* **1995**, 44, 547–551.
- (66) Royston, P. *Statistician* **1993**, 42, 37–43.
- (67) Royston, P. *Statist. Comput.* **1992**, 2, 117–119.
- (68) Royston, J. P. *Appl. Statist.* **1982**, 31, 176–180.
- (69) Hill, I. D. *Appl. Statist.* **1973**, 22, 424–427.
- (70) Wichura, M. J. *Appl. Statist.* **1988**, 37, 477–484.
- (71) Vlachos, P. *StatLib – Applied Statistics algorithms*; Department of Statistics, Carnegie Mellon University: Pittsburgh, PA, 2004; <http://lib.stat.cmu.edu/apstat/>.
- (72) von Neumann, J. *Ann. Math. Stat.* **1941**, 12, 367–395.

CT049844P