JCTC Journal of Chemical Theory and Computation

Accurate QM/MM Free Energy Calculations of Enzyme Reactions: Methylation by Catechol *O*-Methyltransferase

Thomas H. Rod* and Ulf Ryde

Department of Theoretical Chemistry, Chemical Center, Lund University, P.O. Box 124, S-22100 Lund, Sweden

Received April 25, 2005

Abstract: We recently described a method to compute accurate quantum mechanical free energies [Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* **2005**, *94*, 138302]. The method, which we term quantum mechanical thermodynamic cycle perturbation (QTCP), employs a molecular mechanics force field to sample phase space and, subsequently, a thermodynamic cycle to estimate QM/MM free energy changes. Here, we discuss the methodology in detail and test an approach based on a different thermodynamic cycle. We also show that a new way of treating hydrogen link atoms makes the free energy changes converge faster and that extrapolation to higher accuracy can be performed. We finally discuss the quantum mechanical free energy (QM/MM-FE) method in the framework of the QTCP method. All methods considered are applied to the methylation of catecholate catalyzed by catechol *O*-methyltransferase. We compute the free energy barrier for the reaction by computing free energy changes in steps between fixed QM regions along a predetermined reaction pathway. Using the QTCP approach, an extrapolated activation free energy of 69 kJ/mol for the forward reaction and 90 kJ/mol for the reverse reaction are obtained at the level of the B3LYP functional and the 6-311++G(2d,2p) basis set. The value for the forward reaction is in excellent agreement with the experimental value of 75 kJ/mol. Results based on the QM/MM-FE method differ by less than 10 kJ/mol from those values, indicating that QM/MM-FE may be a fairly accurate and cheap alternative to calculate QM/MM free energy changes. Moreover, the results are compared to barriers obtained with a fixed molecular mechanics environment as well as with structures optimized in a vacuum. All the computed free energy barriers are well converged. A major approximation in the current implementation of the QTCP method is that the QM region is fixed. The approximation leads to well-converged free energy barriers, which has been a problem in similar studies.

1. Introduction

The general problem with atomic-scale simulations of enzyme catalysis is that thousands of atoms are involved and chemical bonds are modified. The description of bond breaking or forming calls for fairly accurate quantum mechanical (QM) methods, whereas the huge phase space calls for sampling with a cheaper method like molecular mechanics (MM) or semiempirical QM. The most straightforward way to solve the problem is by sampling phase space using a combined quantum mechanics and molecular mechanics (QM/MM) method where a small subset of atoms are treated by QM and the remaining atoms by MM. This can be pursued either by sampling directly on the Born−Oppenheimer surface or by using a Car-Parrinello approach. The problem is that realistic simulation times are currently limited to a few tens of picoseconds for this approach because of the severe computational load even for small QM systems. This is a rather short simulation time for systems as big as enzymes where many events, besides the reaction catalyzed,

* Corresponding author phone: +46 46 222 4500; e-mail: Thomas.Rod@teokem.lu.se.

10.1021/ct0501102 CCC: $30.25
Published on Web 08/30/2005

occur on time-scales much longer than picoseconds.[1] Various approaches have been proposed and utilized in the literature to solve the problem of computing accurate QM/MM free energies for chemical reactions in solutions[2−12] including enzymatic reactions.[13−23] A basic idea is to use a fast but less accurate method to sample phase space and use this sampling to estimate high-level QM/MM free energies with a modest number of QM/MM calculations. Below we give a summary of methods based on this idea.

In the quantum mechanical free energy (QM-FE) approach by Jorgensen and co-workers,[2−4] a reaction pathway for atoms in the QM region is calculated in a vacuum. Free energies for the interaction between the QM and MM atoms are then calculated along the reaction pathway by performing MM free energy perturbation or thermodynamic integration calculations where electrostatic interactions between the QM and MM atoms are defined via point charge interactions. In the treatment by Jorgensen and co-workers, and later by Kollman and co-workers,[13−15] point charges to represent the QM atoms were derived from calculations in a vacuum, i.e., without an MM region. Jorgensen and co-workers used the method to study organic reactions in solution, and Kollman and co-workers extended the method to that of enzymatic reactions, namely amide hydrolysis in trypsin[13] and methyl transfer by catechol *O*-methyltransferase.[14]

Yang and co-workers applied their QM/MM free energy method (QM/MM-FE) to the enzymes triosephosphate isomerase,[16,17] enolase,[18,17] and 4-oxalocrotonate tautomerase.[19] The QM/MM-FE method is an improvement over the QM-FE method in that a QM/MM optimized reaction pathway and QM energies and point charges derived from QM/MM calculations are used. In this way, polarization of the QM region by the MM region is included. Ishida and Kato employed the same approach to study acylation by serine proteases.[20,21]

An alternative approach is the ab initio QM/MM approach (QM(ai)/MM) by Warshel and co-workers.[5,6,22−24] They sampled phase space by performing molecular dynamics (MD) simulations with a reference potential given by the empirical valence bond (EVB) method.[25] Umbrella sampling ensured that the entire reaction pathway was sampled and made it possible to calculate the potential of mean force (PMF). Free energy changes between the system described by the reference potential and by density functional theory were calculated with free energy perturbation, and in this way a high-level QM/MM PMF can be obtained. The methodology corresponds to using the thermodynamic cycle in Figure 1 (with MM replaced by EVB) and is in principle exact with respect to how the free energy changes are calculated. In practice, the free energy did not converge owing to large fluctuations of the difference between the reference potential and the high-level QM/MM potential, although electrostatic interactions converged. Therefore, Warshel and co-workers used more approximate methods to calculate the free energy difference between the system described by EVB and by a high-level QM/MM method.[6,23]

Like Warshel and co-workers, Wood and co-workers use free energy perturbation in their ab initio/classical free energy perturbation (ABC-FEP) approach, which was used to compute hydration energies of water and the Na^+ and Cl^- ions at different physical conditions.[26−29] In this approach, only solute−solvent interaction energies are perturbed to the QM level. They also studied solute−solvent structural properties as well as water dimer dissociation.[30,31]

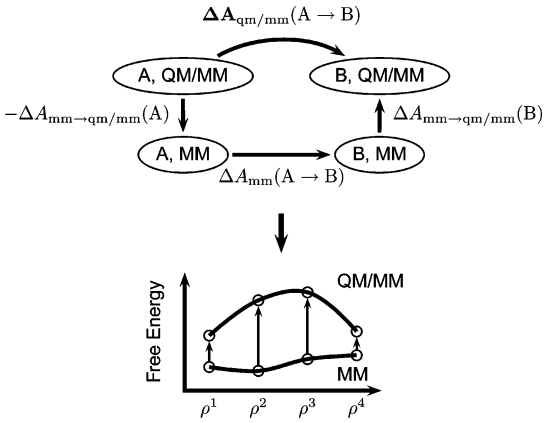

Figure 1. Illustration of the QTCP method, where a thermodynamic cycle is employed to calculate QM/MM free energy changes. The lower figure shows how this can be applied to calculate the free energy barrier along a reaction coordinate defined by ρ. We refer to this particular implementation of the QTCP method by QTCP-U.

Schofield and co-workers and Bandyopadhyay have developed a similar approach termed the molecular mechanics importance based function (MMBIF) method.[7−12] They also use a MM reference potential to sample the phase space and calculate corresponding high-level QM/MM energies for a set of configurations. Based on the two sets of energies, they use a Metropolis-Hastings algorithm to generate a high-level QM/MM canonical ensemble from which QM/MM free energies can be calculated.

Recently, we employed a third approach,[32] which can be considered as a combination of the QM/MM-FE method and the approach by Warshel and co-workers. Like Yang and co-workers as well as Ishida and Kato, we optimized a reaction pathway using QM/MM and selected a number of configurations for the QM region along the reaction pathway. Based on calculated point charges for the QM region, we calculated classical MM-QM interaction free energy changes between subsequent fixed QM configurations along the reaction pathway. Similar to the approach by Warshel and co-workers, we then calculated the MM → QM free energy change for each QM configuration along the reaction pathway, and, in this way, a high-level QM/MM PMF was obtained. With this approach, we obtained a converged PMF for the methyl transfer reaction in catechol *O*-methyltransferase (COMT). Consistent with results by Warshel and co-workers,[22] it shows that the *electrostatic* interaction energies between the QM region and the MM region can be converged to high accuracy. Interactions within the QM region are by definition completely converged in our application since the QM region was fixed in the sampling process.

Here we discuss the approach in a more general framework and study the importance of sampling the phase space of the MM region and ways to limit the computational cost. In

particular, we show that the QM/MM-FE method may be a fair approximation to our approach.

The article is organized as follows. In the next section, we describe the methodology employed, including our treatment of link atoms, and a derivation of the QM-FE and QM/MM-FE approaches. A description of the computational setup follows. Results for the methyl transfer reaction catalyzed by COMT are discussed in section 3. We use two different thermodynamic cycle approaches and demonstrate that free energies can be extrapolated accurately to a different basis set and exchange-correlation functional. The report ends with some concluding remarks in section 4.

2. Method

2.1. Thermodynamic Cycle Approach. The thermodynamic cycle depicted in Figure 1 makes it possible to compute high-level QM/MM free energy changes between two states A and B based on classical sampling. In this approach, the free energy change between A and B described by QM/MM is calculated as the sum of three terms: (1) the negative free energy change between A described by MM and by QM/MM ($-\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{A})$), (2) the free energy change between A and B with both described by the MM potential ($\Delta A_{\mathrm{mm}}(\mathrm{A}\rightarrow\mathrm{B})$), and (3) the free energy change between B described by the MM potential and by QM/MM ($\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{B})$). Hence

$$\Delta A_{\mathrm{qm/mm}}(\mathrm{A}\rightarrow\mathrm{B}) = -\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{A}) + \Delta A_{\mathrm{mm}}(\mathrm{A}\rightarrow\mathrm{B}) + \Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{B}) \quad (1)$$

The methodology is quite general and can be applied to compute free energy changes between different chemical species. For instance, A and B in Figure 1 can be a protonated and deprotonated group in order to compute a pK_a value. The methodology can also be used to compute the free energy along a reaction coordinate as indicated in the lower panel of Figure 1. We term the method the quantum mechanical thermodynamic cycle perturbation (QTCP) method. Because the shape of the thermodynamic cycle in Figure 1 resembles an edged U, we use the acronym QTCP-U when referring specifically to the cycle in Figure 1 (as opposed to the one in Figure 2).

Each of the three terms in the above equation can be calculated by means of free energy perturbation (FEP).[33,34] In the FEP approach, a free energy change, $\Delta A(0\rightarrow 1)$ is calculated as

$$\mathrm{e}^{-\Delta A(0\rightarrow 1)/k_{\mathrm{B}}T} = \langle \mathrm{e}^{-(E_1-E_0)/k_{\mathrm{B}}T}\rangle_0 \quad (2)$$

where k_{B} is Boltzmann's constant and T is the temperature. E_0 and E_1 are the energies of system 0 and 1, respectively, and $\langle ... \rangle_0$ indicates an ensemble average for system 0. Hence, the free energy changes in eq 1 can all be calculated as an ensemble average for the system described by the MM potential, meaning that a QM/MM free energy change can be obtained without sampling the QM/MM potential surface but entirely by sampling the surface of the MM potential.

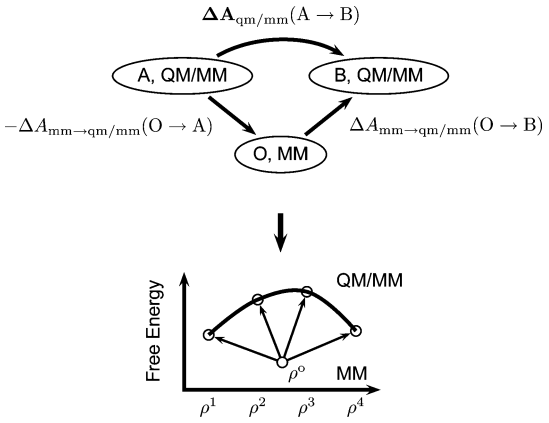

Figure 2. The QTCP method, where a single simulation is used to compute a QM/MM free energy change (QTCP-V). The lower panel shows how this approach can be used to obtain a free energy barrier along a reaction pathway.

Using FEP, the middle term on the right side of eq 1 can be written as

$$\Delta A_{\mathrm{mm}}(\mathrm{A}\rightarrow\mathrm{B}) = -k_{\mathrm{B}}T \ln \langle \mathrm{e}^{-[E^{\mathrm{tot}}_{\mathrm{mm}}(\mathrm{B})-E^{\mathrm{tot}}_{\mathrm{mm}}(\mathrm{A})]/k_{\mathrm{B}}T}\rangle_{\mathrm{mm,A}} \quad (3)$$

and the first and last terms can be written as

$$\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(X) = -k_{\mathrm{B}}T \ln \langle \mathrm{e}^{-[E^{\mathrm{tot}}_{\mathrm{qm/mm}}(X)-E^{\mathrm{tot}}_{\mathrm{mm}}(X)]/k_{\mathrm{B}}T}\rangle_{\mathrm{mm,X}} \quad X = \mathrm{A, B} \quad (4)$$

In the equations above, $E^{\mathrm{tot}}_{\mathrm{qm/mm}}$ and $E^{\mathrm{tot}}_{\mathrm{mm}}$ are the QM/MM and MM potential energies.

The approach above demands an MD simulation per state involved, but the free energy change can in principle be computed from a single simulation. Figure 2 shows how this can be pursued using a single simulation of an intermediate state, O, which we will term an anchor point. In this case, the free energy change, $\Delta A_{\mathrm{qm/mm}}(\mathrm{A}\rightarrow\mathrm{B})$, is calculated as the sum of two terms

$$\Delta A_{\mathrm{qm/mm}}(\mathrm{A}\rightarrow\mathrm{B}) = -\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{O}\rightarrow\mathrm{A}) + \Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{O}\rightarrow\mathrm{B}) \quad (5)$$

where

$$\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{O}\rightarrow X) = -k_{\mathrm{B}}T \ln \langle \mathrm{e}^{-[E^{\mathrm{tot}}_{\mathrm{qm/mm}}(X)-E^{\mathrm{tot}}_{\mathrm{mm}}(X)]/k_{\mathrm{B}}T}\rangle_{\mathrm{mm,O}} \quad (6)$$

Figure 2 shows that the idea also can be applied to compute free energy changes between several points on a reaction pathway. We refer to this particular implementation of the QTCP method by QTCP-V, because the shape of the cycle in Figure 2 resembles a V.

2.2. Definition of Energy Terms. Before discussing the QM-FE and QM/MM-FE methodologies we need to consider the involved potential energy terms in more detail. We denote coordinates of the atoms in the QM and MM region by R_{q} and R_{m}, respectively. We start by defining the MM energy, $E^{\mathrm{tot}}_{\mathrm{mm}}$, in eqs 3, 4, and 6. This term can be separated into three terms

$$E^{\mathrm{tot}}_{\mathrm{mm}} = E^{\mathrm{m}}_{\mathrm{mm}}(R_{\mathrm{m}}) + E^{\mathrm{q/m}}_{\mathrm{mm}}(R_{\mathrm{q}}, R_{\mathrm{m}}) + E^{\mathrm{q}}_{\mathrm{mm}}(R_{\mathrm{q}}) \quad (7)$$

where E^{m}_{mm} describes interactions between MM atoms and $E^{q/m}_{mm}$ describes interactions between the MM and QM regions. The last term, E^{q}_{mm}, describes interactions within the QM region. All these classical terms are defined via a force field, and they can be calculated by standard MM programs such as CHARMM.[35]

To ease the discussion to come, we can further decompose $E^{q/m}_{mm}$ into two terms

$$E^{q/m}_{mm}(R_q, R_m) = E_{ne}(R_q, R_m) + E_{elec}(R_q, R_m) \quad (8)$$

where E_{ne} describes all interactions between the MM region and the QM region besides point charge interactions, which are described by the last term. This term is given by

$$E_{elec}(R_q, R_m) = \sum_{\alpha\in mm}\sum_{\beta\in qm}\frac{Z_\alpha Z_\beta}{|R_\alpha - R_\beta|} \quad (9)$$

where the summations run over MM and QM atoms, and where Z_α and Z_β are point charges of MM and QM atoms, respectively.

In a similar fashion, the total energy, $E^{tot}_{qm/mm}$, can be decomposed into a term that depends solely on the positions of the MM atoms and terms that depend on the positions of both sets of atoms

$$E^{tot}_{qm/mm} = E^{m}_{mm}(R_m) + E_{qm/mm}(R_q, R_m) + E_{corr}(R_q, R_m) \quad (10)$$

where E^{m}_{mm} is the same as in eq 7. E_{corr} is a link atom correction term, which is discussed in the next section.

The term, $E_{qm/mm}$ in eq 10, is composed of a quantum mechanical term, E_{quant}, and a classical interaction term, E_{ne}

$$E_{qm/mm}(R_q, R_m) = E_{quant}(R_q, R_m) + E_{ne}(R_q, R_m) \quad (11)$$

where E_{ne} is the same as in eq 8. In density functional theory notation, the quantum mechanical term is given by

$$E_{quant}(R_q, R_m) = \bar{E}_{vac}(R_q)[n(R_q, R_m)] + \bar{V}_{ext}(R_q, R_m)[n(R_q, R_m)] \quad (12)$$

In this equation, density functionals are indicated by bars, and n denotes the ground-state electron density of the functional $\bar{E}_{vac} + \bar{V}_{ext}$. $\bar{E}_{vac}$ is the density functional for the system in a vacuum, i.e., without external point charges, and $\bar{V}_{ext}$ describes the electrostatic interactions between QM and MM atoms and is given by

$$\bar{V}_{ext}(R_q, R_m)[n] = -\int dr \sum_{\alpha\in mm}\frac{Z_\alpha n(r)}{|R_\alpha - r|} + V_{nuc,\,mm}(R_q, R_m) \quad (13)$$

where the last term defines electrostatic interactions between MM atoms and QM nuclei. Eq 13 is the QM analogue to eq 9. E_{quant} involves polarization of the QM region by the MM region and can be computed by a QM program where external point charges can be included, such as Turbomole.

2.2.1. Link Atom Correction. When calculating E_{quant}, special action is needed when there is a covalent bond between the QM region and the MM region, and several ways to handle such junctions are described in the literature.[36]

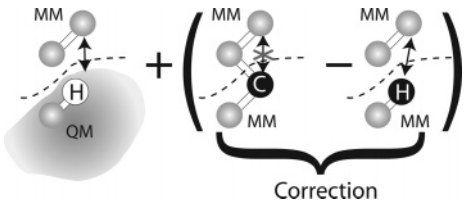

Figure 3. Illustration of the link atom correction employed in the present study. The white ball is a hydrogen link atom, and the black ball represents the corresponding carbon link atom. In the MM system (center) there are no electrostatic interactions between atoms separated by one or two bonds, whereas those interactions exist in the QM/MM system across the MM-QM interface (left). The bond across the interface is removed in the MM calculation to the right, and, hence, the electrostatic interactions across the interface mimic those of the QM/MM calculation.

We use the hydrogen link atom approach, in which the QM region is truncated by hydrogen atoms. The atoms that are converted to hydrogen atoms will be called carbon link atoms. Bond lengths between hydrogen link atoms and adjoining QM atoms are changed such that the relative stretch (or compression) of a given bond from its equilibrium length is preserved compared to that of the bond between the corresponding carbon link atom and QM atom.[37] In our methodology, the coordinates of the link atoms always belong to R_q.

Point charges on MM atoms joining the link atoms are often manipulated to avoid overpolarization of the QM region. In our implementation, we do not manipulate any of the point charges. Instead, we add a correction term, E_{corr}, to eq 11, which is defined as the difference of two classical electrostatic interaction terms

$$E_{corr} = E_{elec} - \tilde{E}_{elec} \quad (14)$$

and which approximately removes unwanted effects from the hydrogen link atoms. E_{elec} is the standard MM electrostatic interaction between the MM region and the QM region (with carbon link atoms) and is defined in eq 9. $\tilde{E}_{elec}$ differs from E_{elec} in that the carbon link atoms are changed to (MM) hydrogen link atoms. This is pursued by deleting the bonds between the link atoms and the MM region and by changing the point charges of the carbon link atoms and reposition them. In that way, $\tilde{E}_{elec}$ models classically the electrostatic interactions, eq 13, which defines the MM-QM electrostatic interactions in E_{quant}.

The above correction modifies only interactions between the MM region and the QM region, whereas interactions entirely within the QM (or MM) region are not changed. Moreover, electrostatic interactions between atoms separated by less than three bonds are approximately removed, and they are scaled if they are between atoms separated by three bonds. This is consistent with the Amber force field employed in the current study. A graphical representation of the correction term is given in Figure 3.

With the correction term, E_{corr}, the free energy between a state described by MM and by QM/MM converges faster because the difference between the MM and QM/MM potential energy surfaces fluctuates less. This is illustrated

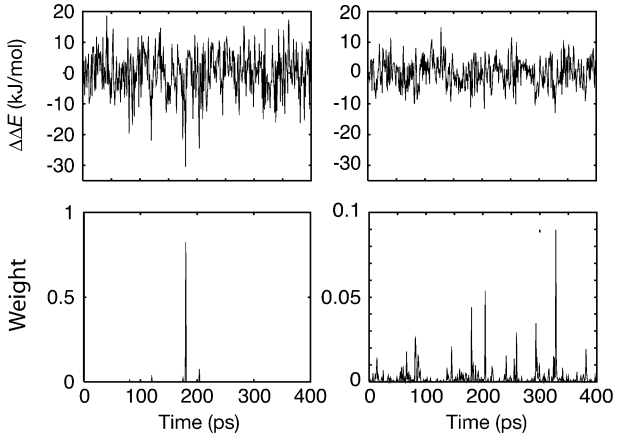

Figure 4. Upper panels show fluctuations of the energy difference between the QM/MM energies and the corresponding MM energies for the simulation with $d_{C-O} = 1.80$ Å. The left panel shows the fluctuations when no link atom correction is employed as opposed to the right panel where a link atom correction is employed. The lower panels show the corresponding contributions to the ensemble average in the free energy perturbation equation.

in Figure 4 where the fluctuations of $(E^{\text{total}}_{\text{qm/mm}} - E^{\text{total}}_{\text{mm}})$ are plotted with and without the correction term included for a simulation of COMT using the setup employed in the current study (section 2.4). The lower panels of Figure 4 show the contributions of the individual configurations to the ensemble averages in the free energy perturbation equations. They demonstrate that many configurations contribute to the free energy if the correction term is employed, whereas the free energy is dominated by a single configuration if the correction is not used.

2.2.2. Free Energy Perturbation Calculations for Fixed QM Regions. In the current application, we use the methodology to compute free energy changes between fixed points along a reaction coordinate as illustrated in Figures 1 and 2. In addition, we fix the entire QM region, such that the free energy depends parametrically of the positions of QM atoms. With that approximation and the definitions in eqs 7−14, eqs 3, 4, and 6 turn into the following formulas

$$\Delta A_{\text{mm}}(R_{\text{q}}^{i} \rightarrow R_{\text{q}}^{i+1}) = -k_{\text{B}}T \ln\langle e^{-[E^{\text{q/m}}_{\text{mm}}(R_{\text{q}}^{i+1},R_{\text{m}}) - E^{\text{q/m}}_{\text{mm}}(R_{\text{q}}^{i},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R_{\text{q}}^{i}} + E^{\text{q}}_{\text{mm}}(R_{\text{q}}^{i+1}) - E^{\text{q}}_{\text{mm}}(R_{\text{q}}^{i}) \quad (15)$$

$$\Delta A_{\text{mm}\rightarrow\text{qm/mm}}(R_{\text{q}}^{i}) = -k_{\text{B}}T \ln\langle e^{-[E_{\text{quant}}(R_{\text{q}}^{i},R_{\text{m}}) - \tilde{E}_{\text{elec}}(R_{\text{q}}^{i},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R_{\text{q}}^{i}} - E^{\text{q}}_{\text{mm}}(R_{\text{q}}^{i}) \quad (16)$$

$$\Delta A_{\text{mm}\rightarrow\text{qm/mm}}(R_{\text{q}}^{\text{o}} \rightarrow R_{\text{q}}^{i}) = -k_{\text{B}}T \ln\langle e^{-[E_{\text{qm/mm}}(R_{\text{q}}^{i},R_{\text{m}}) - E^{\text{q/m}}_{\text{mm}}(R_{\text{q}}^{\text{o}},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R_{\text{q}}^{\text{o}}} - E^{\text{q}}_{\text{mm}}(R_{\text{q}}^{\text{o}}) \quad (17)$$

where R_{q}^{i} and R_{q}^{i+1} are two subsequent points on a reaction pathway. Note, that the terms $\Delta E^{\text{q}}_{\text{mm}}$ cancel in the sum eq 1 and therefore simply can be set to zero. This is of great practical advantage in actual simulations, because no bonded parameters need to be defined within the QM region, i.e., for atoms involved in the chemical reaction.

2.3. QM-FE and QM/MM-FE Approaches. In the QM-FE and QM/MM-FE methods, interactions between the MM and QM regions are described classically. To get to that approximation, starting out from the QTCP method, we consider two different configurations of the MM environment, R_{m}^{0} and R_{m}^{1}, and write eq 12 as

$$E_{\text{quant}}(R_{\text{q}}, R_{\text{m}}^{\alpha}) = \bar{E}(R_{\text{q}}, R_{\text{m}}^{\alpha})[n(R_{\text{q}}, R_{\text{m}}^{\alpha})] = \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R_{\text{m}}^{\alpha})] + \bar{V}^{\alpha}_{\text{ext}}(R_{\text{q}}, R_{\text{m}})[n(R_{\text{q}}, R_{\text{m}}^{\alpha})] \quad \alpha = 0, 1 \quad (18)$$

It follows from the variational principle[38] that

$$E_{\text{quant}}(R_{\text{q}}, R_{\text{m}}^{1}) = \bar{E}(R_{\text{q}}, R_{\text{m}}^{1})[n(R_{\text{q}}, R_{\text{m}}^{1})] = \bar{E}(R_{\text{q}}, R_{\text{m}}^{1})[n(R_{\text{q}}, R_{\text{m}}^{0})] + \text{O}[(n(R_{\text{q}}, R_{\text{m}}^{1}) - n(R_{\text{q}}, R_{\text{m}}^{0}))^2] \quad (19)$$

Hence, provided that the electron density does not fluctuate too much owing to different polarizations by different MM environments, all the QM/MM calculations in the ensemble average can be approximated with nonself-consistent calculations using a frozen density obtained from a single self-consistent calculation with R_{m}^{0}.

At the next level of approximation, we represent the density $n(R_{\text{q}}, R_{\text{m}}^{0})$ by point charges

$$\bar{E}(R_{\text{q}}, R_{\text{m}}^{1})[n(R_{\text{q}}, R_{\text{m}}^{0})] = \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R_{\text{m}}^{0})] + \bar{V}_{\text{ext}}(R_{\text{q}}, R_{\text{m}}^{1})[n(R_{\text{q}}, R_{\text{m}}^{0})] \approx \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R_{\text{m}}^{0})] + E_{\text{elec}}(R_{\text{q}}, R_{\text{m}}^{1}) \quad (20)$$

where E_{elec} is given by eq 9. Thus

$$E_{\text{quant}}(R_{\text{q}}, R_{\text{m}}^{1}) \approx \bar{E}(R_{\text{q}}, R_{\text{m}}^{1})[n(R_{\text{q}}, R_{\text{m}}^{0})] \approx \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R_{\text{m}}^{0})] + E_{\text{elec}}(R_{\text{q}}, R_{\text{m}}^{1}) \quad (21)$$

We restrict the discussion to situations where the QM region is fixed. If the point charges used to represent the QM region in the classical simulations are derived from a QM/MM calculation where the coordinates are given by R_{m}^{0} (so that E_{elec} in eqs 8 and 21 are the same functions), then the above approximation collapses the ensemble average in eq 16 to

$$\Delta A_{\text{mm}\rightarrow\text{qm/mm}}(R_{\text{q}}^{\text{X}}, R_{\text{m}}) \approx \bar{E}_{\text{vac}}(R_{\text{q}}^{\text{X}})[n(R_{\text{q}}^{\text{X}}, R_{\text{m}}^{0})] - E^{\text{q}}_{\text{mm}}(R_{\text{q}}^{\text{X}}) \quad (22)$$

where R_{m}^{0} is the configuration of the MM environment in the calculation used to derive the point charges. With this approximation, and using eq 15, the sum in eq 1 can be written as

$$\Delta A_{\text{qm/mm}}(\text{A} \rightarrow \text{B}) = -k_{\text{B}}T \ln\langle e^{-[E^{\text{q/m}}_{\text{mm}}(R_{\text{q}}^{i+1},R_{\text{m}}) - E^{\text{q/m}}_{\text{mm}}(R_{\text{q}}^{i},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R_{\text{q}}^{i}} + \{\bar{E}_{\text{vac}}(R_{\text{q}}^{\text{B}})[n(R_{\text{q}}^{\text{B}}, R_{\text{m}}^{0})] - \bar{E}_{\text{vac}}(R_{\text{q}}^{\text{A}})[n(R_{\text{q}}^{\text{A}}, R_{\text{m}}^{0})]\} \quad (23)$$

The QM terms, $\bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R_{\text{m}}^{0})]$ for A or B, can be computed by obtaining $n(R_{\text{q}}, R_{\text{m}}^{0})$ from a self-consistent calculation with point charges positioned according to R_{m}^{0},

i.e., by computing $E_{quant}(R_q, R_m^0) = \bar{E}_{vac}(R_q)[n(R_q, R_m^0)] + \bar{V}_{ext}(R_q, R_m^0)[n(R_q, R_m^0)]$. $\bar{E}_{vac}[n(R_q, R_m^0)]$ can then be computed from a nonself-consistent calculation without point charges.

At another level of approximation, $\bar{E}_{vac}[n(R_q, R_m^0)]$ can be computed with $n(R_q, R_m^0)$ represented by point charges in the calculation of $\bar{V}_{ext}(R_q, R_m^0)[n(R_q, R_m^0)]$, such that

$$\bar{E}_{vac}(R_q)[n(R_q, R_m^0)] = E_{quant}(R_q, R_m^0) - \bar{V}_{ext}(R_q, R_m^0)[n(R_q, R_m^0)]$$

$$\approx E_{quant}(R_q, R_m^0) - E_{elec}(R_q, R_m^0) \quad (24)$$

so that eq 23 is modified to

$$\Delta A_{qm/mm}(A \rightarrow B) = -k_B T \ln\langle e^{-[E_{mm}^{q/m}(R_q^{i+1}, R_m) - E_{mm}^{q/m}(R_q^i, R_m)]/k_B T}\rangle_{mm, R_q^i} + (E_{quant}(R_q^B, R_m^0) - E_{elec}(R_q^B, R_m^0)) - (E_{quant}(R_q^A, R_m^0) - E_{elec}(R_q^A, R_m^0)) \quad (25)$$

This form is employed by Yang and co-workers in their QM/MM-FE method.[16,18,19] In practice, it may be a better approximation since the potential terms $\bar{V}_{ext}(R_q, R_m^1)$ and $\bar{V}_{ext}(R_q, R_m^0)$ are treated at the same level of approximation (eqs 21 and 24) and therefore cancelation of errors may occur.

If the point charges are derived from a calculation in a vacuum, such that $n(R_q, R_m^0) = n_{vac}(R_q)$ and $E_{quant} = \bar{E}_{vac}(R_q)[n_{vac}(R_q)]$, then eqs 21 and 25 become identical, and the formalism corresponds to the QM-FE method employed originally by Jorgensen and co-workers[2,3,4] and later by Kollman and co-workers.[13−15]

2.4. Computational Details. We use the methodologies above to compute free energy changes between fixed QM regions along a predetermined reaction pathway for methylation of catecholate catalyzed by COMT. We emphasize that the topic of this article is not the calculation of reaction pathways but rather how to compute free energies once a reaction pathway is known.

2.4.1. Model System. The crystal structure of COMT (pdb code: 1VID) was used as the starting structure.[39] Hydrogen atoms were positioned with the HBuild routine as implemented in CHARMM.[35] The employed QM system is illustrated in Figure 5 and consists of 44 atoms of which five are hydrogen link atoms.[37] The QM system consists of the catecholate molecule, the Mg^{2+} ion coordinated by catecholate, an $S(CH_3)_3^+$ molecule to model *S*-adenosylmethionine, $HCOO^-$, $HCOO^-$, and $HCONH_2$ to mimic the Mg^{2+} ligands: Asp-141, Asp-169, and Asn-170, respectively. In the crystal structure, a water molecule (HOH-400) also coordinates the Mg^{2+} ion, and that is included in the QM system as well. All atoms were allowed to move in the MD simulations, except the 44 QM atoms.

2.4.2. Reaction Pathway. We calculate the free energy along a reaction pathway adopted from another study.[40] Here it is sufficient to give a brief summary. The distance, d_{C-O}, between the acceptor oxygen of catecholate and the carbon atom in the transferred methyl group was used as reaction coordinate. For 10 different values of the reaction coordinate

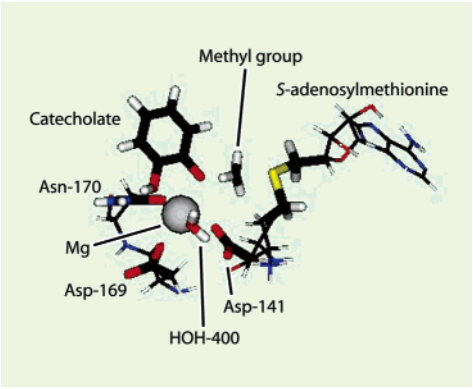

Figure 5. The QM system (bold sticks and sphere) employed in the present study.

with d_{C-O} = 1.47, 1.80, 1.95, 2.00, 2.05, 2.10, 2.13, 2.30, 2.55, and 2.84 Å, the QM region was QM/MM optimized with a constraint on the reaction coordinate and on all the MM atoms. The ComQum program, which in turn is based on the MM software Amber[41] and the QM code Turbomole,[42] was used for these calculations. The transition state was found at 2.13 Å. For each of the points, the RESP protocol[43] as implemented in Amber[41] was used to calculate the point charges in the presence of the MM region using electrostatic potential points obtained by the Merz−Kollman scheme as implemented in Gaussian.[44]

Because a single distance was used as reaction coordinate, it is possible that abrupt changes occur in coordinates orthogonal to the reaction coordinate. This indicates that an improper reaction coordinate is used and that the obtained reaction pathway may be significantly differently than the energy minimum path. It was checked by plotting the progress of other QM degrees of freedom, most noticeably the distance between the transferred methyl group and the donor atom, that this is not the case in the current study, but rather that they progress in a continuous way from the values of the reactant complex to the values of the product complex.

2.4.3. Molecular Dynamics Simulations. Ten molecular dynamics simulations, one for each point along the reaction pathway, were made using the program CHARMM.[35] In these simulations, the QM system was represented by the RESP point charges. The QM region was kept fixed in space during the simulations, and all bond lengths involving hydrogen atoms were constrained by means of the SHAKE algorithm.[45] Atoms of the cofactor and protein that are not part of the QM region were described by the Amber94 force field,[46] and solvents were described explicitly using the TIP3P model.[47] Periodic boundary conditions were employed using an octahedral unit cell and the Particle Mesh Ewald (PME) method[48] with a real-space cutoff of 9 Å to describe the electrostatic interactions. All simulations were performed with a constant volume, and the temperature was kept approximately constant by reassigning the atomic velocities from a Gaussian distribution if the temperature differed by more than 5 K from 298 K.

The 10 simulations were initiated from structures in which the coordinates of the QM system were taken from the QM/MM calculations, whereas the positions of the MM atoms

and the size of the unit cell were taken from the last snapshot of a 200 ps constant pressure (1 atm) and temperature (298 K) simulation for the transition state with $d_{C-O} = 2.13$ Å. The structures were relaxed by 100 steps of steepest-descent minimization with all heavy nonsolvent atoms restrained by a harmonic potential using a mass-weighted force constant of 10 kcal mol^{-1} Å^{-2} amu^{-1}. This was followed by 20 ps of simulation with only the backbone atoms restrained by a harmonic potential using a mass-weighted force constant of 5 kcal mol^{-1} Å^{-2} amu^{-1}. After that, the harmonic restraints were removed. A total of 600 ps of simulations were performed for each of the 10 points defining the reaction pathway.

2.4.4. QM/MM Calculations. For each of the simulations, 400 configurations separated by 1 ps were stored. The configurations were taken from the 200−600 ps time interval of the simulations, and the configurations were used to compute E_{quant} involved in the FEP calculations. In principle, E_{quant} should be computed with periodic boundary conditions employed. However, such calculations are not possible with the QM program employed, and we therefore only include point charges from a single unit cell in the QM calculations as well as in the calculations of $E_{mm}^{q/m}$ in eqs 16 and 17. This is a fair approximation because test calculations for the classical free energy changes defined in eq 3 show that describing the electrostatics in this way rather than with periodic boundary conditions does not change the results significantly.

Density functional theory with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional[49] was employed for the QM calculations. The resolution-of-the-identity (RI) approximation was used for the Coulomb terms,[50] and the one-electron wave functions were expanded in a 6-31G* basis set with an additional set of diffuse s and p functions located on the oxygen and sulfur atoms. We denote this basis set 6-31(+)G*. The calculations were performed with the program TURBOMOLE.[42]

The term, $\bar{E}_{vac}[n(R_q, R_m^0)]$, used in the QM/MM-FE approximation, eq 23, was calculated by making a self-consistent calculation with point charges included. Then, the point charges were removed, and a single iteration (nonself-consistent calculation) was performed using the wave function from the self-consistent calculation.[51]

3. Results and Discussion

3.1. QTCP-U Approach. The results based on the QTCP-U method (Figure 1) are shown in Figure 6 for the enzyme catalyzed reaction. The barrier for forward reaction is 56 kJ/mol and that of the reverse reaction is 78 kJ/mol. The barrier is well converged with standard errors less than 1 kJ/mol.[52] In addition, a difference of 1.6 kJ/mol is obtained for the forward barrier, when individual classical free energy changes (eq 15) based on forward and reverse transitions are used, cf. Table 1.

The contribution to the free energy barrier from the MM environment is indicated by the dashed line. The steepness of the curve shows that the environment has a significant impact on the reaction and emphasizes the importance of a proper description of the environment. If the environment had been omitted, the reaction would occur almost spontanously.

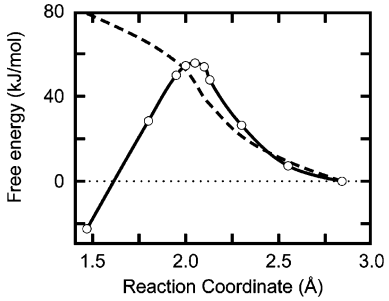

Figure 6. QM/MM free energy barrier at the level of DFT for the methylation of catechol catalyzed by catechol *O*-methyltransferase. The dashed line indicates the contribution to the free energy barrier from the MM environment.

Table 1: Individual FEP Contributions to the Free Energy Barriers Based on the Ensemble Averages in Eqs 15 and 16[a]

			ΔA_{mm}			
d_{C-O}	$\Delta A_{mm\rightarrow qm/mm}$	step	forward	reverse	av	acc
2.84	0.0(0.4)	2.84 → 2.55	9.6(0.2)	9.0(0.2)	9.3	9.3
2.55	−2.1(0.3)	2.55 → 2.30	11.9(0.1)	12.0(0.2)	11.9	21.2
2.30	5.1(0.3)	2.30 → 2.13	14.6(0.2)	15.3(0.2)	15.0	36.2
2.13	11.5(0.5)	2.13 → 2.10	3.1(0.1)	2.9(0.1)	3.0	39.2
2.10	14.7(0.3)	2.10 → 2.05	8.0(0.1)	6.6(0.1)	7.3	46.5
2.05	9.2(0.5)	2.05 → 2.00	5.5(0.1)	6.6(0.1)	6.0	52.5
2.00	1.9(0.7)	2.00 → 1.95	3.5(0.1)	4.1(0.1)	3.8	56.3
1.95	−6.4(0.7)	1.95 → 1.80	9.2(0.1)	8.4(0.2)	8.8	65.1
1.80	−36.8(0.4)	1.80 → 1.47	13.4(0.2)	14.3(0.2)	13.9	79.0
1.47	−101.4(0.4)					

[a] Classical free energy changes are computed for the forward and reverse step, and the averages are listed as well. Accumulated averages (acc) are listed in the last column. All energies are in kJ/mol and reaction coordinates in Å. Statistical standard errors are given in parentheses. Note that $E_{mm}^{q} \equiv 0$, so internal energy changes of the QM system are described exclusively by $\Delta A_{mm\rightarrow qm/mm}$.

Individual contributions from the FEP calculations are listed in Table 1. It can be seen that the barrier is the result of increasingly positive contributions from the classical free energy changes and increasingly negative contributions from the MM → QM/MM calculations. The averages of ΔA_{mm} are used to report the final barrier.

In Figure 7, we investigate how the results vary with the number of data points used in the ensemble averages in eqs 15 and 16. We compare results based on FEP ensemble averages of 20, 50, 100, 200, and 400 data points. It can be seen from the right panel that the barrier height is almost converged already at 20 data points, whereas the left panel shows that additional data points smooth the barrier. However, the transition state at 2.05 Å is not located until 200 data points are used. With 200 data points, the barrier is almost indistinguishable from that using the full data set of 400 points. This shows that it is possible to compute converged free energy barriers with the QTCP approach.

3.2. Extrapolation to Higher Accuracy. The barrier of 56 kJ/mol is lower than the experimental prediction of about 75 kJ/mol.[53] The free energy barrier is converged, but the

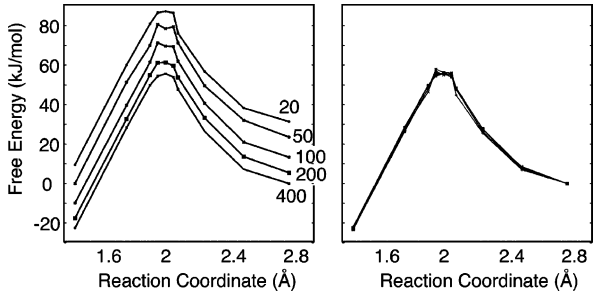

Figure 7. Convergence of free energy barriers with a number of data points utilized. In the left panel, barriers based on 20−400 data points are shifted vertically from one another. In the right panel, the same barriers are superpositioned so that the value at the reactant complex (2.84 Å) is zero.

Table 2: Computed Activation and Reaction Energies (kJ/mol) for Methyl Transfer Using the QM Region Indicated in Figure 5 in a Vacuum[a]

	basis set						
	6-31(+)G*		6-311++G(2d,2p)				
functional	PBE	PBE/RI	PBE	PBE/RI	TPSS/RI	TPSSh	B3LYP
$\Delta E^{\neq}_{fwd}$	12	12	22	22	11	18	27
$\Delta E^{\neq}_{rev}$	111	112	101	101	106	117	115
$\Delta E_{reaction}$	−99	−100	−80	−80	−96	−99	−98

[a] The resolution-of-the-identity (RI) approximation[54] is used in combination with the PBE and TPSS density functionals.

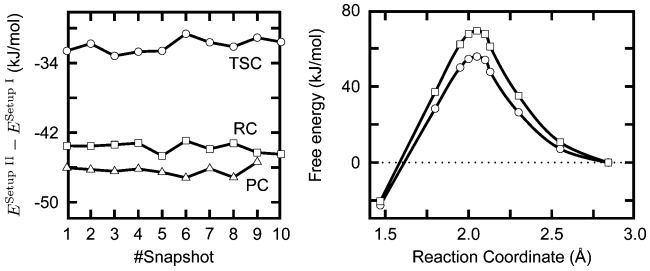

Figure 8. In the left panel, the difference between energies computed with setup II (B3LYP, 6-311++G(2d,2p)) and setup I (PBE, 6-31(+)G*) are plotted for configurations taken from a simulation of the reactant complex (RC), the transition state complex (TSC), and the product complex (PC). A free energy barrier extrapolated from the one based on setup I to one based on setup II (squares) is plotted in the right panel and compared to the one based on setup I (circles).

QM/MM energies depend significantly on the basis set and exchange-correlation functional chosen. This is shown in Table 2, in which the reaction and activation energy for various functionals and basis sets are listed. Energies based on the 6-31(+)G* basis set are not fully converged, and there are also significant changes among the different functionals. On the other hand, energy fluctuations are almost invariant to the basis set and functional employed. This is illustrated in the left panel of Figure 8, where the difference between energies calculated with the original setup (setup I) and a 6-311++G(2d,2p) basis set and the B3LYP hybrid functional (setup II) are plotted for three different simulations. Because the fluctuations are small compared to the shift in average energy along the reaction pathway, it makes sense to extrapolate between the two setups. The result of such an extrapolation is plotted in the right panel in Figure 8.

Using this setup we get a free energy barrier of 69 kJ/mol, which is in excellent agreement with the experimental value of about 75 kJ/mol. The barrier for the reverse reaction is calculated to 90 kJ/mol.

In the extrapolation scheme, a correction term $\Delta E_i^{xpol} = E_i^{setup\ II} - E_i^{setup\ I}$ is added to each point i on the reaction pathway. The strategy which we have chosen and which performs well is one where the same MM environment (i.e. positions and charges of MM atoms) is used to calculate all the ΔE_i^{xpol}. The result shown in Figure 8 are based on energies calculated with the MM environment defined by a configuration taken from the simulation of the transition state, but extrapolation based on calculations with the MM environment turned off, i.e., calculations in a vacuum, performs equally well. The two strategies give almost identical barrier heights; for the forward barrier we compute a barrier of 69 kJ/mol if the TS environment is employed and 71 kJ/mol for the in vacuum corrections. The corresponding barrier heights for the reverse reaction are 90 and 91 kJ/mol, respectively. The agreement between the two strategies validates the extrapolation scheme further.

Cancelation of systematic errors probably occurs when the same MM environment is used to calculate all extrapolated energies along the reaction pathway. In contrast, because the energy difference between setup I and setup II fluctuates somewhat (Figure 8), a rugged free energy barrier is obtained if a different MM environment is used for each point on the reaction pathway. Therefore, we find such a procedure less reliable. A more accurate approach would be to use an average for several energy calculations for each point, but we have not pursued such an approach here.

3.3. CPU Consumption. The total CPU load for the QM/MM calculations using setup I is approximately 8 CPU days on an Intel Xeon 2.4 GHz CPU for the entire free energy barrier based on 4000 QM/MM calculations (∼3 CPU min/calculation). One or two CPU days spend on 10 calculations using setup II should be added if the free energies are extrapolated to higher accuracy (2.5−4 CPU h/calculation).

The fairly small CPU load is due mostly to the RI approach, which saves significant CPU time: a factor of about 6 for the smaller basis set and a factor of about 13 for the larger basis set. With the larger basis set a calculation with the RI approximation takes about 20 CPU min, and, hence, even with the larger basis set, the CPU load does not prohibit convergence of the free energy changes as long as pure density functionals are employed to which the RI approximation applies. We find that the RI approach is sufficiently accurate; the PBE energy changes listed in Table 2 with and without the RI approximation deviate by at most 0.5 kJ/mol from each other.

Finally, the methodology is perfectly suitable for cluster and grid computing since all the single point QM/MM calculations are strictly independent of each other.

3.4. Importance of Sampling. Having established that the environment is important for a proper description of the reaction, it is now interesting to explore the role of sampling. An approach sometimes employed in QM/MM calculations

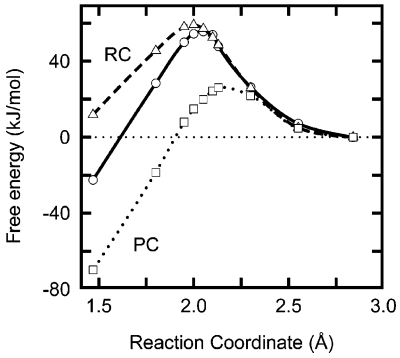

Figure 9. Energy barriers for methyl transfer catalyzed by COMT using a fixed MM enviroment. The curves are averages of 10 configurations taken from a simulation of the reactant complex (RC, dashed) and the product complex (PC, dotted) and compared with the QTCP-U method (solid).

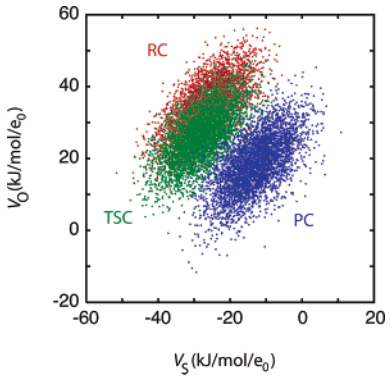

Figure 10. The electrostatic potential caused by the MM environment at the donor atom as a function of that at the acceptor atom in the methyl transfer reaction catalyzed by COMT. The potentials are plotted for a simulation of the reactant (red), transition state (blue), and product (green) complexes.

is to include the environment as a static set of point charges. Figure 9 shows energy barriers computed in this way using the same fixed MM environments (including solvent molecules) for all points on the reaction pathway. Two different configurations for the MM environments are tested. One configuration is taken from a simulation of the reactant complex (dashed curve), and the other is taken from a simulation of the product complex (dotted curve). The former one gives barrier heights of 59 and 47 kJ/mol for the forward and reverse reaction, respectively, and the later one gives barrier heights of 26 and 96 kJ/mol. The difference between the two curves is caused by a change in the polarization of the environment upon going from the reactant complex to the product complex. This can be seen from Figure 10 where the electrostatic potential at the donor atom (V_O) versus that at the acceptor atom (V_S) is plotted for the reactant complex (RC), the transition state complex (TSC), and the product complex (PC). The lack of a significant overlap between the distributions of the product complex and the other complexes also shows that sampling of one complex does not sample important contributions to the free energy of the other complex.[55]

As mentioned previously, omitting the MM environment leads to a very low barrier, if the same QM/MM optimized reaction pathway is employed. A method that has been used extensively in the literature is to omit the MM environment and optimize product, transition state, and the reactant complex in a vacuum. Using this approach with setup II (B3LYP, 6-311+G(2d,2p)) and with a solvent correction, which is based on a continuum model with a dielectric constant of 4, results in a forward barrier of 62 kJ/mol and a barrier of 138 kJ/mol for the reverse reaction. These values should be compared with the extrapolated values of 69 and 90 kJ/mol based on the QTCP-U method (Figure 8). Hence, the forward barrier is fairly well reproduced, whereas the reverse barrier is reproduced less well. The relatively small and large deviations are because the transition state and reactant complexes are similar and polarize the environment in a similar manner as revealed by Figure 10. In contrast, the product and transition state complexes polarize the environment differently in a way that cannot be modeled adequately by the low dielectric continuum model. It should be emphasized that the reaction pathway optimized in a vacuum is different from the QM/MM optimized reaction pathway.

In the calculations in a vacuum, entropy changes and zero-point energies can be estimated from normal mode calculations. Including entropy changes calculated in this way decreases the barrier for forward reaction by 8 kJ/mol and increases that of the reverse reaction by 2 kJ/mol. Entropy changes of the QM region are not included in the results based on the QTCP method, because the QM region is fixed. As a first approximation, the above values can be included. However, the entropy change between the transition-state complex and reactant complex of 8 kJ/mol is probably an overestimation because the orientation of $S(CH_3)_3$ group and the catecholate oxygen in the reactant complex optimized in a vacuum is different from that in the corresponding QM/MM structure. On the other hand, the transition state and product structures are quite similar in the two types of calculations. For these reasons, we do not add these entropy terms to our final results.

3.5. QTCP-V Approach. As described in section 2.1, the actual number of simulations may be reduced if the QTCP-V method (Figure 2) is employed. Here we have tested this approach by computing free energy barriers based on simulations with anchor points at 1.80, 2.05, or 2.55 Å. Using these anchor points, the computed barriers for the forward reaction are found to be 50, 63, or 58 kJ/mol, respectively, and 82, 70 or 61 kJ/mol, respectively, for the reverse reaction. In light of the previous section, these deviations are not surprising since a single simulation does not sample important contributions to all the other points.

We can compute a barrier where each of the three simulations are used only to compute free energy changes in an interval around the point for which the simulation was performed. The three pieces can then be merged together to give the complete free energy barrier. The result from this strategy is plotted in Figure 11 (black line) and compared to the result based on the QTCP-U method in the lower panel. It is seen that the agreement is pretty good. The reaction barrier based on the QTCP-V method is 60 kJ/mol for the

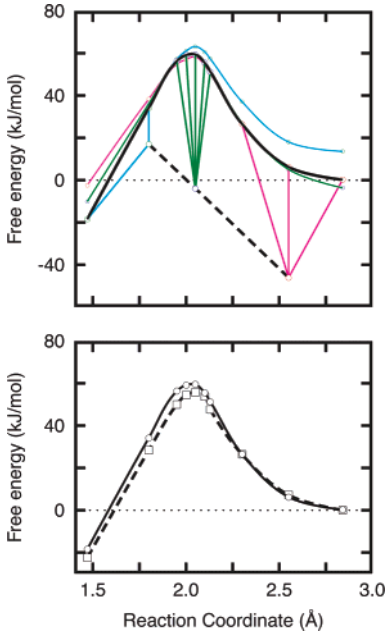

Figure 11. Top: Free energy barriers based on the QTCP-V method using anchor points of $d_{C-O} = 1.80$ Å (cyan), $d_{C-O} = 2.05$ Å (green), or $d_{C-O} = 2.55$ Å (magenta). The bold black curve is a free energy barrier generated by merging pieces from the three other curves together. Bottom: Comparison of the merged black curve from the top panel with the free energy barrier based on the QTCP-U method.

forward reaction and 78 kJ/mol for the reverse reaction, to be compared with 56 and 79 kJ/mol for the QTCP-U method.

The curves based on the 2.55 and 2.05 Å anchor points coincide over a large interval between the two anchor points. In fact, the curve based solely on the 2.55 Å anchor point almost predicts the correct barrier height for forward reaction. This feature is consistent with the scattering plot in Figure 10, which shows that the simulation of the reactant complex (next to the 2.55 Å anchor point) samples important contributions about the transition state complex at 2.05 Å. On the other hand, merging of the curves based on the 2.05 and 1.80 Å anchor points is less well defined. Splicing the two curves either at 1.95 or 1.80 Å gives slightly different results for the reverse barrier height (~4 kJ/mol difference). The solid curve in Figure 11 is based on the average of the two values.

We finally mention that the three curves also can be aligned relative to each other by computing the classical free energy changes between the three anchor points. The single step free energy computations between the anchor points are however associated with great uncertainty. This is revealed by a hysteresis of about 8 kJ/mol for both the 1.80 → 2.05 Å and the 2.05 → 2.55 Å step. Nonetheless, if the averages over the values computed for the forward and reverse transitions are used, almost the same results as that from the splicing of the curves is obtained.

3.6. QM/MM-FE Approach. Figure 12 compares results based on the QM/MM-FE method, eq 23 or 25, with results based on the more exact QTCP-U method. The two QM/MM-FE approaches differ in how the QM energy E_{quant} is approximated (the approximation in eq 24 turns eq 23 into eq 25).

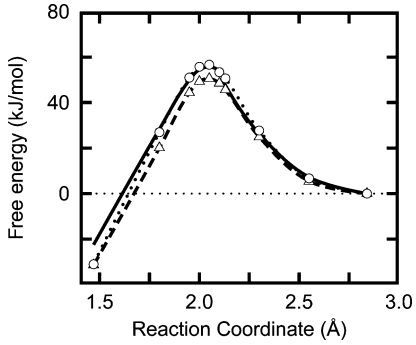

Figure 12. Free energy barrier for the methyl transfer reaction catalyzed by COMT computed by means of the QM/MM-FE approach, eq 23 (circles and dotted line), and compared with the barrier based on the QTCP-U method in Figure 6 (solid line). A free energy barrier based on the QM/MM-FE approach and using the approximation in eq 24 is plotted with the triangles connected with a dashed line.

The results based on eq 23 (dotted line) reproduces the barrier for forward reaction well with a slight overestimation of 1 kJ/mol, whereas the barrier for the reverse reaction is overestimated by 10 kJ/mol. This trend is not surprising, since the point charges employed are derived from a QM/MM calculation where the enzyme is relaxed and the solvent is equilibrated to the reactant complex, and therefore the point charges model the electron density of points close to the reactant complex better than points closer to the product complex.

On the other hand, if eq 25 is used (dashed line), the reaction barrier for forward reaction is underestimated by 5 kJ/mol and that of the reverse reaction by 4 kJ/mol. The similar performance over the entire range might be due to cancellation of errors as discussed in section 2.3.

Nonetheless, with or without the approximation in eq 24 included, the QM/MM-FE approach seems to be a reasonable approximation to the QTCP approach leading to a significant reduction of CPU time consumed.

4. Discussion and Conclusion

We have described a method, QTCP, that makes it possible to compute converged high-level free energy changes. We have used it successfully to compute a free energy barrier for the methyl transfer reaction catalyzed by catechol *O*-methyltransferase. The computations are based on a total of 600 ps simulations per point on a reaction pathway (200 ps equilibration and 400 ps production run) and result in a free energy barrier of 56 kJ/mol using the 6-31(+)G* basis set with the PBE exchange-correlation functional. This value can be extrapolated to 69 kJ/mol using the 6-311++G(2d,-2p) basis set with the B3LYP functional, which is in excellent agreement with an experimental value of about 75 kJ/mol. The results are converged to standard errors to within 1 kJ/mol and with a hysteresis of less than 2 kJ/mol. To obtain such a good convergence we fix the QM region and make a simple link atom correction, which compensates for the improper description of covalent bonds at the interface between the MM and QM region.

An important outcome of our computations is that a significant polarization of the environment occurs as the

reaction progresses. For that reason relaxation of the MM environment is important as obtained through the MD simulations. It is possible that minimization of the total QM/MM energy would give similar results. Electronic polarization of the QM region does not need to be described in great detail. This is illustrated by the fact that energy fluctuations are almost invariant to the exchange-correlation functional and basis set employed. Moreover, using a point charge description of the QM region, as in the QM/MM-FE approach, reproduces the results fairly well.

We have demonstrated that significant CPU time can be saved by using an extrapolation scheme and also that the number of simulations can be reduced. The CPU time can be reduced even further by using the QM/MM-FE approximation.

A major approximation in the current implementation of the QTCP method and in other similar methods is that entropic effects from the internal degrees of freedom of the QM region are ignored or at least considered constant along the reaction pathway. The need for fixing the QM region is caused by the very different potential surfaces for the QM internal degrees of freedom in the QM and MM description of the QM region, which in turn makes the MM → QM FEP calculations difficult to converge, cf. refs 6 and 22. In a first approximation, entropic effects from the QM region can be estimated from normal mode calculations based on optimized QM structures in a vacuum. However, we find such an approach unreliable since the reaction pathway in a vacuum is different from that obtained from the QM/MM reaction pathway. In the current study, this is most noticeable for the forward reaction, since the reaction complex optimized in a vacuum is markedly different from the one optimized by QM/MM calculations. A different approach is to include QM entropic effects by parametrizing the QM region along the reaction pathway and use the parametrized potential surface to describe fluctuations of the QM region.[6,22,56] However, if the parametrization mimics the ab initio QM potential surface well, it should also be possible to obtain converged MM → QM FEP calculations. Indeed, the MM → QM FEP calculations can be considered as a test of a computationally cheaper potential. In this regard, Wood and co-workers find that the FEP calculations can be improved considerably by using better reference potentials. Moreover, the ab initio calculations involved in the MM → QM FEP calculations can be used to optimize the reference potential.[22,26,28,29,31] Warshel and co-workers have also proposed an alternative approach to circumvent the problem of a flexible QM region. The approach is based on the linear response approximation (LRA) and involves sampling of the ab initio potential energy surface in addition to the EVB potential energy surface.[6]

In conclusion, the QTCP and the QM/MM-FE methods seem to be promising methods for computing high-level quantum mechanical free energies. We are currently testing the methods, e.g. for the calculation of proton and electron affinities.

References

(1) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196−1202.

(2) Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974−2975.

(3) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154−163.

(4) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184−189.

(5) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516−17524.

(6) Štrajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333−13343.

(7) Iftimie, R.; Salahub, D.; Wei, D.; Schofield, J. *J. Chem. Phys.* **2000**, *113*, 4852−4862.

(8) Iftimie, R.; Schofield, J. *J. Chem. Phys.* **2001**, *114*, 6763−6773.

(9) Iftimie, R.; Schofield, J. *J. Chem. Phys.* **2001**, *115*, 5891−5902.

(10) Iftimie, R.; Salahub, D.; Schofield, J. *J. Chem. Phys.* **2003**, *119*, 11285−11297.

(11) Iftimie, R.; Schofield, J. *Int. J. Quantum Chem.* **2003**, *91*, 404−413.

(12) Bandyopadhyay, P. *J. Chem. Phys.* **2005**, *122*, 091102.

(13) Stanton, R. V.; Peräkylä, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457.

(14) Kuhn, B.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 2586−2596.

(15) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.

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

(17) Zhang, Y.; Liu, H.; Yang, W. Ab Initio QM/MM and Free Energy Calculations of Enzyme Reactions. In *Macromolecules- -Challenges and Applications*; Schlick, T., Gan, H. H., Eds.; Springer-Verlag's Lecture Series in Computational Science and Engineering, Springer: New York, 2002.

(18) Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2000**, *122*, 6560−6570.

(19) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384−10393.

(20) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035−12048.

(21) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2004**, *126*, 7111−7118.

(22) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293−2301.

(23) Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025−5039.

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

(25) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.

(26) Wood, R. H.; Yezdimer, E. M.; Sakane, S.; Barriocanal, J. A.; Doren, D. J. *J. Chem. Phys.* **1999**, *110*, 1329−1337.

(27) Sakane, S.; Yezdimer, E. M.; Liu, W.; Barriocanal, J. A.; Doren, D. J.; Wood, R. H. *J. Chem. Phys.* **2000**, *113*, 2583−2593.

(28) Sakane, S.; Liu, W.; Doren, D. J.; Shock, E. L.; Wood, R. H. *Geochim. Cosmochim. Acta* **2001**, *65*, 4067−4075.

(29) Liu, W.; Sakane, S.; Wood, R. H.; Doren, D. J. *J. Phys. Chem. A* **2002**, *106*, 1409−1418.

(30) Wood, R. H.; Liu, W.; Doren, D. J. *J. Phys. Chem. A* **2002**, *106*, 6689−6693.

(31) Ming, Y.; Lai, G.; Tong, C.; Wood, R. H.; Doren, D. J. *J. Chem. Phys.* **2004**, *121*, 773−777.

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

(33) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(34) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.

(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(36) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720−1735.

(37) Ryde, U. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153−164.

(38) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−B871.

(39) Vidgren, J.; Svensson, L. A.; Liljas, A. *Nature* **1994**, *368*, 354−358.

(40) Rasmussen, T.; Nilsson, K.; Ryde, U. in preparation.

(41) Case, D. A. et al. AMBER 7, University of California, San Francisco, 2002.

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

(43) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(44) Frisch, M. J. et al. Gaussian 98 (Revision A.1), Gaussian, Inc., Pittsburgh, PA, 11998.

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

(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

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

(49) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(50) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124.

(51) In doing this, it should be ensured that the same grid size is employed in the nonself-consistent calculation as in the last iteration of the self-consistent calculation. In Turbomole notation, we used the grid size 'm3' for the self-consistent calculations, meaning that grid size 3 is used for the very last iteration.

(52) Standard errors were computed using a bootstrap procedure.

(53) Schultz, E.; Nissinen, E. *Biochem. Phamacol.* **1989**, *38*, 3953−3956.

(54) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−289.

(55) We note that the gap between the distributions of the PC and TSC is closed when all simulations are included. This indicates that inclusion of the enzyme and solvent degrees of freedom into the reaction coordinate would not change the barrier height significantly. The lack of a profound hysteresis further supports this point.

(56) Lu, Z.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 89−100.

CT0501102