Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations.

Heimdal, Jimmy; Ryde, Ulf

Published in:
Physical chemistry chemical physics : PCCP

DOI:
10.1039/c2cp41005b

2012

Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations

Jimmy Heimdal & Ulf Ryde *

Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124, SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se, Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

2013-01-29

Abstract

Lately, there has been a great interest of performing free-energy perturbation (FEP) at the combined quantum mechanics and molecular mechanics (QM/MM) level, e.g. for enzyme reactions. Such calculations require extensive sampling of phase space, which typically is prohibitive with density-functional theory or ab initio methods. Therefore, such calculations have mostly been performed with semiempirical QM (SQM) methods, or by using a thermodynamic cycle involving sampling at the MM level and perturbations between the MM and QM/MM levels of theory. However, the latter perturbations typically have convergence problems, unless the QM system is kept fixed during the simulations, because the MM and QM/MM descriptions of the internal degrees of freedom inside the QM system are too dissimilar. We have studied whether the convergence of the MM→QM/MM perturbation can be improved by using a thoroughly parameterised force field or by using SQM/MM methods. As a test case we use the first half-reaction of haloalkane dehalogenase and the QM calculations are performed at with the PBE, B3LYP, and TPSSH density-functional methods. We show that the convergence can be improved with a tailored force field, but only locally around the parameterised state. Simulations based on SQM/MM methods using the MNDO, AM1, PM3, RM1, PDDG-MNDO, and PDDG-PM3 Hamiltonians have slightly better convergence properties, but very long simulations are still needed (~10 ns) and convergence is obtained only if electrostatic interactions between the QM system and the surroundings are ignored.. This casts some doubts on the common practice to base QM/MM FEPs on semiempirical simulations without any reweighting of the trajectories.

Key Words: QM/MM calculations, QTCP, free-energy perturbation, haloalkane dehalogenase, semiempirical calculations

Introduction

During the latest decades, theoretical methods have been established as a powerful complement to experiments for the study of the structure and function of enzymes [1,2,3,4,5,6,7,8]. A great variety of methods have been used, ranging from molecular mechanics (MM), molecular dynamics (MD), and quantum mechanical (QM) calculations, as well as their combinations. For the study of reactions, some type of QM methods or reactive force fields, such as empirical valence bond (EVB) theory and multiconfigurational MM methods, is needed [9,10,11].

Many enzymes have been studied by QM-only cluster methods [1,2,3]. In these, the active site, as well as possibly a few near-by groups (typically ~100 atoms) are cut out of the protein and are studied by QM methods (normally by density functional theory, DFT) in a vacuum or in a continuum solvent. A problem with such an approach is that it is not trivial to know beforehand what groups will be important for catalysis. Therefore, the results may be biased by the selection of the QM system [12].

An alternative method is the combined QM/MM approach [4,5,6,7], in which the active site is still studied by QM, but the QM system is surrounded by a MM model of the protein and solvent. Thereby, the influence of the surroundings is taken into account in a less biased way. However, the QM/MM calculations involve thousands of atoms, which means that there will be an essentially infinite number of local minima available for the system, so it becomes difficult to ensure that all reactants reside in the same local minimum. Therefore, QM/MM energies based on a single minimised structure are quite unstable and may not be representative for the biological reaction. This problem can be reduced by studying several structures, e.g. minimised structures started from several snapshots of a MD simulation [13].

However, a theoretically more attractive approach is to use QM/MM free-energy perturbations (FEP), which in principle should provide the correct results [6,7,8]. Unfortunately, such calculations are extremely expensive, because extensive sampling is needed and the perturbation typically needs to be broken up in several small steps. This means that DFT or high-level QM methods cannot be afforded. Instead, QM/MM FEPs are typically performed at the semiempirical QM (SQM) level, with methods such as AM1 or PM3 [4,6,8,14,15,16]. Unfortunately, SQM methods are fast, but not very accurate. Therefore, the results of the SQM/MM FEPs often need to be extrapolated towards the results of higher-level QM calculations, e.g. by simple single-point energy corrections. This may correct the energies, but it is not certain that the phase space sampled by the approximate SQM methods is the same as with the higher-level methods, i.e. that the entropy calculated at the SQM method is the same as would be calculated with the high-level method. Strictly, the snapshots should be reweighted. This can be accomplished by performing a FEP from SQM/MM to the higher-level QM/MM method [17,18], which can be performed based entirely on the simulations obtained at the SQM/MM level.

Alternatively, the simulations can be performed at the MM level, performing FEPs both between various states at the MM level and from MM to QM/MM, employing the thermodynamic cycle in Figure 1. Such an approach was first suggested by Warshel [19,20,21] and it has then been used by several other groups [22,23,24,25,26], e.g. in the QM/MM thermodynamic cycle perturbation approach, QTCP [27,28]. Unfortunately, such approaches have severe convergence problems, because the MM and QM potentials often are too different. A simple way to solve this problem is to keep the structure of the QM system fixed throughout the MD simulation. Then, the MM→QM/MM perturbation involves only the intermolecular interactions between the QM and MM systems, which show much better convergence properties, but not the internal energy of the QM system [27,28]. On the other hand, this means that the entropy of the QM system is ignored [29,30]. Moreover, it is no longer certain that an increase in the size of the QM system will improve the results [31].

In this paper, we examine whether the convergence of the full MM→QM/MM perturbation can be improved by using a thoroughly parametrised MM method using the

QTCP approach. Moreover, we examine in a similar manner the convergence and quality of SQM/MM→QM/MM perturbations. Thus, we examine if we can perform QTCP calculations in which the QM system is not fixed, but instead treated with a thoroughly parameterised MM method, tailored for the system of interest or at the SQM level. This also gives us the opportunity to test whether a simple extrapolation between SQM and DFT results is valid or if a FEP (reweighting) is needed, and if so, whether the FEP converges.

As a test case we study the first half-reaction in the mechanism of haloalkane dehalogenase (DhlA). This enzyme catalyses the hydrolytic cleavage of C–halogen bonds by a S_N2 mechanism involving a covalent protein intermediate involving an Asp group [32]. This enzyme has been studied theoretically by several groups, using several variants of MM and QM/MM FEP methods [33,34,35,36,37,38,39,40,41,42,43].

Methods

The protein

The DhlA simulations were based on the 1.9 Å-resolution crystal structure 1EDE [44]. The ClH_2CCH_2Cl (1,2-dichloroethane, DCE) substrate was built into this structure by superposing another crystal structure containing this substrate, but obtained at a lower resolution (2DHD) [45]. Protons were added to the protein, assuming standard protonation states at pH 7 for all residues, i.e. all Asp and Glu residues were assumed to be negatively charged and all Lys and Arg residues were positively charged. This assignment was checked with the PROPKA software [46]. The protonation state of the His residues was decided by a detailed study of their local surroundings, solvent exposure, and hydrogen-bond networks. This indicated that DhlA has three doubly protonated His residues (residues 37, 102, and 305) and two His protonated on the $N^{\delta 1}$ atom (residues 54 and 289).

We considered the first step of the reaction of DhlA, starting from a reactant state (RS) consisting of DCE bound to the protein and ending with the formation of the covalent intermediate, $ClH_2CCH_2OAsp + Cl^-$, called the product state (PS), where Asp is the side chain of Asp-124.

MD simulations

Molecular dynamics (MD) simulations were run with the sander module of the Amber 10 software [47], using the Amber 2003 force field for the protein [48]. In all simulations, the SHAKE [49] algorithm was used to constrain the bond length involving hydrogen atoms. The electrostatics were treated with the particle-mesh Ewald method [50,51] with a grid size of 80^3, a fourth-order B-spline interpolation, a tolerance of 10^{-5}, and a real-space cut-off of 8 Å. The temperature was kept constant at 300 K using the Berendsen weak-coupling algorithm [52] with a time constant of 1 ps. The MD time step was 2 fs and the non-bonded pair list was updated every 50 fs.

Three different methods were used for the active site, which consisted of the substrate and the side chain of Asp-124: the general Amber force field (GAFF) [53], an accurate MM force field specifically parametrised for the active site, using the Q2MM approach [54,55,56], and SQM/MM calculations with the six different SQM methods, as implemented in the Amber software [47] (more details are given below).

QM calculations

As the high-level method, we used DFT calculations with the Perdew–Burke–Ernzerhof (PBE) functional [57], together with the def2-SVP basis set [58]. Test calculations were also performed with the B3LYP [59,60] and TPSSH [61] functionals, and the def2-TZVP and

def2-QZVPP basis sets [62]. These calculations were performed with the Turbomole 5.10 package [63]. The structures were optimised until the change in energy between two iterations was below 2.6 J/mol (10^{-6} a.u.) and the norm of the maximum norm of the gradients was below 10^{-3} a.u. Solvation effects were calculated with the COSMO model, implemented in Turbomole, using optimised radii [64,65,66].

QM/MM calculations

QM/MM calculations were carried out with the ComQum program [67,68]. In this approach, the protein and solvent are split into two subsystems: The QM region (system 1) contains the reactive atoms and is relaxed by QM methods, whereas system 2 consists of the rest of the protein and a number of explicitly modelled water molecules. In the QM calculations, system 1 is represented by a wavefunction, whereas all the other atoms are represented by an array of partial point charges, one for each atom, taken from the MM force field. Thereby, the polarisation of the quantum chemical system by the surroundings is included in a self-consistent manner (electrostatic embedding).

When there is a bond between systems 1 and 2 (a junction), the hydrogen link-atom approach was employed: The quantum region is truncated by hydrogen atoms (hydrogen link atoms; HL), the positions of which are linearly related to the corresponding carbon atoms in the full system (carbon link atoms; CL) [67,69].

The QM/MM energy was calculated as:

$$E_{\mathrm{QM/MM}} = E^{\mathrm{HL}}_{\mathrm{QM1+ptch2}} - E^{\mathrm{HL}}_{\mathrm{MM1,noel1}} + E^{\mathrm{CL}}_{\mathrm{MM123,noel1}} \quad (1),$$

where $E^{\mathrm{HL}}_{\mathrm{QM1+ptch2}}$ is the QM energy of the quantum system truncated by HL atoms and embedded in the set of point charges, representing system 2 (but excluding the self-energy of the point charges). $E^{\mathrm{HL}}_{\mathrm{MM1,noel1}}$ is the MM energy of the QM system, still truncated by HL atoms, but without any electrostatic interactions. Finally, $E^{\mathrm{CL}}_{\mathrm{MM123,noel1}}$ is the classical energy of all atoms (including normal CL atoms) and with the charges of the quantum system set to zero (to avoid double-counting of the electrostatic interactions). By using this approach, which is similar to the one used in the Oniom method [70], errors caused by the truncation of the quantum system should cancel out.

The QM/MM geometry optimisations were performed on spherical systems, in which the full DhlA protein was solvated in a cap of water molecules with a radius of 40 Å, in total ~6350 water molecules or ~24 000 atoms (the protein consists of 310 residues). The QM system consisted of the DCE group and an acetate model of Asp-124. Following earlier studies [33,34], we use the anharmonic combination of the C–O and C–Cl bond lengths as the reaction coordinate:

$$\xi = r_{\mathrm{C\text{-}Cl}} - r_{\mathrm{C\text{-}O}} \quad (2)$$

where the C and Cl atoms belong to DCE, whereas O is the OD1 atom of Asp-124. A total of 28 states were optimised with ξ restrained to values along this reaction coordinate from –2.0 to 2.0. The harmonic restraint was 9376 kJ/mol/Å^2, which ensured that the reaction coordinate after optimisation was within 0.01 Å of the expected value.

Standard QTCP calculations

The QTCP approach (QM/MM thermodynamic cycle perturbation) is a method to calculate free energies with a high-level QM/MM method, using sampling at only the MM

level [27,28]. It employs the thermodynamic cycle in Figure 1 and calculates the free energy from three separate terms:

$$\Delta G_{QTCP}(\mathrm{R}\rightarrow\mathrm{P}) = -\Delta G_{\mathrm{MM}\rightarrow\mathrm{QM/MM}}(\mathrm{R}) + \Delta G_{\mathrm{MM}}(\mathrm{R}\rightarrow\mathrm{P}) + \Delta G_{\mathrm{MM}\rightarrow\mathrm{QM/MM}}(\mathrm{P}) \qquad (3)$$

where R and P are the reactant and product states. The standard QTCP calculations were performed as has been described before [27,28]: First each state of interest was optimised by QM/MM, keeping system 2 fixed at the starting crystal structure. Then, the protein was further solvated in an octahedral box of TIP3P water molecules [71], extending at least 9 Å from the spherical QM/MM system, in total ~12 200 water molecules or ~40 500 atoms.

For one state, the system was first subjected to a 100-step minimisation, keeping all heavy atoms, except those in water molecules, restrained towards their positions in the QM/MM structure with a force constant of 418 kJ/mol/Å^2. Then, two 20 ps MD simulations were run with heavy atoms still restrained. The first used a constant volume, the second a constant pressure. Finally, the box size was equilibrated by a 50-ps MD simulation with a constant pressure and only the heavy atoms in the QM system restrained. The final structure of this simulation was then used as the starting structure for the simulations of all the other states (to ensure that the volume of the simulated box was the same for all states). Finally, an equilibration of 200 ps and a production of 400 ps were run with a constant volume. During the production run, 400 snapshots were collected every 1 ps. Separate simulations were performed for each value of the reaction coordinate.

Then, three sets of FEPs were performed on these snapshots as is shown in Figure 1. First, FEPs were performed at the MM level in the forward and reverse direction along the reaction coordinate by changing the charges and coordinates of the QM system to those of the QM/MM calculations for the previous and subsequent steps. The charges were obtained by a fit to the QM electrostatic potential calculated in a number of points around the QM system, sampled according to the Merz–Kollman scheme [72], but with a higher-than-default density of points (~2000 points/atoms). Second, MM→QM/MM FEPs were performed, as has been described before for QTCP [27,28], keeping everything else fixed. All FEPs were performed with the local software calcqtcp (available from the authors upon request). Further details of the QTCP calculations are available in http://teokem.lu.se/~ulf/Methods/qtcp.html.

QTCP-free calculations

QTCP calculations with a flexible QM system were performed in the following way: No QM/MM optimisation was performed. Instead, the protein was set up as for the MD simulations in the QTCP-fixed calculations. Then, the same minimisation and five MD simulations were run as for QTCP-fixed, but keeping only the reaction coordinate fixed, not the whole QM system. In the final step, 200 snapshots were collected every 1 ps. Again, both MM and MM→QM/MM FEP calculations were performed on these snapshots.

The MM→QM/MM or SQM/MM→QM/MM FEP calculations were performed by calculating the total energy with the two methods. Since the force field is identical for all atoms outside the QM system, this means that we sample the differential internal energy of the QM system, as well as its interactions with the surroundings. Thus, the FEP energy correction is calculated from

$$\Delta G_{\mathrm{low}\rightarrow\mathrm{high}} = -kT \ln \left\langle \exp -\left(\frac{E_{\mathrm{high}} - E_{\mathrm{low}}}{kT}\right)\right\rangle_{\mathrm{low}} \qquad (4)$$

where the angular brackets indicate an average over the snapshots from the MD simulation performed with the low-level method (MM or SQM/MM), k is Boltzmann's constant and T is

the absolute temperature.

In practice, details of this perturbation can be varied. For the MM → QM/MM perturbation, we tried three different approaches. The simplest one uses mechanical embedding, implying that the electrostatic interactions between systems 1 and 2 is calculated at the MM level. For the free-energy difference in Eqn. 4, this simply means that all energy components outside system 1 are identical between MM and QM/MM so that

$$\Delta E_{hl} = E_{\text{high}} - E_{\text{low}} = E^{\text{HL}}_{\text{QM1}} - E^{\text{HL}}_{\text{MM1}} \quad (5)$$

i.e. the energy difference of the isolated QM system calculated with QM or MM.

At the intermediate level of sophistication, the QM/MM energy is calculated by electrostatic embedding, i.e. by Eqn. 1. Then ΔE_{hl} will depend also on the difference in the electrostatic interaction energy between systems 1 and 2, calculated either by QM (i.e. by the point-charge model included in the QM calculation) or by MM (the normal electrostatic MM term),

$$\Delta E_{hl} = E^{\text{HL}}_{\text{QM1+ptch2}} - E^{\text{HL}}_{\text{MM1}} - E^{\text{CL}}_{\text{MM-el(1-2)}} \quad (6)$$

where the first two terms on the right-hand side are the same as in Eqn. 1, whereas the third one is the MM electrostatic interaction between systems 1 and 2.

A problem with such an approach is that the electrostatic interactions in the QM calculations are obtained with HL atoms and includes all possible interactions, whereas at the MM level, the calculations involve CL atoms and excludes interactions of atoms that are separated by one or two bonds, and scales down interactions between atoms separated by three bonds. In standard QTCP, this is corrected by performing the MM electrostatic calculation exactly the same way as in the QM calculation (i.e. with HL atoms and no exclusion, $E^{\text{HL}}_{\text{el(1-2)}}$) [28], called the QTCP correction [73]:

$$\Delta E_{hl} = E^{\text{HL}}_{\text{QM1+ptch2}} - E^{\text{HL}}_{\text{MM1}} - E^{\text{HL}}_{\text{el(1-2)}} \quad (7)$$

The calculation of this term involves the use of ESP charges for the QM system and as the geometry of the QM system changes every snapshot, these ESP charges are also recalculated each snapshot.

The challenge for the QTCP-free approach is that the perturbation in Eqn. 4 must converge in a single step if the very time-consuming simulations at the high level should be avoided. We monitored the convergence by calculating the standard error of the free-energy estimate (the standard deviation divided by the number of snapshots), as well as the weight of the maximum component in the average in Eqn. 4 (i.e. how much the snapshot with the most negative difference ΔE_{hl} contributes to the exponential average). We used the method of statistical inefficiency and block averaging to ensure that the sampled data were independent (i.e. that the sampling time was longer than the correlation time) [74,75]. FEP free energies and their uncertainties were estimated by cumulant expressions [74].

Q2MM parametrisation

Specific and accurate MM parametrisations for the reactant, product, and transition states were performed by the ideal iterative method of Norrby and Liljefors [54,55,56], as implemented for Amber [76]. In this approach, parameters for the bond, angle, and dihedral terms are fitted to reproduce the optimised QM structure and the Hessian matrix of each state, by minimising the deviation of geometries and Hessian elements between the DFT and MM

data using a penalty function that gives different weights to different kinds of data. The geometries were described as lists of all bonds, angles, and dihedral angles. The weight factors of the various data types were 100 $Å^{-1}$ for bonds, 2 $degree^{-1}$ for angles, 1 $degree^{-1}$ for torsions, and 0.01–0.1 mole $Å^2$/kcal for Hessian elements (0.01 for elements involving interactions of an atom with itself, 0.02 for atoms bound to each other, 0.04 for atoms connected by two bonds, 0.1 for atoms connected by three bonds, and 0.01 for all other elements) [54].

Van der Waals parameters were not changed from the GAFF (DCE) or Amber-03 (Asp-124) force fields, whereas charges were obtained with the RESP method, using electrostatic potentials sampled with the Merz–Kollman scheme at the HF/6-31G* level. The same RESP charges were used also in the simulations with the GAFF force field. In a second parametrisation, we also tried to use the ESP charges calculated for each snapshot in the QTCP-free calculations (based on the first Q2MM parameterisation), averaged over the 200 snapshots, but this did not improve the QTCP-free results significantly. For the transition state, the imaginary eigenvalue of the Hessian matrix was adjusted to a large positive value (1.0 a.u.) before use, to enable us to model the transition state as a minimum [55]. Hydrogen atoms of the CH_2 groups in DCE and Asp-124 shared the same parameters. Starting parameters for this iterative approach were obtained by the method of Seminario [77], based on the information in the QM Hessian matrix, using the Hess2FF software [78].

The atom types and charges are described in Table 1 and the final force field for bonds, angles, and dihedrals for the three states are given in Table 2.

SQM/MM calculations

The SQM/MM calculations were performed with the Amber 10 software [47], using the SQM method PDDG-PM3 [79,80]. We also tested five additional SQM methods available in the Amber software, AM1 [81], PM3 [82], RM1 [83], MNDO [84], and PDDG -MNDO [79]. The QM system was the same as in the standard QM/MM calculations. The simulation parameters were the same as for the MD simulations in standard QTCP (i.e. with an octahedral periodic box and particle-mesh Ewald treatment of electrostatics). For the SQM/MM→QM/MM perturbations, only the mechanical embedding and uncorrected electrostatic embedding approaches were tested. In the former case, this means that the QM and SQM energies of the isolated QM system are compared.

Result and Discussion

QM results

In this investigation, we study the first half-reaction of haloalkane dehalogenase (DhlA), viz. the simple S_N2 reaction between DCE and Asp-124, forming a covalent intermediate, ClH_2CCH_2O–Asp and a free Cl^- ion. We started the investigation by performing standard QM-only cluster calculations (using an acetate model of Asp-124) of the reactant state (RS; $ClH_2CCH_2Cl + CH_3COO^-$), the transition state (TS), and the product state (PS) ($ClH_2CCH_2OCOCH_3 + Cl^-$). Optimised structures at the PBE/def2-SVP level are shown in the upper part of Figure 2 and the corresponding energies at various levels of theory are given in Table 3, including energies obtained in a COSMO continuum solvent with a dielectric constant of $\varepsilon = 4$ or 80 (a protein-like environment and water).

Following earlier studies [33,34], we use the anharmonic combination of the C–O and C–Cl bond lengths as the reaction coordinate (ξ, defined in Eqn. 2). In the RS, the C–Cl distance is 1.83 Å and the C–O distance is 3.04 Å ($\xi = -1.21$) . The corresponding distances in the TS are 2.33 Å and 2.05 Å ($\xi = 0.28$), and in the PS 3.17 Å and 1.46 Å ($\xi = 1.71$), respectively. From Figure 2, it can be seen that the TS is more similar to the product than to the reactant

state in vacuum.

From Table 3, it can be seen that the activation energy is 71 kJ/mol with the PBE/def2-SVP method in vacuum. It is insensitive to thermal corrections (including entropy and the zero-point energy) and solvation effects, whereas it decreases to 59 kJ/mol if the basis set is extended. B3LYP gives a 6–8 kJ/mol higher activation energy than PBE. On the other hand the TPSSH functional gives a 4 kJ/mol lower barrier.

The reaction energy is –6 kJ/mol with the PBE/def2-SVP method. Thermal effects make it 8 kJ/mol more positive, whereas solvation effects make it much more negative (–44 kJ/mol in water solution), owing to the improved solvation of the Cl^- ion. Extended basis sets (–21 kJ/mol) and the B3LYP functional (–33 kJ/mol) also makes it more negative, as does the TPSSH functional (–34 kJ/mol).

The full reaction path in vacuum is shown in Figure 3. It can be seen that the highest point along the ξ reaction coordinate is close to the true transition state, both in terms of the reaction coordinate (ξ = 0.25) and energies (71 kJ/mol). Consequently, we can employ the ξ reaction coordinate in the QM/MM studies.

Optimum geometries obtained with the B3LYP and TPSSH functionals are closely similar to those of PBE, with ξ = –1.25, 0.30, and 1.75 with B3LYP and –1.22, 0.25, and 1.70 with TPSSH for the RS, TS, and PS, respectively. Consequently, single-point energy calculations on the PBE/def2-SVP structure give energies that are closely similar to those obtained for fully optimised structures (Table 3) with differences of less than 0.2 kJ/mol if the same basis set is used (def2-SVP) and up to 2 kJ/mol with the def2-TZVP basis set.

On the other hand, the SQM PDDG-PM3 energies and structures are quite different. The activation energy, 125 kJ/mol, is much higher than for any other method in this investigation, whereas the reaction energy of –41 kJ/mol is more similar to the other results. However, if the energies are calculated without optimisation on the PBE/def2-SVP structures, the results are very different, 25 and –19 kJ/mol for the activation and reaction energies, respectively. This is also reflected in the geometries, which give ξ = –1.51, 0.19, and 1.68 for the RS, TS, and PS, respectively. This makes the use of SQM structures and extrapolation for QM/MM FEPs somewhat suspicious.

QM/MM results

Next, we studied the reaction inside the DhlA protein with the QM/MM approach. The resulting QM/MM reaction curves are shown in Figure 4. The calculations were performed both at the PBE/def2-SVP and B3LYP/def2-SVP levels of theory. It can be seen that the protein has only a restricted influence on the reaction. With PBE, the activation energy decreased to 62 kJ/mol (72 kJ/mol with B3LYP) and the TS moved to ξ = 0.4 (more product-like). The PS is destabilised by 11 kJ/mol to 5 kJ/mol (1 kJ/mol with B3LYP) and it appears at ξ = 1.7 (PBE) or 1.8 (B3LYP).

The effect of the protein can be understood by decomposing the QM/MM energy into various terms, as is shown in Table 4 [85]. Let us start from the reaction energy in vacuum, E_{vac}, –6 kJ/mol (also shown in Table 3). This energy is directly comparable to the QM energy without the point-charge model of the QM system optimised in the protein with QM/MM, E_{QM1}, which gives a reaction energy of –18 kJ/mol (this energy is also included in Figure 4). The difference, $\Delta E_{QM1} = E_{QM1} - E_{Vac}$, –12 kJ/mol, shows that the change in geometry of the QM system when bound to the protein favours the PS. The reason for this is that the geometry of the RS changes more than that of the PS, as can be seen in Figure 2. In fact, the ΔE_{QM1} energy of the RS is 30 kJ/mol and that of the PS is 17 kJ/mol.

Next, we add the point-charge model of the protein in the QM calculation, i.e. the $E^{HL}_{QM1+ptch2}$ energy in Eqn. 1. It gives a reaction energy of 8 kJ/mol, i.e. the point-charge model more than compensate for the effect of the change in geometry. The net effect is 27 kJ/mol, which represents the direct energetic effect of the surrounding protein and solvent (

$\Delta E_{\mathrm{Ptch}} = E^{\mathrm{HL}}_{\mathrm{QM1+ptch2}} - E_{\mathrm{QM1}}$). The sign of this effect is quite unexpected, because it favours the RS, although the importance of hydrogen bonds of Trp-125 and Trp-175 to the Cl^- ion in the PS have often been emphasized [44,45].

This term can be further understood by decomposing it into contributions of all residues in the protein, as well as of all water molecules. This can be done at the MM level, if the QM system is represented by a point-charge model, obtained from the wavefunction. Of course, this is an approximation (the same used in mechanical embedding QM/MM) that does not give exactly the same results as the QM calculation (the difference is 7 kJ/mol), but it is accurate enough to see the trends. Interestingly, the decomposition shows that the protein actually favours the PS by 67 kJ/mol. The largest contributions come from Trp-175, Hid-289, Lys-176, Glu-56, Lys-224, Asp-260, Gly-55, Asp-178, Lys-221, Arg-220, Glu-94, and Trp-125 in this order (28–10 kJ/mol, with Glu-56, Gly-55, Asp-178, and Glu-94 giving negative contributions). However, this effect is more than compensated by interactions with solvent water molecules, stabilising the RS by 86 kJ/mol, owing to the larger charges of the carboxylate atoms of Asp-124. In fact, a single water molecule, hydrogen bonded to the side chain of Asp-124 contributes by 39 kJ/mol. This shows that the observed energies are sums of a large number of mainly compensating contributions and that it is hard to interpret the energetics based only on structures.

Finally, we can go from the $E^{\mathrm{HL}}_{\mathrm{QM1+ptch2}}$ energy to the full $E_{\mathrm{QM/MM}}$ energy, by adding the two MM terms in Eqn. 1. From Table 4, it can be seen that the effect of the MM terms on the reaction energy ($\Delta E_{\mathrm{MM}} = E_{\mathrm{QM/MM}} - E^{\mathrm{HL}}_{\mathrm{QM1+ptch2}} = E^{\mathrm{CL}}_{\mathrm{MM123,noel1}} - E^{\mathrm{HL}}_{\mathrm{MM1,noel1}}$) is minor, only 3 kJ/mol. This is expected, because the surroundings are fixed, implying that ΔE_{MM} is only the difference in van der Waals energy between the QM system and surroundings for the two states.

A similar analysis can be performed also on the activation energy. From Table 4, it can be seen that ΔE_{QM1} is even larger (more negative) for the activation energy, because that the TS is less strained by the protein (12 kJ/mol) than both RS and PS, indicating that the protein is designed to bind the TS. This effect is partly compensated by the point charges (ΔE_{Ptch} = 10 kJ/mol), a term that again is dominated by effects from water molecules (–44 kJ/mol) and with an opposing effect of the protein. The latter are dominated by contributions from Trp-175, Glu-56 (opposing), Hid-289, Lys-176, Lys-224, Asp-260, and Trp-125 (16–8 kJ/mol in this order). The ΔE_{MM} term is minor.

Considering the large effect of the protein and solvent molecules, as well as the fact that the surroundings have been fixed at the same positions in these QM/MM optimisations, it is of great interest to study the effect of dynamics for the surroundings, using QTCP calculations.

Standard QTCP results

Therefore, we performed standard QTCP calculations. We started from the QM/MM structures (obtained with the surrounding protein and water molecules fixed). We solvated the protein in an extended periodic octahedral box and performed MD simulations with the QM system fixed. This means that the MM→QM/MM perturbation only involves the difference in the interaction energy between the active site and the surroundings, not the internal energy of the QM system. The results are presented in Figure 5, showing calculations with two DFT functionals, PBE and B3LYP (i.e. both the QM/MM and the QTCP calculations were performed at these levels) and the def2-SVP basis set. For both methods, we run two independent sets of calculations (run1 and run2), to estimate the statistical uncertainty.

It can be seen that the results change somewhat between QM/MM and QTCP. At the PBE/def2-SVP level, the activation free energy is ~53 kJ/mol, which is 10 kJ/mol lower than

the QM/MM results. The effect of the surrounding protein is only 2 kJ/mol (column ΔG_{MM} cumul in Table 5), i.e. 8 kJ/mol lower than in QM/MM. Likewise, the reaction free energy is 0 kJ/mol, 5 kJ/mol lower than the QM/MM result. The effect of the protein is ~23 kJ/mol, 4 kJ/mol less than with QM/MM. Thus, in this case, there is no indication that QM/MM should overestimate electrostatic effects in proteins, contrary to what has been observed for several other enzymes [28,31].

The results are statistically well converged, as can be seen in Table 5: The standard error of the MM perturbations is always less than 0.2 kJ/mol (average 0.07 kJ/mol) and the weight of the maximum term in the exponential average is 0.01 on average (i.e. more than 100 terms contribute significantly to the average). However, the hysteresis (difference between the results of the forward and backward MM perturbations) is appreciably larger, up to 4 kJ/mol, with an average of 0.9 kJ/mol. The two independent runs give maximum differences of 1.5 kJ/mol for the MM free energies and mean absolute differences of 0.3 kJ/mol. This gives a better estimate of the precision of the calculations.

For the MM→QM/MM perturbation, the standard error is 0.1–0.3 kJ/mol and the weight of the maximum term in the exponential average is up to 0.12 (average 0.03). The difference between the two independent simulations is up to 3 kJ/mol (0.7 kJ/mol on average). Thus, we can conclude that standard QTCP gives converged results for the DhlA reaction. However, the problem with QTCP is that the QM system needs to be fixed, implying that the entropy of the QM system is ignored. This entropy can estimated from QM frequencies in vacuum calculated at the reactant, product and TS structures, i.e. using the difference between the first two lines in Table 3. In the present case, these corrections are small, 2–8 kJ/mol, making interpolations along the reaction coordinate reasonable. However, this will still ignore the coupling between the QM and MM systems.

Figure 5 shows that the activation free energy is 9 kJ/mol higher for B3LYP than for PBE, whereas the reaction free energy is ~7 kJ/mol lower. This is in reasonable agreement (within 2 kJ/mol) with the difference between the PBE and B3LYP/def2-SVP results in Table 3. In addition, both methods predict the TS to be at ξ = 0.4 Å and the PS to be at ξ = 1.6 Å.

It is common to extrapolate QM calculations to better methods and basis sets. The current small QM system gives us the opportunity to test whether such an extrapolation is valid or whether a reweighting of the snapshots (via FEP) is needed. This was done by running the MM→QM/MM calculations to both the PBE/def2-SVP and B3LYP/def2-TZVP levels (using the same PBE/def2-SVP QM/MM structures and the same MD simulations (run 1) for both perturbations). These results are also included in Figure 5. It can directly be seen that the values in Table 3 cannot be used to correct the QTCP free energies: Table 3 shows that for the optimised vacuum geometries, B3LYP/def2-TZVP predicts a 4 kJ/mol lower activation free energy than PBE/def2-SVP, whereas in Figure 5, B3LYP/def2-TZVP predicts a 2 kJ/mol *higher* activation free energy. For the reaction free energy, the sign of the correction is the same in vacuum and with QTCP, but the size of the correction is nearly half as big with QTCP (–12 kJ/mol) than in vacuum (–22 kJ/mol). This shows that it is quite inaccurate base the extrapolation on vacuum structures. Single-point vacuum energies on the original QM/MM structures give similarly poor results, but with errors in the opposite direction, viz. corrections of 9 and –5 kJ/mol.

This shows that accurate corrections need to be based on the actual QTCP structures, including the point-charge model. A plain average for the difference between the PBE/def2-SVP and B3LYP/def2-TZVP calculations over the 400 snapshots gives corrections of 3.2 ± 0.1 and -14.6 ± 0.1 kJ/mol for the activation and reaction free energies, respectively with ranges of 15 and 18 kJ/mol for the individual snapshots. This shows that proper extrapolations could give reasonable results if enough structures are included in the averages, but for accurate results a full FEP (i.e. with exponential averaging) is needed.

QTCP-free with a GAFF reference

Next, we tried to run QTCP calculations in which the QM system was allowed to move in the MD simulations (only the reaction coordinate was fixed). This implies that the MM→QM/MM perturbation in Eqn. 3 involves the internal energy of the QM system, in addition to the electrostatic interaction between the QM system and the surroundings. We started with using a standard MM force field, viz. the GAFF force field. The results of these calculations are collected in Table 6.

It can be seen that the results are far from converged. The standard error of the MM→QM/MM perturbation is 7–8 kJ/mol around the RS ($\xi = -1.1$), but it increases to 40 kJ/mol around the PS, indicating that ~100 times longer simulations are needed to converge the results to an precision of ~4 kJ/mol. This is confirmed by the weight of the maximum term in the exponential average, which is 0.3–1.0. This illustrates the problem with QTCP-free. The free-energy curve is very noisy, but an activation free energy of ~37 kJ/mol can be estimated, although the PS cannot be identified.

It is notable that the forward and backward MM perturbations also have an appreciably worse precision than for standard QTCP, with standard errors (SE) of 0.1–9 kJ/mol (average 1.7 kJ/mol), i.e. over 20 times worse than for standard QTCP. The reason for this is that these perturbations now include also the internal degrees of freedom of the QM system, which are quite large energy terms with extensive fluctuations. Still, the precision of the net free energies is limited by the MM→QM/MM perturbations.

The results in Table 6 are based on the GAFF force field for the RS. If we instead use the force field for the PS, the results are even worse: The standard error of the MM→QM/MM perturbation is 9–313 kJ/mol (101 kJ/mol on average), and the weight of the maximum element in the exponential average is 0.4–1.0 (average 0.79). These results are essentially useless and it is not even possible to identify the PS and TS, so that the two curves could be combined.

As mentioned in the Methods section, we have tested three methods to estimate the MM→QM/MM free energies. The results in Table 6 were obtained with electrostatic embedding and the QTCP correction in Eqn. 7. However, the results without the correction or with mechanical embedding are similar. For example, the average SE for the MM→QM/MM perturbations with the RS force field changes from 25 to 23 kJ/mol with mechanical embedding and the average maximum weight changes from 0.65 to 0.64.

We have also performed the MM→QM/MM perturbation for a simulation of the unconstrained RS and PS, where the employed GAFF force fields can be expected to be most accurate. This gave SEs of 14 and 26 kJ/mol for the RS and PS, respectively, and maximum weights of 0.99 and 0.88, respectively. These results are slightly worse than the best ones for the full reaction curves. The reason this is probably that the reaction coordinate was fixed in the calculations in Table 6, leading to more restricted dynamics, which is somewhat easier to describe by the MM potential.

QTCP-free with a Q2MM reference

Next, we tested whether a tailor-made force field may improve the convergence. To this end, we employed the ideal procedure, developed by Norrby and Liljefors (Q2MM) [54,55]. Given a certain functional form of the force field (we used the diagonal harmonic force field of Amber), a set of van der Waals parameters, and the RESP atomic charges, this procedure provides an ideal fit of the force-field parameters to the optimised QM structure and the corresponding Hessian matrix. This essentially provides the best possible force field, within a certain functional form [86]. We have performed Q2MM parameterisations for three points along the DhlA reaction, the RS, TS, and PS.

The results of the QTCP calculations based on the Q2MM force field for the RS are

collected in Table 7. It can be seen that around the RS, the results are slightly better than for the GAFF force field, with SEs for the MM→QM/MM perturbation of 5–7 kJ/mol and maximum weights down to 0.13. However, around the PS, the results are appreciably worse, with SEs of 97–209 kJ/mol. The force field for the PS gave worse results, as for the GAFF force field, and the force field for the TS gave completely useless results with an average maximum weight of 0.96.

The reason for this quite disappointing behaviour is probably that the Q2MM force field is strongly biased to accurately reproduce the structure of the parameterised state. It is possible that the results could be somewhat improved by increasing the weight factors for the Hessian elements or by including energies and gradients of structures along the reaction coordinate. However, it seems unlikely that the results will be so much improved that they will be useful in routine use of QTCP.

QTCP-free with a SQM/MM reference

Therefore, we instead tested to use SQM/MM simulations with different Hamiltonians, starting with the PDDG-PM3 method [79,80]. These calculations take approximately 3–4 longer time than the MM simulation. The results are collected in Table 8 and are illustrated in Figure 6. It can be seen the PDDG-PM3 method gives standard errors of 3–21 kJ/mol for the SQM/MM→QM/MM extrapolation, with an average of 10 kJ/mol. Although this is still not fully satisfactorily, it is at least promising, indicating that with up to 25 times longer simulations, the precision could be improved to ~4 kJ/mol. However, the weight of maximum element in the FEP average is still 0.1–0.9 for the SQM/MM→QM/MM perturbation, with an average of 0.5, indicating that the results strongly depend on a few snapshots and therefore are quite unstable. This problem is illustrated in Figure 6, which shows that the free energy curve is quite noisy. Still, the curve allows us to estimate activation and reaction free energies of 75 and –56 kJ/mol.

The SQM/MM simulations themselves also allow us to estimate activation and reaction free energies at the PDDG-PM3 level, 115±4 and –112±4 kJ/mol ($\Delta G_{\mathrm{SQM/MM}}$ cumul column in Table 8, but with RS at $\xi = -1.1$). The former is fairly similar to what is found for the isolated QM system with the PDDG-PM3 method in vacuum, 125 kJ/mol (Table 3), whereas the latter estimate is much more negative (–41 kJ/mol in vacuum). Extrapolating from these data and the corresponding vacuum PBE/def2-SVP energies in Table 3, gives activation and reaction free energies of 61 and –75 kJ/mol, which is 13–21 kJ/mol from the QTCP-free results. Exactly the same result is obtained if we assume that the entropy calculated at the PDDG-PM3 level can be added to the energies obtained at the PBS/def2-SVP levels [15,16]. This shows the problem with extrapolations.

These results were obtained with the MM→QM/MM perturbation performed with mechanical embedding (i.e. estimated from the energy difference of the QM system estimated by PBE and PDDG-PM3). If we instead use electrostatic embedding, the precision becomes much worse, with SEs of 54–182 kJ/mol (average 104 kJ/mol) and weights of the maximum element of 0.3–1.0 (average 0.8). This makes the results practically useless and shows that, the SQM/MM and PBE/MM methods give too different results for the electrostatic interactions between the QM and MM systems (whereas the internal energies of the isolated QM system are similar enough to allow for a single-step perturbation). This also cast some doubts on the trajectories generated by the SQM/MM method.

We have also tested five other SQM methods, available in the Amber software, AM1, MNDO, PDDG-MNDO, PM3, and RM1 [79,81,82,83,84]. From Table 9, it can be seen that the results (obtained with mechanical embedding) are quite similar to those obtained with the PDDG-PM3 method, with average SEs for the SQM/MM→QM/MM perturbation of 7–11 kJ/mol and average maximum weights of 0.5. All methods except MNDO (and AM1 for ΔG^0)

allow the estimation of the activation and reaction free energies. It can be seen that the activation free energies are quite similar, 53–75 kJ/mol, with a precision of 10–24 kJ/mol. The reaction free energies are more varying, –54 to –90 kJ/mol, although the precision is similar. It is notable that the reaction free energy is much more negative with QTCP-free than with standard QTCP (~0 kJ/mol). This is also confirmed by the GAFF results in Table 6, indicating that the reaction free energy is significantly negative, <–50 kJ/mol. Consequently, the results indicate that the PS with its free Cl^- ion is hard to describe with a fixed QM/MM structure and that its estimated free energy may decrease if it is allowed to move in MD simulations. In other words, the entropy of the Cl^- ion is significant and it is hard to estimate with a harmonic approximation.

Conclusions

In this paper, we have studied whether we can release the approximation of a fixed QM system in QM/MM free-energy perturbations and improve the convergence of the MM→QM/MM perturbation in QTCP and similar approaches. We have tested a standard MM force field, a thoroughly parametrised MM force field, and six different SQM methods. As a test case, we have used the first half-reaction of DhlA, i.e. the S_N2 reaction in which the Asp-124 side-chain in DhlA replaces a Cl group in DCE.

The results in Tables 6 and 7 show that a thoroughly parameterised force field gives slightly better results than a standard force field around the parametrised state. However, the results become worse around the other states of the reaction and both the standard and tailored force fields give a poor precision for the MM→QM/MM perturbation. Simulations of at least 20 ns length are needed for a proper precision.

Calculations with a SQM/MM reference give somewhat better results, with SEs for the SQM/MM→QM/MM perturbation of 3–21 kJ/mol. This indicates that ~5 ns simulations are needed for converged results. Unfortunately, such a precision is only obtained with mechanical embedding, i.e. when the SQM/MM→QM/MM perturbation only involves energies of the isolated QM system. If the electrostatic interactions between the QM and MM systems are also included in the perturbation, the precision becomes so poor that the results become useless.

These results show that it is quite problematic to obtain accurate and precise results for QM/MM FEPs of enzyme reactions if you want to avoid sampling at the QM/MM level. Typically, the overlap between distributions sampled by the QM and MM or SQM methods is so poor that very extensive sampling is needed to converge the single-step perturbation. It should be mentioned that alternative methods to perform QM/MM FEPs have been suggested, e.g. using a linear-response approximation [30] or MM-guided enhanced sampling [24,26,87,88,89,90,91]. However, such approaches require simulations or extensive sampling at the QM/MM level. Another alternative is to use specific reaction parameters, i.e. SQM methods parametrised for a specific reaction [92,93,94].

Interestingly, the present results cast strong doubts on the widely use of SQM/MM simulations to obtain QM/MM free energies [4,6,8,56], typically in combination of SQM→QM extrapolations to correct the quite poor SQM energies, or the assumption that the entropy obtained at the SQM/MM level is applicable also at the QM/MM level [15,16]. An extrapolation is an approximation to the more strict FEP calculations employed in this paper. If the SQM/MM→QM/MM perturbation does not converge, it indicates that the phase space sampled by the two approaches is dissimilar and a simple extrapolation scheme is not valid (although it will give the impression of converged results, because the perturbation is never attempted). Instead, a reweighting of the snapshots is needed. Therefore, the results of a simple single-point extrapolation of SQM→QM results cannot be accepted unless it can be shown that the SQM/MM→QM/MM perturbation actually converges.

Acknowledgements

This investigation has been supported by grants from the Swedish research council (project 2010-5025) and by computer resources of Lunarc at Lund University.

References

1 P. E. M. Siegbahn, T. Borowski, *Acc. Chem. Res*. 2006, **39**, 729-738.
2 F. Himo, *Theor. Chim. Acta* 2006, **116**, 232
3 P. E. M. Siegbahn, F. Himo, *J. Biol. Inorg. Chem*. 2009, **14**, 643-651.
4 Lin H.; Truhlar D. G. *Theor. Chem. Acc.*, 2007, **117**, 185-199
5 U. Ryde (2009), in Computational inorganic and bioinorganic chemistry, E. I. Solomon, R. B. King, R. A. Scott, eds. J. Wiley, Sons, Ltd. Chichester, UK, pp. 33-42.
6 H. M. Senn, Thiel W. *Angew. Chem. Int. Ed.* 2009, **48**. 1198-1229.
7 R.Lonsdale, Ranaghan, K.E.; Mulholland, A. J., *Chem. Comm.* 2010, **46**, 2354-2372
8 H. Hu and W. Yang, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573-601.
9 A. Warshel *J. Am. Chem. Soc.* 1980, **102**, 6218
10 Y. Kim, Corchado J. C., Villa J., Xing J., D. G. Truhlar, *J. Chem. Phys.* 2000, **112**, 2718
11 Jensen, F.; Norrby, P.-O.; Theor. Chim. Acta 2003, **109**, 1-7.
12 L. Hu, J. Eliasson, J. Heimdal, U. Ryde (2009) J. Phys. Chem. A, **113**, 11793–11800
13 Clayssens, F.; Ranaghan, K. E.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J.; Chem. Commun. 2005, 5068-5070.
14 Riccardi, D.; Schaefer, P.; Cui, Q.; J. Phys. Chem. B 2005, **109**, 17715-17733.
15 F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel and H.-J. Werner, Angew. Chem., Int. Ed., 2006, **45**, 6856–6859.
16 R. Lonsdale, S. Hoyle, D. T. Grey, L. Ridder, A. J. Mulholland, Biochemistry 2012, **51**, 1774-1786.
17 Allen, M. P.; Tildesley, D. J. Computer Simulation of Liquids; Oxford University Press, New York, 1987.
18 Shirts, M. R.; Mobley, D. L; Chodera, J. D.; Pande, V. S.; J. Phys. Chem. B 2007, **111**, 13052-13063.
19 Muller, R. P.; Warshel, A. *J. Phys. Chem.* 1995**,** ***99***, 17516-17524.
20 Strajbl, M.; Hong, G.; Warshel, A. J. Phys. Chem. B 2002, **106**, 13333
21 Olsson, M. H. M.; Hong, G. Y.; Warshel, A.; J. Am. Chem. Soc. 2003, **125**, 5025-5039.
22 Wood, R. H.; Yezdimer, E. M.; Sakane, S.; Barriocanal, J. A.; Doren, D. J. *J. Chem. Phys.* 1999, ***110***, 1329-1337.
23 Wood, R. H.; Liu, W.; Doren, D. J. *J. Phys. Chem. A* 2002, ***106***, 6689-6693.
24 Iftimie, R.; Salahub, D.; Wei, D.; Schofield, J. *J. Chem. Phys.* 2000, ***113***, 4852-4862.
25 Iftimie, R.; Schofield, J. *Int. J. Quantum Chem.* 2003, ***91***, 404-413.
26 Bandyopadhyay, P. *J. Chem. Phys.* 2005, ***122***, 091102.
27 Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* 2005**,** ***94***, 138302.
28 Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* 2005**,** ***1***, 1240-1251.
29 Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A.; J. Phys. Chem. B 2005, **109**, 15645-15650.
30 Rosta, E.; Klähn, M.; Warshel, A.; J. Phys. Chem. 2006, **110**, 2934-2941.
31 M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde (2008) J. Chem. Theory Comput. **4**, 985-1001
32 Keuning, S; Janssen, D. B.; Witholt, B. *J. Bacteriol.* 1985, **163,** 635–639.
33 Maulitz, A. H.; Lightstone, F.; Zheng, Y.-J.; Bruice, T. C.; Proc. Natl. Acad. Sci. USA

1997, **94**, 6591-6595.
34 Lightstone, F.; Zheng, Y.-J.; Maulitz, A. H.; Bruice, T. C.; Proc. Natl. Acad. Sci. USA 1997, **94**, 8417-8420.
35 Damborský, J.; Kutý, M.; Nemec, M.; Koca, J.; J. Chem. Inf. Comput. Sci. 1997, **37**, 562-568.
36 Lightstone, F.; Zheng, Y.-J.; Bruice, T. C.; J. Am. Chem. Soc. 1998, **120**, 5611-5621.
37 Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C.; Proc. Natl. Acad. Sci. USA 2000, **97**, 9934-9942.
38 Shurki, A.; Strajbl, M.; Villa, J.; Warshel, A.; J. Am. Chem. Soc. 2002, **124**, 4097-4107.
39 Nam, K.; Prat-Resina, X.; Garcia-Viloca, M.; Devi-Kesavan, L. S.; Gao, J.; J. Am. Chem. J. Am. Chem. Soc. 2004, **126**, 1369-1376.
40 Devi-Kesavan, L. S.; Gao, J.; J. Am. Chem. Soc. 2004, **126**, 1532-1540.
41 Olsson, M. H. M.; Warshel, A.; J. Am. Chem. Soc. 2004, **126**, 15167-15179.
42 Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G.; Chem. Rev. 2006, **106**, 3188-3209.
43 Valero, R.; Song, L.; Gao, J.; Truhlar, D. G.; J. Chem. Theory Comput. 2009, **5**, 1-22; 2191.
44 Verschueren, K.H., Franken, S.M., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W. J Mol Biol 1993, **232**, 856-872
45 Verschueren, K. H; Seljee, F.; Rozeboom, H. J.; Kalk, K.H.; Dijkstra, B. W. Nature 1993, **363**, 693-697.
46 Li, H.; Robertson, A. D.; Jensen, J. H. Proteins, Struct. Funct. Bioinf. 2005, **61**, 704-721.
47 D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C.Walker,W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin,and P.A. Kollman (2008), AMBER 10, University of California, San Francisco.
48 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee.; J. Comput. Chem. 2003, **24**, 1999-2012.
49 Ryckaert, J. P, Ciccotti, G., Berendsen, H. J. C. (1977) J. Comput. Phys. **23**, 327-341
50 Darden, T., York, D., Pedersen, L. J. Chem. Phys. **98**, 10089-10092, 1993.
51 Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G. A smooth particle mesh Ewald potential. J. Chem. Phys. **103**, 8577-8592, 1995.
52 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 1984, **81**, 3684-3690.
53 Wang, J. M.; Wolf, R. M.; Caldwell, K. W.; Kollman, P. A.; Case, D. A. J Comput Chem 2004, **25**, 1157-1174
54 Norrby, P.-O.; Liljefors, T.; J. Comput. Chem. 1998, **19**, 1146-1166.
55 Norrby, P.-O.; J. Mol. Struct. (Theochem) 2000, **506**, 9
56 Norrby, P.-O.; Brandt, P. Coord. Chem. Rev. 2001, **212**, 79
57 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
58 F. Weigend & R. Ahlrichs, Phys. Chem. Chem. Phys. **7** (2005) 3297-3305.
59 Becke, A. D. *J. Chem. Phys.* 1993, ***98***, 1372-1377.
60 Hertwig R. H.; Koch, W. *Chem. Phys. Lett.* 1997***, 268***, 345-351.
61 Staroverov, V.N., Scuseria, G.E, Tao, J., Perdew, J.P., J. Chem. Phys. **119** (2003), 12129.
62 F. Weigend & R. Ahlrichs, Phys. Chem. Chem. Phys. 7 (2005) 3297-3305.
63 Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* 1995, ***102***, 346-354.
64 Klamt, A.; Schüürmann, J. *J Chem Soc Perkin Trans 2* 1993***, 5***, 799-805.

65 Schäfer, A.; Klamt A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* 2000***, 2***, 2187-2193.
66 Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem.*, 1998***, 102***, 5074-5085.
67 Ryde, U. *J. Comput.-Aided Mol. Design* 1996, ***10***, 153-164.
68 Ryde, U.; Olsson, M. H. M. *Int. J. Quantum Chem.* 2001, ***81***, 335-347.
69 Reuter, NI.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem.* 2000, ***104***, 1720-1735.
70 Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* 1996, ***100***, 19357-19363.
71 W.L. Jorgensen, J. Chandrasekhar, J. Madura & M.L. Klein. J. Chem. Phys. **79**, 926-935 (1983).
72 B. H. Besler, K. M. Merz, P. A. Kollman, *Atomic Charges Derived from semiempirical Methods*, J. Comput. Chem. 1990, **11**, 431-439.
73 L. Hu, P. Söderhjelm, U. Ryde, J. Chem. Theory Comput., 2011, **7**, 761-777;
74 Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* 2006, ***2***, 452-461.
75 Yang, W.; Bitetti-Putzer. R.; Karplus, M. J. Chem. Phys. 2004, 120, 2618-2628.
76 P. Rydberg, L. Olsen, P.-O. Norrby & U. Ryde (2007) J. Chem. Theory Comput., **3**, 1765-1773
77 Seminario, J. M. *Int. J. Quantum Chem.* 1996, **60**, 1271-1277.
78 K. Nilsson, D. Lecerof, E. Sigfridsson & U. Ryde (2003), Acta Crystallogr. D, **59**, 274-289
79 Repasky, M.P.; Chandrasekhar, J.; Jorgensen, W.L. J. Comput. Chem., 2002, **23**, 1601–1622
80 Acevedo, O.; Jorgensen, W. L.; Acc. Chem. Res. 2010, **43**, 142-151.
81 Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. J. Am. Chem. Soc. , 1985 , **107** , 3902–3909.
82 Stewart, J. J .P. J. Comput. Chem. , 1989 , **10** , 209–220.
83 Rocha, G.B.; Freire, R.O.; Simas, A.M.; Stewart, J.J.P. J. Comp. Chem. , 2006 , **27** , 1101–1111.
84 Dewar, M.J.S.; Thiel, W. J. Am. Chem. Soc. , 1977 , **99** , 4899–4907.
85 K. P. Jensen & U. Ryde (2005) J. Am. Chem. Soc., **127**, 9117-9128
86 L. Hu, U. Ryde J. Chem. Theory Comput., 2011, 7, 2452-2463
87 Iftimie, R.; Salahub, D.; Schofield, J. *J. Chem. Phys.* 2003, ***119***, *11285*.
88 Hetényi, B.; Bernacki, K.; Berne, B. J.; J. Chem. Phys. 2002, **117**, 8203.
89 Gelb, L. D.; J. Chem. Phys. 2003, **118**, 7747.
90 Woods, C. J.; Manby, F. R.; Mulholland, A. J.; J. Chem. Phys. 2008, **128**, 014109.
91 K. E. Shaw, C. J. Woods, A. J. Mulholland, *J. Phys. Chem. Lett*. 2010, **1**, 219-223.
92 A. Gonzalez-Lafont, T. N. Truong and D. G. Truhlar, *J. Phys. Chem*. 1991, **95**, 4618-4627.
93 C. Alhambra, M. L· Sánchez, J. C. Corchado, J. Gao, D. G. Truhlar, *Chem. Phys. Lett.*, 2002, **355**, 388-394.
94 L. Masgrau, K. E. Ranghan, N. S. Scrutton, A. J. Mulholland, M. J. Sutcliffe, *J. Phys. Chem. B*, 2007, **111**, 3032-3047.

Table 1. Atom types, charges, and van der Waals parameters used in the Q2MM parameterisations. CT is a standard Amber-03 atom types, whereas the other atom types are new.

Atom	Atom	Charge			vdW	
	type	RS	TS	PS	*r*	ε
Asp-124						
CA	CT	0.269298	0.329169	0.383543	1.9080	0.1094
CB	ca	-0.401152	-0.559226	-0.665104	1.9080	0.1094
HB	ha	0.054904	0.114773	0.169148	0.6000	0.0157
CG	cb	0.925607	0.899620	0.868184	1.9080	0.1094
OD1	o1	-0.794519	-0.669865	-0.504719	1.6612	0.2100
OD2	o2	-0.787266	-0.702838	-0.589223	1.6612	0.2100
DCE						
C1	c1	-0.094266	0.157562	0.381159	1.9080	0.1094
CL1	l1	-0.264135	-0.646706	-0.906159	2.4700	0.1000
H1	h1	0.150669	0.118521	0.025500	0.6000	0.0157
C2	c2	-0.086148	-0.023985	-0.116303	1.9080	0.1094
CL2	l2	-0.263987	-0.225729	-0.239766	2.4700	0.1000
H2	h2	0.149909	0.094903	0.106744	0.6000	0.0157

Table 2. Q2MM force field parameters for the RS, TS, and PS states in the DhlA reaction. Parameters involving the CT atom type are standard Amber parameters.

Bonds	Force constant			Equilibrium value		
	RS	TS	PS	RS	TS	PS
ha-ca	362.483	116.642	404.646	1.0956	1.1105	1.0857
ca-cb	147.814	44.952	206.873	1.5780	1.7026	1.5495
cb-o1	551.904	43.790	412.203	1.2591	1.3997	1.3193
cb-o2	583.110	68.969	950.219	1.2569	1.2950	1.1907
c1-l1	111.466	275.141		1.7960	2.3656	
c1-h1	366.209	142.031	426.099	1.0879	1.0714	1.0770
c1-c2	76.687	7.261	238.337	1.4580	1.6059	1.4982
c1-o1		258.499	136.824		2.0730	1.4136
c2-l2	89.964	31.223	186.036	1.7879	1.7814	1.7995
c2-h2	367.269	122.453	424.274	1.0864	1.0936	1.0764
CT-ca	310.000	310.000	310.000	1.5260	1.5260	1.5260
Angles	Force constant			Equilibrium angle		
	RS	TS	PS	RS	TS	PS
ha-ca-ha	35.446	1.933	39.530	107.991	108.575	108.602
ha-ca-cb	54.106	30.379	74.101	110.798	113.755	110.922
ca-cb-o1	6.370	13.628	76.820	115.587	115.792	111.617
ca-cb-o2	62.814	13.607	89.208	116.698	120.669	123.938
o1-cb-o2	124.795	22.030	142.545	128.111	124.933	123.166
cb-o1-c1	21.425	114.310		118.656	119.822	
h1-c1-l1	103.835	23.993		106.330	90.917	
h1-c1-c2	18.975	3.525	42.898	109.754	119.980	109.616
h1-c1-h1	36.894	23.635	42.745	110.595	118.930	108.746
c2-c1-l1	212.880	18.490		112.144	100.998	
o1-c1-l1		51.867			162.580	
c2-c1-o1		0.267	108.136		97.312	112.093
h1-c1-o1		48.708	57.886		87.385	107.638
c1-c2-l2	243.620	39.110	295.537	112.198	110.419	111.306
c1-c2-h2	8.422	1.765	9.720	109.783	110.638	110.552
l2-c2-h2	69.402	29.071	77.593	106.021	104.235	105.639
h2-c2-h2	43.005	25.168	46.389	110.329	108.990	110.428
ca-CT-C	63.000	63.000	63.000	111.100	111.100	111.100
H1-CT-ca	50.000	50.000	50.000	109.500	109.500	109.500
N-CT-ca	80.000	80.000	80.000	109.500	109.500	109.500
CT-ca-ha	50.000	50.000	50.000	109.500	109.500	109.500
CT-ca-cb	63.000	63.000	63.000	111.100	111.100	111.100
Dihedrals	Period	Phase		Force constant		
				RS	TS	PS
ha-ca-cb-o1	3	0		0.062	0.795	0.407
ha-ca-cb-o2	3	180		0.143	0.067	0.000
ca-cb-o1-c1	2	180			0.312	4.504
o2-cb-o1-c1	2	180			0.438	7.305
cb-o1-c1-h1	3	0			0.056	0.011
cb-o1-c1-c2	3	0			0.566	0.002
cb-o1-c1-l1	3	180			0.429	
l1-c1-c2-l2	6	180		1.498	0.037	
l1-c1-c2-h2	6	180		0.337	0.005	
h1-c1-c2-l2	6	180		0.886	0.000	0.186

h1-c1-c2-h2	6	180	0.001	0.001	0.040
o1-c1-c2-l2	6	180		0.000	38.225
o1-c1-c2-h2	6	180		0.082	0.000
cb-ca-CT-C	3	0	1.4	1.4	1.4
ha-ca-CT-C	3	0	1.4	1.4	1.4
H1-CT-ca-ha	3	0	1.4	1.4	1.4
H1-CT-ca-cb	3	0	1.4	1.4	1.4
CT-ca-cb-o1	2	0	0.0	0.0	0.0
CT-ca-cb-o2	2	0	0.0	0.0	0.0
N-CT-ca-ha	3	0	1.4	1.4	1.4
N-CT-ca-cb	3	0	1.4	1.4	1.4

Table 3. Results of QM calculations of the active-site model of DhlA (kJ/mol), using the PBE, B3LYP, TPSSH, and PDDG-PM3 methods with three different basis sets either for the energy or the geometry calculations, and either in vacuum or in a continuum solvent with a dielectric constant of 4 (ε = 4) or 80 (ε = 80). Δ*G* includes the zero-point energy and thermal corrections to the Gibbs free energy, obtained from a frequency calculation at the PBE-def2-SVP level.

Energy	Geometry	$\Delta G^{\#}$	ΔG^{0}
PBE/def2-SVP	PBE/def2-SVP	71.3	-5.7
Δ*G*		73.5	2.2
ε = 4		71.6	-30.4
ε = 80		71.9	-43.7
PBE/def2-TZVP		60.7	-15.7
PBE/def2-QZVPP		59.5	-21.4
B3LYP/def2-SVP		79.3	-14.4
B3LYP/def2-TZVP		66.9	-26.3
B3LYP/def2-QZVPP		65.4	-32.5
TPSSH/def2-SVP		65.1	-21.8
TPSSH/def2-QZVPP		55.6	-34.0
PDDG-PM3		25.0	-19.2
B3LYP/def2-SVP	B3LYP/def2-SVP	79.5	-14.5
B3LYP/def2-TZVP	B3LYP/def2-TZVP	66.9	-28.1
TPSSH/def2-SVP	TPSSH/def2-SVP	65.2	-22.0
PDDG-PM3	PDDG-PM3	125.4	-40.9

Table 4. Decomposition of the QM/MM energies for the reaction and activation energies (kJ/mol). The various energy terms are explained in the text and they are obtained at the PBE/def2-SVP level.

	E_{vac}	E_{QM1}	$E_{QM1+ptch2}$	$E_{QM/MM}$	ΔE_{QM1}	ΔE_{Ptch}	ΔE_{MM}
$\Delta G^{\#}$	71.3	53.0	62.9	62.2	-18.3	9.8	-0.7
ΔG^{0}	-5.7	-18.2	8.3	5.2	-12.5	26.5	-3.1

Table 5. Results from the standard QTCP calculations (PBE/def2-SVP, run 1). The free energies are in kJ/mol. SE is the standard error of the free energies. Weight is the weight of the maximum term in the exponential average. ΔG is the final QTCP free-energy estimate. Note that the $\Delta G_{MM\rightarrow QM/MM}$ applies to given point of ξ, whereas the various ΔG_{MM} values are the differences between the given point and the previous point. ΔG_{MM} cumul is the cumulative average MM free energy. Both this and the QM/MM energy are calculated relative to the free energy at $\xi = -1.0$.

ξ	$\Delta G_{MM\rightarrow QM/MM}$			ΔG_{MM} forward			ΔG_{MM} backward			Hyst-	ΔG_{MM}	ΔG
	Energy	SE	weight	Energy	SE	weight	Energy	SE	weight	eresis	cumul	
-2.0	9.0	0.1	0.04								8.9	17.9
-1.9	7.0	0.1	0.04	0.1	0.1	0.01	0.8	0.0	0.01	0.9	8.5	15.5
-1.8	6.1	0.1	0.01	-0.5	0.1	0.01	0.8	0.0	0.01	0.3	7.9	14.0
-1.7	5.6	0.1	0.01	-0.9	0.0	0.01	0.7	0.1	0.01	0.1	7.1	12.7
-1.6	4.0	0.1	0.04	-0.9	0.0	0.01	0.9	0.0	0.01	0.1	6.2	10.2
-1.5	2.6	0.1	0.03	-0.8	0.1	0.01	0.7	0.0	0.01	0.1	5.4	8.0
-1.4	2.1	0.1	0.02	-0.9	0.0	0.01	1.0	0.0	0.01	0.2	4.5	6.6
-1.3	0.5	0.1	0.02	-0.9	0.0	0.01	0.9	0.0	0.01	0.0	3.6	4.1
-1.2	-0.4	0.1	0.05	-1.0	0.0	0.01	1.0	0.0	0.00	0.0	2.6	2.2
-1.1	-0.9	0.1	0.02	-1.1	0.0	0.01	1.0	0.0	0.01	0.1	1.6	0.6
-1.0	0.0	0.1	0.01	-1.1	0.0	0.01	2.0	0.1	0.01	0.8	0.0	**0.0**
-0.8	1.0	0.1	0.02	-1.0	0.0	0.01	0.8	0.1	0.01	0.2	-0.9	0.1
-0.6	7.4	0.1	0.02	-1.7	0.1	0.02	-0.2	0.1	0.01	1.9	-1.6	5.7
-0.4	17.9	0.1	0.02	-0.6	0.1	0.02	-0.5	0.1	0.01	1.1	-1.7	16.2
-0.2	30.2	0.1	0.03	0.0	0.1	0.01	-0.9	0.1	0.02	1.0	-1.3	29.0
0.0	41.5	0.1	0.01	0.5	0.1	0.01	-1.2	0.1	0.01	0.6	-0.4	41.1
0.2	49.9	0.1	0.02	0.7	0.1	0.01	-1.0	0.1	0.01	0.3	0.4	50.3
0.4	51.1	0.1	0.02	1.0	0.1	0.02	-1.9	0.1	0.02	0.9	1.8	**52.9**
0.6	45.5	0.1	0.05	1.5	0.1	0.02	-2.4	0.1	0.03	0.9	3.8	49.3
0.8	32.2	0.1	0.04	1.5	0.1	0.01	-4.1	0.1	0.02	2.7	6.6	38.8
1.0	11.8	0.2	0.03	3.4	0.1	0.02	-5.7	0.1	0.02	2.3	11.1	22.9
1.2	-2.6	0.1	0.02	4.7	0.2	0.02	-4.6	0.1	0.02	0.2	15.7	13.1
1.4	-15.6	0.2	0.04	4.2	0.1	0.01	-4.1	0.1	0.02	0.0	19.9	4.3
1.6	-23.3	0.3	0.03	4.0	0.1	0.01	-2.3	0.0	0.01	1.7	23.1	**-0.3**
1.7	-24.1	0.3	0.06	4.8	0.1	0.02	-2.1	0.0	0.01	2.7	26.5	2.4
1.8	-24.9	0.3	0.12	2.3	0.1	0.01	-2.6	0.0	0.01	0.3	28.9	4.0
1.9	-24.2	0.3	0.05	2.2	0.1	0.01	-1.9	0.1	0.01	0.3	31.0	6.8
2.0	-22.8	0.3	0.12	2.3	0.1	0.01	2.0	0.1	0.01	4.3	31.1	8.3

Table 6. Result of the QTCP-free perturbation, based on MD simulations with the GAFF force field for the RS. $\Delta G_{MM \rightarrow QM/MM}$ was estimated with electrostatic embedding and the QTCP correction in Eqn. 7. The entries are the same as in Table 5.

ξ	$\Delta G_{MM \rightarrow QM/MM}$		ΔG_{MM} forward		ΔG_{MM} backward		Hyst-	ΔG
	SE	weight	SE	weight	SE	weight	eresis	
-2.0	10.9	0.85	0.0	0.00	0.0	0.00	0.0	28.9
-1.9	19.2	0.25	1.7	0.16	2.2	0.29	0.4	8.3
-1.8	10.4	0.72	4.1	0.55	2.8	0.65	0.1	24.3
-1.7	10.7	0.73	3.2	0.14	2.7	0.31	0.2	20.8
-1.6	8.9	0.70	3.5	0.43	3.3	0.23	0.5	11.1
-1.5	7.1	0.34	3.5	0.21	2.3	0.23	0.1	18.5
-1.4	8.9	0.87	2.6	0.18	2.3	0.25	0.6	7.2
-1.3	7.4	0.33	2.3	0.28	3.1	0.13	0.1	10.8
-1.2	8.2	0.66	2.7	0.28	2.3	0.26	0.0	4.5
-1.1	8.3	0.43	2.0	0.32	2.5	0.65	0.3	**0.0**
-1.0	7.1	0.49	1.5	0.37	1.3	0.39	0.0	1.1
-0.8	6.3	0.36	4.7	0.98	8.5	0.46	0.1	-2.2
-0.6	7.4	0.57	2.4	0.53	3.4	0.80	0.1	0.7
-0.4	8.9	0.63	1.1	0.16	1.5	0.44	1.4	4.1
-0.2	10.4	0.33	0.6	0.35	0.8	0.35	2.7	19.1
0.0	6.8	0.95	0.7	0.12	0.7	0.34	0.5	**37.3**
0.2	12.0	0.98	0.7	0.37	0.5	0.09	0.3	34.5
0.4	13.6	0.86	0.6	0.10	0.7	0.09	2.5	32.7
0.6	13.7	0.44	1.0	0.31	0.8	0.10	0.5	29.2
0.8	17.9	0.74	1.1	0.12	0.8	0.45	0.5	10.9
1.0	24.4	1.00	1.1	0.87	1.0	0.18	0.2	-12.1
1.2	28.0	0.98	0.6	0.09	0.7	0.08	0.7	-28.1
1.4	29.9	0.82	0.7	0.19	0.6	0.10	0.4	-31.2
1.6	39.9	0.53	0.6	0.11	0.7	0.20	0.6	-61.4
1.7	39.4	0.70	0.2	0.03	0.2	0.02	0.3	-52.8
1.8	37.5	0.39	0.1	0.03	0.2	0.03	0.1	-53.7
1.9	49.6	0.68	0.4	0.13	0.4	0.07	0.3	-68.0
2.0	234.2	0.97	0.2	0.05	0.2	0.03	0.1	-229.2

Table 7. Result of the QTCP-free perturbation, based on MD simulations with the Q2MM force field for the RS. $\Delta G_{MM \rightarrow QM/MM}$ was estimated with electrostatic embedding and the HL atom correction in Eqn. 7. The entries are the same as in Table 5.

ξ	$\Delta G_{MM \rightarrow QM/MM}$		ΔG_{MM} forward		ΔG_{MM} backward		Hyst-	ΔG
	SE	weight	SE	weight	SE	weight	eresis	
-2.0	30.1	0.87	0.0	0.00	0.0	0.00	0.0	-29.2
-1.9	4.3	0.13	2.5	0.23	2.0	0.58	0.8	9.6
-1.8	5.0	0.53	3.8	0.53	3.9	0.47	0.9	9.8
-1.7	4.8	0.41	4.5	0.66	4.2	0.26	0.2	13.4
-1.6	5.6	0.52	5.7	0.75	4.4	0.48	1.0	6.5
-1.5	6.1	0.30	4.8	0.66	3.6	0.27	0.3	1.7
-1.4	4.7	0.45	4.3	0.29	3.0	0.39	0.4	2.8
-1.3	5.5	0.27	3.2	0.53	2.5	0.35	0.3	-0.2
-1.2	7.1	0.78	2.3	0.18	2.5	0.23	1.2	-6.0
-1.1	4.7	0.56	2.3	0.46	3.0	0.38	1.2	0.0
-1.0	11.0	0.62	2.2	0.48	3.6	0.51	0.4	-13.6
-0.8	6.1	0.48	17.4	0.29	11.4	0.99	0.5	2.0
-0.6	6.6	0.79	4.9	0.75	4.0	0.66	0.1	6.6
-0.4	6.4	0.19	1.8	0.12	4.1	0.22	0.1	22.0
-0.2	10.1	0.44	1.9	0.17	3.3	0.20	0.3	37.4
0.0	15.2	1.00	2.3	0.54	2.4	0.12	0.2	53.4
0.2	17.3	0.98	2.5	0.26	3.2	0.22	0.2	77.0
0.4	16.7	0.86	3.2	0.83	2.6	0.17	0.2	109.2
0.6	22.3	0.21	3.3	0.95	4.6	0.65	0.8	138.1
0.8	71.4	1.00	4.2	0.31	8.2	0.60	0.1	115.5
1.0	87.5	0.96	5.8	0.51	5.4	0.33	0.2	144.3
1.2	96.7	1.00	1.9	0.18	1.6	0.49	1.7	170.5
1.4	160.5	1.00	1.0	0.28	2.0	0.20	1.4	135.8
1.6	156.4	0.91	1.4	0.57	1.2	0.16	1.1	168.8
1.7	142.6	1.00	0.3	0.08	0.2	0.06	0.6	183.0
1.8	125.5	0.95	0.2	0.04	0.2	0.03	0.5	228.2
1.9	208.7	0.86	0.3	0.07	0.5	0.10	1.1	145.4
2.0	118.1	0.82	0.3	0.06	0.3	0.04	0.2	229.4

Table 8. Results of the PDDG-PM3/MM QTCP-free simulations. $\Delta G_{MM\rightarrow QM/MM}$ was estimated with mechanical embedding. The entries are the same as in Table 5.

ξ	$\Delta G_{SQM/MM\rightarrow QM/MM}$			$\Delta G_{SQM/MM}$ forward			$\Delta G_{SQM/MM}$ backward			Hyst-	$\Delta G_{SQM/MM}$	ΔG
	Energy	SE	weight	Energy	SE	weight	Energy	SE	weight	eresis	cumul	
-2.0	17.0	11.0	0.81	0.0	0.0	0.00	0.0	0.0	0.00	0.0	39.9	56.9
-1.9	12.4	15.1	0.45	-8.0	5.1	0.42	4.4	1.7	0.20	3.6	33.7	46.1
-1.8	2.7	20.7	0.59	-10.8	8.9	0.53	6.4	2.1	0.15	4.4	25.1	27.8
-1.7	6.5	19.4	0.57	-7.1	7.1	0.27	5.9	1.9	0.30	1.2	18.6	25.1
-1.6	7.7	19.5	0.44	-6.1	6.7	0.52	2.3	2.9	0.76	3.8	14.4	22.1
-1.5	17.0	13.9	0.23	-9.8	7.5	0.26	-0.5	3.0	0.48	10.3	9.8	26.8
-1.4	12.5	13.9	0.85	-11.6	8.3	0.18	-0.2	2.9	0.21	11.9	4.1	16.5
-1.3	17.3	11.3	0.50	-8.2	6.8	0.39	2.3	2.2	0.27	5.9	-1.2	16.1
-1.2	0.0	18.0	0.37	-0.5	3.7	0.64	-2.9	3.1	0.26	3.3	0.0	**0.0**
-1.1	19.0	9.9	0.48	-5.5	5.0	0.50	-0.8	1.8	0.27	6.3	**-2.4**	16.6
-1.0	11.9	12.6	0.30	3.2	2.0	0.44	-5.9	2.0	0.33	2.7	2.2	14.1
-0.8	10.9	9.7	0.78	0.6	13.6	0.52	-9.6	5.1	0.36	9.0	7.3	18.2
-0.6	-12.7	14.9	0.46	21.4	2.4	0.33	-20.4	2.8	0.24	1.1	28.2	15.5
-0.4	-18.5	12.9	0.66	25.1	1.0	0.07	-25.8	1.0	0.50	0.7	53.6	35.1
-0.2	-31.0	13.3	0.52	28.2	0.5	0.31	-29.7	0.7	0.12	1.6	82.6	51.6
0.0	-30.6	8.5	0.60	23.7	0.7	0.17	-21.6	0.6	0.09	2.1	105.2	**74.6**
0.2	-41.5	9.9	0.83	5.5	0.9	0.09	-10.0	0.9	0.19	4.6	**112.9**	71.5
0.4	-40.1	11.7	0.77	-9.7	1.0	0.16	11.2	1.3	0.29	1.5	102.5	62.4
0.6	-11.3	6.4	0.80	-30.4	1.7	0.36	32.6	1.2	0.24	2.2	71.0	59.7
0.8	6.3	6.9	0.44	-40.5	1.4	0.20	50.9	1.6	0.30	10.3	25.3	31.6
1.0	29.8	5.3	0.14	-45.9	1.6	0.34	59.9	1.0	0.12	14.0	-27.6	2.2
1.2	46.1	4.6	0.49	-36.7	0.5	0.08	45.1	0.6	0.11	8.5	-68.5	-22.4
1.4	56.6	5.0	0.81	-22.9	0.5	0.12	30.3	0.4	0.06	7.4	-95.0	-38.5
1.6	59.5	4.2	0.37	-10.3	0.4	0.08	17.1	0.5	0.16	6.9	-108.7	-49.3
1.7	63.5	3.3	0.23	-3.8	0.2	0.03	3.6	0.2	0.03	0.2	-112.4	-48.9
1.8	58.3	6.1	0.72	-1.9	0.2	0.06	1.4	0.2	0.03	0.5	-114.0	**-55.8**
1.9	60.1	3.9	0.25	0.8	0.2	0.07	1.5	0.2	0.05	2.3	**-114.4**	-54.3
2.0	63.5	3.2	0.50	0.8	0.2	0.03	0.2	0.2	0.04	1.0	-114.1	-50.6

Table 9. Results of the QTCP-free calculations with various SQM methods, presenting the average (over the 28 different states along the ξ reaction coordinate) SE and maximum weights for the three perturbations in Figure 1. In addition, estimated activation and reaction free energies are presented, when the curves allowed a proper definition of the RS, TS, and PS.

	$\Delta G_{MM \rightarrow QM/MM}$		ΔG_{MM} forward		ΔG_{MM} backward		$\Delta G^{\#}$	ΔG^{0}
	SE	weight	SE	weight	SE	weight		
PDDG-PM3	10.5	0.53	3.3	0.27	1.6	0.23	75	-56
AM1	9.0	0.53	2.2	0.28	1.4	0.28	63	
MNDO	7.9	0.55	2.3	0.27	1.4	0.24		
PDDG-MNDO	6.8	0.52	1.8	0.24	1.2	0.27	53	-90
PM3	11.6	0.53	3.6	0.31	1.8	0.34	71	-54
RM1	8.9	0.54	3.1	0.36	1.8	0.28	69	-65

Figure 1. The thermodynamic cycle that is the basis of QTCP and related approaches. R and P are two states in a reaction mechanism.

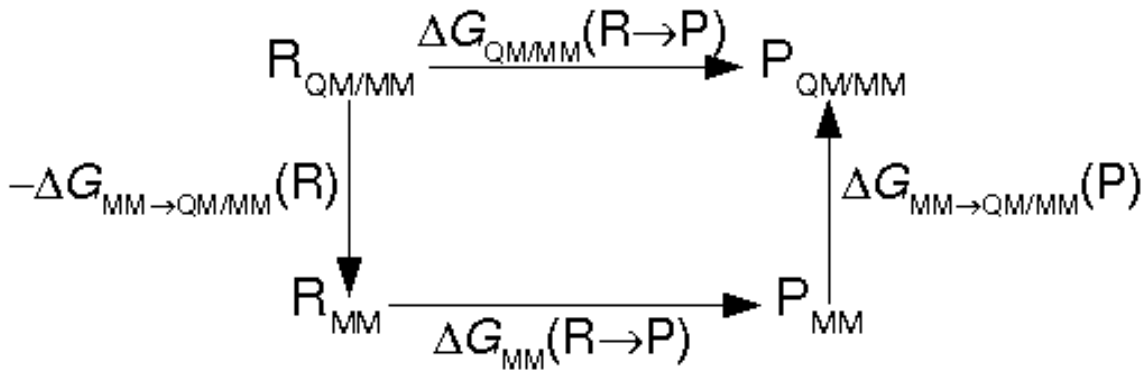

Figure 2. Geometries of the RS (left), TS (middle), and PS (right) of the first half-reaction of DhlA obtained in vacuum (top) and with QM/MM, using the PBE/def2-SVP method.

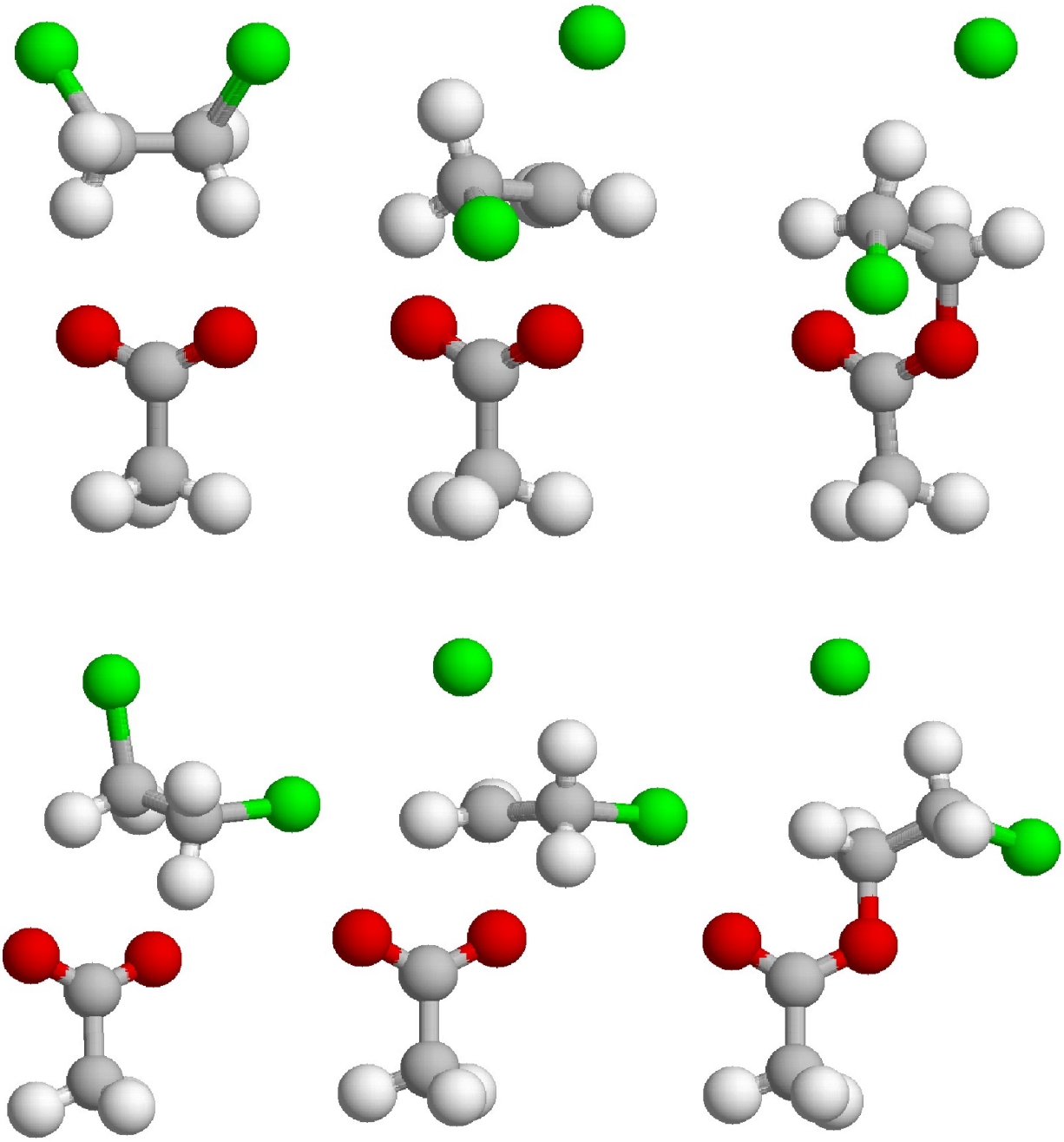

Figure 3. The reaction along the ξ reaction coordinate in vacuum, calculated with PBE/def2-SVP method.

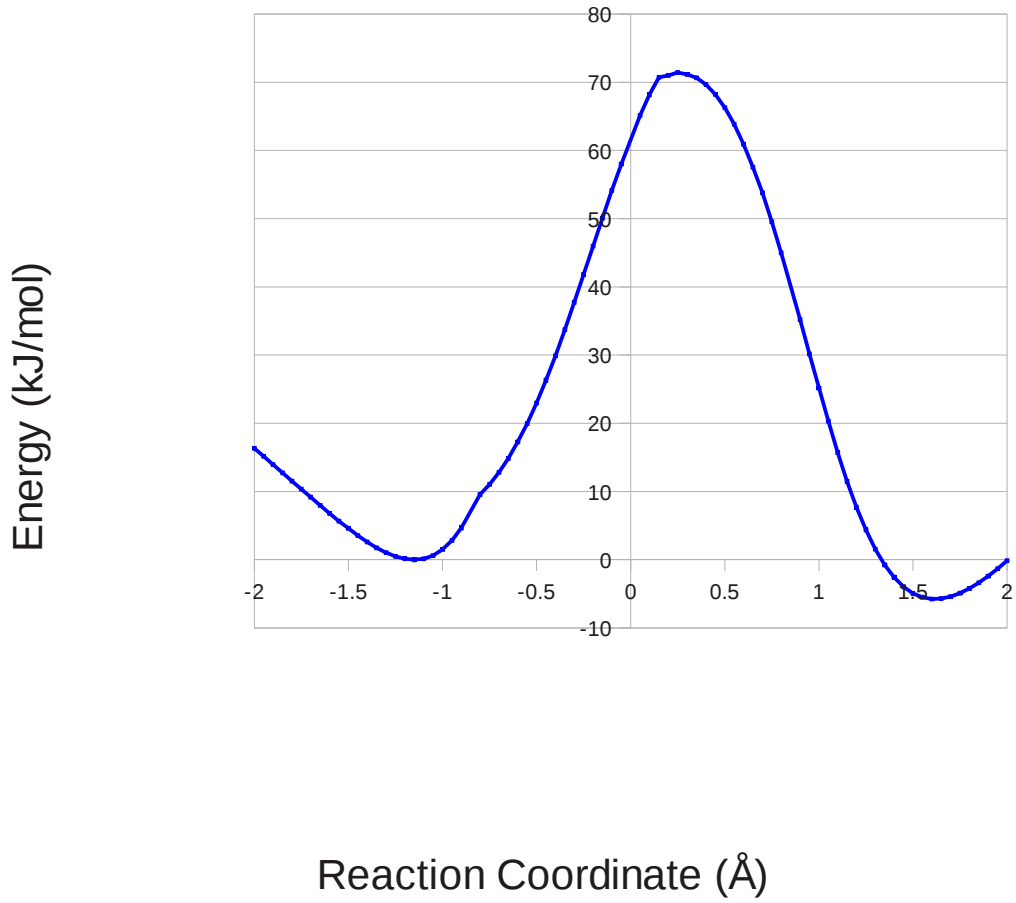

Figure 4. Reaction energies at the QM/MM level. Results obtained with the PBE and B3LYP methods are presented, both calculated with the def2-SVP basis set and full geometry optimisations. For the PBE method, the QM energies without the point-charge model (E_{QM1}) are also shown.

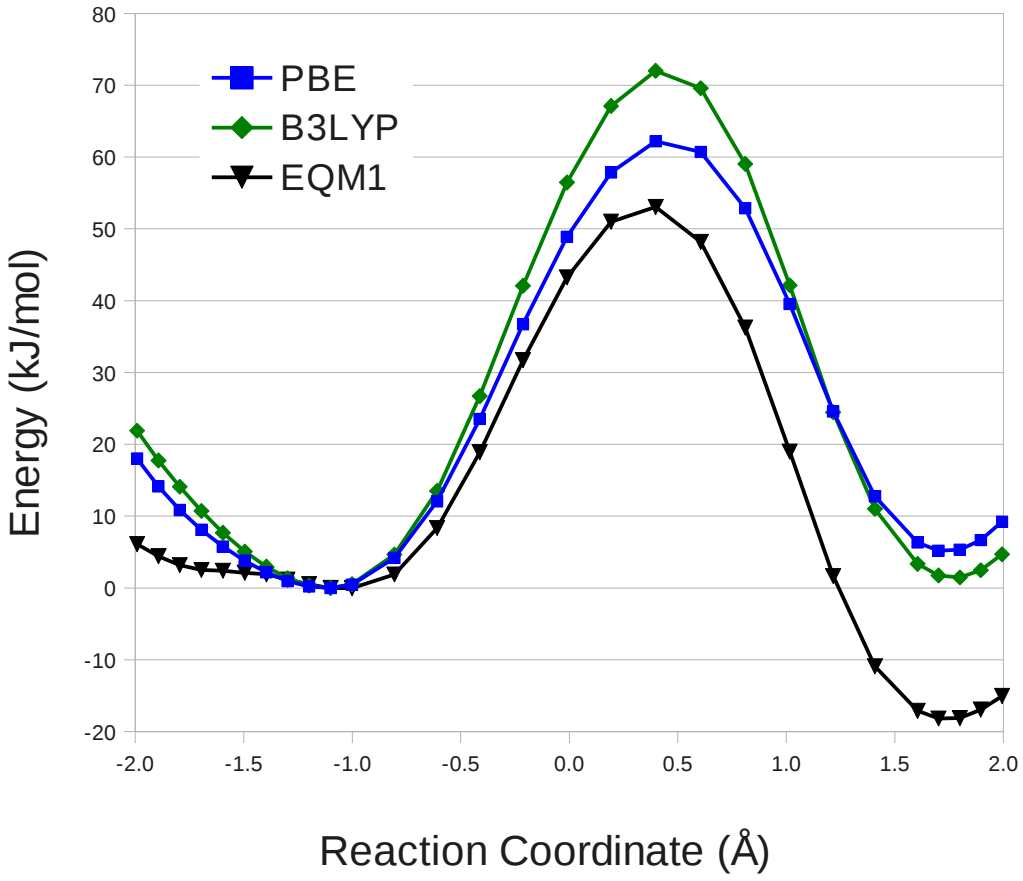

Figure 5. Free-energy curves obtained with standard QTCP-fixed method with two DFT functionals, PBE and B3LYP and two independent runs. In the B3LYP/TZVP calculation, the PBE/def2-SVP QM/MM structures are used in a QTCP calculation using the B3LYP/def2-TZVP method.

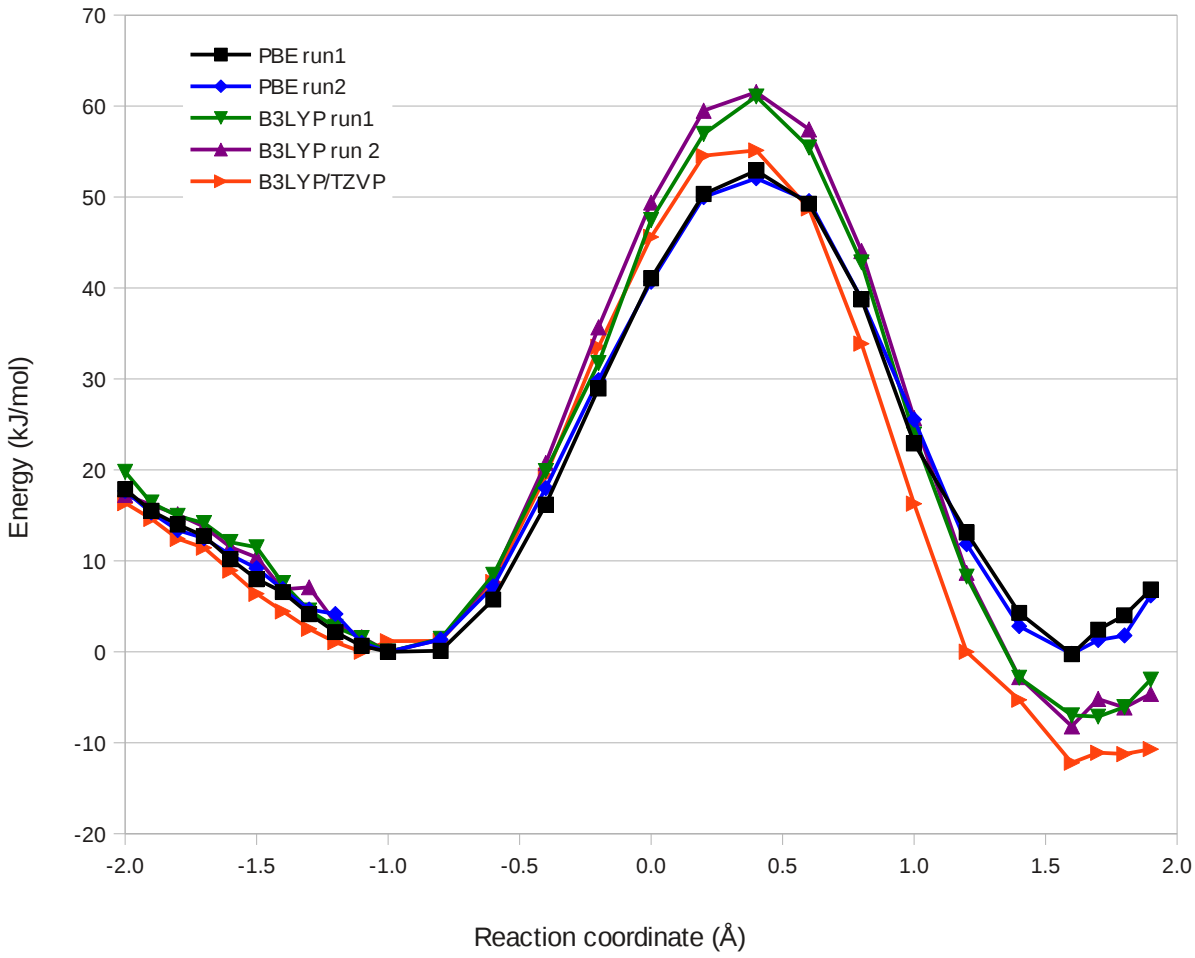

Figure 6. Free-energy curves obtained by QTCP-free, based on PDDG-PM3/MM simulations with mechanical embedding in the SQM/MM→QM/MM perturbation.