High Accuracy Computation of Reaction Barriers in Enzymes

Frederik Claeyssens, Jeremy N. Harvey, Frederick R. Manby, Ricardo A. Mata, Adrian J. Mulholland, Kara E. Ranaghan, Martin Schuetz, Stephan Thiel, Walter Thiel, Hans-Joachim Werner

QM/MM geometry optimisations of transition states and corresponding reactants were performed using B3LYP/6-31G*/CHARMM\textsuperscript{[1]} and B3LYP/TZVP/GROMOS\textsuperscript{[2]} for CM (16 pathways) and PHBH (10 pathways) respectively. The QM regions were treated by Jaguar\textsuperscript{[3]} and Turbomole\textsuperscript{[4]}, and the QM/MM coupling by QoMMMa\textsuperscript{[5]} and ChemShell,\textsuperscript{[6]} for CM and PHBH respectively. The final single-point B3LYP, LMP2, and LCCSD(T) calculations were carried out with the MOLPRO package of ab initio programs.\textsuperscript{[7]}

For CM, the initial geometries were sampled from AM1/CHARMM\textsuperscript{[8, 9]} and PM3/CHARMM\textsuperscript{[10]} QM/MM molecular dynamics (MD) simulations of Bacillus subtilis CM restrained to the transition state region (for details see\textsuperscript{[11, 12]}). Reaction pathways were obtained by adiabatic mapping, using as a reaction coordinate the difference in length between the breaking C–O and forming C–C bonds.

For PHBH, snapshots from GROMOS and AM1/GROMOS MD runs served as starting structures for AM1/GROMOS geometry optimisations. The resulting AM1/GROMOS transition structures were refined at the B3LYP/GROMOS level, by re-optimising the QM region and all surrounding residues within a distance of 5 Å. The B3LYP/GROMOS transition structures were then relaxed towards the associated reactants by careful stepwise energy minimizations, making sure that they are connected by a continuous path. The mapping coordinate used here was the difference in length between the breaking O–O and forming C–O bonds. Fig. S1 shows an optimized structure of the transition state within the active site, with a hydrogen bond between the transferring OH moiety and the backbone carbonyl of Pro 293.
The definition and preparation of the QM/MM models has been described elsewhere.\textsuperscript{[11, 12, 13]} Briefly, the CM system was comprised of a sphere of radius 25 Å centered on the substrate, containing 4192 protein atoms, the 24-atom substrate, and 947 water molecules (7057 atoms in total). The QM region contained the 24 atoms of the substrate, which does not form any covalent bonds with the protein. The PHBH system consisted of 7004 atoms (enzyme, substrate and cofactor) and an aqueous solvent layer (ca. 23 000 atoms in total). The QM region consisted of 49 atoms comprising p-hydroxybenzoate and FADHOOH with R=Me replacing the ribityl side chain of the cofactor (see Scheme 2 of the paper) which gave rise to one link atom at this covalent QM/MM junction. Coupling between the QM and MM regions was introduced by including a point charge representation of the MM region in the QM Hamiltonian, and by computing van der Waals non-bonded terms between QM and MM.
atoms. For the QM atoms, standard MM parameters for analogous MM atoms were used for the corresponding van der Waals well depths and radii.

The computed activation enthalpies for PHBH in Table 1 include a zero-point energy (ZPE) correction of $-1.1 \text{ kcal mol}^{-1}$ and an enthalpic temperature correction of $-0.2 \text{ kcal mol}^{-1}$ which were determined from AM1 gas-phase calculations for the QM region. Entropic corrections computed using AM1 on a gas-phase model amount to $1.2 \text{ kcal mol}^{-1}$. The entropic corrections for the whole system can be estimated by comparing the directly computed free energies of activation from 10 AM1/GROMOS TI runs\textsuperscript{[13]} with the corresponding AM1/GROMOS energy barriers, which leads to an average entropic contribution of $0.4 \text{ kcal mol}^{-1}$.

ZPE corrections for CM were computed by performing B3LYP frequency calculations on six different optimized reactant and TS structures of the substrate within the active site consisting of 2 water molecules within 3 Å of chorismate and 4 hydrogen terminated amino-acid side chains (Arg7, Arg63, Arg90 and Glu78). The water molecules and the amino acids were frozen during the reactant and TS optimisations. Rows and columns of the Hessian corresponding to the frozen atoms were deleted prior to diagonalisation. The values of the ZPE and enthalpic temperature correction are very similar for the six structures used; on average, the ZPE correction to the barrier height is $-1.5 \text{ kcal mol}^{-1}$ and the average enthalpic temperature correction is $-0.1 \text{ kcal mol}^{-1}$. Free energy corrections for the CM reaction paths are computed from density functional tight binding (SCC-DFTB/CHARMM) QM/MM umbrella sampling\textsuperscript{[14]} MD calculations along the reaction coordinate ($k_{\text{umbrella}}=200 \text{ kcal mol}^{-1} \text{ Å}^{-2}$).\textsuperscript{[15]}

All \textit{ab initio} wave function calculations used density fitting (DF) approximations\textsuperscript{[16, 17]} which greatly reduce the computational effort. The DF-HF and DF-LMP2 methods are described in Refs.\textsuperscript{[18, 19]}; the DF-LCCSD(T0) calculations were performed using a newly developed program.\textsuperscript{[20, 21]} The errors arising from density fitting are very small and typically less than 0.05 \text{ kcal mol}^{-1} for energy differences.

The basis set convergence was established at the canonical DF-MP2 level by comparing the barrier heights for one pathway in each system for different basis sets. For CM, the [aug]-
cc-pVTZ set (aug-cc-pVTZ for O and cc-pVTZ for all other atoms), which was used in the final DF-LCCSD(T0) calculations, was compared with the full aug-cc-pVTZ, aug-cc-pVQZ and aug-cc-pV5Z sets. For PHBH, a similar comparison was made with full aug-cc-pVTZ and aug-cc-pVQZ sets. In addition, for both systems a recently developed DF-MP2-F12(loc) method\cite{22, 23} was used to check the basis set convergence. As has been demonstrated in Ref.\cite{23}, the DF-MP2-F12(loc) method yields results very close to the basis set limit even with the aug-cc-pVTZ orbital basis. In fact, as shown in Table 1, the MP2 barrier heights converge with increasing basis set towards the MP2-F12(loc) values for both systems. From these results it can be concluded that the computed MP2/[aug]-cc-pVTZ barrier heights are within 0.5 kcal mol$^{-1}$ of the complete basis set (CBS) limits. It is known from small-molecule calculations that the basis set effect is usually somewhat smaller at the LCCSD(T0) level than at the MP2 level, and therefore it can be assumed that the LCCSD(T0) results are also within about 0.5 kcal mol$^{-1}$ of the basis set limit.

<table>
<thead>
<tr>
<th>Method</th>
<th>CM</th>
<th>PHBH</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF-MP2/[aug]-cc-pVTZ\textsuperscript{a}</td>
<td>12.3</td>
<td>11.8</td>
</tr>
<tr>
<td>DF-MP2/aug-cc-pVTZ</td>
<td>11.8</td>
<td>11.7</td>
</tr>
<tr>
<td>DF-MP2/aug-cc-pVQZ</td>
<td>12.2</td>
<td>12.1</td>
</tr>
<tr>
<td>DF-MP2/aug-cc-pV5Z</td>
<td>12.3</td>
<td>–</td>
</tr>
<tr>
<td>DF-MP2-F12/aug-cc-pVTZ\textsuperscript{c}</td>
<td>12.3</td>
<td>12.2</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Without zero-point corrections
\textsuperscript{b} Diffuse functions only on O atoms
\textsuperscript{c} Using MP2-F12/2*A(loc) corrections, see Ref. 23

Local correlation calculations used domains determined through the Boughton-Pulay procedure\cite{24} with a completeness criterion of 0.98. In order to guarantee a balanced treatment of reactants and transition state and to ensure smooth potentials, the domains of the reactants
and transition state should be merged and then kept fixed. This was done for CM, but not for the larger PHBH system. We tested the impact of this approximation by computing a single PHBH barrier height using merged domains, and found that this differed from the more efficient calculation by only 0.2 kcal mol\(^{-1}\).

The domain approximation was tested by comparing the barrier heights of DF-MP2 and DF-LMP2 calculations. In LMP2 no weak or distant pair approximations were made, and therefore the difference of DF-MP2 and DF-LMP2 solely reflects the effect of restricting the excitations to domains. It was found for both systems that LMP2 underestimates the correlation contribution by 0.3–0.4 kcal mol\(^{-1}\). From previous studies\cite{25} it is known that the errors caused by the domain approximation are very similar at the LMP2 and LCCSD(T) levels, and therefore it can be assumed that these error bounds apply to LCCSD(T) as well.

In the LCCSD(T) calculations additional weak pair and triples approximations\cite{26, 27} were made. Furthermore, the (T0) approximations was used, in which couplings between different orbital triples are neglected.\cite{27} For one pathway of each system the latter approximation has been tested by comparing with the more accurate (T1) approximation,\cite{28} and the errors were found to be negligible. In order to save computer time, the triples correction was computed using the [aug]-cc-pVDZ basis for most pathways of PHBH; tests on three pathways show that the errors compared to the larger basis are less than 0.3 kcal mol\(^{-1}\). The convergence with respect to the weak pair and triples approximations has been established for each system by extending the pair and triples lists until the barrier heights were converged to within 0.3 kcal mol\(^{-1}\) (or better). The final distance criteria\cite{26, 27} for close and weak pairs were \(R_{\text{close}} = 1\) bohr / 1 bohr and \(R_{\text{weak}} = 5\) bohr / 7 bohr for CM / PHBH, respectively. From the convergence of the results with respect to these parameters it was observed that the weak pair and triples approximations lead to a slight overestimation of the correlation contribution. This effect partly cancels with the error due to the domain approximation. Such cancellations are well known from other studies,\cite{25, 26} and it can therefore be assumed that the total error caused by the local approximations should be less than 0.5 kcal mol\(^{-1}\). Together with the basis set error it can be concluded that the LCCSD(T0) results are likely to be converged to within 1 kcal mol\(^{-1}\) of the full CCSD(T)/CBS limits.
References


