

High Accuracy Computation of Reaction Barriers in Enzymes**

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

Reasons for publication in *Angewandte Chemie*

With the advent of combined quantum mechanics / molecular mechanics (QM/MM) methods, enzymatic reactions have become accessible to theoretical modeling in recent years. QM/MM calculations on enzymes have generally been restricted to semiempirical or density functional QM treatments, which are realistic but of limited accuracy and cannot be improved in a systematic manner. In this work it has for the first time been possible to apply high-level coupled-cluster *ab initio* electronic structure methods to enzymatic catalysis, in a QM/MM framework. Excellent agreement is obtained between the computed and the experimentally determined activation barriers for two quite different enzymatic reactions, a Claisen rearrangement in chorismate mutase and an electrophilic aromatic substitution in para-hydroxybenzoate hydroxylase. This agreement between experimental and converged theoretical results has broader implications concerning the role of specific dynamic effects in enzyme catalysis that are currently under debate: at least for the two enzymes studied here, such dynamic effects must be small since standard transition state theory describes the reactivity quantitatively.

We believe that this paper will be of obvious interest to researchers in the fields of biochemistry and enzymology as well as computational chemistry because it reports a breakthrough towards unprecedented accuracy in QM/MM studies and thereby provides strong evidence that transition state theory is adequate for treating enzymatic reactivity. More generally, the paper should appeal to the broad audience of *Angewandte Chemie* because it demonstrates significant advances in high-level *ab initio* QM/MM methodology that can be applied to many complex and chemically relevant systems.

High Accuracy Computation of Reaction Barriers in Enzymes**

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

[*] Dr. F. Claeysens, Dr. J. N. Harvey, Dr. F. R. Manby, Dr. A. J. Mulholland, K. E. Ranaghan
School of Chemistry
University of Bristol
Cantocks Close
Bristol, BS8 1TS
UK
Phone: (+44) 117 928 9097
Fax: (+44) 117 925 1295
Email: Adrian.Mulholland@bristol.ac.uk

R. A. Mata, Prof. H.-J. Werner
Institut für Theoretische Chemie
Universität Stuttgart
Pfaffenwaldring 55
D-70569 Stuttgart
Germany
Phone: (+49) 711/685-64401
Fax: (+49) 711/685-64442
Email: werner@theochem.uni-stuttgart.de

Prof. M. Schütz
Institut für Physikalische und Theoretische Chemie
Universität Regensburg
Universitätsstr. 31
D-93040 Regensburg
Germany

Dr. S. Thiel, Prof. W. Thiel
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1
D-45470 Mülheim an der Ruhr
Germany
Phone: (+49) 208/306-2150
Fax: (+49) 208/306-2996
Email: thiel@mpi-muelheim.mpg.de

[**] The authors are grateful for financial support from the Royal Society (FRM), EPSRC (FRM, AJM, JNH, KER), BBSRC (AJM), IBM (AJM, FC), Deutsche Forschungsgemeinschaft (Leibniz program and SFB 706, HJW), Fonds der Chemischen Industrie (WT, HJW), Volkswagenstiftung (WT), and Fundação para a Ciência e Tecnologia (RAM).

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the authors.

The accurate prediction of enzyme kinetics from first principles is one of the central goals of theoretical biochemistry. There is considerable current debate^[1, 2, 3] about the applicability of transition state theory (TST) to compute rate constants of enzyme-catalyzed reactions. Classical TST is known to be insufficient in some cases, but corrections for dynamical recrossing and quantum mechanical tunneling can be included.^[1, 2, 4] Many effects that go beyond the framework of TST have been proposed, particularly focusing on the possible role of protein dynamics and conformational effects on the enzyme activity. Unfortunately, the overall importance of these effects for the effective reaction rate is difficult (if not impossible) to determine experimentally. However, if one could compute the quasi-thermodynamical free energy of activation with chemical accuracy (i.e. 1 kcal/mol), comparison with the effective measured free energy of activation would show the importance of other effects directly.

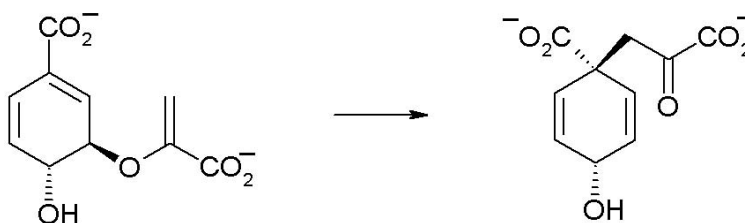
Combined quantum mechanics / molecular mechanics (QM/MM) methods have become an important tool for computational modelling of enzyme-catalysed reactions. Only the substrate(s) and relevant residues in the active site are treated quantum mechanically, the rest of the protein being described at the empirical MM level. This lowers the computational expense, making it possible to treat large enzyme systems and surrounding solvent, and to sample phase space. Nevertheless, the number of QM atoms is still relatively large, and until now only low levels of QM theory, such as semi-empirical methods or density functional theory (DFT), have been feasible. Semi-empirical methods, though applicable to large systems, are generally not accurate enough, since computed free energies of activation may be in error by 10 or more kcal/mol. DFT (especially with the B3LYP functional^[5]) offers improved accuracy, but still lacks key physical interactions (e.g., dispersion). Often, DFT underestimates barrier heights by several kcal/mol, and this cannot be systematically improved. Thus, where theoretical barriers do not agree with experiment, it is not clear whether the discrepancy arises from deficiencies in the electronic structure theory and the sampling, in the experimental observations, or in the underlying theoretical framework of QM/MM and TST.

Consequently, there is a need for high-level electronic structure calculations for reliable predictions of enzyme reactivity. The *ab initio* electron correlation methods MP2 (Møller-Plesset second-order perturbation theory), CCSD (coupled-cluster theory with single and

double excitations) and CCSD(T) (CCSD with a perturbative treatment of triple excitations) provide a well established hierarchy for converging reliably to high accuracy, and rate constants of gas-phase reactions involving only a few atoms can be predicted with error bars comparable to those in experiment.^[6, 7] However, the computational expense of these methods very rapidly increases with the number of atoms (doubling the system size increases cost of CCSD(T) by two orders of magnitude), and this has restricted applications of such methods to small molecules.

The steep increase of the computational cost is mainly a consequence of the delocalized character of the molecular orbitals from Hartree-Fock theory. However, in covalent molecules dynamic electron correlation is a short-ranged phenomenon, and by localizing the molecular orbitals it is possible to introduce a hierarchy of approximations^[8, 9] that lead to linear scaling of all computational resources with system size. Recent breakthroughs in the development of such methods^[10, 11, 12, 13, 14] now make it possible to treat systems one order of magnitude larger with comparable levels of accuracy. In the present study, we use these new high-level local correlation methods in QM/MM calculations on two biochemically well characterized enzymes, chorismate mutase (CM) and *para*-hydroxybenzoate hydroxylase (PHBH). Combining the *ab initio* results with thorough sampling we predict activation enthalpies that approach chemical accuracy.

The CM enzyme catalyzes the Claisen rearrangement of chorismate to prephenate, a key conversion in the shikimic acid pathway that produces aromatic amino acids (Scheme 1).

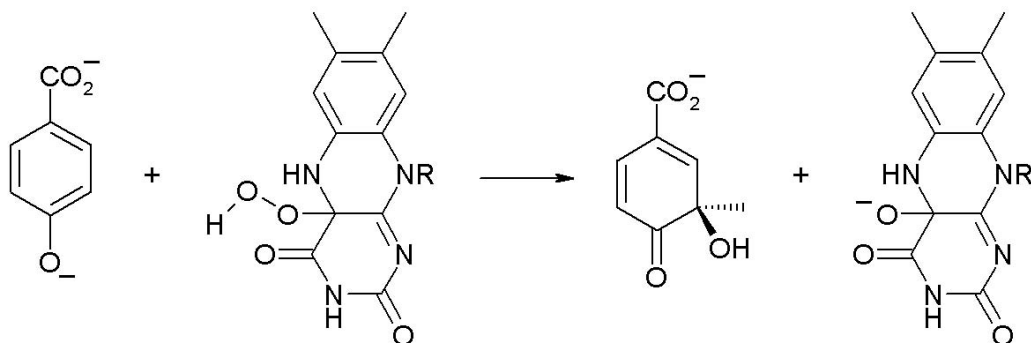


Scheme 1

This enzyme has been the object of extensive experimental^[15, 16] and computational^[17, 18, 19] research, which has conclusively shown that catalysis proceeds without covalent binding of

the substrate to the enzyme. The chemical step is believed to be largely rate-limiting in *Bacillus subtilis* CM, and the uncatalyzed reaction in water proceeds by the same mechanism. This makes CM a particularly convenient target for QM/MM studies,^[17, 18, 19] which have focused on aspects such as the structure of the enzyme-substrate complex, reaction pathways, and the role of active site residues in transition state stabilization. Calculations demonstrate that the key catalytic interactions are primarily electrostatic and are well described in the QM/MM model^[17, 18]. However, almost all of this previous reaction modelling has been carried out using semi-empirical or DFT methods, which do not predict barrier heights with chemical accuracy.

PHBH is a flavoprotein monooxygenase which catalyzes the hydroxylation of the substrate *p*-hydroxybenzoate (Scheme 2).



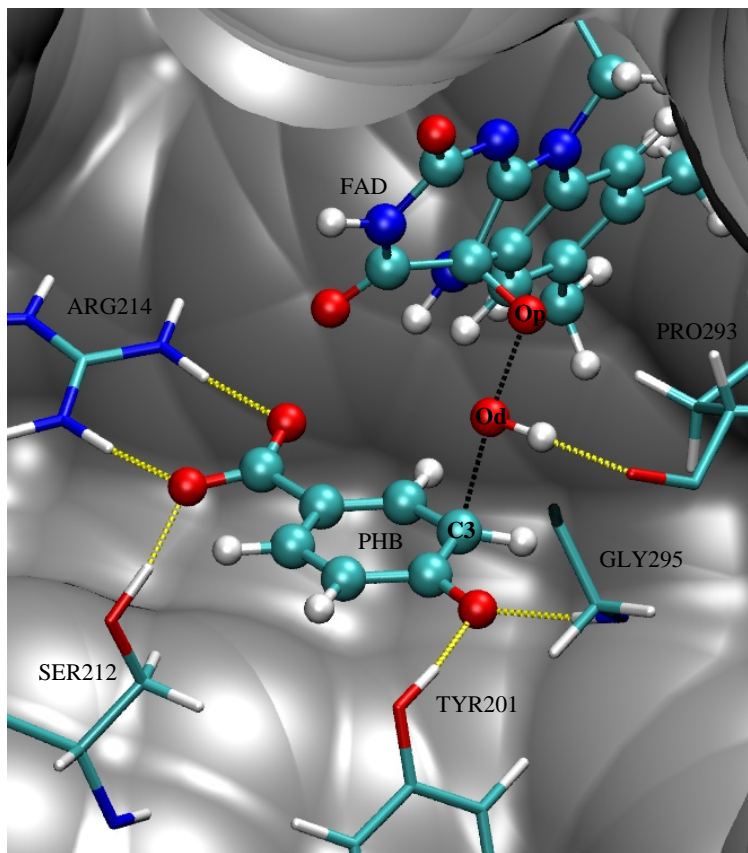
Scheme 2

It plays a crucial role in the oxidative degradation of aromatic compounds, such as lignin, by soil bacteria. The active hydroxylation agent is a flavin hydroperoxide (FADHOOH) that hydroxylates *p*-hydroxybenzoate in the meta position.

The extensive experimental work on PHBH has been reviewed,^[20] and the mechanistically relevant results for the hydroxylation step have been summarized recently.^[21] There is consensus that the hydroxylation involves an electrophilic aromatic substitution at the dianionic substrate and at around pH 8 where the enzyme is most active, the hydroxylation is considered to be the rate-determining step.^[22, 23] An activation enthalpy of 12 kcal/mol has

been obtained at pH 8 from temperature-dependent measurements of the overall rate between 277 and 298 K,^[24] while the measured rates of hydroxylation at pH 6.5^[23, 25] and the reported turnover rates at pH 8^[22, 23] indicate free energies of activation of 14 and 15 kcal/mol,^[21] respectively. On the theoretical side, the hydroxylation mechanism in PHBH has been the subject of several QM/MM studies,^[21, 26, 27] where the QM region has been described by the semi-empirical AM1 method (in one case also at the Hartree-Fock (HF) level^[26]). As with CM, tests showed that the key interactions are well captured in the QM/MM model.^[21, 26] Fig. 1 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.

Figure 1: PHBH active site



In the current work MP2, LMP2 and LCCSD(T0) calculations on CM and PHBH have been carried out (the L in the acronyms indicates that local approximations were used, and T0 is an approximate triples correction^[10, 11]). These are certainly by far the largest coupled cluster calculations ever performed (for PHBH correlating 284 electrons using 1294 basis functions). To account for the effects of conformational fluctuations the results were averaged over multiple pathways (16 for CM and 10 for PHBH). Initial structures were sampled from semi-empirical QM/MM dynamics, and reaction pathways were optimized using B3LYP/MM. To establish the reliability with respect to all other approximations, extensive convergence studies were performed. Using an extended QM region for CM (including Arg7, Arg63, Glu78 and Arg90 side chains and three water molecules in addition to the substrate) changes the B3LYP barrier by only around 0.7 kcal/mol. Calculations at the MP2 level with very large basis sets (up to quintuple zeta) alter the MP2 barrier heights by less than 0.5 kcal/mol. This excellent convergence with respect to basis has been confirmed by calculations using the recently developed explicitly correlated method MP2-F12(loc),^[14] which essentially delivers the MP2 basis set limit. Further extensive tests, as fully described in the Supporting Information, have shown that the effect of local approximations is smaller than 0.5 kcal/mol. These studies indicate that the computed LCCSD(T0) barrier heights are likely to agree with the full CCSD(T) values at the basis set limit to within 1 kcal/mol. This is an unprecedented accuracy which previously was difficult to achieve even for small molecules.

An underlying assumption in these calculations is that the CCSD(T) method in the basis set limit would predict barrier heights accurately. Small molecule calculations show that the error is unlikely to exceed 1 kcal/mol. Furthermore it is assumed that the B3LYP geometries are reliable. This is justified by the fact the B3LYP and LCCSD(T) transition states appear at almost the same reaction coordinate value along the B3LYP pathways.

The averaged activation enthalpies are summarized in Table 1, and the energy profiles along one typical pathway in CM are shown in Fig. 2.

Table 1: Computed^a activation enthalpies ΔH^\ddagger (300K) for CM and PHBH in kcal/mol

Method	CM	PHBH
HF	28.3	36.7
B3LYP	10.2	8.4
LMP2	9.5	10.7
LCCSD(T0) ^b	13.1	13.3
Experiment	12.7 ^c	12.0 ^d

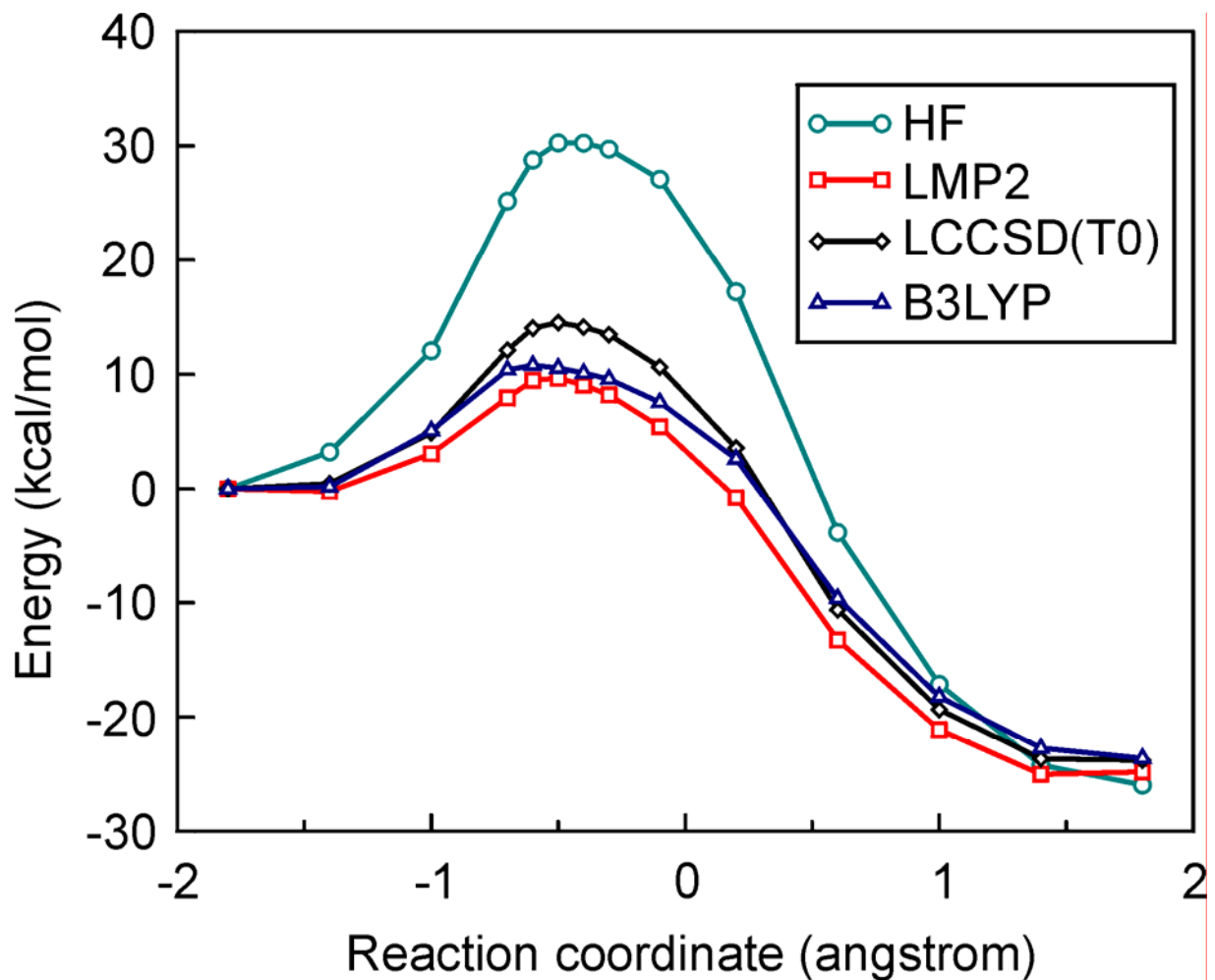
^aActivation enthalpies were derived from averages of energy differences from single-point QM/MM calculations for the reactant complex and the TS on different adiabatic pathways. The aug-cc-pVTZ basis was used on oxygen and cc-pVTZ on all other atoms, and a point-charge representation of the MM environment was included in the QM calculations. Starting points for adiabatic pathways were derived from snapshots from molecular dynamics (MD) sampling described elsewhere.^[21, 28] The geometries along the adiabatic pathways were obtained from separate QM/MM optimizations, using B3LYP for the QM part. Zero-point energy (ZPE) and 300 K thermal corrections were derived from QM-only calculations on cluster models. For CM, the cluster included chorismate and surrounding residues and was treated with B3LYP; for PHBH, the QM region was used, and treated with AM1. Full details can be found in Supporting Information.

^b LCCSD(T0) denotes local coupled cluster with an approximate treatment of the (T) triples correction in which certain couplings are neglected.^[10, 11]

^cReference^[15]

^dReference^[24]

Figure 2: QM/MM potential energy curves (without thermal corrections) for a single representative reaction pathway in CM. The reaction coordinate is defined as the difference between the length of the breaking C–O bond and the forming C–C bonds. See Table 1 for details of methods used.



By comparison of the HF and LCCSD(T0) results it can be seen that electron correlation effects lower the activation enthalpies by about a factor of 3. This is due to the fact that near the transition state more electrons are spatially close, leading to a significant increase in the correlation energy relative to the reactants. LMP2 and B3LYP overestimate the correlation effect and predict barriers that are too low by 3–5 kcal/mol. Only the LCCSD(T0) results are

in close agreement with experiment. These results clearly demonstrate that a high-level electron correlation treatment such as LCCSD(T0) is necessary to make quantitative predictions of barrier heights or other energetic properties in enzymes. It is likely that the limiting factor for the overall accuracy is now the QM/MM approximation and not the QM treatment as in previous semi-empirical or DFT studies. It should be noted in this context that the barrier height has been found to be insensitive to the size of the chosen QM region.^[18,19]

Reaction rates depend on activation free energies rather than on the enthalpies shown in Table 1. The difference between these two quantities can be estimated using low-level QM/MM methods by comparing mean activation enthalpies with activation free energies. The free energies are computed by umbrella sampling (CM)^[29] or thermodynamic integration (PHBH).^[21] For CM the entropic contribution at 300K is computed to be 2.5 kcal/mol, in good agreement with the experimental value of 2.7 kcal/mol.^[15] For PHBH the computed entropic contribution is 0.4 kcal/mol, which together with the best activation enthalpy computed here gives $\Delta G^\ddagger(300\text{ K}) = 13.7\text{ kcal/mol}$, in good agreement with the experimental values of 14–15 kcal/mol.^[22, 23, 25]

Our high-level QM/MM calculations provide near-quantitative results for the activation enthalpies and free energies of the reactions catalyzed by CM and PHBH. The agreement with experiment indicates that transition state theory provides a good general framework for understanding the rates of such enzyme-catalysed reactions. Our results show that it is now possible to perform electronic structure calculations on large systems approaching chemical accuracy, thus allowing quantitative studies of reaction mechanisms in enzymes. This development opens new horizons for theoretical biochemistry and many other areas of computational chemistry.

Keywords: enzyme catalysis, ab initio calculations, QM/MM calculations, transition state theory, reaction mechanisms

References

- [1] A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.
- [2] M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **2004**, *303*, 186–195.
- [3] W. Min, B. P. English, G. Luo, B. J. Cherayil, S. C. Kov, X. S. Xie, *Acc. Chem. Res.* **2005**, *38*, 923–931.
- [4] L. Masgrau, A. Roujeinikova, L. O. Johannissen, P. Hothi, J. Basran, K. E. Ranaghan, A. J. Mulholland, M. J. Sutcliffe, N. S. Scrutton, D. Leys, *Science* **2006**, *312*, 237–241.
- [5] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [6] D. G. Truhlar, B. C. Garrett, S. J. Klippenstein, *J. Phys. Chem.* **1996**, *100*, 12771–12800.
- [7] T. Wu, H.-J. Werner, U. Manthe, *Science* **2004**, *306*, 2227–2229.
- [8] P. Pulay, *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- [9] C. Hampel, H.-J. Werner, *J. Chem. Phys.* **1996**, *104*, 6286–6297.
- [10] M. Schütz, H.-J. Werner, *Chem. Phys. Lett.* **2000**, *318*, 370–378.
- [11] M. Schütz, *J. Chem. Phys.* **2000**, *113*, 9986–10001.
- [12] M. Schütz, H.-J. Werner, *J. Chem. Phys.* **2001**, *114*, 661–681.
- [13] H.-J. Werner, F. R. Manby, P. Knowles, *J. Chem. Phys.* **2003**, *118*, 8149–8160.
- [14] F. R. Manby, H.-J. Werner, T. B. Adler, A. J. May, *J. Chem. Phys.* **2006**, *124*, 094103 (6 pages).
- [15] P. Kast, M. Asif-Ullah, D. Hilvert, *Tetrahedron Lett.* **1996**, *37*, 2691–2694.
- [16] S. K. Wright, M. S. DeClue, A. Mandal, L. Lee, O. Wiest, W. W. Clieland, D. Hilvert, *J. Am. Chem. Soc.* **2005**, *127*, 12957–12964.
- [17] B. Szefczyk, A. J. Mulholland, K. E. Ranaghan, W. A. Sokalski, *J. Am. Chem. Soc.* **2004**, *126*, 16148–16159.
- [18] F. Claeysens, K. E. Ranaghan, F. R. Manby, J. N. Harvey, A. J. Mulholland, *Chem. Comm.* **2005**, 5068–5070.

- [19] A. Crespo, M. A. Marti, D. A. Estrin, A. E. Roitberg, *J. Am. Chem. Soc.* **2005**, *127*, 6940–6941.
- [20] B. Entsch, W. J. H. van Berkel, *FASEB* **1995**, *9*, 476–483.
- [21] H. M. Senn, S. Thiel, W. Thiel, *J. Chem. Theory Comp.* **2005**, *1*, 494–505.
- [22] M. Husain, B. Entsch, D. P. Ballou, V. Massey, P. J. Chapman, *J. Biol. Chem.* **1980**, *255*, 4189–4197.
- [23] B. Entsch, B. A. Palfey, D. P. Ballou, V. Massey, *J. Biol. Chem.* **1991**, *266*, 17341–17349.
- [24] W. J. H. van Berkel, F. Müller, *Eur. J. Biochem.* **1989**, *179*, 307–314.
- [25] M. Ortiz-Maldonado, L. J. Cole, S. M. Dumas, B. Entsch, D. P. Ballou, *Biochemistry* **2004**, *43*, 1569–1579.
- [26] L. Ridder, J. N. Harvey, I. M. C. M. Rietjens, J. Vervoort, A. J. Mulholland, *J. Phys. Chem. B* **2003**, *107*, 2118–2126.
- [27] S. R. Billeter, C. F. W. Hanser, T. Z. Mordasini, M. Scholten, W. F. Van Gunsteren, *Phys. Chem. Chem. Phys.* **2001**, *3*, 688–695.
- [28] K. E. Ranaghan, B. Szefczyk, W. A. Sokalski, J. C. Hermann, A. J. Mulholland, *Org. Biomol. Chem.* **2004**, *2*, 968–980.
- [29] L. Ridder, I. M. C. M. Rietjens, J. Vervoort, A. J. Mulholland, *J. Am. Chem. Soc.* **2002**, *124*, 9926–9936.

Computational Enzymology

F. Claeysens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, H.-J. Werner

High Accuracy Computation of Reaction Barriers in Enzymes

Modelling enzyme catalysis: High-level ab initio QM/MM calculations yield activation enthalpies and free energies for chorismate mutase and para-hydroxybenzoate hydroxylase that are in excellent agreement with experiment. Enzyme reactivity is described quantitatively by transition state theory.



Supporting Information

High Accuracy Computation of Reaction Barriers in Enzymes

Frederik Claeysens, 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^[1] and B3LYP/TZVP/GROMOS^[2] for CM (16 pathways) and PHBH (10 pathways) respectively. The QM regions were treated by Jaguar^[3] and Turbomole^[4], and the QM/MM coupling by QoMMMa^[5] and ChemShell,^[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.^[7]

For CM, the initial geometries were sampled from AM1/CHARMM^[8, 9] and PM3/CHARMM^[10] QM/MM molecular dynamics (MD) simulations of *Bacillus subtilis* CM restrained to the transition state region (for details see^[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.

The definition and preparation of the QM/MM models has been described elsewhere.^[11, 12, 13] Briefly, the CM system was comprised of the 24 atoms of the substrate in the QM region and 7033 MM atoms. 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. 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 kcal/mol and an enthalpic temperature correction of -0.2 kcal/mol 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 kcal/mol. 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^[13] with the corresponding AM1/GROMOS energy barriers, which leads to an average entropic contribution of 0.4 kcal/mol.

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 \AA 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 kcal/mol and the average enthalpic temperature correction is -0.1 kcal/mol. Free energy corrections for the CM reaction paths are computed from density functional tight binding (SCC-DFTB/CHARMM) QM/MM umbrella sampling^[14] MD calculations along the reaction coordinate ($k_{umbrella} = 200 \text{ kcal/\AA}$).^[15]

All *ab initio* wave function calculations used density fitting (DF) approximations^[16, 17] which greatly reduce the computational effort. The DF-HF and DF-LMP2 methods are described in Refs.^[18, 19]; the DF-LCCSD(T0) calculations were performed using a newly developed program.^[20, 21] The errors arising from density fitting are very small and typically

less than 0.05 kcal/mol 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^[22, 23] was used to check the basis set convergence. As has been demonstrated in Ref. ^[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 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 of the basis set limit.

Table 1: Basis set convergence of DF-MP2 barrier heights^a

Method	CM	PHBH
DF-MP2/[aug]-cc-pVTZ ^b	12.3	11.8
DF-MP2/aug-cc-pVTZ	11.8	11.7
DF-MP2/aug-cc-pVQZ	12.2	12.1
DF-MP2/aug-cc-pV5Z	12.3	–
DF-MP2-F12/aug-cc-pVTZ ^c	12.3	12.2

a) Without zero-point corrections

b) Diffuse functions only on O atoms

c) Using MP2-F12/2*A(loc) corrections, see Ref. 23

Local correlation calculations used domains determined through the Boughton-Pulay procedure^[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.

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. From previous studies^[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^[26, 27] were made. Furthermore, the (T0) approximations was used, in which couplings between different orbital triples are neglected.^[27] For one pathway of each system the latter approximation has been tested by comparing with the more accurate (T1) approximation,^[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. 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 (or better). The final distance criteria^[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,^[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. 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 of the full CCSD(T)/CBS limits.

References

- [1] A. D. MacKerell, *et al.*, *J. Chem. Phys. B* **1998**, *102*, 3586.
- [2] W. F. van Gunsteren, *et al.*, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Biomos b.v., Zürich and Groningen, VdF Hochschulverlag, ETH Zürich, Zürich, **1996**.
- [3] Jaguar 4.1, Schrödinger Inc., Portland, OR, **1991-2000**.
- [4] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165.
- [5] J. N. Harvey, *Faraday Discuss.* **2004**, *127*, 165.
- [6] P. Sherwood, *et al.*, *J. Mol. Struct. (Theochem)* **2003**, *632*, 1.
- [7] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, *et al.*, MOLPRO, Version 2006.1, a Package of Ab Initio Programs, Cardiff (UK), **2006**, see <http://www.molpro.net>.
- [8] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *207*, 3902.
- [9] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700.
- [10] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209.
- [11] K. E. Ranaghan, *et al.*, *Org. Biomol. Chem.* **2004**, *2*, 968.
- [12] F. Claeysens, K. E. Ranaghan, F. R. Manby, J. N. Harvey, A. J. Mulholland, *Chem. Comm.* **2005**, 5068.
- [13] H. M. Senn, S. Thiel, W. Thiel, *J. Chem. Theory Comp.* **2005**, *1*, 494.
- [14] Q. Cui, M. Elstner, E. Kaxirasa, T. Frauenheim, M. Karplus, *J. Phys. Chem. B* **2001**, *105*, 569.
- [15] K. E. Ranaghan, Ph. D. Thesis, University of Bristol, 2006.
- [16] E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41.
- [17] J. L. Whitten, *J. Chem. Phys.* **1973**, *58*, 4496.
- [18] M. Schütz, H.-J. Werner, R. Lindh, F. R. Manby, *J. Chem. Phys.* **2004**, *121*, 737.
- [19] R. Polly, H.-J. Werner, F. R. Manby, P. J. Knowles, *Mol. Phys.* **2004**, *102*, 2311.
- [20] M. Schütz, F. R. Manby, *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349.
- [21] H.-J. Werner, M. Schütz, *to be published*.
- [22] H.-J. Werner, F. R. Manby, *J. Chem. Phys.* **2006**, *124*, 054114.

- [23] F. R. Manby, H.-J. Werner, T. B. Adler, A. J. May, *J. Chem. Phys.* **2006**, *124*, 094103.
- [24] J. W. Boughton, P. Pulay, *J. Comput. Chem.* **1993**, *14*, 736.
- [25] H.-J. Werner, K. Pflüger, *Ann. Rev. Comp. Chem.*, **2006**, in press.
- [26] C. Hampel, H.-J. Werner, *J. Chem. Phys.* **1996**, *104*, 6286.
- [27] M. Schütz, H.-J. Werner, *Chem. Phys. Lett.* **2000**, *318*, 370.
- [28] M. Schütz, *J. Chem. Phys.* **2000**, *113*, 9986.