Computational Enzymology

DOI: 10.1002/ange.200602711

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 Schütz, Stephan Thiel, Walter Thiel,* and Hans-Joachim Werner**

The accurate prediction of enzyme kinetics from first principles is one of the central goals of theoretical biochemistry. Currently there is considerable debate[1–3] 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^{-1}), comparison with the effective measured free energy of activation would directly show the importance of other effects.

Combined quantum mechanical/molecular mechanical (QM/MM) methods have become an important tool for computational modeling of enzyme-catalyzed reactions. Only the substrate(s) and relevant residues in the active site are treated quantum mechanically, the rest of the protein is described at the empirical MM level. This lowers the computational expense, making it possible to treat large enzyme systems and the 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 semiempirical methods or density functional theory (DFT), have been feasible. Semiempirical methods, though applicable to large systems, are generally not accurate enough because computed free energies of activation may have an error of ten or more kcal mol^{-1}. 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^{-1}, which cannot be systematically improved. Thus, when theoretical barriers do not agree with those from 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 that converge reliably to give high accuracy, and rate constants of gas-phase reactions involving only a few atoms can be predicted with error bars comparable to those found experimentally.[6,7] However, the computational expense of these methods increases very rapidly with the number of atoms (doubling the system size increases the cost of CCSD(T) by two orders of magnitude), and this has restricted the applications of such methods to small molecules.

The steep increase in the computational cost is mainly a consequence of the delocalized character of the molecular orbitals from the Hartree–Fock (HF) 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 developments[10–14] have made it possible to treat systems that are one order of magnitude larger with a

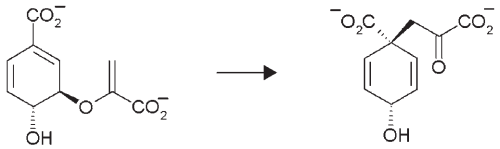
[*] Dr. F. Claeyssens, Dr. J. N. Harvey, Dr. F. R. Manby, Dr. A. J. Mulholland, K. E. Ranaghan
School of Chemistry
University of Bristol
Cantocks Close, Bristol, BS8 1TS (UK)
Fax: (+44) 117-925-1295
E-mail: adrian.mulholland@bristol.ac.uk
Dr. S. Thiel, Prof. W. Thiel
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr (Germany)
Fax: (+49) 208-306-2996
E-mail: thiel@mpi-muelheim.mpg.de
R. A. Mata, Prof. H.-J. Werner
Institut für Theoretische Chemie
Universität Stuttgart
Pfaffenwaldring 55, 70569 Stuttgart (Germany)
Fax: (+49) 711-685-64442
E-mail: werner@theochem.uni-stuttgart.de
Prof. M. Schütz
Institut für Physikalische und Theoretische Chemie
Universität Regensburg
Universitätsstrasse 31, 93040 Regensburg (Germany)

[**] The authors are grateful for financial support from the Royal Society (F.R.M.), EPSRC (F.R.M., A.J.M., J.N.H., K.E.R.), BBSRC (A.J.M.), IBM (A.J.M., F.C.), Deutsche Forschungsgemeinschaft (Leibniz program and SFB 706, H.J.W.), Fonds der Chemischen Industrie (W.T., H.J.W.), Volkswagenstiftung (W.T.), and Fundação para a Ciência e Tecnologia (R.A.M.).

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

comparable level 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.

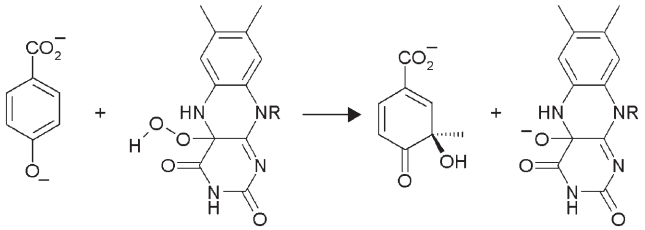
The CM enzyme catalyzes the Claisen rearrangement of chorismate to prephenate, a key conversion in the shikimic acid pathway that produces aromatic amino acids (Scheme 1).



Scheme 1. Claisen rearrangement in CM.

This enzyme has been the object of extensive experimental[15,16] and computational[17–19] research, which have 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,[17–19] 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 modeling has been carried out by using semiempirical or DFT methods, which do not predict barrier heights with chemical accuracy.

PHBH is a flavoprotein monooxygenase that catalyzes the hydroxylation of the substrate *p*-hydroxybenzoate (Scheme 2). It plays a crucial role in the oxidative degradation



Scheme 2. Hydroxylation reaction in PHBH.

of aromatic compounds, such as lignin, by soil bacteria. The active hydroxylation agent is a flavin hydroperoxide (FAD-HOOH) 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 a consensus that the hydroxylation involves an electrophilic aromatic substitution on the dianionic substrate, and at around pH 8 (pH at which the enzyme is most active) the hydroxylation is considered to be the rate-determining step.[22,23] An activation enthalpy of 12 kcal mol^{-1} has been obtained at pH 8 from temperature-dependent measurements of the overall rate between 277 and 298 K.[24] 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^{-1},[21] respectively. On the theoretical side, the hydroxylation mechanism in PHBH has been the subject of several QM/MM studies[21,26,27] in which the QM region has been described by the semiempirical AM1 method (in one case also at the HF level[26]). As with CM, tests showed that the key interactions are well captured in the QM/MM model.[21,26]

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 by 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 semiempirical QM/MM dynamics, and reaction pathways were optimized by using B3LYP/MM. To establish the reliability with respect to all other approximations, extensive convergence studies were performed. The use of an extended QM region for CM (including Arg 7, Arg 63, Glu 78, and Arg 90 side chains and three water molecules in addition to the substrate) changes the B3LYP barrier by only around 0.7 kcal mol^{-1}. 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^{-1}. This excellent convergence with respect to basis has been confirmed by calculations that use 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^{-1}. 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^{-1}. This accuracy is unprecedented and was previously 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^{-1}. Furthermore, it is assumed that the B3LYP geometries are reliable. This is justified by the fact that 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 Figure 1.

Through 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

Table 1: Computed[a] activation enthalpies $\Delta H^{\neq}$ (300 K) for CM and PHBH in kcal mol^{-1}.[b]

Method	CM	PHBH
HF	28.3 (2.1)	36.7 (2.6)
B3LYP	10.2 (1.8)	8.4 (1.4)
LMP2	9.5 (1.0)	10.7 (1.2)
LCCSD(T0)[c]	13.1 (1.1)	13.3 (1.5)
experimental	12.7[d]	12.0[e]

[a] Activation 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 sampling described elsewhere.[21,28] The geometries along the adiabatic pathways were obtained from separate QM/MM optimizations by using B3LYP for the QM part. Zero-point energy 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 the Supporting Information. [b] In parentheses: root-mean-square deviations for 16 (CM) and 10 (PHBH) snapshots. [c] LCCSD(T0) denotes local coupled cluster with an approximate treatment of the (T) triples correction in which certain couplings are neglected.[10,11] [d] Reference [15]. [e] Reference [24].

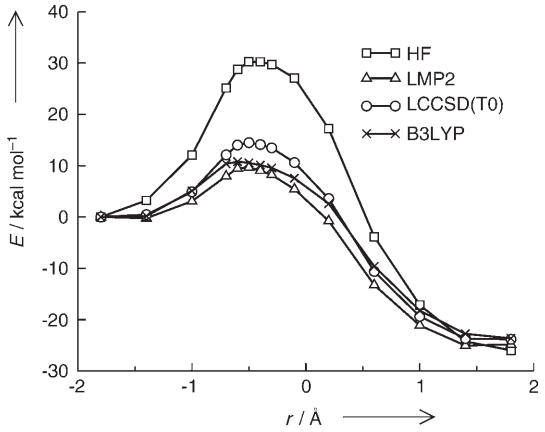


Figure 1. QM/MM potential-energy curves (without thermal corrections) for a single representative reaction pathway in CM. The reaction coordinate *r* is defined as the difference between the length of the breaking C–O bond and the forming C–C bonds. See Table 1 for details of methods used.

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^{-1}. Only the LCCSD(T0) results are in close agreement with experimental results. These results indicate 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 with the use of 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 300 K is computed to be 2.5 kcal mol^{-1}, in good agreement with the experimental value of 2.7 kcal mol^{-1}.[15] For PHBH, the computed entropic contribution is 0.4 kcal mol^{-1}, which, together with the best activation enthalpy computed herein, gives $\Delta G^{\neq}(300\ \mathrm{K}) = 13.7$ kcal mol^{-1}. This is in good agreement with the experimental values of 14–15 kcal mol^{-1}.[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 experimental results indicates that transition-state theory provides a good general framework for understanding the rates of such enzyme-catalyzed 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.

Received: July 7, 2006
Published online: September 22, 2006

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

[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. Claeyssens, K. E. Ranaghan, F. R. Manby, J. N. Harvey, A. J. Mulholland, *Chem. Commun.* **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 J.* **1995**, *9*, 476 – 483.
[21] H. M. Senn, S. Thiel, W. Thiel, *J. Chem. Theory Comput.* **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.

