Large-scale Density Functional Theory Transition State Searching in Enzymes**

Draft May 7, 2014

Greg Lever,[†] Daniel J. Cole,[*,†,‡] Richard Lonsdale,[¶] Kara E. Ranaghan,[¶] David J. Wales,[§] Adrian J. Mulholland,[¶] Chris-Kriton Skylaris,[‖] and Mike C. Payne[†]

Theory of Condensed Matter group, Cavendish Laboratory, 19 JJ Thomson Ave, Cambridge CB3 0HE, UK, Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, United States, Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK, University Chemical Laboratory, Lensfield Rd, Cambridge CB2 1EW, UK, and School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

E-mail: daniel.cole@yale.edu

[*]To whom correspondence should be addressed
[†]Cavendish Laboratory, University of Cambridge
[‡]Yale University
[¶]University of Bristol
[§]Chemical Laboratory, University of Cambridge
[‖]University of Southampton

**The authors thank Louis Lee (Cavendish Laboratory) and Karl Wilkinson (University of Southampton) for methodological improvements. Computational resources were provided by the Cambridge HPC Service, funded by EPSRC Grants EP/F032773/1 and EP/J017639/1, by the IRIDIS3 supercomputer of the University of Southampton, and by the HECToR national supercomputer, funded by EPSRC Grant EP/F038038/1 (UKCP consortium). The authors are grateful for funding provided by the EPSRC (GL, RL, KER, DJW and AJM), specifically RL and AJM for support under the CCP-BioSim project (grant EP/J010588/1; www.ccpbiosim.ac.uk) and AJM as an EPSRC Leadership Fellow (grant EP/G007705/01). The authors also acknowledge a Marie Curie International Outgoing Fellowship within the 7$^{\text{th}}$ European Community Framework Programme (DJC) and the Royal Society (C-KS).

Supporting information (SI) for this article is available.

Abstract

Linear-scaling quantum mechanical density functional theory calculations have been applied to study the rearrangement of chorismate to prephenate in large-scale models (up to around 2,000 atoms) of the *Bacillus subtilis* chorismate mutase enzyme. By treating all atoms at the same quantum mechanical level of theory, we obtain an unbiased, parameter-free description of the transition state geometry and energetics. The activation energy barrier is calculated to be lowered by 10.5 kcal/mol in the enzyme, compared with the equivalent reaction in water, which is in good agreement with experiment. Natural bond orbital analysis identifies a number of active site residues that are important for this catalytic rate enhancement. This benchmark study demonstrates that linear-scaling density functional theory techniques are now at a stage where they can play a much wider role in computational enzymology.

Accurate computational prediction of activation energy barriers and mechanisms of rate enhancement in catalysis could contribute significantly to pharmaceutical inhibitor development and biomimetic catalyst design. In this context, combined quantum mechanical/molecular mechanical (QM/MM) simulations, in which the substrate is treated with a QM method and the surrounding enzyme and water molecules are treated with classical point charges, are an invaluable tool. Indeed, recent advances in the treatment of electron correlation provide a well-established methodological hierarchy, which has been shown to reproduce activation energy barriers to chemical accuracy (within 1 kcal/mol) as part of the QM/MM framework.[1] However, with these advances it is becoming clear that an implicit limiting factor in computational enzymology is the QM/MM methodology itself, rather than the level of QM theory.[1,2] These inaccuracies may take a number of forms: charge transfer between the substrate and protein that cannot be incorporated into the classical force field; the substrate may be affected by the so-called electron leakage effect, whereby point charges over-polarize the electron density;[3] the classical force field may simply provide an incorrect description of the electrostatic environment[4] or intra-molecular strain; or, in cases of covalent bonding between the QM and MM regions, the method may be sensitive to the link

algorithm employed. In contrast, state-of-the-art linear-scaling density functional theory (LS-DFT) codes allow a consistent quantum mechanical treatment of systems comprising thousands of atoms, naturally accounting for charge transfer and polarization, and have the potential to provide a much improved description of the enzyme. Thus, there is a pressing need for a demonstration and benchmarking of transition state searching in enzymes using the LS-DFT formalism. Here we employ the LS-DFT code, ONETEP,[5] to perform fully quantum mechanical transition state searching calculations in large-scale models (up to 2,000 atoms) of the chorismate mutase (CM) enzyme and in water.

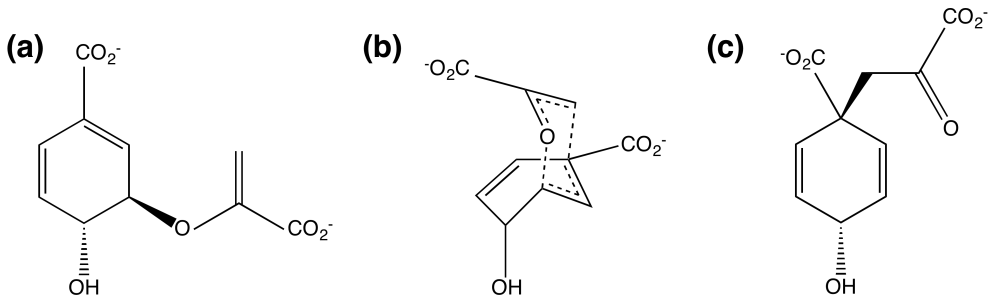


Figure 1: The rearrangement of (a) chorismate, via (b) the transition state, to (c) prephenate.

Situated at a branch point in the shikimate pathway, which is crucial for generating aromatic amino acids, CM catalyzes the Claisen rearrangement of chorismate to prephenate (Figure 1). Due to the lack of covalent bonding between the substrate and enzyme active site,[6] the fact that the reaction also occurs in water with a similar mechanism, and the availability of experimental data,[6–8] this has become a model system for benchmarking computational methodologies.[1,9,10] Previous computations have demonstrated that the catalytic effect of the enzyme is dominated by electrostatic interactions that should be well-described by QM/MM.[1,9] Nevertheless the observation that the reaction energy may vary by more than 5 kcal/mol upon enlargement of the QM region in CM,[9] along with similar observations in other enzymes,[2] motivates an investigation into the use of large-scale QM calculations in which the need to define a QM/MM boundary is removed.

In the current work, we have performed LS-DFT calculations, with the PBE exchange-

Table 1: Convergence of energies (kcal/mol) of activation ($\Delta^{\ddagger}E$) and reaction (ΔE) in CM. Mobile atoms are allowed to relax during optimization and TS searching, while frozen atoms are fixed to the B3LYP/MM geometry.

Mobile	Frozen	$\Delta^{\ddagger}E$	ΔE
98	901	13.4	-7.7
211	788	13.5	-8.0
98	1901	13.3	-7.9

correlation functional, on CM to demonstrate the feasibility of performing large-scale transition state searching in enzymes, with every atom treated at the QM level. Initial reactant state (RS) and product state (PS) structures, both in CM and in water, were generated via semi-empirical QM/MM molecular dynamics, followed by B3LYP/MM optimization,[10] from which spherical clusters were extracted and re-optimized using the ONETEP LS-DFT code (See SI).[5,11] Transition state (TS) searching was initially performed according to a modified linear synchronous transit (LST) and quadratic synchronous transit (QST) pathway approach,[12] in combination with conjugate gradient optimization. Extensive convergence tests were performed to determine the reliability of the results. Refinement of the transition state of the chorismate to prephenate rearrangement in vacuum generated by the LST/QST algorithm using the more rigorous gradient-only version of hybrid eigenvector-following[13] reduced the activation energy barrier by less than 0.1 kcal/mol (See SI Section 2.2). Table 1 demonstrates convergence of the activation energy barrier and reaction energy with respect to the size of the system and the number of mobile and frozen atoms for clusters comprising up to 1999 atoms. The maximum change in energy is just 0.3 kcal/mol. Similar results were obtained for the reaction in water, although a larger mobile region was required (See SI).

Our convergence tests indicate that, for CM, a computational model comprising a $\sim$100 atom mobile region within a larger $\sim$1000 atom cluster, in vacuum, provides an accurate description of the energetics of this system. This model was therefore used to compute the energetics of the chorismate to prephenate rearrangement in CM, and the results were averaged over five different pathways to account for temperature-induced fluctuations. Figure 2

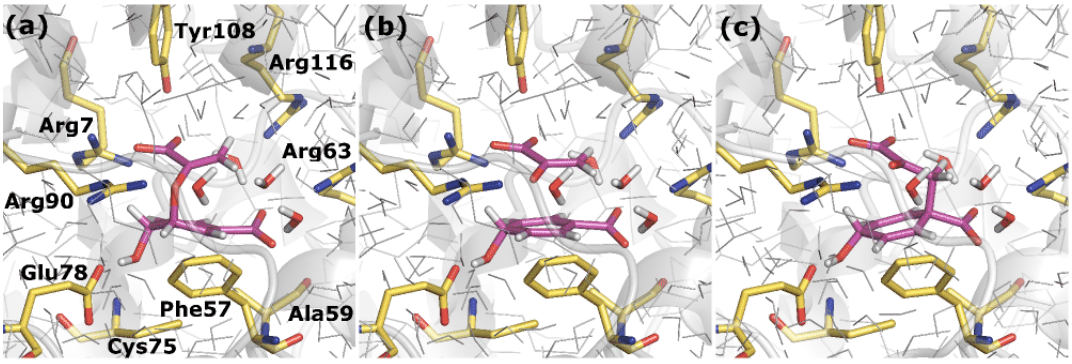


Figure 2: Rearangement of the substrate (magenta) from chorismate to prephenate within the CM active site (yellow) and surrounding protein (grey). The (a) reactant, (b) transition and (c) product state conformations as obtained from LS-DFT calculations.

Table 2: Comparison of averaged energies (kcal/mol) of activation ($\Delta^{\ddagger}E$) and reaction (ΔE) in CM and water. A description of the substrate in water computational model is given in the SI.

		$\Delta^{\ddagger}E$	ΔE
Enzyme	ONETEP	13.6±1.3	-7.8±0.5
	Experiment[8]	12.7±0.4	–
Water	ONETEP	24.1±1.1	-9.4±2.2
	Experiment[6]	20.71±0.35	-13.2±0.5

shows exemplar geometries of the optimized RS and PS structures, as well as the TS generated from LS-DFT, all of which are in excellent agreement with B3LYP/MM geometries.[10] Table 2 shows the final averaged energies of activation and reaction for the chorismate to prephenate rearrangement in both CM and water environments. The barrier height in the enzyme is in good agreement with experiment, while in water both the activation and reaction energies are too positive by ∼4 kcal/mol. The change in the activation barrier in CM is predicted to be −10.5 kcal/mol, while the reaction energy is similar in the two environments (difference of 1.6 kcal/mol). The former is in agreement with experiment (−8.0 kcal/mol) while the lack of experimental data regarding the reaction enthalpy in CM precludes any conclusions to be drawn from the latter result. The remaining discrepancies between computation and experiment are expected to be dominated by inaccuracies in the treatment of electron exchange and correlation – here, we have employed the PBE functional, augmented by damped London energy expressions to account for dispersion interactions.[14] However, we emphasize that the methods presented here serve as a template for future studies that will employ more rigorous treatments of electron exchange and correlation.[15]

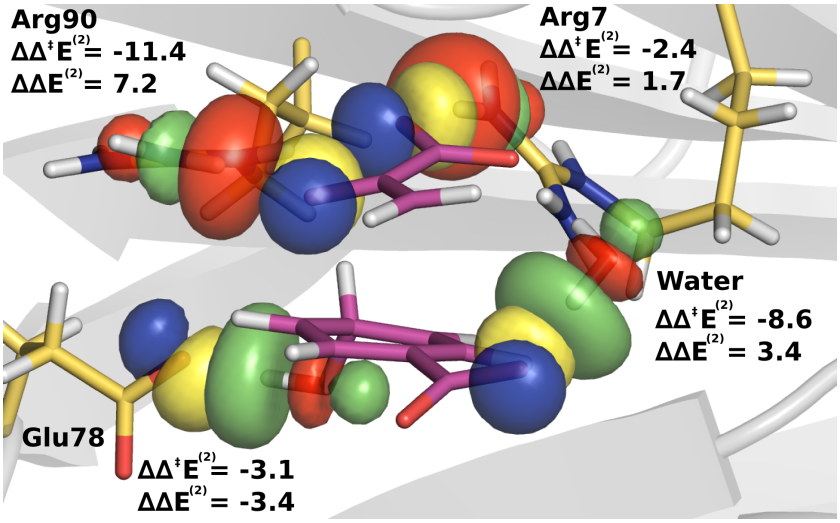


Figure 3: NBOs participating in enzyme–substrate interactions at the TS. Lone pairs (n) are shown as blue/yellow isosurfaces, while anti-bonding (σ^*) orbitals are in red/green. Electronic delocalization energetic contributions (kcal/mol) to the stabilization of the TS ($\Delta\Delta^{\ddagger}E^{(2)}$) and the destabilization of the PS ($\Delta\Delta E^{(2)}$), relative to the RS, are also shown.

While a standard LS-DFT calculation has the potential to improve the accuracy of TS searching in enzymes, it is not immediately clear how to determine the contribution of individual active site residues to TS stabilization. In this respect, it is instructive to transform the NGWFs that are used to describe the active site of CM into a set of natural bond orbitals (NBOs).[16,17] The resulting NBOs may then be categorized into chemically-intuitive Lewis-type bonding and lone pair orbitals, as well as their anti-bonding counterparts. Electronic delocalization from filled to vacant NBOs causes a variational lowering of the total energy, which can be estimated from second order perturbation theory ($\Delta E^{(2)}$, see SI). This stabilization energy is strongly correlated with the strengths of hydrogen bonds in simple systems,[17,18] and we can therefore use it as a qualitative estimate of the importance of active-site residues in the catalytic rate enhancement in CM. Figure 3 shows the four pairs of NBOs that are responsible for the strongest enzyme–substrate charge transfer interactions. The residues involved are Arg 90 (in agreement with mutagenesis experiments[19] following theoretical predictions[20,21] and experimental proposals[8,22]), Glu 78 (in agreement with other computational studies[23]), Arg 7 (in agreement with previous calculations[24]), and a water molecule. In all cases, electronic delocalization acts to stabilize the TS and (apart from Glu 78) destabilize the PS. This analysis suggests the structure of the active site has evolved to provide optimal orbital overlap between charged residues and the substrate at the TS, thus lowering the activation barrier.

This communication demonstrates the feasibility of performing transition state searching in enzymes by quantum mechanical modelling of the substrate and a large proportion of its environment. The structure obtained with LS-DFT and the LST/QST transition state searching algorithm is in excellent agreement with previous studies using B3LYP/MM, and changes insignificantly following more accurate refinement using eigenvector-following. The energies of activation and reaction are in good agreement with experiment and, more importantly, are shown to be converged with respect to system size and virtually independent of computational parameters. Furthermore, natural bond orbital analysis provides a good

predictive measure of active site residues that are important for catalytic rate enhancement. Treating entire proteins with full QM has the potential to circumvent inaccuracies associated with the QM/MM approach, and the methods presented here will increase not only the accuracy, but also the range, of problems open to investigation in fields from small molecule therapeutics to molecular biology, enzymology and biomimetics.

References

(1) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H. J. *Angew. Chem. Int. Ed.* **2006**, *45*, 6856.

(2) Solt, I.; Kulhànek, P.; Simon, I.; Winfield, S.; Payne, M. C.; Csànyi, G.; Fuxreiter, M. *J. Phys. Chem. B* **2009**, *113*, 5728.

(3) Neugebauer, J. *Chem. Phys. Chem,* **2009**, *10*, 3148.

(4) Lee, L. P.; Cole, D. J.; Skylaris, C. K.; Jorgensen, W. L.; Payne, M. C. *J. Chem. Theory Comput.* **2013**, *9*, 2981.

(5) Skylaris, C. K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. *J. Chem. Phys.* **2005**, *122*, 084119.

(6) Kast, P.; Tewari, Y. B.; Wiest, O.; Hilvert, D.; Houk, K. N.; Goldberg, R. N. *J. Phys. Chem. B* **1997**, *101*, 10976.

(7) Andrews, P. R.; Smith, G. D.; Young, I. G. *Biochemistry* **1973**, *12*, 3492.

(8) Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691.

(9) Crespo, A.; Scherlis, D. A.; Marti, M. A.; Ordejon, P.; Roitberg, A. E.; Estrin, D. A. *J. Phys. Chem. B* **2003**, *107*, 13728.

(10) Claeyssens, F.; Ranaghan, K. E.; Lawan, N.; Macrae, S. J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. *Org. Biomol. Chem.* **2011**, *9*, 1578.

(11) Hine, N. D. M.; Robinson, M.; Haynes, P. D.; Skylaris, C. K.; Payne, M. C.; Mostofi, A. A. *Phys. Rev. B* **2011**, *83*, 195102.

(12) Govind, N.; Petersen, M.; Fitzgerald, G.; King-Smith, D.; Andzelm, J. *Comp. Mat. Sci.* **2003**, *28*, 250.

(13) Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969–3980.

(14) Hill, Q.; Skylaris, C. K. *Proc. R. Soc. A* **2009**, *465*, 669.

(15) Dziedzic, J.; Hill, Q.; Skylaris, C.-K. *J. Chem. Phys.* **2013**, *139*, 214103.

(16) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. *NBO 5*; Theoretical Chemistry Institute: University of Wisconsin, Madison, 2001.

(17) Lee, L. P.; Cole, D. J.; Payne, M. C.; Skylaris, C. K. *J. Comp. Chem.* **2013**, *34*, 429.

(18) Mohajeri, A.; Nobandegani, F. F. *J. Phys. Chem. A* **2008**, *112*, 281.

(19) Cload, S. T.; D, L., R; Pastor, R. M.; Schultz, P. G. *J. Am. Chem Soc.* **1996**, *118*, 1787.

(20) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *17*, 11345.

(21) Guimaraēs, C. R. W.; Udier-Blagović, M.; Tubert-Brohman, I.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2003**, *125*, 6892.

(22) Gorisch, H. *Biochemistry* **1978**, *3700*, 17.

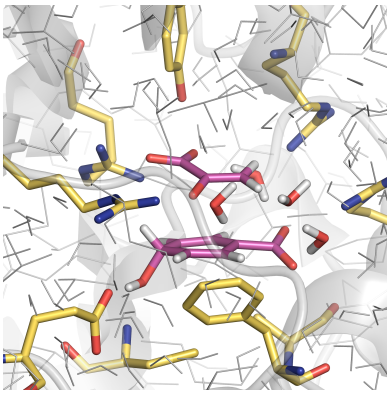
(23) Worthington, S. E.; Roitberg, A. E.; Krauss, M. *J. Phys. Chem. B* **2001**, *103*, 7087.

(24) Lee, Y. S.; Worthington, S. E.; Krauss, M.; Brooks, B. R. *J. Phys. Chem. B* **2002**, *106*, 12059.

Graphical TOC Entry

All-atom quantum mechanical modelling of a significant proportion (∼2,000 atoms) of the *Bacillus subtilis* chorismate mutase enzyme is presented. The computed lowering of the activation energy barrier in the enzyme is in good agreement with experiment and is attributed to strong overlap between orbitals on the substrate and charged active site residues.

Ab initio calculations, Computational chemistry, Density functional calculations, Enzyme catalysis, Linear-scaling



A short text for the Table of Contents of the issue (up to 450 characters) and image of 55 x 52.5 mm and a maximum of five keywords in alphabetical order should be included as the last page of the manuscript. At least two of the keywords should be taken from the "Keyword Catalogue" (see the complete Notice to Authors on the homepage). The text for the Table of Contents should (ideally with the help of a graphic, color is free here) arouse curiosity. Repetition or a paraphrase of the title and presentation of experimental details should be avoided.