Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2011, **9**, 1578

www.rsc.org/obc

PAPER

Analysis of chorismate mutase catalysis by QM/MM modelling of enzyme-catalysed and uncatalysed reactions

Frederik Claeyssens,[a,b] Kara E. Ranaghan,[a] Narin Lawan,[a,c] Stephen J. Macrae,[a,d] Frederick R. Manby,[a] Jeremy N. Harvey[a] and Adrian J. Mulholland*[a]

Received 7th September 2010, Accepted 18th November 2010
DOI: 10.1039/c0ob00691b

Chorismate mutase is at the centre of current controversy about fundamental features of biological catalysts. Some recent studies have proposed that catalysis in this enzyme does not involve transition state (TS) stabilization but instead is due largely to the formation of a reactive conformation of the substrate. To understand the origins of catalysis, it is necessary to compare equivalent reactions in different environments. The pericyclic conversion of chorismate to prephenate catalysed by chorismate mutase also occurs (much more slowly) in aqueous solution. In this study we analyse the origins of catalysis by comparison of multiple quantum mechanics/molecular mechanics (QM/MM) reaction pathways at a reliable, well tested level of theory (B3LYP/6-31G(d)/CHARMM27) for the reaction (i) in *Bacillus subtilis* chorismate mutase (BsCM) and (ii) in aqueous solvent. The average calculated reaction (potential energy) barriers are 11.3 kcal mol^{-1} in the enzyme and 17.4 kcal mol^{-1} in water, both of which are in good agreement with experiment. Comparison of the two sets of reaction pathways shows that the reaction follows a slightly different reaction pathway in the enzyme than in it does in solution, because of a destabilization, or strain, of the substrate in the enzyme. The substrate strain energy within the enzyme remains constant throughout the reaction. There is no unique reactive conformation of the substrate common to both environments, and the transition state structures are also different in the enzyme and in water. Analysis of the barrier heights in each environment shows a clear correlation between TS stabilization and the barrier height. The average differential TS stabilization is 7.3 kcal mol^{-1} in the enzyme. This is significantly higher than the small amount of TS stabilization in water (on average only 1.0 kcal mol^{-1} relative to the substrate). The TS is stabilized mainly by electrostatic interactions with active site residues in the enzyme, with Arg90, Arg7 and Glu78 generally the most important. Conformational effects (*e.g.* strain of the substrate in the enzyme) do not contribute significantly to the lower barrier observed in the enzyme. The results show that catalysis is mainly due to better TS stabilization by the enzyme.

Introduction

Understanding how enzymes achieve their catalytic rate accelerations is not only of fundamental interest, but also increasing practical importance. The origins of enzyme catalysis are hotly debated. Detailed analysis of enzymic and equivalent uncatalysed reactions, for example by computational modelling, has the potential to identify and quantify the causes of enzyme catalysis.[1–6]

Chorismate mutase (CM) has become a crucial system in current arguments about enzyme catalysis.[7–22] CM catalyses the pericyclic Claisen rearrangement of chorismate to prephenate (Fig. 1). This reaction is part of the shikimate pathway which produces aromatic amino acids in plants, fungi and bacteria, making chorismate mutase a potential target for herbicides, fungicides and antibiotics. Here, we study the *Bacillus subtilis* enzyme (BsCM).[23] This enzyme

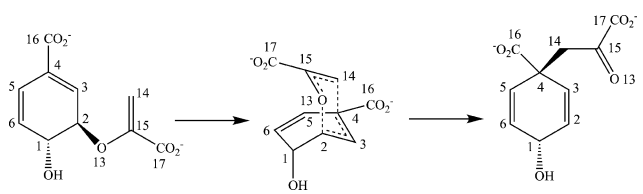

Fig. 1 Schematic depiction of the Claisen rearrangement of chorismate to prephenate *via* a chair-like transition state, showing the atom numbering used in the text.

[a]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, UK, BS8 1TS. E-mail: Adrian.Mulholland@bristol.ac.uk; Fax: +44 (0)117 925 0612; Tel: +44 (0)117 928 9097*
[b]*University of Sheffield, Engineering Materials Department, Biomaterials and Tissue Engineering Group, Kroto Research Institute, Broad Lane, Sheffield, UK, S3 7HQ*
[c]*Computational Simulation and Modelling Laboratory (CSML), Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand, 50200*
[d]*United Utilities, Thirlmere House, Lingley Mere, Lingley Green Avenue, Great Sankey, Warrington, UK WA5 3LP*

is an excellent system for testing theories of catalysis, because the reaction does not involve any covalent bonding between the enzyme and the substrate, and because the same (uncatalysed) reaction also occurs in aqueous solution by the same mechanism. The activation free energy $\Delta^{\ddagger}G = 15.4$ kcal mol^{-1} ($\Delta^{\ddagger}H = 12.7$ kcal mol^{-1}) in the (BsCM) enzyme is much lower than the $\Delta^{\ddagger}G$ = 24.5 kcal mol^{-1} ($\Delta^{\ddagger}H$ = 20.7 kcal mol^{-1}) for the uncatalysed reaction in aqueous solution at 25 °C.[24] This translates to a rate acceleration of 10^6 by the enzyme ($\Delta\Delta^{\ddagger}G = 9.1$ kcal mol^{-1}). Earlier studies indicated TS stabilization by the enzyme *via* electrostatic stabilization.[7–12] The first QM/MM study (at the AM1/CHARMm) level,[7] found a (potential energy) barrier of 17.8 kcal mol^{-1}, and showed transition state (TS) stabilization by Arg90, and by Glu78. Substrate destabilization (substrate strain, *i.e.* binding of a higher energy reactive conformer to the enzyme) was also suggested to contribute to catalysis because the bound structure was distorted from the gas phase geometry towards a more TS-like geometry. Subsequent investigations (*e.g.* applying semiempirical QM/MM methods in molecular dynamics simulations) have also highlighted substrate conformational effects as potential contributors to catalysis,[8,11,14] although these may be a consequence of TS stabilization rather than a distinct effect in their own right.[12,13,17] The enzyme-bound conformation of chorismate is significantly different from that in solution, more closely resembling the TS.[13–17] It is thought that the enzyme binds the pseudo-diaxial form of chorismate while the pseudo-diequatorial structure is the global minimum in water.[25–27]

Bruice and Hur[18–20] have controversially argued that catalysis in CM does not involve TS stabilization, and that it arises instead almost entirely from the selection of a reactive conformation, described as a near-attack conformation (NAC). They have proposed that this is of wide importance in enzyme catalysis. Several different definitions of a 'NAC' have been suggested. Most commonly, a 'NAC' has been defined as a conformer in which the reacting atoms are within the van der Waals distance (*e.g.* a carbon–carbon distance ≤3.7 Å in chorismate mutase) and the angle of approach is within 30° of the corresponding angle adopted at the TS. Bruice and Hur[19] propose that once the enzyme has bound (stabilized) this more TS-like conformation, almost no extra stabilization of the TS is necessary for catalysis. They propose that 'NAC's are 'turnstiles' through which the reactants must pass through in order to reach the TS. They initially estimated the free energy for 'NAC' formation from the mole fraction of NACs found in unrestricted molecular dynamics (MD) simulations of chorismate in *Escherichia coli* chorismate mutase (EcCM), and in solution. The results suggested that the observed catalytic effect of the enzyme is 90% due to the ability of the enzyme to support NACs, with transition state stabilization playing only a minor role. The calculation of populations through unrestrained molecular dynamics simulations in this way has been criticized as being unreliable.[28] Subsequently, they used a combination of MM and AM1/CHARMM simulations to estimate the free energy of NAC formation in EcCM, BsCM, R90Cit-BsCM E52A-EcCM and catalytic antibody 1F7, though some details of the approaches used in this work are not made clear.[29] More recent work of Bruice's group applied the SCCDFTB/CHARMM method to EcCM[30] and chorismate mutase from the thermophile *Thermus thermophilus* (TtCM).[31] From these studies, 90% and 87% of the catalysis in EcCM and TtCM was attributed to 'NAC' formation.

There is no unique definition of a 'NAC'. Different definitions of a 'NAC' lead to different estimates of the catalytic benefit of forming a 'NAC': some 'predict' catalytic effects larger than that actually observed.[31] Instead of trying to create a definition to fit the effect of interest, it is more useful to determine the catalytic benefit to the enzyme of binding the conformation of the substrate observed in the Michaelis complex. With this clear and objective definition, a 'NAC' is simply the conformation of the substrate bound to the enzyme. The cost of forming this conformation in solution (*i.e.* the maximum catalytic contribution of the 'NAC effect') can be calculated reliably *via* free energy perturbation (FEP) methods. Independent calculations by FEP MD methods yield a free energy cost of 3.8–4.6, or 5 kcal mol^{-1} (by AM1/CHARMM22 QM/MM[17] or empirical valence bond[13] methods, respectively), *i.e.* only accounting for 40–55% of the total $\Delta\Delta^{\ddagger}G$ between enzyme and solvent. This indicates that catalysis must involve TS stabilization relative to the bound conformation (*i.e.* significantly more stabilization of the TS, relative to the substrate, in the enzyme than in solution), in agreement with QM/MM calculations.[10]

Guimaraes *et al.*[16] studied the chorismate to prephenate reaction in BsCM using a combined AM1/TIP3P QM/MM Monte Carlo/free energy perturbation (MC/FEP) approach and found that the free energy of conversion from non-NACs to NACs in both enzyme and water environments is negative, and thus provides no free energy contribution to catalysis. This study also indicates that conformational compression contributes ~70% of the calculated lowering of the free energy barrier and that this is enthalpic in origin.

Replacement of the active site residue Arg90 by citrulline in BsCM[32] (*i.e.* R90Cit-BsCM) increases the apparent free energy barrier by 5.9 kcal mol^{-1}. This has been attributed to poorer TS stabilization in the mutant than in the wild type enzyme. These observations have been corroborated computationally by Guimaraes *et al.*[33] who report an increase of the free-energy barrier to reaction by 3.3 kcal mol^{-1} upon mutagenesis, and attribute this to inferior TS stabilization by the mutant. Ishida's recent results for the effect of several point mutations in BsCM, using a combination of *ab initio* QM/MM calculations and MD FEP simulations, also support TS stabilization as important in BsCM catalysis.[22]

Here we extend significantly beyond our earlier work,[34] and compare the reaction in the enzyme to the reaction in aqueous solution. We apply well-validated QM/MM methods using the B3LYP hybrid density functional for the QM region, which provides a reliable description of the reaction.[35] Lower level methods that have previously been used (*e.g.* semiempirical or *ab initio* Hartree–Fock methods) have significant limitations for this reaction. We also calculate multiple reaction pathways. Comparison of the reaction in the enzyme with that in the solvent environment enables us to investigate the origins of catalysis by this important enzyme. Detailed structural and electronic analysis of the reaction elucidates the fundamental basis of catalysis. The results we obtain are in excellent agreement with experiment.

Methods

To model the enzyme-catalysed reaction, we used the structure of *Bacillus subtilis* chorismate mutase[23] taken from the protein databank (PDB code 2CHT), containing Bartlett's inhibitor

(a TS analogue)[36] bound in the active site. Chorismate was substituted for the TS analogue as described in references 9 and 10. This initial enzyme structure was hydrogenated and solvated (with a cut-off radius of 2.8 Å; the CHARMM[37,38] version of the TIP3P model[39] was used for water molecules). Multiple structures were then generated by semiempirical QM/MM molecular dynamics simulations[40] of the BsCM complex (at the SCCDFTB/CHARMM22[41] level) with the chorismate QM region restrained to be close to the TS (to a reaction coordinate value of $r = -0.3$ Å).[34,35] The reaction coordinate, r, was defined as the difference in length between the breaking C2–O13 bond and the forming C4–C14 bond. This coordinate has been used in our[9,10,34,35] and other previous studies, and is a good choice, as shown *e.g.* by comparison to the path generated by the replica path method,[42] and to fully optimized TS structures.[9] Structures from a 60 ps dynamics run (after 500 ps equilibration) were saved at regular intervals of 4 ps, giving 16 different TS complexes. For the reaction in solution, we used structures from a 50 ps dynamics run, again with the chorismate conformation restrained to $r = -0.3$ Å, generated by semiempirical QM/MM simulations (SCCDFTB/CHARMM22[41]). In water, AM1/CHARMM22 and PM3/CHARMM22 dynamics simulations were not stable at the TS. Calculations of the reaction barrier in the enzyme taken from SCCDFTB/CHARMM22 starting points give structurally slightly different results to those from AM1(PM3)/CHARMM22 starting structures (which were used in our previous studies).[34,35] Some hydrogen bonds in network around the substrate were found to break during AM1(PM3)/CHARMM22 dynamics simulations, most notably the formation of an interaction between Arg90 and Glu78 is often observed. This was not observed in the SCCDFTB/CHARMM22 derived structures used in this study.

From these starting points we calculated 16 different adiabatic reaction pathways in the enzyme, and 24 for the uncatalysed reaction in water, using B3LYP density functional theory QM/MM methods. For this, Jaguar[43] and Tinker,[44] linked by our program QoMMMa,[45] were used for QM and MM calculations, with electronic coupling between the two regions treated by including MM charges in the QM Hamiltonian. Earlier studies have shown that the QM/MM approach is reliable for this system.[9,10,34,35] CHARMM Lennard-Jones parameters (for standard CHARMM atom types) were used to describe QM/MM van der Waals interactions.[10] The QM region (chorismate/TS/prephenate only (24 atoms)) was treated at the hybrid density functional B3LYP/6-31G(d) level of theory, which gives a good description of the reaction.[9,10,34,46] The barrier at higher levels of theory (LCCSD(T) with large basis sets) is very similar,[35] showing that B3LYP is a good choice for modelling this reaction. Including active site residues in the QM region does not greatly influence the reaction barrier, indicating that the QM/MM model used in this study is appropriate.[46,47] Quantum chemical decomposition analysis reinforces this conclusion.[12] The enzymic model consisted of 7077 atoms and the solution model contained 7218 atoms. The MM region comprised an approximate 25 Å radius sphere of protein and solvent, treated with the CHARMM27 force field.[38,48] The outer 5 Å was fixed (3324 atoms in the enzyme model and 3535 in the solution model), with all other atoms free to move. The set-up of the model is described in detail elsewhere.[9,10,17,34] As there is no evidence for large-scale conformational changes during the reaction[23,49,50] this approach should give a representative sample of reactive conformations in the enzyme. Each structure was fully optimized at the B3LYP/6-31G(d)/CHARMM27 QM/MM level, while restraining the reaction coordinate (r) to –0.3 Å with a harmonic force constant of 500 kcal mol^{-1} Å^{-2}, to generate starting structures. Reaction pathways were generated by restrained optimizations in both directions along the reaction coordinate, towards the reactant and the product, in steps of 0.2 Å (0.1 Å around the TS), with both the MM and QM systems fully and consistently optimized at each step. Energy profiles were calculated from $r = -2.2$ Å to 2.2 Å for the reaction in the enzyme, and from $r = -2.6$ Å to 2.2 Å in water, to identify the reactant and product minima. Reoptimization of the reactant complex without restraints, both in the enzyme and solution, gave structures very similar to the lowest energy restrained structures (and energy differences less than 1 kcal mol^{-1}). It should be noted that the two sets of reactant structures (in enzyme and water) differ significantly from each other, as will be discussed below.

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}} + E_{\mathrm{QM\text{-}MM}} + E_{\mathrm{MM}} \quad (1)$$

$$E_{\mathrm{INTERACTION}} = E_{\mathrm{QM\text{-}MM}} + E_{\mathrm{MM}} = E_{\mathrm{QM/MM}} - E_{\mathrm{QM}} \quad (2)$$

The total energy of the QM/MM system ($E_{\mathrm{QM/MM}}$) was decomposed into different components so as to examine the relative stabilization provided by the enzyme or water environments along the reaction paths. To enable this decomposition, the QM energy of the reacting substrate alone, at the corresponding QM/MM optimized geometry, was calculated for each point along each reaction path. This is E_{QM}. We can then write formally that $E_{\mathrm{QM/MM}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM\text{-}MM}}$ (eqn (1)). E_{MM} is the MM energy of the MM part, and $E_{\mathrm{QM\text{-}MM}}$ is the interaction energy between the QM and MM regions, including electrostatic and polarization effects as well as an MM term which includes van der Waals interactions between QM and MM atoms. $E_{\mathrm{INTERACTION}} = E_{\mathrm{MM}} + E_{\mathrm{QM\text{-}MM}} = E_{\mathrm{QM/MM}} - E_{\mathrm{QM}}$ (eqn (2)), is the interaction energy of the QM system with its environment. We define the stabilization energy for a given structure along a given reaction path as the difference between $E_{\mathrm{INTERACTION}}$ for that structure, and $E_{\mathrm{INTERACTION}}$ for the reactant complex of the corresponding reactant species on the same reaction path; *i.e.* the stabilization energy is the amount by which a structure is stabilized by the environment relative to the reactant.

Results

QM Energy profiles, reactant and TS structures

A summary of the energy profiles for reaction in the two environments (enzyme and water) is given in Fig. 2 (which also shows structures of typical reactant, transition state and product conformations). The average reaction barrier in the enzyme is 11.3 ± 1.8 kcal mol^{-1} and in water the barrier is 17.4 ± 1.9 kcal mol^{-1}. The average potential energy barriers are in good agreement with the experimental enthalpy barriers (12.7 ± 0.4 kcal mol^{-1} and 20.7 ± 0.4 kcal mol^{-1} in enzyme and water,[24] respectively; see the discussion below on the contribution of conformational changes to the observed energy barrier in solution). The average potential energy barrier in the enzyme in this work is 0.7 kcal mol^{-1} lower than the reaction barrier we previously reported at the same level of theory,[34] this is due to the structural differences mentioned above in

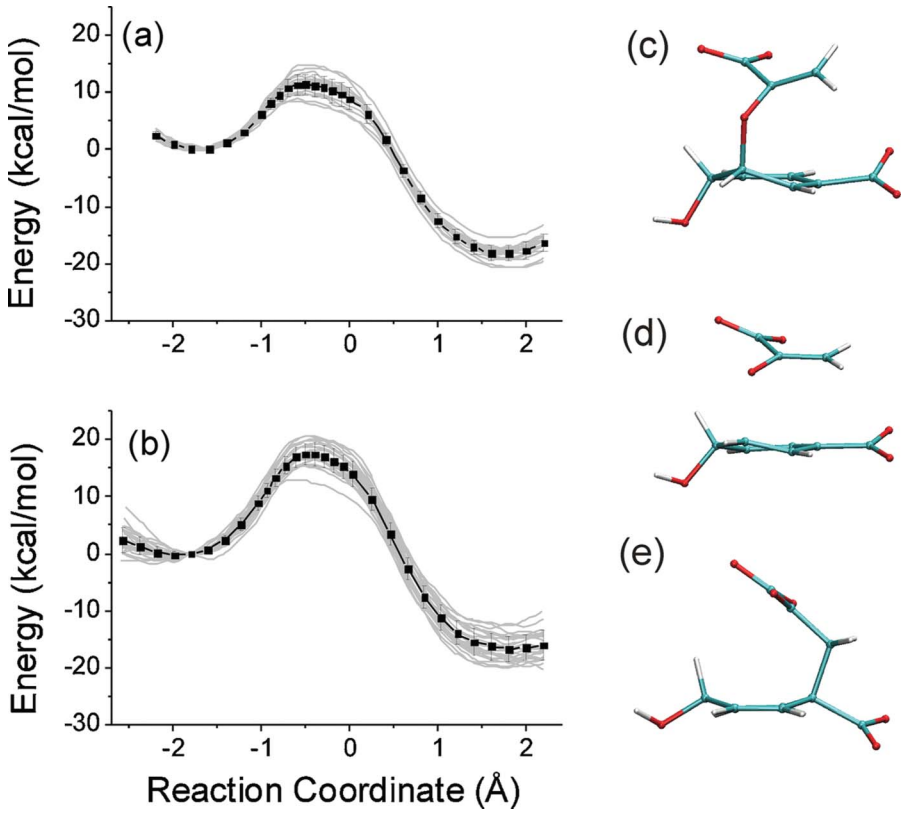

Fig. 2 B3LYP/6-31G(d)/CHARMM27 QM/MM energy profiles for reaction in (a) BsCM and (b) water; (c)–(e) typical structures for the reactant (top), transition state (middle) and product (bottom) in the enzyme. The individual pathways are shown in grey. All curves are plotted relative to the lowest point on the reactant (substrate) side. The average energy relative to the reactant at each point is given by black dots with error bars showing the standard deviation of the different pathways.

the starting structures for the QM/MM optimization which were obtained at a different level of theory (SCCDFTB/CHARMM22 here *versus* AM1(PM3)/CHARMM22 previously). In previous work we have found that the contributions of the thermal energy, the $p\Delta V$ term and the zero point energy (ZPE) corrections to the enthalpy barrier are small.[35] For the reaction in the enzyme, ZPE and thermal energy are estimated to reduce the barrier by 1.5 kcal mol^{-1} and 0.1 kcal mol^{-1}, respectively. Recently, Senn *et al.* have calculated finite temperature effects for the chorismate to prephenate reaction to be ~ 9 kJ mol^{-1} (~2.2 kcal mol^{-1}) at a temperature of 300 K.[51] Additionally, it is known from experiment that the entropy contribution to the free energy barrier is small and similar in both environments ($\Delta^{\ddagger}S = -12.1 \pm 0.4$ cal mol^{-1} K^{-1} in solution and ($\Delta^{\ddagger}S = -9.1 \pm 1.2$ cal mol^{-1} K^{-1} in BsCM),[24] showing that catalysis an enthalpic effect.[24,51] Although the calculated and experimental enthalpy barriers in water, (17.4 ± 1.9 kcal mol^{-1} from our calculations and 20.7 ± 0.4 kcal mol^{-1} for experiment[24]) are within the typical error of B3LYP calculations (3 kcal mol^{-1}[52]), this difference between experiment and theory suggests an additional enthalpic barrier in the solution related to formation of a reactive conformation. The optimized reactant structure in both environments is a pseudo-diaxial conformation[27,34] of chorismate, but the structures in solution differ from those in the enzyme as described below. It is probable that the reactant conformation found here is representative of the global minimum in the enzyme, but is only a local (free energy) minimum in water. The global minimum energy structure in solution has been found previously to be the pseudo-diequatorial conformation.[25–27] This energy, estimated as 0.9–3.6 kcal mol^{-1},[25–27] probably also contributes to the overall barrier to the (non-enzymic) reaction in solution. Thus, the overall value of $\Delta^{\ddagger}H = 18.8$–$21.0$ (± 1.9) kcal mol^{-1} for enthalpy barrier to the reaction in solution is the sum of the 17.4 kcal mol^{-1} from the reaction pathway calculations and this 0.9–3.6 kcal mol^{-1} estimate of this conformational contribution (see Table 2).

Geometrical details of the reactant and TS structures are summarized in Table 1. The energy profiles have similar shapes for the two different environments. As expected from the Hammond postulate, the TS (energy maximum) for this exothermic reaction is early in both environments. Despite the calculated difference in exothermicity, however, the TS is found at similar reaction coordinate values in both environments. The reaction coordinate values for the TS range from $r = -0.7$ to -0.4 Å in the enzyme, and from $r = -0.6$ to -0.4 Å in water. This is earlier than in semiempirical QM/MM studies, which position the TS typically at $r = -0.3$ Å.[9-11] For the average reaction path in the enzyme (found by averaging all the energies relative to the substrate), the TS lies at $r = -0.5$ Å, with the (bond breaking) C–O distance 2.14 ± 0.05 Å and (bond forming) C–C distance 2.64 ± 0.05 Å. For water, the TS for the average reaction path lies also at $r = -0.5$ Å, and the bond breaking C–O distance is 2.04 ± 0.05 Å (bond forming C–C distance is 2.54 ± 0.05 Å). All the TS structures for the different pathways in the enzyme are very similar to one another. The TS structures in water are also all very similar to one other (see below). However, the TS structures in the enzyme are clearly different

Table 1 Selected QM/MM B3LYP/6-31G(d)/CHARMM27 average structural parameters for the substrate (RS) and TS, for the reaction in the enzyme and in water[a]

	Enzyme			Water		
	RS*[b]	RS	TS	RS	RS*[b]	TS
C2–O13/Å	1.50 (0.01)	1.52 (0.01)	2.14 (0.05)	1.47 (0.01)	1.48 (0.01)	2.04 (0.05)
C4–C14/Å	3.29 (0.01)	3.13 (0.01)	2.64 (0.05)	3.46 (0.02)	3.28 (0.02)	2.54 (0.05)
C16–C17/Å	5.07 (0.15)	5.04 (0.16)	4.84 (0.1)	5.17 (0.11)	5.14 (0.10)	4.94 (0.07)
'NAC angle'/°	15.4 (1.9)	11.7 (0.5)	1.7 (1.7)	17.5 (2.1)	13.7 (1.9)	–0.6 (1.3)
Dihedral 1:						
C3–C2–O13–C15/°	60.9 (3.6)	59.7 (3.3)	57.3 (2.7)	58.4 (5.4)	64.5 (3.5)	52.3 (3.2)
Dihedral 2:						
C2–O13–C15–C14/°	–84.0 (3.0)	–91.1 (2.9)	–80.5 (2.5)	–96.7 (4.8)	–115.5 (3.5)	–77.9 (3.0)

[a] See Fig. 1 for atom labels. The 'NAC angle' is defined here as the angle between the line perpendicular to the plane defined by C5–C4–C3 and the line connecting C4–C14.[18] The standard deviation is given in parentheses. [b] The minimum energy substrate conformation (reactant state, RS) lies at reaction coordinate values of $r = -1.6$ Å and –2.0 Å for the enzyme and water environments, respectively. To aid comparison of the two environments in subsequent analyses a common reaction coordinate value (RS*) of $r = -1.8$ Å is used for both environments. The TS lies at a value of $r = -0.5$ Å in both environments.

Table 2 Experimental ($\Delta^{\ddagger}H$) and calculated potential energy barriers in kcal mol^{-1} (calculated barriers from the average of the potential energy profiles here)[a]

	Experiment[24]	B3LYP/6-31G(d)/CHARMM27	With conformational correction
Uncatalysed reaction	20.7 ± 0.4	17.4 ± 1.9	18.3–21.0 ± 1.9
BsCM	12.7 ± 0.4	11.3 ± 1.8	—

[a] The standard deviation of the 16 pathways in the enzyme and the 24 pathways in water is also shown. The third column (With conformational correction) adds an estimate (0.9–3.6 kcal mol^{-1}[25–27]) of the enthalpy difference between the ground state conformation in water (the pseudo-diequatorial conformation) and the pseudo-diaxial reactant conformation. ZPE is not included in the computed barriers here (estimated to reduce the barrier by ~1.5 kcal mol^{-1}).[35] It should also be noted that the B3LYP method underestimates the barrier to this reaction by around 3 kcal mol^{-1}.[35]

from those in water: the distance between the two carboxylate carbon atoms (C16–C17) is substantially smaller in the enzyme than in water. There is a substantial spread in the barrier in both environments. As will be shown below, this spread in both the enzyme and water is largely due to changes in the environment and does not involve significant differences in the structure of the TS.

The average (potential) energy of reaction is –18.2 ± 1.3 kcal mol^{-1} in the enzyme, and –16.7 ± 2.2 kcal mol^{-1} in water. The reaction is clearly exothermic. The calculated value of the potential energy of reaction in water at 25 °C is in good agreement with the experimental free energy and enthalpy of reaction in solution of: $\Delta_r G = -13.4$ kcal mol^{-1} and $\Delta_r H = -13.2 \pm 0.5$ kcal mol^{-1}, respectively:[53] it is important to note that the calculated energy of reaction in water is not calculated from the ground state conformation in water, so the energy for conversion from the pseudo-diaxial to the pseudo-diequatorial conformation should be taken into account. With this correction (estimated as 0.9–3.6 kcal mol^{-1},[25–27] see discussion below), the calculated reaction energy (between –15.3 and –13.1 kcal mol^{-1}) agrees well with experiment. The experimental figures show clearly that the reaction is enthalpy driven. Both the theoretical and experimental figures indicate that the reaction is effectively irreversible. The stability of the product is also demonstrated by the fact that prephenate is known to bind to the enzyme and form a stable complex. Indeed, crystal structures of the prephenate–enzyme complex have been crystallized and characterized.[50]

A second notable difference between the reactant structure in water from that in the enzyme is the orientation of the hydroxyl hydrogen (see also Fig. 2(c–e)). The hydroxyl OH on C1 has a definite preferred orientation in the enzyme but not in water. The orientation of the OH group can be defined by the H–O–C1–C4 torsion: a value of ~180° indicates that the hydroxyl hydrogen points outwards (OH_{out}) while a value of ~0° indicates a conformation with the hydrogen atom pointing inwards (OH_{in}). The average value for the torsion is 218.6 ± 4.6° in the enzyme and 167.2 ± 126.1° in water. Although the average values indicate an outward orientation in both environments, the standard deviation shows there is much bigger variation in water than in the enzyme. As discussed below, this is due to specific hydrogen bonding of the hydroxyl hydrogen to Glu78 in the enzyme, while no such specific interaction exists in water.

The reactant (minimum energy) conformation lies at $r = -1.6$ Å and –2.0 Å for the enzyme and water environments, respectively (this is the lowest energy point on the average path along the calculated profiles in each case; see Table 1). The geometrical parameters indicate that, in both environments, the reactant structures are so-called 'NAC's, according to definitions that have been used previously. Several definitions of a NAC have been given, including (i) Hur *et al.*[18] define a NAC as a conformation where the 'attack angle' is lower than 30° and the C–C bond forming distance is lower than 3.7 Å (sum of van der Waals radii); (ii) the same study[18] also defines a NAC as a conformation in which C3–C2–O13–C15 is ~50° and C2–O13–C15–C14 is ~–100°; and (iii) Guimaraes *et al.*[16] defined a NAC as a conformation in which the C4–C14 distance is less than 3.7 Å. The definitions (i) and (ii) are geometrically equivalent to each other only if it is assumed that the chorismate molecule is in the pseudo-diaxial conformation. Definition (iii) is a looser definition of a NAC than definition (i). As can be concluded from Table 1, all of these criteria are fulfilled in all substrate structures both in the enzyme and in water, showing that these NAC definitions do not provide a useful discrimination between the two environments. All the reactant structures in both environments are NACs. The NAC angle and the dihedral angles

do not differ significantly between the different environments. The results show a significant difference in barrier height between the enzyme and water (Table 2), despite the fact that the substrate is a so-called 'NAC' in each case. The observed differences in reaction barrier are therefore not due to the so-called NAC effect.

The difference between the calculated barriers in solution and in the enzyme is ~6.1 kcal mol^{-1}. This is less than the observed catalytic effect ($\Delta\Delta^{\ddagger}G$ = 9.1 kcal mol^{-1}). This suggests that there is an additional conformation contribution to catalysis. The free energy difference between the pseudo-diequatorial and pseudo-diaxial conformations of chorismate in water is estimated as 0.9–3.6 kcal mol^{-1}.[25–27] This is believed to be largely an enthalpic effect, so to a first approximation these figures can be taken as the ΔH between these conformations. To compare the calculated reaction barrier in water with the experimental value, this energy for the conformational change between the pseudo-diequatorial and pseudo-diaxial conformations should be added to the computed reaction barrier. These adjusted values are given in column 3 of Table 2, and are in excellent agreement with the experimental data. The calculated barriers given in Table 2 do not include zero-point effects, which are expected to lower the barrier slightly,[35] to (approximately) the same extent in the enzyme and in solution. It should also be pointed out that the B3LYP method underestimates the barrier to this reaction by around 3 kcal mol^{-1}.[35] Altogether, given these considerations, our results appear to be in excellent agreement with experiment.

Analysis of differences in energy barriers

To examine stabilization during the reaction, single-point calculations on the isolated QM region were carried out for the QM/MM optimized geometries along the different reaction paths (*i.e.* calculating the energy of the reacting system without the effects of the protein or water environment). These *in vacuo* reaction pathways for both environments have similar shapes, but begin at slightly different values of the reaction coordinate: the reactant in the enzyme is later on the reaction coordinate than in solution (closer to the TS). To clarify the analysis, and separate the different energetic contributions, the barriers (and TS stabilization) here are calculated relative to the same point on the reaction coordinate, r = –1.8 Å, which is in between the reactant (substrate) minimum for the enzyme and water environments (–1.6 Å and –2.0 Å, respectively). The energy change due to the shift in position along the reaction coordinate is small: only 0.1 and 0.8 kcal mol^{-1} for the enzyme and water environment, respectively. The TS structure is taken to lie at r = –0.5 Å, the typical value for the reaction in both the enzyme and water environment.

The absolute *in vacuo* QM energies (plotted relative to the value at r = –1.8 Å on the reaction pathway in water (–837.1312 Hartrees) for ease of viewing) are shown in Fig. 3. The average *in vacuo* barrier for the enzyme pathways is 19.2 ± 1.3 kcal mol^{-1} and is 18.4 ± 2.5 kcal mol^{-1} for the reaction in water. The difference between the *in vacuo* and QM/MM energy gives the relative stabilization of the reacting system by the environment (see eqn (1)). In Fig. 4, the stabilization energy along the reaction coordinate (relative to the reactant) is plotted for the different optimized pathways. In the enzyme (Fig. 4(a)), the TS is stabilized more than the reactant and the product is destabilized relative to the reactant. In water (Fig. 4(b)) the region around the transition state is also stabilized by a small amount in most reaction pathways. The product is also destabilized (relative to the substrate) in water, and the destabilization of the product is on average similar in the two environments. Substrate conformations in water earlier on the reaction coordinate than r = –1.8 Å are in general stabilized less than the r = –1.8 Å reactant conformation. Two important observations can be made from the *in vacuo* calculations: (i) the *in vacuo* reaction barrier is similar for both aqueous and enzymatic TS paths; and (ii) both environments, on average, stabilize the TS, but the enzyme stabilizes the TS significantly more than the water environment does.

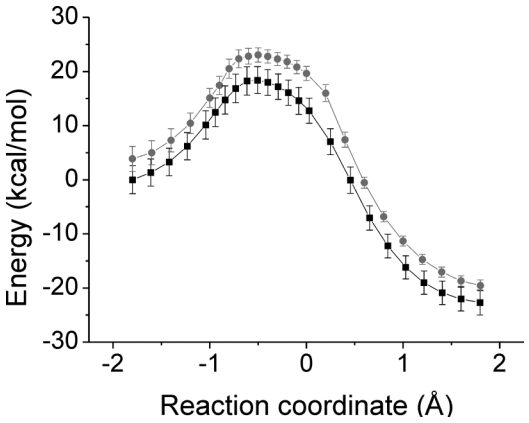

Fig. 3 Comparison of average QM energies *in vacuo* for paths from water (black squares) and enzyme (grey circles) environments. Error bars indicate the standard deviation from the average value at each point. The energies are plotted relative to the value at r = –1.8 Å on the reaction pathway in water (which has an absolute value of –837.1312 Hartrees) to aid comparison.

The TS stabilization correlates linearly with the computed barrier height in both environments, as shown in Fig. 4(c). Higher TS stabilization equates to a lower barrier. On average, the enzyme stabilizes the TS by 7.3 kcal mol^{-1} more than it stabilizes the reactant. In water, the TS is stabilized by only 1.0 kcal mol^{-1} on average. The gradient of the linear relationship fitted to both datasets is –0.95 (R value of 0.96), and the intercept is 18.3 kcal mol^{-1} (*i.e.* when the stabilization energy is 0 kcal mol^{-1}); this latter value is the predicted 'intrinsic' barrier from the substrate conformation in the absence of TS stabilization and is in good agreement with the gas phase barriers in both environments. The variation in the amount of the TS stabilization is due to structural differences in the environment. In water and in the enzyme, there are conformations of the environment that better stabilize the TS than others. The structural origin of these differences will be discussed below. The TS stabilization varies between 11.4 and 3.2 kcal mol^{-1} for the enzyme, showing that in all cases the TS is stabilized relative to the substrate in the enzyme. For water the extreme values are 3.9 and –2.0 kcal mol^{-1}, the latter indicating that the water environment destabilizes the TS (relative to the substrate) in some pathways. Importantly, the maximal stabilization does not necessarily occur at the TS reaction coordinate value (at r = –0.5 Å) for each individual pathway. However, in the enzyme environment the point of maximal stabilization is always close to the TS (typically the TS is found at r = –0.5 Å in the enzyme), indicating that the active site is built specifically to stabilize the TS. The stabilization found in solution is not very specific for a

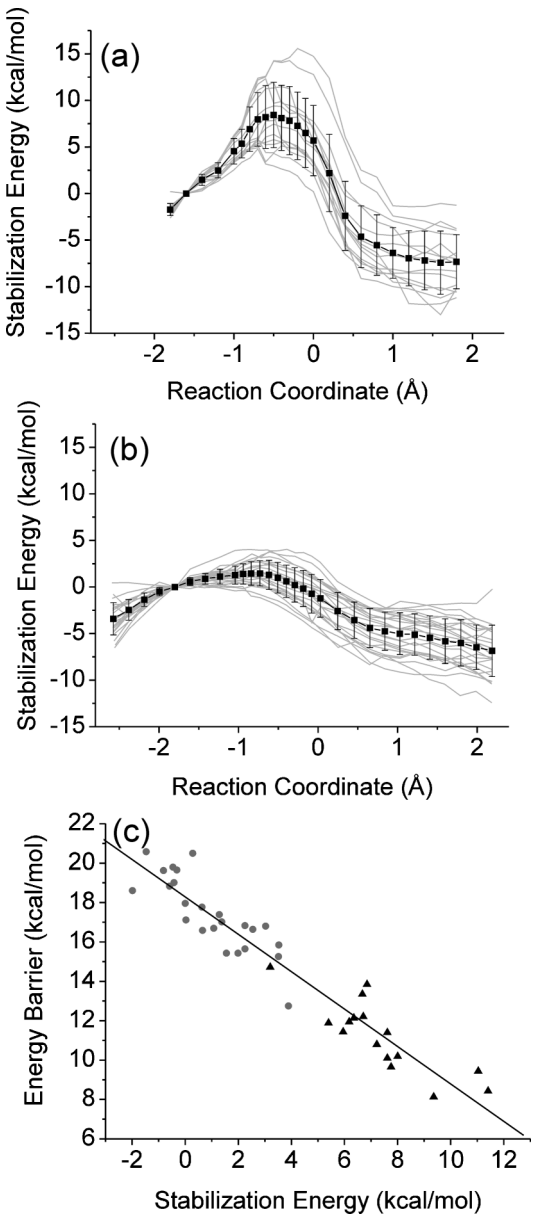

Fig. 4 Stabilization energy along the reaction coordinate (relative to the reactant point at $r = -1.8$ Å) (a) in the enzyme and (b) in water, the interaction energies for the individual pathways are shown in grey, while the average (with standard deviations as error bars) is shown in black. This is the interaction energy between the reacting (QM) system and the MM protein/water environment. A positive value indicates stabilization of the reacting system by the environment (*e.g.* the stabilization energy reaches a maximum for the transition state (TS) in the enzyme. In (c) the reaction barrier height *versus* stabilization energy at the TS (at $r = -0.5$ Å) for both environments (at the B3LYP/6-31G(d)/CHARMM27 level of QM/MM theory) is shown (grey circles: reaction in water environment; black triangles: reaction in the enzyme).

particular structure and clearly the solvent environment is not organized specifically for TS stabilization. The water environment generally stabilizes the region around the TS, but not specifically the TS itself: the TS in water is typically found around $r = -0.5$ Å, but the stabilization by the water environment does not change significantly between $r = -1.5$ Å and $r = -0.5$ Å. In both environments, the observed TS stabilization is electrostatic in nature; a detailed discussion follows below.

The average *in vacuo* energy profiles plotted in Fig. 3 also shows that the average QM reaction profile in the enzyme lies higher in absolute energy than for water. The energy difference between the enzyme and water profiles along the reaction path is constant along the reaction profile [average of all points on reaction coordinate: 5.1 ± 1.4 kcal mol^{-1}; 3.9 kcal mol^{-1} at $r = -1.8$ Å, 4.7 kcal mol^{-1} $r = -0.5$ Å, and 3.2 at $r = 1.8$ Å] and indicates a constant energy contribution along the reaction coordinate due to substrate destabilization, or strain, by the enzyme. An important indicator of this strain can be found in the distance between the two carboxylate carbons (C16–C17) of chorismate. The values for the C16–C17 distance over the whole reaction profile are on average 0.1 Å smaller in the enzyme environment compared to water (*e.g.* the C16–C17 distances for the reactant, TS and product are 5.07, 4.84 and 4.75 Å in the enzyme environment and 5.14, 4.94 and 4.85 Å in the water environment, respectively). In the enzyme, the electrostatic repulsion between the substrate carboxylate groups is higher than in solution, thus resulting in a higher absolute *in vacuo* energy for the substrate. The shorter substrate carboxylate–carboxylate distance in the enzyme compared to that in water is solely due to electrostatic stabilization of the compressed structure by the active site, *e.g.* the positively charged residues Arg7, Arg90 and Arg63, the negatively charged residue Glu78 and the overall hydrogen bonding network in the active site. The enzyme appears to force the carboxylate moieties closer together than is energetically allowed in the analogous solution reactant state. It is important to note that this substrate compression does not contribute to the lowering of the overall reaction barrier in the enzyme environment, compared to that within the water environment. Compression of the substrate by the enzyme was observed in the first QM/MM study of CM.[7] A compression of the carboxylate–carboxylate distance in the enzyme environment compared to water was also observed in the study of Štrajbl *et al.*,[13] and was shown to be a result of electrostatic stabilization of the TS. Guimaraes *et al.* found compression to be important in CM catalysis.[16]

The total difference in potential energy barrier between the enzyme and solution reactions can be split into two contributions: (i) a difference in TS stabilization (6.3 kcal mol^{-1}) and (ii) a small compression term (0.5 kcal mol^{-1}) due to the different position of the reactant state on the reaction coordinate. From this set of profiles we can conclude that there is no substantial energy contribution from substrate compression to catalysis. These data can be compared with, and contradict, the study of Guimaraes *et al.*[16] which calculated a conformational compression contribution of ~70% of the lowering of the free-energy barrier by the enzyme over that in aqueous solution. It should be noted that Guimaraes *et al.* used the AM1 method, which overestimates substrate compression in CM.[17] The thermodynamic cycles used in the analysis by Guimaraes *et al.*[16] have also been criticized[13] and may overestimate the contribution of this effect somewhat.

It is important to stress the implications of this kind of effect on the validity of the NAC proposal. This proposal rests on defining a common structural 'turnstile' conformation through which the substrate must pass in both solution and enzyme before proceeding to the TS. Given the clear structural differences between the solution and enzymic reactions, this idea does not seem to apply.

Table 3 Average transition state stabilization energies for the enzyme environment relative to the substrate[a]

	Electrostatic stabilization energy (kcal mol^{-1})	MM stabilization energy (kcal mol^{-1})
Enzyme model	9.9 (2.4)	–2.6 (1.7)
Arg90	6.0 (0.8)	
Arg7	3.1 (0.5)	
Glu78	2.1 (0.3)	
Cys75	1.3 (0.3)	
Tyr108	1.1 (0.2)	
Arg116	1.0 (0.9)	
Phe57	0.8 (0.3)	
Arg63	–0.6 (0.5)	

[a] Positive values show a stabilization of the TS relative to the substrate. The values given in brackets are the standard deviations obtained from the 16 different paths. The stabilization energy can be divided into an electrostatic (QM/MM) and a molecular mechanics (MM) term: the latter includes QM/MM van der Waals interactions and the small energy change due to changes in the structure of the protein itself (see eqn (1) and 2).

Although it may have been tempting to propose that such a state exists, given the overall similarity between the enzyme and solution mechanisms, subtle differences in structure can cause substantial energetic effects, and must be carefully considered. There is no unique reactive conformation of the substrate common to both environments, and the TS structures are also significantly different in the enzyme and in water. It is therefore probably futile to attempt to define a meaningful 'near attack conformation' for this reaction.

Stabilization by active site residues and structure of the active site in the enzyme

As shown in Fig. 4, both environments in general stabilize the TS and destabilize the product relative to the substrate, but by significantly different amounts. Transition state stabilization (relative to the substrate; sometimes called differential transition state stabilization (DTSS[12])) can be divided into an electrostatic term and a term relating to changes in the MM environment (Table 3). The electrostatic term is due to the interaction of the MM partial charges with the QM wavefunction, which includes polarization of the latter. The MM term relates to changes in MM energy, including terms corresponding to QM/MM van der Waals interactions. The total TS stabilization (relative to the substrate) provided by the enzyme is 7.3 ± 2.0 kcal mol^{-1} (average value from all paths ± standard deviation), with 9.9 ± 2.4 kcal mol^{-1} stabilization arising from QM/MM electrostatic interactions and –2.6 ± 1.7 kcal mol^{-1} from MM contributions. The electrostatic stabilization is the dominant term in TS stabilization, in agreement with previous findings.[7,12,13] The contributions of individual residues to the electrostatic transition state stabilization were also analysed. On average, the most stabilizing residue is Arg90 (6.0 ± 0.8 kcal mol^{-1}), followed by Arg7 (3.1 ± 0.5 kcal mol^{-1}) and Glu78 (2.1 ± 0.3 kcal mol^{-1}). Other active site residues (Cys75, Tyr108, Arg116 and Phe57) have a small stabilizing effect and Arg63 has on average a destabilizing effect on the TS (–0.6 ± 0.5 kcal mol^{-1}). The standard deviations of the stabilization energies for the individual active site residues are large, which is not translated to a correspondingly large standard deviation for the total transition state stabilization energy. This suggests that when different active site microstructure leads to a loss of TS stabilization due to a given residue, other residues compensate by providing enhanced transition state stabilization. The active site is apparently well organized to stabilize the TS specifically, with a number of groups able to provide the stabilization.

Detailed hydrogen bond analysis (based on standard structural criteria[10]) was performed and showed that there is no noticeable increase of number of hydrogen bonds at the TS compared to the reactant. The hydrogen bond between the Arg90-(N)H and O13 (*i.e.* between the reacting substrate ether oxygen atom and the residue that contributes most to TS stabilization) occurs in all pathways and its distance is ~0.05 Å shorter at the transition state than in the reactant state (1.75 ± 0.02 Å at reactant compared to 1.70 ± 0.02 Å at TS) in all paths.

Analysis of the chorismate to prephenate reaction in water

A typical structure of the first coordination sphere of water around the reactant, TS and product (with a cut-off of 6 Å from the centre of the substrate) is shown in Fig. 5. Overlaying the structures obtained for reactant, TS and product shows there is no major rearrangement in the first coordination sphere during reaction. There is also no evidence for an increase in hydrogen bonding in the TS compared to the reactant. Comparing the hydrogen bond network in water around the solute in the reactant and the TS reveals that, on average, in both (i) six water molecules are hydrogen bonded to each of the carboxylate groups, (ii) three water molecules are bound to the hydroxyl group and (iii) one water is hydrogen bonded to the ether oxygen. In Fig. 5(d) the effect of the surrounding water molecules on TS stabilization is shown, from calculating the barrier to reaction while including an increasingly larger water sphere around the substrate. To make direct comparison possible between the water and enzyme environments, the reactant conformation here is taken to be at r = –1.8 Å and the TS to be at r = –0.5 Å in water. This has only a small effect on the energy values (0.8 kcal mol^{-1}). The light grey lines depict the dependence of the barrier height on water sphere radius for each of the 24 pathways, and the average (with standard deviation as error bar) is given in black. The effect of including the water sphere on the barrier height is highly variable, if one compares the individual reaction pathways. Including the nearest water molecules, up to ~7 Å from the centre of the substrate, can have a stabilizing or destabilizing effect on the TS (relative to the reactant) for an individual pathway, although on average the inclusion of this solvation sphere has a net stabilizing effect of the TS relative to the substrate. The effect of increasing the size of the water sphere on the barrier height varies to a distance of ~15 Å; above 15 Å the barrier seems to have reached its asymptotic value of 16.7 kcal mol^{-1}. This is similar to the corresponding average value of 16.9 kcal mol^{-1} for the energy difference between points at r –0.5 Å and –1.8 Å along the reaction coordinate in water.

Charge analysis

The charge distribution of the QM region was studied *via* NBO (Natural Bond Orbital) analysis.[54] These NBO charges were compared to the Mulliken charges obtained from the calculations, and important differences were observed. While any method of assigning charges to atoms has limitations, it is generally found

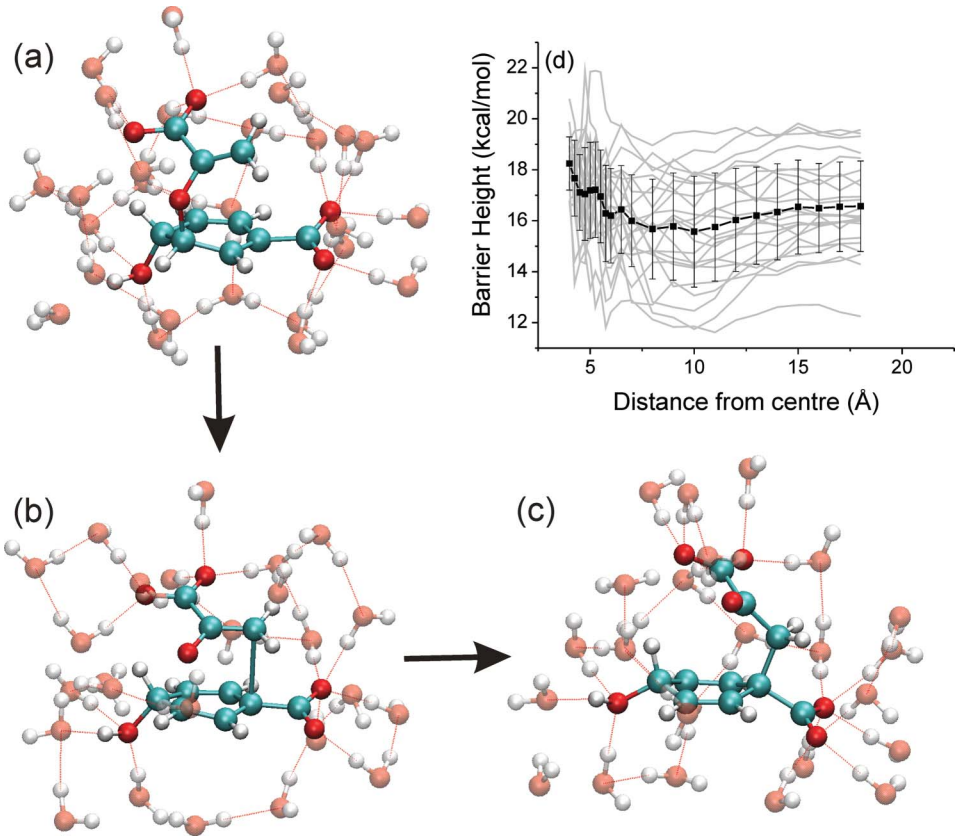

Fig. 5 Representative structures of (a) chorismate (at $r = -1.8$ Å), (b) the TS (at $r = -0.5$ Å) and (c) prephenate (at $r = 1.8$ Å) for reaction in water, in the first coordination sphere of water (waters shown are less than 6 Å from the centre of the substrate). (d) Influence on the barrier height of including an increasingly larger sphere of water molecules in the computation: the grey lines are for the individual pathways, and the black line is the average barrier height, with the standard deviation as error bars.

that NBO charges are more reasonable and more chemically meaningful than Mulliken charges. Analyses of this reaction based on Mulliken charges[30] are likely to be misleading. Here we only discuss the charges obtained *via* NBO analysis.

Average NBO charges for the atoms involved in the pericyclic reaction (C15, C2, C4, C3, C14 and O16) are given in Fig. 6 for the reaction in the enzyme and in solution. Both sets of charges show that there is important charge redistribution during the reaction. The most significant change in charge, in both environments, occurs for C2 which becomes more negatively charged (from 0.04 to –0.25 *e*), and C15 which becomes more positively charged as the reaction proceeds (from 0.19 to 0.50 *e*). For both C2 and C15, the changes in chemical environment explain the charge changes during reaction (see also Fig. 1), and this charge redistribution is largely independent of the environment. For C2, the bonding environment changes from a formal σ bond to carbon and a σ bond to oxygen to a σ and a π bond to carbon; both changes (sp^3 to sp^2 rehybridization, and C–O bond breaking) result in an increase in negative charge on C2 during the reaction (Fig. 6). For C15 the bonding changes from a formal π bond to carbon to a formal π bond to oxygen, *i.e.* changing from an ether carbon to a carbonyl carbon, this results in more positive charge on C15 in the product. Crespo *et al.*[46] presented an analysis based on Mulliken charges at the PBE/DZVP level of theory which were significantly different in magnitude, but show similar trends during the reaction.

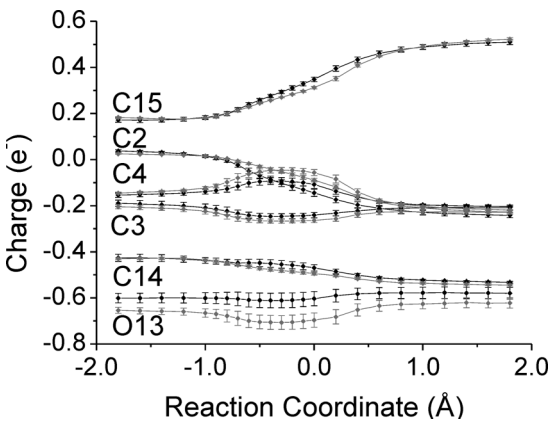

Fig. 6 Average B3LYP/6-31G(d)/CHARMM27 NBO atomic charges on the reacting atoms of (chorismate/prephenate) in the enzyme (grey diamonds) compared to those for the same reaction in solution (black diamonds). Error bars indicate the standard deviation from the average value. The transition state for the enzyme reaction lies on average at a reaction coordinate value of $r = -0.5$ Å.

The average charges on C3, C4, C14 and O13 all go through an extremal value near the TS in the enzyme and in solution. Fig. 6 shows the charges on O13, C3, C4 and C14 during reaction in the enzyme. The fact that the point at which the extremal values for the charges on these atoms are observed coincides with the

TS is consistent with our finding that TS stabilization is mainly electrostatic in both environments. The active site is evolved to recognize the unique electrostatic properties of the TS, which are significantly different from those of the substrate,[12] and thus to stabilize the TS specifically. The difference in the magnitude of the charges in the enzyme and in solution is greatest in the region around the TS, particularly for O13 which may be a consequence of the strengthening of electrostatic interactions with the active site at the TS.

Discussion

Multiple QM/MM adiabatic reaction pathways in the BsCM enzyme and in water produce well-defined estimates of the potential energy barrier to reaction in these environments. The results agree well with experiment. The use of an appropriate level of quantum chemical QM/MM theory (B3LYP/CHARMM27) allows reliable conclusions to be drawn. The calculated barrier heights differ substantially (11.3 kcal mol^{-1} in enzyme compared to 17.4 kcal mol^{-1} in water). Entropic, zero-point and thermal contributions to the barrier are relatively small for the chorismate to prephenate reaction,[24] and importantly are probably similar in the enzyme and in water. Catalysis by chorismate mutase can therefore be understood by comparison of potential energy barriers.

The activation enthalpies for the enzymic and water reactions at 25 °C are $\Delta^{\ddagger}H = 12.7 \pm 0.4$ kcal mol^{-1} and $\Delta^{\ddagger}H = 20.7 \pm 0.4$ kcal mol^{-1}, respectively, a reduction of $\Delta\Delta^{\ddagger}H = 8 \pm 0.4$ kcal mol^{-1} by the enzyme (compare to the experimental $\Delta\Delta^{\ddagger}G$ of 9.1 kcal mol^{-1}).[24] Of the experimentally observed $\Delta\Delta^{\ddagger}H$ of 8 kcal mol^{-1}, approximately 0.9–3.6 kcal mol^{-1} (estimated from experimental[26] and computed values[25,27]) is due to the binding of the pseudo-diaxial conformation of chorismate by the enzyme. The lower estimate of 0.9–1.4 kcal mol^{-1} was obtained experimentally by Copley and Knowles from the temperature dependence of ^{1}H NMR coupling constants at 25 °C.[26] However, Campbell *et al.*[55] were unable to detect the pseudo-diaxial form of chorismate in solution using transferred nuclear Overhauser effects, which may suggest that the abundance of chorismate in the pseudo-diaxial form is not as high as indicated in the study of Copley and Knowles.[26] The upper limit of 3.6 kcal mol^{-1} comes from the MC/FEP study of chorismate in solution by Carlson and Jorgensen.[27] The estimates of the free energy of NAC formation from QM/MM FEP[17] and EVB[13] studies (3.8–5 kcal mol^{-1}) have not been included in this estimate as this energy is attributed to more than just the energy difference between pseudo-diequatorial and pseudo-diaxial forms and as such would be an over-estimate of this conformational contribution. Further contributions to catalysis can be analyzed by comparison of the potential energy barriers found here. The calculated average difference in potential barrier between the enzyme and solution reactions is 6.1 kcal mol^{-1}. This can be divided into two contributions: (i) better TS stabilization by the enzyme (6.3 kcal mol^{-1}); and (ii) a small effect (0.5 kcal mol^{-1}) due to compression of the distance between the C4 and C14 (the atoms which form a bond during the reaction) from ~3.5 Å in water to ~3.3 Å in the enzyme. The latter effect can be described as strain of the substrate. The first QM/MM investigation of chorismate mutase[7] suggested that catalysis is due to a combination of substrate conformational effects and TS stabilization, and it is gratifying to see these early proposals borne out by detailed analysis using reliable QM methods; the contribution of strain appears small, however.

From these calculations, and from previous FEP studies,[13,17] we can conclude that TS stabilization contributes ~69% of the $\Delta\Delta^{\ddagger}G$ of 9.1 kcal mol^{-1}, additionally, there is a compression term contributing ~5% to the $\Delta\Delta^{\ddagger}G$, while binding of the pseudo-diaxial conformation also contributes ~26%. These results are in line with previous calculations,[13,17] but not with the more approximate (and indirect) estimates reported by Bruice *et al.*[18,20,29–31] who attributed ~90% or more of the total $\Delta\Delta^{\ddagger}G$ to the so-called NAC effect in chorismate mutases from all the species studied. TS stabilization by BsCM has been found in most previous studies at different levels of theory.[7,9–11,17,28,34] The importance of TS stabilization is also consistent with experimental mutagenesis results.[32] It should also be pointed out that the binding of the pseudo-diaxial form, and compression of the substrate, are the result of the complementarity of the enzyme active site for the TS, so are consequences of the high affinity of the enzyme for the TS.

The barriers calculated here for the reaction in both environments show excellent agreement with experiment, further indicating that the B3LYP/CHARMM27 level of QM/MM theory gives an accurate picture of the chemical reaction. The barrier and the energy of reaction can be compared with previous QM/MM studies of BsCM.[7,10,12,15,42,46,47,56] From these studies it can be concluded that both AM1 and HF[7,10,12,15,42,46,47,56] overestimate the reaction barrier severely, while standard non-hybrid density functional techniques (*e.g.* PBE[46] and SCC-DFTB) underestimate the reaction barrier. The energy of reaction of –18.2 (± 1.3) kcal mol^{-1} for the enzyme reaction in our study is comparable to values reported in other studies at similar levels of theory.[10,42] The reaction energy in water found here is quite different from that calculated by Crespo *et al.*,[46] (–16.7 (± 2.2) kcal mol^{-1} *versus* –22.0 kcal mol^{-1}), due probably to the different levels of QM theory used (B3LYP *vs.* PBE). It is important to remember that B3LYP has some limitations. For example, dispersion is not represented, which can be important for some enzyme-catalysed reactions.[57] B3LYP energies for enzyme reactions can also be significantly incorrect in some cases; they can be tested against high level correlated *ab initio* methods.[58] For CM, comparison with LCCSD(T) results shows that B3LYP underestimates the barrier to reaction somewhat, by approximately 3 kcal mol^{-1}.[35] This underestimation is probably very similar for the reaction in either enzyme or solvent.

No results based on low-level (*e.g.* semiempirical QM) modelling can be considered to be definitive, unless the methods have been specifically developed and tested for that specific application. Similarly, no definitive conclusions on catalysis can be drawn without detailed analysis of transition states. It is important to note that methods which calculate significantly incorrect barriers may well also give estimates of TS stabilization and conformational energies that are incorrect. Other potentially important effects to consider are the need to study multiple pathways when calculating energy profiles for enzymic reactions, and the choice of the crystal structure used for modelling.[2,5,6,9,59,60]

Analysis of the effects of the environment on the reaction pathways shows that both environments reduce the barrier, compared to the same reaction pathways in vacuum, by TS stabilization. However, the enzyme environment is much better at stabilizing the TS than water (on average TS stabilization

energy of 7.3 *versus* 1.0 kcal mol^{-1}). The TS stabilization is linearly correlated with the barrier height for both environments. In each environment, the reaction pathways are structurally similar (though not identical). The variation in barrier height in either environment is due to differences in the structure of the environment, and depends on the degree of TS stabilization. When the effect of the environment is completely removed, the pathways for the solution and enzymic reactions show similar intrinsic reaction barriers. However, the reaction pathway in the two environments is slightly different. This is exemplified by the larger distance between the two carboxylate moieties of the substrate in water throughout the reaction. This indicates that electrostatic repulsion between the two carboxylate moieties is the origin of this strain effect, as proposed previously.[13,16] The compressed conformation is stabilized in the enzyme by positively charged groups in the active site (Arg90, Arg7, Arg63). This compression equates to a destabilization (strain) of the reactant in the enzyme compared to the water environment (see Fig. 3) but this destabilization remains constant during the reaction, and does not contribute significantly to catalysis. It should be noted also that the product conformation is also compressed (strained) compared to that in solution. This may assist the overall catalytic process by reducing the binding affinity for the product somewhat, thus speeding product release. Štrajbl *et al.* also found that the compression of the distance between the two carboxylate groups plays an important role in catalytic activity.[13] When excluding this electrostatic interaction, by setting the charges on the carboxylate groups to zero, the free energy barriers for reaction within water and the enzyme converge to a similar value. Also the electrostatic free energy to go from the open pseudo-diequatorial form to the pseudo-diaxial rises by 3 kcal mol^{-1} for the reaction in water, while it drops by 10 kcal mol^{-1} for the reaction in the enzyme (see Figure 9 in reference[13]). Compression effects were mentioned by Crespo *et al.* [46] who report a distance between the C4–C14 atoms of ~3.1 Å in the enzyme and ~3.6 Å in water in QM/MM MD simulations at the PBE/DZVP QM level. Both the absolute distances (average values in minimized structures of ~3.3 Å and ~3.5 Å in the enzyme and solution, respectively) and degree of compression found here differ somewhat, which is probably due mostly to the different level of QM treatment (B3LYP here). However, in the same study the C4–C14 distances in the TS are reported to be ~2.6 and ~2.5 Å in the enzyme and in water, respectively, which agrees very well with the values reported in this study (average ~2.6 and ~2.5 Å, respectively). Guimaraes *et al.*[16] calculated a conformational compression contribution of ~70% of the lowering of the free energy barrier by the enzyme over aqueous solution. However, these results were found with the AM1 method, which overestimates substrate compression in chorismate mutase.[7,9,10]

All the substrate conformations found here, both in water and in the enzyme, qualify as 'near attack conformers' by all previous definitions. The differences in reaction barrier found here therefore are not due to the 'NAC effect'. The observed structural differences in reaction pathways between the two environments are also an argument against the notion of a NAC, which is supposed to be a 'turnstile' through which the system must pass for reaction.[19] There is no common 'turnstile' here: not only is the conformation of the substrate in the enzyme different from that in solution, the TS structures are different also. There is no unique definition of a NAC. While perhaps further definitions could be proposed to fit the observations here, this does not seem to be a useful route to understanding catalysis.

An additional important difference between the structures found here in the enzyme and in water is the directionality of the hydroxyl hydrogen. Marti *et al.* found two possible different pathways in the enzyme, depending on which direction the hydroxyl O–H points.[11] When the OH points inwards (towards the centre of the ring, OH_{in}), they found that the reaction follows a higher energy pathway than when the OH points outwards (to the enzyme environment, OH_{out}). This is because of different hydrogen bonding networks in the two cases: when the OH points inwards, the hydroxyl group forms a hydrogen bond to Cys75; while when the OH group points outwards, it donates a hydrogen bond to the side-chain of Glu78. Marti *et al.*[11] report a free energy difference between OH_{in} and OH_{out} pathways of 7.7 kcal mol^{-1}. The hydroxyl group in our reaction pathways in water has no preferred orientation. In the enzyme we find it always points outwards (hydrogen bonding to Glu78). This appears to be the favored orientation in the enzyme. Possible alternative orientations of this hydroxyl group have also been observed in *ab initio* QM/MM modelling.[10]

In water, the relatively small stabilization of the TS is mostly due to the closest water molecules, as expected, but water molecules up to 15 Å from the centre of the substrate have some effect. The amount of TS stabilization differs substantially for different water conformations. The structure of the first coordination sphere does not change significantly from reactant to TS. Analysis of hydrogen bonds for the reaction in water shows, on average, (i) six water molecules hydrogen bonded to both of the carboxylates, (ii) three water molecules bound to the hydroxyl group and (iii) one water hydrogen bonded to the ether oxygen at the TS. These results are generally similar to the results of Carlson *et al.*[27] who also found six hydrogen bonds to both carboxylates and three to the hydroxyl oxygen, but 2 H-bonds to the ether oxygen, from AM1/TIP3P Monte Carlo FEP calculations.

The TS stabilization in the enzyme is mainly electrostatic. This finding is in line with previous, lower-level studies.[7,9,10,12,13,28,42,46] Arg90 makes the largest contribution to TS stabilization (6.0 ± 0.8 kcal mol^{-1}). Arg7 and Glu78 both stabilize the TS (relative to substrate) in most of the pathways, and are on average the second and third most important residues in TS stabilization (3.1 ± 0.5 kcal mol^{-1} and 2.1 ± 0.3 kcal mol^{-1}, respectively). There are minor, variable contributions to TS stabilization from Cys75, Arg116, Tyr108 and Phe57. Arg63 on average slightly destabilizes the TS relative to the substrate. In crystal structures of BsCM (PDB codes 2CHT[23] and 1COM[50]) this residue is only bound to the substrate in ~1/3 of active sites, remaining solvent exposed in the others. Woodcock *et al.*[59] calculated reaction paths with and without Arg63 bound to the substrate and found little difference in barrier height at the SCCDFTB/CHARMM22 level (6.0 kcal mol^{-1} compared to 5.7 kcal mol^{-1} Arg63 bound and unbound, respectively), suggesting that Arg63 is not important for catalysis.

Calculation of multiple pathways (*i.e.* for multiple active site configurations) provides insight into the factors affecting TS stabilization. It is important to ensure that the structure of the active site remains stable and agrees with available experimental data throughout a simulation. The available crystal structures of BsCM[23,50,61] and NMR data[49] suggest that there is a hydrogen bond formed between Glu78 and Met79 in all structures. We have

found in previous AM1/CHARMM22 and PM3/CHARMM22 MD simulations that this hydrogen bond can be broken, allowing Glu78 to form a hydrogen bond with Arg90 for which there is no experimental evidence. Structural changes of this type are not observed in the SCCDFTB/CHARMM22 MD simulations which provided the starting structures for the adiabatic mapping calculations discussed here. However, this minor structural difference has a relatively small effect on the calculated barrier heights (~0.7 kcal mol^{-1}; average barrier here ~11.3 kcal mol^{-1} and ~12.0 kcal mol^{-1} in previous modelling[34]).

NBO population analysis of structures from the two different environments show that the charge on the O13 atom is overall more negative in the enzyme than in water and that the negative charge on this atom is greatest in the region of the TS. This finding is in contrast to the conclusion of Zhang *et al.*[30] that the charge for O13 is relatively constant throughout the reaction and similar in the two environments, ruling out the stabilization of any increasing negative charge on this atom by Arg90. Zhang *et al.* based their analysis on Mulliken and ESP charges at the SCCDFTB/MM level of theory.[30] The NBO charges used here are generally recognized to be more accurate than Mulliken charges, in that they reflect expected chemical behaviour more reliably.[62]

The hydrogen bond between Arg90-(NH) and O13 shortens by ~0.05 Å at the TS in all pathways. Woodcock *et al.*[42] also found a small shortening of the Arg90(NH)-O13 H-bond (from 1.86 to 1.80 Å) in a single reaction pathway at a similar level of theory. The observed 0.05 Å decrease in the length of the Arg90-NH-O13 hydrogen bond found here is smaller than at the AM1/CHARMM22[9] or RHF/6-31G(d)/CHARMM22 levels,[10] which may be related to the unrealistically high barriers to reaction given by these lower levels of theory. Similarly, MP2/6-31G(d) calculations on the RHF/6-31G(d)/CHARMM structures indicate very large TS stabilization relative to the gas phase.[12] As should be apparent, it may be important to use methods that give accurate barriers and structures to draw quantitative conclusions about energetic contributions to catalysis. It is heartening, however, that the key findings of TS stabilization by the enzyme, and the central role of Arg90 in TS stabilization, are found consistently at a variety of different levels of theory.[7–13,22,33]

Conclusions

Multiple QM/MM reaction pathways, in the enzyme and in water, give barriers for the chorismate to prephenate rearrangement in very good agreement with experiment. The use of a reliable level of QM/MM theory allows conclusions about catalysis to be drawn with confidence. There are important differences between the reaction pathways in the enzyme and those in water. This is a consequence of compression of chorismate by the enzyme (in particular, the carboxylate groups of the substrate being closer to one another). This distortion does not significantly affect the intrinsic barrier for reaction, but destabilizes the enzyme-bound substrate compared to that in solution by an equal amount (~5 kcal mol^{-1}) throughout the whole reaction. This is not a distortion towards a common TS: the TS structures in the enzyme are also significantly different from those in solution. Catalysis cannot be understood in terms of simple arguments based on substrate structure alone. Both environments lower the reaction barrier, compared to their corresponding *in vacuo* values, but the enzyme stabilizes the TS (relative to substrate) significantly more than water (average TS stabilization of 7.3 kcal mol^{-1} *versus* 1.0 kcal mol^{-1}). The most important residues for TS stabilization in the enzyme are Arg90, Arg7 and Glu78, which stabilize the TS *via* electrostatic interactions. Binding to the enzyme is tightest for the TS: the enzyme active site is complementary to the TS. The observed differences in substrate conformation when bound to the enzyme are therefore due in all probability to the complementarity of the enzyme to the TS. CM is a good example of an enzyme for which TS stabilization is central to catalysis. TS stabilization is the major contributor to lowering the potential energy reaction barrier. Catalysis in CM is therefore mainly due to TS stabilization.

Acknowledgements

AJM is an EPSRC Leadership fellow and also thanks IBM High Performance Computing Life Sciences Outreach Program (with FC), BBSRC, and EPSRC (with KER) for support. JNH thanks EPSRC for an Advanced Research Fellowship. FRM thanks the Royal Society. FC thanks EPSRC for a Postdoctoral Research Fellowship at the Life Sciences Interface.

Notes and references

1 M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, *Science*, 2004, **303**, 186–195.
2 K. E. Ranaghan and A. J. Mulholland, *Int. Rev. Phys. Chem.*, 2010, **29**, 65–133.
3 A. J. Mulholland, *Drug Discovery Today*, 2005, **10**, 1393–1402.
4 M. H. M. Olsson, W. W. Parson and A. Warshel, *Chem. Rev.*, 2006, **106**, 1737–1756.
5 R. Lonsdale, K. E. Ranaghan and A. J. Mulholland, *Chem. Commun.*, 2010, **46**, 2354–2372.
6 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
7 P. D. Lyne, A. J. Mulholland and W. G. Richards, *J. Am. Chem. Soc.*, 1995, **117**, 11345–11350.
8 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *J. Phys. Chem. B*, 2000, **104**, 11308–11315.
9 K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Mol. Phys.*, 2003, **101**, 2695–2714.
10 K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Org. Biomol. Chem.*, 2004, **2**, 968–980.
11 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Theor. Chem. Acc.*, 2001, **105**, 207–212.
12 B. Szefczyk, W. A. Sokalski, K. E. Ranaghan and A. J. Mulholland, *J. Am. Chem. Soc.*, 2004, **126**, 16148–16159.
13 M. Štrajbl, A. Shurki, M. Kato and A. Warshel, *J. Am. Chem. Soc.*, 2003, **125**, 10228–10237.
14 H. Guo, Q. Cui, W. N. Lipscomb and M. Karplus, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 9032.
15 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Chem.–Eur. J.*, 2003, **9**, 984–991.
16 C. R. W. Guimaraes, M. P. Repasky, J. Chandrasekhar, J. Tirado-Rives and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2003, **125**, 6892–6899.
17 K. E. Ranaghan and A. J. Mulholland, *Chem. Commun.*, 2004, 1238–1239.
18 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 5964–5972.
19 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 1472–1473.
20 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 10540–10542.
21 T. Ishida, *J. Chem. Phys.*, 2008, **129**, 125105-1-14.
22 T. Ishida, *J. Am. Chem. Soc.*, 2010, **132**, 7104–7118.
23 Y. Chook, H. Ke and W. Lipscomb, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 8600–8603.
24 P. Kast, M. Asif-Ullah and D. Hilvert, *Tetrahedron Lett.*, 1996, **37**, 2691–2694.
25 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *THEOCHEM*, 2003, **632**, 197–206.
26 S. D. Copley and J. R. Knowles, *J. Am. Chem. Soc.*, 1987, **109**, 5008–5013.

27 H. A. Carlson and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1996, **118**, 8475–8484.
28 A. Shurki, M. Štrajbl, J. Villà and A. Warshel, *J. Am. Chem. Soc.*, 2002, **124**, 4097–4107.
29 S. Hur and T. C. Bruice, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12015–12020.
30 X. D. Zhang, X. H. Zhang and T. C. Bruice, *Biochemistry*, 2005, **44**, 10443–10448.
31 X. H. Zhang and T. C. Bruice, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 18356–18360.
32 A. Kienhöfer, P. Kast and D. Hilvert, *J. Am. Chem. Soc.*, 2003, **125**, 3206–3207.
33 C. R. W. Guimaraes, M. Udier-Blagovic, I. Tubert-Brohman and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2005, **1**, 617–625.
34 F. Claeyssens, K. Ranaghan, F. Manby, J. Harvey and A. Mulholland, *Chem. Commun.*, 2005, 5068–5070.
35 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.
36 P. A. Bartlett and C. R. Johnson, *J. Am. Chem. Soc.*, 1985, **107**, 7792–7793.
37 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.
38 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
39 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
40 M. J. Field, *Mol. Phys.*, 1997, **91**, 835–845.
41 Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2001, **105**, 569–585.
42 H. L. Woodcock, M. Hodoscek, P. Sherwood, Y. S. Lee, H. F. I. Shaefer and B. R. Brooks, *Theor. Chem. Acc.*, 2003, **109**, 140–148.
43 *Jaguar 4.0*. 1996–2001, Schrödinger, Inc.: Portland, Oregon.
44 J. W. Ponder, *TINKER: Software Tools for molecular design*. 2003: Saint Louis, MO, http://dasher.wustl.edu/tinker/.
45 J. N. Harvey, *Faraday Discuss.*, 2004, **127**, 165–177.
46 A. Crespo, D. A. Scherlis, M. A. Martí, P. Ordejon, A. E. Roitberg and D. A. Estrin, *J. Phys. Chem. B*, 2003, **107**, 13728–13736.
47 Y. S. Lee, S. E. Worthington, M. Krauss and B. R. Brooks, *J. Phys. Chem. B*, 2002, **106**, 12059–12065.
48 N. Foloppe and A. D. MacKerell, *J. Comput. Chem.*, 2000, **21**, 86–104.
49 A. Eletsky, A. Kienhofer, D. Hilvert and K. Pervushin, *Biochemistry*, 2005, **44**, 6788–6799.
50 Y. Chook, J. Gray, H. Ke and W. Lipscomb, *J. Mol. Biol.*, 1994, **240**, 476–500.
51 H. M. Senn, J. Kästner, J. Breidung and W. Thiel, *Can. J. Chem.*, 2009, **87**, 1322–1337.
52 P. E. M. Siegbahn, *J. Biol. Inorg. Chem.*, 2006, **11**, 695–701.
53 P. Kast, Y. Tewari, O. Wiest, D. Hilvert, K. N. Houk and R. N. Goldberg, *J. Phys. Chem. B*, 1997, **101**, 10976–10982.
54 J. P. Foster and F. Weinhold, *J. Am. Chem. Soc.*, 1980, **102**, 7211–7218.
55 A. P. Campbell, T. M. Tarasow, W. Massefski, P. E. Wright and D. Hilvert, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 8663–8667.
56 T. Ishida, D. G. Fedorov and K. Kitaura, *J. Phys. Chem. B*, 2006, **110**, 1457–1463.
57 R. Lonsdale, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. Lett.*, 2010, **1**, 3232–3237.
58 M. W. van der Kamp, J. Zurek, F. R. Manby, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. B*, 2010, **114**, 11303–11314.
59 H. L. Woodcock, M. Hodoscek and B. R. Brooks, *J. Phys. Chem. A*, 2007, **111**, 5720–5728.
60 A. Lodola, M. Mor, J. Zurek, G. Tarzia, D. Piomelli, J. N. Harvey and A. J. Mulholland, *Biophys. J.*, 2007, **92**, L20–L22.
61 J. E. Ladner, P. Reddy, A. Davis, M. Tordova, A. J. Howard and G. L. Gilliland, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2000, **56**, 673–683.
62 K. B. Wiberg and P. R. Rablen, *J. Comput. Chem.*, 1993, **14**, 1504–1518.

