ARTICLE

OBC
www.rsc.org/obc

Transition state stabilization and substrate strain in enzyme catalysis: *ab initio* QM/MM modelling of the chorismate mutase reaction

Kara E. Ranaghan,[a] Lars Ridder,[c] Borys Szefczyk,[b] W. Andrzej Sokalski,[b] Johannes C. Hermann[d] and Adrian J. Mulholland*[a]

[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] *Wroclaw University of Technology, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland. E-mail: sokalski@mml.ch.pwl.wroc.pl; Fax: +48-71-3203364; Tel: +48-71-320245*

[c] *Molecular Design & Informatics, N.V. Organon, P.O. Box 20, 5430 BH Oss, The Netherlands*

[d] *Institut für Pharmazeutische Chemie, Heinrich-Heine-Universität, 40 225 Düsseldorf, Germany*

Received 30th October 2003, Accepted 10th February 2004
First published as an Advance Article on the web 3rd March 2004

To investigate fundamental features of enzyme catalysis, there is a need for high-level calculations capable of modelling crucial, unstable species such as transition states as they are formed *within* enzymes. We have modelled an important model enzyme reaction, the Claisen rearrangement of chorismate to prephenate in chorismate mutase, by combined *ab initio* quantum mechanics/molecular mechanics (QM/MM) methods. The best estimates of the potential energy barrier in the enzyme are 7.4–11.0 kcal mol^{-1} (MP2/6-31+G(d)//6-31G(d)/CHARMM22) and 12.7–16.1 kcal mol^{-1} (B3LYP/6-311+G(2d,p)//6-31G(d)/CHARMM22), comparable to the experimental estimate of $\Delta H^{\ddagger} = 12.7 \pm 0.4$ kcal mol^{-1}. The results provide unequivocal evidence of transition state (TS) stabilization by the enzyme, with contributions from residues Arg90, Arg7, and Arg63. Glu78 stabilizes the prephenate product (relative to substrate), and can also stabilize the TS. Examination of the same pathway in solution (with a variety of continuum models), at the same *ab initio* levels, allows comparison of the catalyzed and uncatalyzed reactions. Calculated barriers in solution are 28.0 kcal mol^{-1} (MP2/6-31+G(d)/PCM) and 24.6 kcal mol^{-1} (B3LYP/6-311+G(2d,p)/PCM), comparable to the experimental finding of $\Delta G^{\ddagger} = 25.4$ kcal mol^{-1} and consistent with the experimentally-deduced 10^6-fold rate acceleration by the enzyme. The substrate is found to be significantly distorted in the enzyme, adopting a structure closer to the transition state, although the degree of compression is less than predicted by lower-level calculations. This apparent substrate strain, or compression, is potentially also catalytically relevant. Solution calculations, however, suggest that the catalytic contribution of this compression may be relatively small. Consideration of the same reaction pathway in solution and in the enzyme, involving reaction from a 'near-attack conformer' of the substrate, indicates that adoption of this conformation is not in itself a major contribution to catalysis. Transition state stabilization (by electrostatic interactions, including hydrogen bonds) is found to be central to catalysis by the enzyme. Several hydrogen bonds are observed to shorten at the TS. The active site is clearly complementary to the transition state for the reaction, stabilizing it more than the substrate, so reducing the barrier to reaction.

Introduction

Chorismate mutase is an important enzyme in the testing and development of theories of catalysis, and is at the centre of much current debate in enzymology. Chorismate mutase catalyses the Claisen rearrangement of chorismate to prephenate (Fig. 1), a rare example of a biochemically-catalysed pericyclic reaction. An activation free energy of $\Delta G^{\ddagger} = 15.4$ kcal mol^{-1} has been found for *Bacillus subtilis* chorismate mutase (BsCM), compared to $\Delta G^{\ddagger} = 24.5$ kcal mol^{-1} for the uncatalyzed reaction in aqueous solution (the reaction is believed to be a unimolecular pericyclic reaction in both cases), a rate acceleration by the enzyme of 10^6.[1] The enzymic reaction is part of the shikimate pathway for the biosynthesis of aromatic amino acids in plants, fungi, and bacteria. As a consequence, chorismate mutase is a potential target for the development of new herbicides, fungicides and antibiotics. It has been the subject of many theoretical[2–14] and experimental[1,15–20] studies over the years, but the source of the catalytic efficiency of the enzyme remains a matter of debate. Until recently, it has been widely accepted that stabilization of the transition state (through electrostatic interactions with the enzyme) is central to catalysis in this enzyme.[2,3,6,9,11–14,19,20] However, Bruice and co-workers[21,22] have recently proposed that catalysis in this enzyme (and many others) is instead due to the ability of the enzyme to bind 'near attack conformers' of the substrate (NACs). NACs are defined as conformations of the substrate in which the bond-forming atoms are distances apart less than or equal to the sum of their van der Waals radii (*e.g.* a carbon–carbon

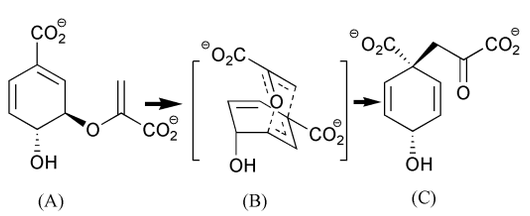

Fig. 1 The Claisen rearrangement of chorismate to form prephenate, showing (A) diaxial chorismate, (B) the chair-like TS and (C) prephenate.

DOI: 10.1039/b313759g

distance ≤ 3.7 Å in chorismate mutase) (other definitions have also been proposed). Bruice *et al.* suggest that once the enzyme has bound (stabilized) this more TS-like conformation, almost no extra stabilization of the transition state is necessary for catalysis. Based on structures obtained from MD simulations, the mole fractions of NACs in the enzyme and in solution were calculated and used to estimate a free energy for NAC formation in *E. coli* chorismate mutase (EcCM). Their results suggest that the observed catalytic effect of the enzyme is 90% due to the ability of the enzyme to support NACs, compared to the very low concentrations of NACs in solution, with transition state stabilization playing only a minor role. This proposal has been the subject of considerable debate and controversy. The calculation of potential catalytic contributions through unrestrained molecular dynamics simulations in this way has been criticized as being unreliable.[23] In contrast to the NAC proposal, earlier QM/MM modelling of the enzyme reaction (with semiempirical methods, see below) and mutagenesis studies, indicate significant transition state (TS) stabilization in chorismate mutase.[2,13,14,17,20] Calculations and experiments dating back many years have also suggested a role for substrate destabilization/strain/binding of a reactive conformer in catalysis.[2,8,13,24] Detailed comparison of the reaction in solution and in the enzyme, with semiempirical QM/MM methods, indicates that both substrate preorganization and transition state stabilization contribute to catalysis in the enzyme.[7–10,14] Recently Jorgensen and co-workers[25] have studied solvent effects on the reaction, and their AM1 QM/MM Monte Carlo/Free Energy Perturbation (MC/FEP) study suggests that the conversion of non-NACs to NACs provides no free energy contribution to catalysis. Chorismate mutase is thus an important system for the testing and development of hypotheses on the fundamental nature of enzyme catalysis.

We have modelled the reaction in chorismate mutase by *ab initio* QM/MM methods. We have applied high-level *ab initio* and density-functional theory approaches to calculate corrections to the QM/MM energy profiles, leading to good-quality estimates of the activation barrier in the enzyme, which are in agreement with experiment. These calculations go beyond our[2,14] and other previous studies[7–10,14] in using higher levels of QM/MM theory. We study a large, fully solvated, enzyme model. The model, and details of the reaction pathway calculations, have been extensively tested by previous lower-level modelling.[14] We compare the results in detail with previous studies. Contributions of individual residues to lowering the activation energy have been identified. The energetics of the same reaction pathway in solution have been studied, using various continuum solvation models, using the same *ab initio* QM levels of theory applied in the QM/MM studies of the enzyme reaction. Comparison of the reaction in the enzyme with that in solution shows that the major contribution to catalysis arises through better stabilization of the transition state (relative to the substrate) by the enzyme. The active site is organized to stabilize the transition state specifically. Comparison with the solution reaction is essential in analyzing catalysis.[26] The results, at reliable levels of quantum chemical theory, provide a detailed picture of the structural and energetic features of the reaction. They demonstrate, unequivocally, the central importance of TS stabilization in chorismate mutase. They also indicate a strain or distortion of the substrate by the enzyme, towards a more TS-like geometry with a reduced carbon–carbon distance, in agreement with earlier modelling.[2,13,14,25] The degree of compression of chorismate by the enzyme is, however, smaller than predicted by lower-level methods. This strain/compression/substrate destabilization may also play a role in catalysis, as indicated by an empirical relationship,[27] and suggested by earlier QM/MM results.[2,24] Comparison here with results for the reaction in solution indicates that this may be only a small contribution to catalysis.

Methods

We have modelled the reaction within the enzyme (a large, fully solvated model) at the *ab initio* RHF/6-31G(d) QM/MM level using CHARMM (version 27a1)[28] interfaced with GAMESS-US.[29,30] The substrate (chorismate) was treated at the QM level, including the effects of the rest of the system *via* a molecular mechanical force field, in this case the CHARMM22 all-atom MM protein parameters.[31] The method allows geometry optimization of the system as a whole at every step of the calculation, and the QM atoms are polarized by the MM atoms at each stage. Extensive testing of these QM/MM techniques has been carried out. They have been applied recently to several different enzymes.[29,32,33] Adiabatic mapping along a defined reaction coordinate was used to model the reaction, a method tested fully in our previous calculations at the AM1 level of QM/MM theory.[14] *Ab initio* calculations in the gas phase (B3LYP/6-311+G(2d,p) and MP2/6-31+G(d)) have been used to correct the RHF/6-31G(d)/CHARMM QM/MM potential energy barrier for the effects of electron correlation (see below). The levels of theory used here (MP2/6-31G(d)//6-31G(d) and B3LYP/6-311+G(2d,p)) provide good descriptions of the TS structure and energetics of the reaction.[14]

The model was prepared according to the method described and tested in our previous (lower level) study.[14] The active-site at the interface between chains A and B of the BsCM trimer (from the transition state analogue[34] (TSA) complex structure 2CHT[35]) was chosen as the model system. This structure has been used in some previous theoretical studies.[2,6,11,14] The structure of chorismate optimized in the gas-phase at the RHF/6-31G(d) level[14] (with electrostatic potential fit derived atomic charges (for MM solvation, see below), and atom types consistent with the CHARMM22 all-atom parameter set[31]) was superimposed on to the TSA within the enzyme using a rigid body least squares fit, and then the TSA was deleted.

Hydrogen atoms were built onto the enzyme structure (including the crystallographic water molecules), and their positions optimized (500 steps steepest descents (SD)) with CHARMM.[28] The enzyme was then solvated by superimposing the structure on a 30 Å cube of 8000 pre-equilibrated CHARMM TIP3P water molecules. The enzyme–substrate complex was centred within the box and all water molecules and protein residues outside a 25 Å sphere centred on C5 of chorismate (see Fig. 2 for atom numbering used here) were deleted, while any residue with at least one heavy atom within the 25 Å sphere was retained. Water molecules having their oxygen atom within 2.8 Å of any heavy atom were deleted. The energy of the water molecules was minimized by MM, keeping all other atoms fixed (500 steps SD, 1500 steps Adopted Basis Newton Raphson (ABNR)), and then equilibrated by 85 ps of stochastic boundary molecular dynamics with a friction coefficient of $\beta = 62$ ps^{-1} applied to the oxygen atoms of the water molecules.[36] The water was restrained to remain within the simulation system by a spherical deformable potential of radius 25 Å.[37] The central point for the application of the water boundary potential was also C5 of chorismate. The positions of the water molecules (only) were again optimized (500 steps SD, 904 steps ABNR). A gradient tolerance convergence criterion of 0.01 kcal mol^{-1} was used in all calculations (MM and QM/MM). A non-bonded cut-off of 25 Å was also used in all MM calculations. The model comprised 4211 protein atoms, 24

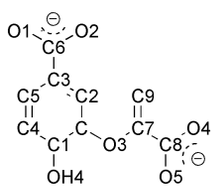

Fig. 2 Chorismate showing the atom numbering used here.

atoms of the substrate and 947 TIP3 water molecules (including 144 crystallographically observed), altogether 7076 atoms.

For MM calculations, atoms 21–25 Å from the centre of the model were restrained using force constants scaled to increase from the inside to the outside of the region.[14] Charged groups in this region were made artificially neutral (retaining atomic charges designed to show realistic H-bonding behaviour[14]), making the total charge of the system (including chorismate) $-2e$.

The enzyme was then optimized by MM with the substrate fixed (1000 steps SD followed by 1143 steps ABNR). The energy of the whole system was then optimized at the RHF/6-31G(d) QM/MM level, treating chorismate QM. Diffuse functions were not used, because the 6-31G(d) basis set gives good results, similar to 6-31+G(d), for this reaction,[14] and the use of diffuse functions can cause SCF convergence problems in QM/MM calculations.[32] The energy was minimized by the ABNR algorithm (250 steps). As only a limited amount of minimization can be carried out at the *ab initio* QM/MM level atoms further than 16 Å away from the chorismate were held fixed. Water molecules in the region between 10 Å and 16 Å away from the chorismate were restrained by applying a small harmonic force to the oxygen atoms (5 kcal mol^{-1} Å^{-2}). A deformable boundary potential was not applied in the QM/MM calculation. The effects of these restraints on the barrier were tested at the semiempirical level (AM1/CHARMM). The results showed a 3.0 kcal mol^{-1} increase in barrier height for the total energy, compared to the barriers found with fewer restraints.[14] An approximate reaction coordinate was defined as the difference of two bond lengths—the breaking (C–O3) and forming (C3–C9) bonds. This was found to be a suitable choice in test calculations.[14] A series of minimizations was run at reaction coordinate values from −1.8 to 1.8 Å at 0.3 Å intervals. A harmonic restraint (k = 5000 kcal mol^{-1}) was applied to maintain the desired reaction coordinate values. The structures were minimized sequentially for 10 steps using the SD algorithm and 40 steps of ABNR. The *ab initio* QM/MM calculations are extremely computationally demanding, and so only limited minimization is possible. Comparison of the QM geometries with fully optimized gas-phase structures[14] indicates that the QM system is properly optimized. After initial analysis of the energy profiles, two further points around the energy maximum (−0.4 and −0.2 Å on the reaction coordinate) were chosen and minimized by 10 steps of SD and 40 steps of ABNR to obtain a smoother profile.

Structures were analyzed for important interactions with the protein. A hydrogen bond was defined as a hydrogen to acceptor distance of ≤ 2.6 Å (2.8 Å if the donor is sulfur) and a donor to acceptor distance of ≤ 3.5 Å (4.0 Å if the donor is sulfur) and the angle between atoms forming the hydrogen bond must also be greater than 90°.[38]

To test the effects of electron correlation, the energies of the QM atoms alone (*i.e.* not including the effects of the MM atoms) along the RHF/6-31G(d) QM/MM optimized pathway were calculated using Møller–Plesset perturbation theory (MP2/6-31+G(d)) and density functional theory (B3LYP/6-311+G(2d,p)) with GAUSSIAN98.[39] The energies obtained were used to correct the RHF/6-31G(d) QM/MM energy profile, as in previous work.[32,33] This approach is necessary because correlated QM/MM calculations are not currently practical, due to their extremely high computational demands. It is suitable because of the similarity of the substrate and TS geometries at the different levels.[14] The energy profiles were corrected by subtracting the RHF/6-31G(d) QM energy from the RHF/6-31G(d) QM/MM energy and then adding the QM energy calculated at the higher level. Energy profiles generated at the AM1/CHARMM level[14] were also corrected in the same way for comparison.

To compare TS stabilization in the enzyme and in aqueous solution, the reaction in solvent was modelled using solvent continuum models in the GAUSSIAN98[39] and JAGUAR[40] programs. As with the higher-level corrections, the geometry of the substrate was taken from the structures generated by the QM/MM RHF/6-31G(d)/CHARMM22 pathway calculations at each point on the reaction coordinate, allowing direct comparison of the enzyme and solution profiles. Single point energy calculations using the polarized continuum model (PCM),[41] and the isodensity polarized continuum model (IPCM)[42] (in GAUSSIAN98[39]), and the polarized continuum model in JAGUAR[40] (PCM(J)[43,44]) were carried out. The PCM model was used with the RHF/6-31+G(d), MP2/6-31+G(d) and B3LYP/6-311+G(2d,p) *ab initio* methods, and the IPCM model was used with the RHF/6-31+G(d) method. The Gaussian98 default settings for the PCM and IPCM models (water as the solvent, dielectric constant of 78.39; cut-off of 0.0004 for IPCM) were used in these calculations. The PCM(J) model was used in calculations at the B3LYP/6-31G(d) level.

Table 1 H-Bonds after initial RHF/6-31G(d) QM/MM minimisation[a]

	A–H/Å	A–D/Å	AHD/°
C75 (N/HN)–O	2.084	2.888	155.2
R63 (NH1/HH11)–O1	1.667	2.652	154.2
SOLV7102 (OH2/H1)–O2	1.629	2.619	169.4
R90 (NH2/HH21)–O3	1.968	2.875	128.5
XSOL124 (OH2-H1)–O3	1.808	2.777	163.3
R7 (NH1/HH12)–O4	1.692	2.699	161.9
R7 (NH2/HH22)–O4	2.477	3.190	134.5
XSOL116 (OH2/H2)–O4	1.636	2.567	150.1
R7 (NH2/HH22)–O5	1.690	2.704	163.0
R90 (NE/HE)–O5	1.644	2.644	162.1

[a] A means acceptor, D means donor, and H means hydrogen.

Results

Structure of the enzyme–substrate complex

The geometry of the substrate changed significantly from its gas-phase geometry on QM/MM minimization. The bond forming distance is reduced and the length of the breaking bond increased, giving a more TS-like structure. In comparison to the gas-phase, the C3–C9 (forming) bond length at the RHF/6-31G(d) level was reduced by 0.3 Å to 3.43 Å, and the C–O3 (breaking) bond length increased by 0.02 Å to 1.45 Å. The total energy and QM/MM energy decreased significantly during geometry optimization, with the total energy of the system being reduced by 57.5 kcal mol^{-1} and the QM/MM energy reduced by 61.0 kcal mol^{-1}. The geometry optimization of the system destabilizes the substrate, shown by an increase in QM energy of 9.7 kcal mol^{-1} (RHF/6-31G(d)).

Table 1 shows the hydrogen bonds between the substrate and protein residues in the *ab initio* QM/MM minimized enzyme–substrate complex. The substrate forms hydrogen bonds (as defined above) with: Arg7 (NH1 & NH2), Arg63 (NH1), Arg90 (NH2 & NE), and Cys75 (N). Three water molecules were also observed to form hydrogen bonds with the substrate: XSOL 116 and 124 (crystallographically observed water molecules), and SOLV7102 (a water added during the solvation procedure).

Reaction pathway calculations

Figure 3 shows the energy profile for the reaction calculated at the RHF/6-31G(d)/CHARMM level of QM/MM theory. The total energy barrier at this level is 36.6 kcal mol^{-1}. The approximate transition state occurs at a value of −0.3 Å on the reaction coordinate. The QM/MM energy barrier (*i.e.* from the energy of the QM system plus QM/MM electrostatic interactions) is 27.0 kcal mol^{-1}. The energy barrier for the substrate alone (*i.e.* the QM energy) was 41.4 kcal mol^{-1} (RHF/6-31G(d)). These energy barriers are overestimates, due mostly to the well-known limitations of Hartree–Fock methods. Table

Table 2 Energy barriers to reaction in the enzyme

	Total energy barrier[a]/kcal mol^{-1}	QM/MM energy barrier [b]/kcal mol^{-1}	Position on reaction coordinate/Å	ΔE of reaction/ kcal mol^{-1}
MP2/6-31+G(d)//6-31G(d)/CHARMM22	11.0	5.0	−0.6	−35.6
B3LYP/6-311+G(2d,p)//6-31G(d)/CHARMM22	16.1	10.0	−0.6	−30.9
MP2/6-31+G(d)+AM1/CHARMM22 [c]	12.3	10.8	−0.4	−30.7
B3LYP/6-311+G(2d,p) +AM1/CHARMM22[c]	15.2	14.0	−0.6	−26.1

[a] Total energy = MM + QM/MM energy. [b] QM/MM energy = QM energy including electrostatic interactions with the MM atoms. [c] Similar to barriers reported in,[14] except calculated with the same restraints used in RHF/6-31G(d) pathway calculations (see text).

Table 3 Energy barriers to reaction in the gas-phase

	Energy barrier[a]/kcal mol^{-1}	Position on reaction coordinate/Å	ΔE of reaction/kcal mol^{-1}
RHF/6-31G(d)	49.1	−0.3	−16.0
MP2/6-31+G(d)	22.4	−0.3	−17.5
B3LYP/6-311+G(2d,p)	24.8	−0.6	−12.8

[a] Single point calculations on geometries taken from the 6-31G(d)/CHARMM22 enzyme path.

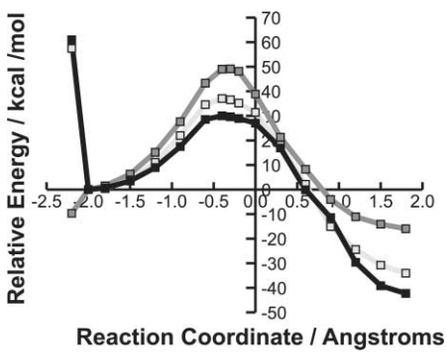

Fig. 3 RHF/6-31G(d)/CHARMM22 QM/MM energy profiles (relative to bound substrate). Light grey: total energy of the system (*i.e.* MM + QM/MM energy); black: QM/MM energy; medium grey: QM energy. The first point is the energy of the model with chorismate in its gas-phase optimized geometry,[14] and is shown for comparison.

2 shows the corrected values for the total energy barrier and the QM/MM energy barrier, and Fig. 4 shows the corrected energy profiles. MP2/6-31+G(d) corrections give a total energy barrier of 11.0 kcal mol^{-1}, while the B3LYP/6-311+G(2d,p) corrected total energy barrier is 16.1 kcal mol^{-1}. The (approximate) TS occurs earlier on the corrected profiles than on the uncorrected profile: in both of the corrected profiles the TS is at −0.6 Å on the reaction coordinate. The energy change for the reaction predicted by the B3LYP/6-311+G(2d,p) calculations is less than that predicted by RHF/6-31G(d) or MP2/6-31+G(d) methods. Entropic effects are relatively small in the enzyme environment ($\Delta S^{\ddagger}_{act}$ = −9.1 cal mol^{-1} K^{-1} at 298 K[1]), so the calculated energies show good agreement with the experimental

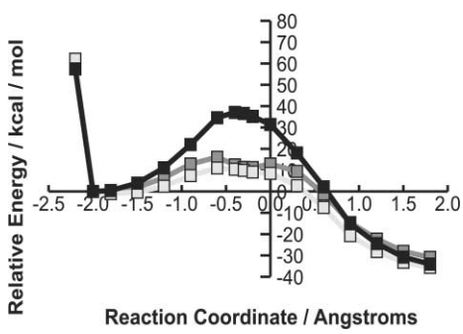

Fig. 4 Corrected total energy profiles for the reaction in the enzyme, relative to bound substrate. Black: RHF/6-31G(d) total energy profile; medium grey: B3LYP/6-31+G(2d,p)//6-31G(d)/CHARMM22; light grey: MP2/6-31+G(d)//6-31G(d)/CHARMM22. The first point is again the energy of the model with chorismate in its gas-phase optimized geometry,[14] and is shown for comparison.

values for the thermodynamic parameters for this enzyme ($\Delta G^{\ddagger}$ = 15.4 kcal mol^{-1} and $\Delta H^{\ddagger}$ = 12.7 kcal mol^{-1}).[1] The experimental value is an upper limit for the rearrangement of chorismate to prephenate as the reaction rate may also be partly limited by diffusion.[45]

Table 3 and Fig. 5 show the results of the MP2/6-31+G(d) and B3LYP/6-311+G(2d,p) gas phase calculations. The MP2/6-31+G(d) and B3LYP/6-311+G(2d,p) energy barriers (22.4 kcal mol^{-1} and 24.8 kcal mol^{-1}, respectively) are approximately half that given by the RHF/6-31G(d) method. As expected from previous gas phase calculations, the energy barrier is clearly overestimated at the RHF/6-31G(d) level.[14] The MP2/6-31+G(d) and B3LYP/6-311+G(2d,p) barriers are very similar, although the barrier occurs at different points on the reaction coordinate: the location of the maximum at the MP2/6-31+G(d) level is the same (−0.3 Å) as RHF/6-31G(d) (gas phase and enzyme), whereas B3LYP/6-311+G(2d,p) predicts an earlier TS (−0.6 Å), also earlier than in the enzyme. Gas phase optimization at the B3LYP/6-31+G(d) level shows the TS to be at a value of −0.53 Å on the reaction coordinate.[14]

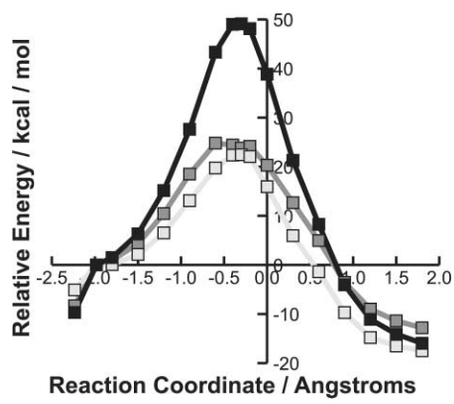

Fig. 5 *Ab initio* QM energy profiles (*i.e.* for QM atoms only in the gas phase) along the RHF/6-31G(d)/CHARMM22 QM/MM reaction path (energies relative to bound substrate). Black: RHF/6-31G(d) energies, medium grey: B3LYP/6-311+G(2d,p) energies, and light grey: MP2/6-31+G(d) energies. Energies are the result of single point calculations of the QM atoms in their RHF/6-31G(d)/CHARMM22 optimized geometries at each point on the reaction coordinate. The first point is the energy of the model with chorismate in its gas-phase optimized geometry,[14] and is shown for comparison.

Analysis of the QM/MM reaction path: TS stabilization

The difference between the QM energy and the QM/MM energy provides an indication of the electrostatic stabilization provided by the enzyme. An electrostatic stabilization of the TS

of 19.7 kcal mol^{-1} (relative to the bound, optimised substrate) is found at the RHF/6-31G(d)/CHARMM22 QM/MM level. The transition state is clearly significantly stabilized, relative to the substrate, by its interactions with the enzyme. Figure 6 shows how the electrostatic stabilization varies along the reaction path. The stabilization increases to a peak at the point on the reaction coordinate corresponding to the barrier in the total energy profile. The amount of stabilization drops off after the TS before increasing again towards the product end of the reaction coordinate, indicating stabilization of the product (relative to the substrate).

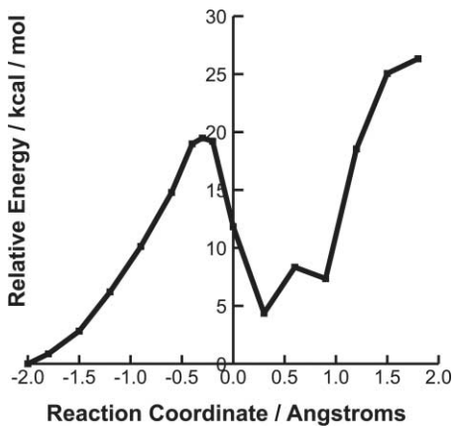

Fig. 6 QM/MM electrostatic stabilization, relative to substrate, along the RHF/6-31G(d) path. Note: TS stabilization may be underestimated here due to the absence of the interaction with Glu78 until 1.2 Å on the reaction coordinate. The appearance of this interaction later in the reaction may relate to the increased stabilization of the product.

Stabilization of the TS relative to the reactant is achieved through strengthening of hydrogen bonds with charged active site residues. Figure 7 shows the variation in hydrogen to acceptor distances between the reacting system and residues Arg90, Arg7, and XSOL124 along the reaction coordinate. The hydrogen bonds with these residues shorten along the reaction path, reaching a minimum at the TS (−0.3 Å on the reaction coordinate). This indicates an increase in strength of the hydrogen bonds, showing that they stabilize the transition state. The change is most significant for the hydrogen bonds between Arg90 and O3 (shortened by 0.18 Å at the TS), and between XSOL124 and O3 (0.1 Å). Figure 8 shows the active site structure at the transition state. XSOL124 and Arg90 help stabilize the oxygen of the ether bond as the C–O bond breaks, and arginines 7, 63, and 90 interact with the carboxylate groups.

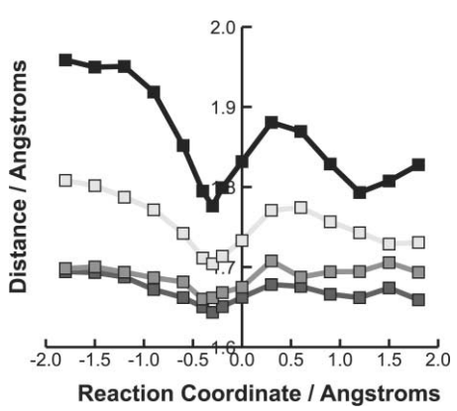

Fig. 7 Variation in H-bond lengths along the RHF/6-31G(d)/CHARMM22 reaction path. Darkest to lightest: Arg90(HH21)–O3, Arg7(HH22)–O5, Arg7(HH12)–O4, and XSOL124(H1)–O3.

There is also a change in the number of hydrogen bonds along the path. Initially, the interactions are the same as those observed in the enzyme–substrate complex (Table 1). An additional hydrogen bond between the sulfur of Cys75 and H4 of chorismate was formed at an early stage on the reaction coordinate (−0.9 Å). At −1.2 Å on the reaction coordinate, a hydrogen bond forms between O2 of chorismate and Arg63. The number of interactions remains unchanged until a value of 0.6 Å on the reaction coordinate when the interaction of H4 with SG of Cys75 is lost (the H-bond with backbone of Cys75 (N) remains throughout). In the QM/MM minimized enzyme–substrate complex the donor (O) to acceptor (SG) distance is 3.61 Å and the acceptor (SG) to hydrogen (H4) distance is 2.86 Å, close to the definition of a (OH. . .S) hydrogen bond used here. At the TS the hydrogen to acceptor distance is shorter at 2.80 Å, making the interaction a hydrogen bond. However, in the product the acceptor to donor distance is significantly longer at 4.22 Å (3.82 Å A–H). At a value of 1.2 Å on the reaction coordinate, a hydrogen bond forms between H4 of the reacting system and Glu78 (*i.e.* the OH group of prephenate/chorismate rotates). In the product, the acceptor (OD2) to donor (O) distance is 2.67 Å and the acceptor (OD2) to hydrogen (H4) distance is 1.70 Å, making it a strong interaction. This change in hydrogen bonding is reflected in the calculated stabilization along the path (Fig. 6) between reaction coordinate values of 0 and 1.0 Å. The hydrogen bond with Glu78 stabilizes the product relative to the bound substrate. This interaction would also stabilize the TS, as shown by analysis of semi-empirical QM/MM results.[2,14] The stabilization found here for the TS is an underestimate, because the hydrogen bond with Glu78 is not formed in the TS complex, and instead is only formed at a later stage in the reaction. This is probably because of the limited minimization possible at the *ab initio* QM/MM level. Formation of this hydrogen bond was observed after more extensive QM/MM optimization at the (computationally cheaper) AM1/CHARMM22 level.[14] The additional stabilization of the TS (relative to the substrate) that the interaction with Glu78 would provide can be estimated from QM calculations as up to 3.6 kcal mol^{-1}.[46] All possible conformations of the hydroxyl group of chorismate were analysed at the MP2/6-31G(d) level of theory, indicating that the maximal reduction in activation barrier is achieved when the proton of the hydroxyl group is directed towards Thr174 and Cys75 in the substrate complex and towards Glu78 at the TS.[47] This would increase the total TS stabilization by the enzyme, relative to the reactants, to 23.3 kcal mol^{-1}, comparable to the calculated stabilization of the product.

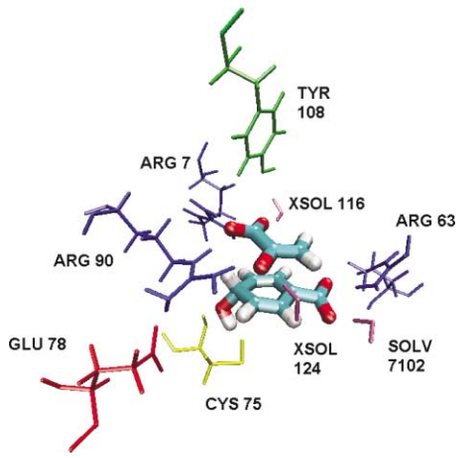

Fig. 8 The transition state at the enzyme active-site, showing important residues (QM atoms shown as thick tubes).

The reaction in solution

Figure 9 shows the energy profiles and Table 4 shows the energy barrier to the reaction in solution using continuum solvent models, with structures taken from the QM/MM path in the enzyme. This allows direct comparison of the stabilization

Table 4 The energy barriers to reaction in solution

	Energy barrier/kcal mol^{-1}	Position on reaction coordinate/Å	ΔE reaction/kcal mol^{-1}
RHF/6-31G(d)/PCM	53.0	−0.2	15.9
RHF/6-31+G(d)/PCM	53.4	−0.3	15.2
RHF/6-31+G(d)/IPCM	36.0	−0.4	23.3
MP2/6-31+G(d)/PCM	28.0	−0.2	18.5
B3LYP/6-311+G(2d,p)/PCM	24.6	−0.2	12.2
B3LYP/6-31G(d)/PCM(J)[a]	21.9	−0.4	21.1

[a] PCM(J) denotes the solvent model of the Jaguar program.[40]

Table 5 Relative energies of gas-phase (−2.2 Å) and enzyme bound (−2.2 Å) optimized structures of chorismate with different solvation models and *ab initio* methods

Ab initio method	Energy of structure at −2.2 Å on the reaction coordinate (relative to −2.0 Å on the reaction coordinate)
HF/6-31+G(d)/IPCM	−12.5
HF/6-31+G(d)/PCM	1.8
MP2/6-31+G(d)/PCM	5.0
B3LYP/6-311+G(2d,p)/PCM	2.1
B3LYP/6-31G(d)/PCM(J)[a]	2.2

[a] PCM(J) means the solvent model of the Jaguar program.[40]

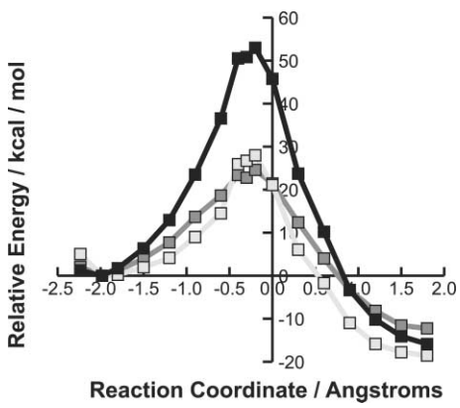

Fig. 9 Energy profiles for the reaction in solution. Black: PCM/RHF/6-31G(d) energy profile; light grey: PCM/MP2/6-31+G(d); medium grey: PCM/B3LYP/6-31+G(2d,p). The first point is the energy of chorismate in its gas-phase optimized geometry,[14] and is shown for comparison.

provided by the enzyme with that in solution, along the reaction path. As expected, the Hartree–Fock *ab initio* methods significantly overestimate the barrier to reaction in solution. There is also some considerable variation in the barrier depending on the solvent model used. Different continuum solvent models give very different estimates of these relatively subtle effects. To investigate this further, variations in the parameters of the continuum models could be tested. Ideally, detailed simulations with explicit solvent would be preferable, but these are not feasible at the *ab initio* QM/MM level.

The B3LYP/6-311+G(2d,p)/PCM barrier is in very good agreement (24.6 kcal mol^{-1}) with the experimental value of $\Delta G^{\ddagger}$ = 24.5 kcal mol^{-1} for the solution reaction. The MP2/6-31+G(d)/PCM (28.0 kcal mol^{-1}) result is also in reasonable agreement (Table 4 & Fig. 9). Note that the energies calculated in this study are electronic energies and the experimental values are free energies; however, entropic effects are relatively small,[1] making the comparison here meaningful. In all these cases, the maximum occurred at the same point (−0.2 Å) on the reaction coordinate, slightly later than that in the enzyme. The PCM(J) model at the B3LYP/6-31G(d) level gave a barrier in solution of 21.9 kcal mol^{-1}, slightly earlier (−0.4 Å) on the reaction coordinate. Davidson *et al.*[5] reported an energy barrier in solution of 41.0 kcal mol^{-1}, at the HF/6-31G(d) level using the self consistent isodensity polarized continuum model (SCIPCM). It is clear that continuum solvent models have limitations, but the use of many explicit water molecules is too computationally expensive at this level of theory. Another significant finding is that in all cases the PCM and PCM(J) models showed the gas phase optimized structure of chorismate in solution to be higher in energy than the structure at −2.0 Å on the reaction coordinate (*i.e.*, the structure after QM/MM minimization in the enzyme), whereas the IPCM model shows the gas-phase structure to be significantly *lower* in energy (Table 5). All but one of the calculations therefore show that the minimum energy structure of chorismate in the enzyme is not energetically unfavourable in solution. If this result is indeed correct, strain/compression of the substrate is expected to be similar in solution and in the enzyme, and so would not play an important part in catalysis. If, however, the IPCM result is more realistic, strain/compression could be an important contributor to catalysis in the enzyme. Different continuum solvent models give very different estimates of this relatively subtle effect, and it should be investigated with more detailed solvent models, or by the use of non-default values for the cut-offs used in the calculations (especially the IPCM model).

Comparison of the energy profile for reaction in solution with that for the same system in the gas phase shows the stabilization provided by the solvent (compare Figs. 5 and 9). The barriers to reaction are somewhat lower in solution than in the gas phase (by 5–10 kcal mol^{-1}). However, this reduction of the barrier is due largely to stability of a more 'compressed' structure in solution, rather than transition state stabilization by the solvent (this contributes approximately 5–10 kcal mol^{-1} to the reduction of the barrier). Relative to this distorted/compressed conformation (which could be described as a NAC), which is calculated to be the minimum energy form of chorismate in the enzyme, the barrier to reaction is not lowered significantly by solvent—*i.e.* it is similar to that in the gas phase. This indicates that the NAC effect is small: if the NAC effect were large, the barrier to reaction from a NAC in solution would be similar to that in the enzyme. The barrier to reaction in solution is higher because the solvent does not significantly stabilize the TS (relative to the substrate), according to the solvent models applied here. In all cases the solvent stabilizes the system until just before the transition state (a maximum of approximately 5 kcal mol^{-1} with all methods) and then the solvent actually becomes destabilizing with respect to the gas-phase profile. Also, the product is not stabilized more than the substrate by solvent (except for the MP2/6-31+G(d)/PCM

Table 6 Comparison of barriers to reaction in the enzyme and in solution (energies in kcal mol^{-1})

Method	Barrier in enzyme[a]	Barrier in solution[b]	Energy difference	Rate acceleration
RHF/6-31G(d)	35.7	53.0	17.3	4.6×10^{12}
MP2/6-31+G(d)	10.1	28.0	17.9	1.3×10^{13}
B3LYP/6-311+G(2d,p)	15.2	24.6	9.4	7.5×10^{6}
Experiment	15.4	24.5	9.1	4.5×10^{6}

[a] Barriers in enzyme corrected for entropic effects and the contribution of Glu78 (see text). [b] Solution barriers from PCM calculations (see text).

results with 1.0 kcal mol^{-1} stabilization), in contrast to the significant product stabilization in the enzyme.

From transition state theory, the energy barriers to the rearrangement of chorismate in the enzyme and in solution can be used to predict the rate acceleration in the enzyme environment. Formally, the calculated barriers in solution are free energy barriers; as intramolecular entropic, zero-point and thermal terms are likely to be small, and similar in enzyme and in solution, it is appropriate to add an entropic correction to the calculated barriers in the enzyme (for BsCM $\Delta S^{\ddagger}_{act} = -9.1$ cal mol^{-1} K^{-1} at 298 K,[1] *i.e.* $T\Delta S^{\ddagger}_{act} = 2.7$ kcal mol^{-1}) for comparison. The likely additional TS stabilization from Glu78 (estimated as 3.57 kcal mol^{-1} [46]) should also be considered. The difference in energy barriers between the solution and enzyme reactions at the RHF/6-31G(d) level (17.3 kcal mol^{-1}) corresponds to a rate enhancement of 4.6×10^{12} (Table 6). At the MP2/6-31+G(d) level the energy difference of 17.9 kcal mol^{-1} corresponds to a rate acceleration of 1.3×10^{13} over the rate of the reaction in solution at the same level. The B3LYP/6-311+G(2d,p) energy barriers predict a rate acceleration of 7.5×10^{6} over the reaction in solution. The B3LYP/6-311+G(2d,p) results are in good agreement with the experimental findings of a rate acceleration of 4.5×10^{6} over the reaction in solution, whereas the MP2/6-31+G(d) and HF/6-31G(d) results apparently overestimate this enhancement significantly. The chemical step is probably not entirely rate-limiting to the enzymic reaction, so the larger rate accelerations are not inconsistent with experiment (the chemical step could be faster than the observed experimental rate for the enzyme).

Comparison with lower-level QM/MM modelling

We compare here the present *ab initio* QM/MM results with our recent lower-level study,[14] which used the same structural models and MM parameters, thus differing only in the QM level of the calculations (and the amount of minimization along the path). (We do not attempt here to compare with all previous semiempirical QM/MM studies, as detailed comparisons have been made in our previous work,[14] and by others[13]).

The semiempirical (AM1/CHARMM22) and *ab initio* (RHF/6-31G(d)/CHARMM22) methods predict similar structures of chorismate bound to the enzyme, and show that the bound structure is significantly different to the gas phase structure. Both methods therefore show the structure of chorismate bound to the enzyme to be distorted/compressed, being further along the reaction coordinate than the gas-phase structure, and therefore more like the transition state. However, the methods differ in the degree of distortion predicted for enzyme-bound chorismate. The distance between the reacting atoms (C3 and C9) is predicted to be considerably shorter by the lower level calculations. For a comparable model, AM1/CHARMM22 calculations give a C3–C9 distance of 3.25 Å,[14] notably shorter than the 3.43 Å predicted here at the *ab initio* RHF/6-31G(d)/CHARMM22 level. Other AM1/MM calculations predict relatively short C3–C9 distances also.[2,14,24] In considering the possible catalytic contribution of substrate compression,[24,48] it will be important to take into account that AM1/MM methods may significantly overestimate the distortion of chorismate by the enzyme.

One significant difference in structure at the two levels of theory is the orientation of the hydroxyl group of chorismate. At the AM1/CHARMM22 level, the hydroxyl group rotated (to point away from the ring), forming a hydrogen bond with Glu78.[14] At the RHF/6-31G(d)/CHARMM22 level the hydroxyl group remains pointing towards the ring until a value of 1.2 Å on the reaction coordinate. This difference is probably due to the limited amount of minimization possible at the RHF/6-31G(d)/CHARMM22 QM/MM level (and is reflected in the calculated stabilization along the path, as discussed above).

Pathway calculations were carried out at the AM1/CHARMM22 level of theory with the same set-up and restraints as the RHF/6-31G(d)/CHARMM22 model. The total energy barrier at the AM1/CHARMM22 level was 28.4 kcal mol^{-1} and the QM/MM energy barrier was 27.0 kcal mol^{-1}. For comparison, the AM1/CHARMM22 energy profile was also corrected (using single point energies from the *ab initio* QM/MM path (Table 4), as described above). When corrected with the MP2/6-31+G(d) energies, the total energy barrier (*i.e.* AM1/CHARMM + MP2/6-31+G(d)) to reaction was reduced to 12.3 kcal mol^{-1}. The AM1/CHARMM22+B3LYP/6-311+G(2d,p) energy barrier is 15.2 kcal mol^{-1}. These results are in good agreement with the corrected *ab initio* QM/MM barriers found here.

The amount of TS stabilization (relative to the substrate) predicted by the RHF/6-31G(d)/CHARMM22 method is greater than that found at the AM1/CHARMM22 level of theory (19.7 kcal mol^{-1} RHF/6-31G(d)/CHARMM22, and 14.4 kcal mol^{-1} AM1/CHARMM22). The TS stabilization at the RHF/6-31G(d) level may be underestimated as Glu78 (a major contributor to TS stabilization in AM1/CHARMM22 models) does not form a hydrogen bond with the hydroxyl group of chorismate until after the TS is reached.

The hydrogen bond distances found in the RHF/6-31G(d)/CHARMM22 model are shorter than the equivalent distances in AM1/CHARMM22 profiles. At the transition state, the hydrogen to acceptor distance for the interaction between Arg90 and O3 is 1.78 Å in the *ab initio* model compared to 2.11 Å in a model treated at the AM1 level of theory. Similarly, the acceptor to donor distance for the interaction between XSOL124 and O3 is 1.91 Å in an AM1 model and 1.70 Å in the *ab initio* model. It has been found previously that the *ab initio* (RHF/6-31G(d)/CHARMM) QM/MM method overestimates the stabilization of a (QM) transition state by a (MM) hydrogen bond acceptor.[33] On the other hand, hydrogen bond energies may be underestimated at the AM1 QM/MM level, as AM1 itself is known to do.[29] However, in both the *ab initio* and AM1 models the changes in hydrogen bond lengths between the starting structure and the transition state are similar. For Arg90–O3 this change is 0.18 Å with both methods and for XSOL124–O3 the change is 0.06 Å (AM1 model) and 0.1 Å for the *ab initio* model.

Discussion

To analyze fully an enzymic reaction, or indeed any chemical reaction, it is essential to study the properties of the transition state, and ideally other unstable species along the reaction pathway, as well as the reactants and products. *Ab initio*

quantum chemical methods can provide good descriptions of TSs, for example, it has been shown that the levels of *ab initio* theory used here treat the TS for the Claisen rearrangement of chorismate to prephenate well.[14] This makes *ab initio* QM/MM modelling an excellent choice for the investigation of the mechanism of, and catalysis in, chorismate mutase. This enzyme is a key test of fundamental theories of enzyme catalysis. The first QM/MM study of chorismate mutase[2] was restricted to a small model without full solvation, and the semiempirical AM1 QM method with the CHARMM19 (united-atom) MM parameters. Advances in technology since then have allowed us to carry out calculations on larger, fully solvated models at the AM1/CHARMM22 QM/MM level,[14] and now to perform *ab initio* QM/MM calculations on the same model. Semiempirical methods, which have been used for most previous studies, have some well-known limitations. We have therefore used *ab initio* QM/MM calculations to test conclusions of previous work and to examine the nature of catalysis in this fundamentally important enzyme.

Adiabatic mapping along an approximate reaction coordinate was used at the *ab initio* RHF/6-31G(d)/CHARMM22 QM/MM level to obtain a potential energy profile for the conversion of chorismate to prephenate in BsCM. This method has been applied successfully at the AM1[2,48,49] and *ab initio* levels,[12,32,33] showing good correlation with experiment. This method has limitations—no configurational averaging is taken into account; however, previous lower-level modelling has shown that the adiabatic mapping approach is suitable for chorismate mutase.[14] RHF/6-31G(d)/CHARMM22 QM/MM pathway calculations here give a barrier to reaction in the enzyme of 36.6 kcal mol^{-1}. Corrections along the path were calculated at the MP2/6-31+G(d) and B3LYP/6-311+G(2d,p) *ab initio* levels. As expected, the higher *ab initio* methods predicted energy barriers significantly lower than that obtained at the RHF/6-31G(d) level (Table 3). The resulting QM/MM total energy barriers of 11.0 kcal mol^{-1} (with MP2/6-31+G(d) corrections) or 16.1 kcal mol^{-1} (B3LYP/6-211+G(2d,p) corrections) are comparable to the experimental finding of $\Delta H^{\ddagger}$ = 12.7 ± 0.4 kcal mol^{-1}.[1] As the RHF/6-31G(d)/CHARMM22 method may overestimate the QM/MM interactions, the AM1/CHARMM22 QM/MM energy profile[14] was also corrected with QM energies calculated at the higher levels of theory, giving barriers of 12.3 kcal mol^{-1} (MP2/6-31+G(d)/AM1/CHARMM22) and 15.2 kcal mol^{-1} (B3LYP/6-311+G(2d,p)/AM1/CHARMM22). These values are the upper estimates of the barrier to the reaction: the relatively tight restraints on the protein boundary used in generating the RHF/6-31G(d) path were shown to increase the barrier by approximately 3 kcal mol^{-1} at the AM1 level. QM/MM optimizations at higher levels may also reduce the barrier to reaction. The barrier could be further reduced (*e.g.* by 3.6 kcal mol^{-1} [46]), by a reorientation of Glu78 (see below).

Analysis of the RHF/6-31G(d) QM/MM path shows significant stabilization of the TS (by 19.7 kcal mol^{-1} relative to the substrate). This may be an underestimate of the stabilization as Glu78 (found to be a major contributor to TS stabilization at the AM1/CHARMM22 level[14]) does not form a hydrogen bond with the substrate until after the TS has been reached. The presence of this hydrogen bond after a value of 1.2 Å on the reaction coordinate may explain the significant increased stabilization of the product (Fig. 6). Glu78 was also found to stabilize the product at the AM1/CHARMM QM/MM level.[14] It is clear that the enzyme stabilizes the product, and the TS, relative to the substrate. Analysis of the structures along the path shows hydrogen bonds with Arg7, Arg63, Arg90, XSOL116, XSOL124, and SOLV7102. Several of these shorten significantly at the TS (*e.g.* Arg90(HH21)–O3, Arg7(HH22)–O5, Arg7(HH12)–O4, and XSOL124(H1)–O3). This indicates a strengthening of these hydrogen bonds at the TS. TS stabilization by these residues has previously been shown at the AM1/CHARMM22 level.[14] Detailed analysis of the physical nature of transition state stabilization in these structures of BsCM by nonempirical quantum chemical techniques[50,51] shows that electrostatic interactions are dominant in this active site,[46] and the stabilization predicted here reproduces with excellent agreement that found by the more detailed calculations. The nonempirical analysis shows significant TS stabilization originating mainly from the charged residues: Arg90, Arg7, Glu78, Arg116, Arg63, and water molecule XSOL124. Typically for hydrogen-bonded systems, the dominant interaction component is electrostatics (contributions calculated at the first order Heitler–London level agree very well with MP2 results, with a correlation coefficient of 0.92), whereas other terms (correlation, delocalisation, and exchange) cancel each other. A qualitatively correct description is already obtained at the electrostatic multipole level (with a correlation coefficient of 0.85). The static catalytic field derived from the structures of the substrate and TS complexes agrees with the positioning of the charged residues (arginines and glutamate) found in the QM/MM model, proving the reliability of the model. Moreover, the agreement with the catalytic field shows that the enzyme has evolved to fit the electrostatic pattern formed by the reacting system.[46] The interaction found here between Arg90(HH21) and O3 shows the most significant shortening at the TS, in agreement with other QM/MM modelling[2,12,14] and experiment[17,19,20] that Arg90 is the most important residue in TS stabilization. This is consistent with mutagenesis experiments that show no measurable activity when Arg90 is mutated to alanine[17,19] or a 10^4-fold reduction in k_{cat}/K_m for an Arg90Lys mutant.[17] Mutation of Arg7 to alanine results in a 10^6-fold reduction in k_{cat}/K_m.[17]

The rearrangement of chorismate to prephenate occurs by the same unimolecular pericyclic mechanism in the enzyme and in solution, making direct comparison of the important features of the catalyzed and uncatalyzed reaction possible. Comparison of the same reaction in the enzyme and in solution is important when considering the source of the catalytic efficiency of the enzyme.[23,26] We have therefore compared the same reaction pathway in the two environments, at the same levels of *ab initio* theory, using continuum models to represent the effects of solvent. The aim is not to model the reaction as it actually occurs in solution (although the reaction is likely to be similar), but to compare the stabilization provided to the same structures by water and by the enzyme. At the MP2/6-31+G(d)/PCM level the barrier to reaction in solution was 28.0 kcal mol^{-1}, whereas the B3LYP/6-311+G(2d,p)/PCM method gave a lower barrier of 24.6 kcal mol^{-1}. The solvent is unable to provide TS stabilization with as much specificity as the enzyme, hence the higher barrier in solution.

The difference in energy between the barrier in solution and in the enzyme environment corresponds to a rate acceleration over the reaction in solution of 1.3×10^{13} at the MP2/6-31+G(d) level and 7.5×10^6 at the B3LYP/6-311+G(2d,p) level. The reaction may also be partly diffusion-controlled,[45] and as a result both these values are consistent with the experimental value of a 10^6-fold rate acceleration. Solvent continuum models have their limitations, as shown here by some significant differences in results from different polarized continuum models. The use of explicit solvent molecules, combined with configurational sampling, may give a better estimate of the barrier to reaction in solution, although it is not yet feasible to carry out the necessary large-scale simulations at the *ab initio* QM/MM level.

Substrate strain or conformational compression has also been proposed as a factor in catalysis in chorismate mutase.[2] Initial QM/MM minimization of the model system caused the C–O3 breaking-bond to lengthen slightly, and the C3–C9 forming-bond to shorten, resulting in a more TS-like structure. These changes in geometry resulted in a 9.7 kcal mol^{-1} increase in the energy of the substrate, potentially catalytically signifi-

Table 7 Comparison of structures with the earlier work of Hall *et al.*[6] (all distances in Å)

	Hall *et al.*[6]			RHF/6-31G(d)/CHARMM22 (this work)		
	Substrate	TS	Product	Substrate	TS	Product
C–O3	1.44	1.98	3.07	1.45	2.28	3.38
C3–C9	3.63	2.63	1.74	3.44	2.58	1.78
Y108(HH)–O4	1.68	1.68	1.68	4.47	4.62	4.64
R7(HH12)–O4	2.62	1.73	1.76	1.69	1.66	1.69
R7(HH22)–O5	1.74	1.74	1.70	1.69	1.64	1.66
R90(HH22)–O5	1.80	1.80	1.80	3.00	3.03	3.05
R90(HH21)–O3	1.93	1.83	1.74	3.49	3.23	1.83
E78(OD2)–H4	1.78	1.68	1.70	4.75	4.15	1.70
C75(HG)–O	2.01	1.92	1.97	2.08	2.01	1.87
K60(NH)–O2	1.90	3.27	3.50	2.81	3.14	2.95
A59(NH)–O1	1.79	6.14	6.15	5.05	5.51	4.29
R116(HE)–O1	10.50	1.70	1.69	5.09	5.63	4.53

cant. However, the calculations in solution show that the minimum energy structure of chorismate in the enzyme is not higher in energy in solution, suggesting that strain may play only a minor role in catalysis.

The relationship between the compression of the bond-forming distance and barrier height in BsCM has been explored previously by Menger and co-workers.[27] They carried out calculations (B3LYP/6-31G(d)) on a series of Claisen rearrangements in the gas phase. The results predict a linear relationship between the calculated C–C bond-forming distance and activation barrier. Using the compression found in the first AM1/CHARMM QM/MM study of BsCM[2] (0.45 Å, from 3.298 Å in the gas-phase to 2.849 Å in the enzyme), they predicted that the compression may contribute as much as 10 kcal mol^{-1} to lowering the barrier to reaction. Using our *ab initio* results, a compression of 0.3 Å (from 3.73 Å in the gas-phase to 3.43 Å in the enzyme) would correspond to a 5.6 kcal mol^{-1} reduction in barrier height. The compression found here at the *ab initio* QM/MM level is less than that found previously by lower-level modelling. The current RHF/6-31G(d)/CHARMM22 result is likely to be a better description of the degree of substrate distortion in chorismate mutase. Estimates of the catalytic contribution of substrate compression based on lower-level QM/MM modelling may need to take this into account.

Shurki *et al.*[52] recently proposed a method for calculating the effect of compression or 'NAC' effect using an MD/Free Energy Perturbation (FEP) approach. Preliminary results from FEP simulations at the AM1/CHARMM22 level indicate a contribution of 2–5 kcal mol^{-1}.[53] At the AM1 QM/MM (MC/FEP) level, Jorgensen and co-workers[54] have found that compression of the C3–C9 distance from 3.53 Å in water to 2.53 Å in the enzyme accounts for ~70% of the observed $\Delta\Delta G^{\ddagger}$ (7.9 kcal mol^{-1}), suggesting that preferential stabilization of the TS plays only a secondary role in catalysis in chorismate mutase. Again, this study is at the AM1 QM/MM level of theory and as a result the compression may be overestimated. Warshel and co-workers[55] have very recently calculated the upper limit of the apparent NAC effect in BsCM to be 5 kcal mol^{-1}, in good agreement with our findings, using a quite different modelling technique (the empirical valence bond (EVB) method). They have also shown that the apparent NAC effect is predominantly electrostatic in nature and conclude that the NAC effect is a result of TS stabilization, and not the major factor in catalysis.

Comparison with previous QM/MM modelling

As chorismate mutase has been the focus of much research in the past two decades, it is important to compare the results obtained from different approaches to modelling the reaction. Here we compare our *ab initio* QM/MM study with some previous studies. We have extensively reviewed previous studies in our earlier work.[14] Hall *et al.*[6] used a combination of the Gaussian 94[56] QM package with AMBER[57] for QM/MM calculations, in which the QM and MM regions are optimized 'asynchronously and self-consistently'. The QM region was held fixed whilst purely MM (not QM/MM) optimizations were carried out and then QM atoms were optimized in a frozen field of static MM atoms. Varying MM/CHELPG atomic charges of the QM atoms at each point along the reaction path were used to optimize the MM region. The structures of chorismate, the TS and prephenate were optimized and characterized as stationary structures at the B3LYP/6-31G(d) level. The model system was taken from the same crystal structure of BsCM as used here (2CHT[35]), and comprised 24 QM atoms and 4117 MM atoms. The initial C–O3 bond length in chorismate agrees with the results for our model but the C3–C9 (bond-forming) distance is significantly shorter in our model. Our RHF/6-31G(d) model differs from the results reported by Hall *et al.* in the orientation of some protein residues, possibly due to choice of active site and QM/MM method used in optimization of structures, and hence some very different interatomic distances were obtained. The hydroxyl group of Tyr108 is in a different orientation making the distance from O5 greater than 4 Å in our model, compared to the 1.68 Å reported by Hall *et al.* In our model the water molecule XSOL116 bridges between Tyr108 and O5 of chorismate. Arg90 is also oriented differently: in our model, NE of Arg90 acts as a hydrogen bond donor to both O3 and O5, not the NH2 group shown by Hall *et al.* This NH2 group can interact with O3 but hydrogen bond formation with O5 is not possible. Ala59 and Arg116 also do not show the same behaviour as reported by Hall *et al.* The positions of these residues do not change by as much here as reported by Hall *et al.* (Table 7). The position of the barrier reported by Hall *et al.* corresponds to a value of −0.65 Å on the reaction coordinate used here, similar to the location of the maximum on the corrected pathways. A barrier of 1.4 kcal mol^{-1} was reported for the enzymic reaction. This is probably a QM/MM barrier (*i.e.* no MM energy included) rather than a total energy barrier. It is lower than that found here and in most other studies.[11,12,58] A similarly low barrier may be expected by the inclusion of electron correlation to correct the 16.6 kcal mol^{-1} to 19.7 kcal mol^{-1} (RHF/4-31G(d)/CHARMM22) found by Lee *et al.*[12] (see comparison below). However, Woodcock *et al.*[58] report an activation enthalpy of 33.4 kcal mol^{-1} at the same RHF/4-31G(d) level. It has been stated that kinetic data show that the rearrangement itself is not the rate-limiting step, making product release the limiting factor.[16,59] More recent studies suggest that the chemical step is rate-limiting[60,61] or that the reaction rate is partially diffusion-controlled.[45] The barriers found in the present work are in good agreement with the 11.3 kcal mol^{-1} (MP2 corrected SBK/4-31G) barrier reported by Worthington *et al.*[11] (see below), and with our corrected semi-empirical barriers.[14] Chorismate mutase has also been a test-case for the new replica path method,[58] where an activation enthalpy of 14.9 kcal mol^{-1} was predicted from a B3LYP/6-31G(d) energy analysis of an RHF/4-31G optimized replica

Table 8 A comparison of interatomic distances with those reported by Worthington *et al.*[a],[11] (all distances in Å)

	Worthington *et al.*		RHF/6-31G(d)/CHARMM	
	Chorismate	TS	Chorismate	TS
C–O3	1.49	2.50	1.45	2.28
C3–C9	3.59	2.63	3.44	2.58
C1–O	1.44	1.44	1.41	1.41
O–H4	0.98	0.97	0.95	0.95
R90(HH21)–O3	2.04	2.07	1.97	1.77
E78(OD2)–H4	2.18	2.29	4.75	4.15
Y108(HH)–O5	1.93	2.07	6.31	6.43

[a] Worthington *et al.* carried out optimizations at the HF/4-31G SBK level using effective fragment potentials for the enzyme residues.

path. After this work was completed, a DFT/MM approach was applied to chorismate mutase also indicating significant TS stabilization by the enzyme.[62] Martí *et al.*[10] report a somewhat higher AM1/CHARMM free energy barrier, corrected with B3LYP/6-31G(d) gas phase energies, of 20.6 kcal mol^{-1}.

Worthington *et al.*[11] studied the enzyme reaction using a different (MD/QM) approach. The AMBER[57] program was used to carry out dynamics simulations from which active site 'snapshots' were taken for QM optimization. QM calculations were carried out with GAMESS-US,[29] using effective fragment potentials to model the interaction of the active site with the substrate. Optimized structures of the enzyme–substrate complexes were obtained at the Hartree–Fock level using the 4-31G SBK basis set. MP2 corrections to the energies were also obtained. In their study, the complete chorismate mutase trimer was used, making it a much larger model system than that studied here. The model of the enzyme studied by Worthington *et al.* was built from a mixed template of 2CHT[35] and 1DBF[63] crystal structures. Residues Arg90, Glu78 and Tyr108 were identified as important in catalysis. Table 8 shows a comparison of our results with those reported by Worthington *et al.* As found by Hall *et al.*[6] (see above), Tyr108 was found to be in a different orientation by Worthington *et al.* to that in our model, probably due to the presence of XSOL116 in our model (see above). The hydroxyl group to O5 distance is greater than 5 Å in our model. The interaction quoted here is with O5 whereas Hall *et al.* gave distances for the interaction of Tyr108 with O4. This distance is smaller, but still not within hydrogen bonding range. Arg90 is also closer to O3 in our model, and shows a decrease in H-bond distance at the transition state, whereas Worthington *et al.* report a slight increase in the Arg90–O3 distance at the transition state. Glu78 is also further away in our model.

The analysis of the dynamics simulation carried out by Worthington *et al.* may provide insight into why Arg116 was not identified as a hydrogen bond donor in our models. The study by Worthington *et al.* shows that Arg116 is a special case. Two of the three active sites show a hydrogen bond between Arg116 and O4, whereas this hydrogen bond is absent in the third active site. They also report that there is some variation in hydrogen bonding with Arg63 in the different active sites. It is likely that our model for this study and our previous study[14] are based on the active site that shows no interaction of the TS analogue with Arg116.

The C–O3 and C3–C9 bond lengths from our RHF/6-31G(d) model for chorismate are in quite good agreement with the values reported by Worthington *et al.* at a different level of theory, but the C3–C9 distance is 0.13 Å shorter in our model. The values given for their transition state correspond to −0.13 Å on the reaction coordinate, later than observed here. A barrier to reaction corrected with MP2 of 11.3 kcal mol^{-1} was reported,[11] which is in good agreement with our MP2/6-31+G(d)//6-31G(d)/CHARMM22 finding (Table 2).

Lee *et al.*[12] have recently carried out an *ab initio* QM/MM study of chorismate mutase also using the CHARMM[28] interface with GAMESS-US.[29,30] A solvated trimer of chorismate mutase from the 2CHT[35] crystal structure was modelled with prephenate in the active site, treating prephenate QM in the A/C active site only. Reaction pathway calculations were carried out at the RHF/4-31G/CHARMM22 level using a similar method to that described here, with 200 cycles of energy minimization (ABNR) carried out at each step on the path. Initially, only prephenate was treated QM for the pathway calculations, but for refinement of the reactant, TS, and product, Glu78 and Arg90 were included in the QM region. An energy barrier of 19.7 kcal mol^{-1} was reported, reduced to 16.6 kcal mol^{-1} when the side chains of Glu78 and Arg90 were included in the QM region. These are significantly lower than the 36.6 kcal mol^{-1} found here at the comparable RHF/6-31G(d)/CHARMM22 level. The study by Lee *et al.* was carried out using a smaller basis set (4-31G), and as a result more cycles of minimization were possible at each point along the path. Table 9 shows a comparison of the protein–substrate interactions found by Lee *et al.* at the RHF/4-31G/CHARMM22 level with those obtained here (RHF/6-31G(d)/CHARMM22). On the whole there is good agreement between the interactions observed at the RHF/4-31G/CHARMM22 and the RHF/6-31G(d) levels of theory. The interaction between Glu78 and the hydroxyl group of the substrate is significantly longer in our RHF/6-31G(d) structures due to the fact that the hydroxyl group does not rotate to form this interaction until 1.2 Å on the reaction coordinate. However, once formed this interaction is shorter than that observed by Lee *et al.* (A–D 2.67 Å RHF/6-31G(d) compared to 2.87 Å RHF/4-31G, A–H 1.70 Å RHF/6-31G(d) compared to 1.91 Å RHF/4-31G). This difference could be due to the fact that Glu78 was treated QM by Lee *et al.* Arg90 is closer to the substrate, probably also due to the fact that Arg90 was also treated QM in some minimizations. The distance between Arg7 and the substrate is shorter in our RHF/6-31G(d)/CHARMM22 structures, and shows more variation along our pathway. The A–D distance decreases by 0.03 Å and the A–H distance decreases by 0.04 Å at the TS, but in the study by Lee *et al.* the distances decrease by 0.01 Å and 0.02 Å respectively. As observed in our previous AM1 study[14] (and in contrast to Hall *et al.*,[6] Worthington *et al.*,[11] and Lee *et al.*[12]) Tyr108 is too far away to form any hydrogen bonds with the substrate in our model.

Lee *et al.* also carried out an electrostatic energy decomposition analysis for the active site residues. This was done by comparing the energy of the system when a particular residue has its full MM charge and the energy of the system when the charges of the residue are set to zero. Arg90, Glu78, and Arg7 were identified as the major contributors to electrostatic stabilization of the TS, with Arg90 as the major contributor. The critical role of Arg90 has recently been demonstrated experimentally.[20] The conclusion of this analysis was that all direct ionic interactions with the substrate are catalytically important in addition to their role in binding. The energetic analysis supports the results of our hydrogen bond analysis that shows a decrease in acceptor to hydrogen distance at the TS for Arg90 and Arg7, with the decrease being most significant for Arg90. The work by Lee *et al.* is also in agreement with the findings of our previous AM1/CHARMM studies[2,14] that TS stabilization is a significant factor in catalysis in BsCM.

The interactions of chorismate with the protein in our model are more similar to those found by Lee *et al.*[12] and Hall *et al.*[6] than those reported by Worthington *et al.*, probably due to the fact that Worthington *et al.*[11] used more than one crystal structure of chorismate mutase in creating their model (Hall *et al.* and Lee *et al.* used the same structure as here (2CHT[35])). It should be remembered also that there are differences in the structure of the three different active sites of the trimer in this 2CHT structure. The position of Tyr108 is a major difference

Table 9 A comparison of interatomic distances with those reported by Lee *et al.*[a],[12] (Upper entry is the acceptor–donor distance; lower entry is the acceptor–hydrogen distance; all distances in Å)

	Chorismate		TS		Product	
	Lee *et al.*	RHF/6-31G(d)/CHARMM	Lee *et al.*	RHF/6-31G(d)/CHARMM	Lee *et al.*	RHF/6-31G(d)/CHARMM
E78OE2–O	2.85	4.25	2.82	3.82	2.87	2.67
	1.89	4.76	1.85	4.15	1.91	1.70
C75NH–O	2.88	2.88	2.83	2.83	2.78	2.85
	1.88	2.07	1.84	2.00	1.80	1.87
C75SH–O	3.45	3.61	3.55	2.68	3.64	3.82
	2.46	2.86	2.62	3.09	2.74	2.78
R7HH12–O4	2.74	2.71	2.73	2.68	2.75	2.70
	1.77	1.70	1.75	1.66	1.77	1.69
R7HH21–O5	2.71	2.71	2.70	2.67	2.72	2.68
	1.73	3.36	1.71	3.31	1.73	3.32
		(1.70 HH22)		(1.64 HH22)		(1.66 HH22)
R90HE–O3	2.83	3.26	2.81	3.28	3.03	3.50
	2.06	2.68	2.01	2.65	2.28	2.82
R90HE–O5	2.82	2.64	2.86	2.68	2.84	2.69
	1.95	1.65	1.99	1.69	1.92	1.69
R90HH22–O3	2.83	2.86	2.60	2.74	2.65	2.78
	1.69	3.48	1.64	3.33	1.68	3.33
		(1.96 HH21)		(1.77 HH21)		(1.83 HH21)
R63HH12–O2	2.90	3.22	2.72	3.04	2.71	3.13
	1.76	3.03	1.76	2.85	1.77	2.93
		(2.65 HH11)		(2.49 HH11)		(2.57 HH11)
Y108OH–O4	2.79	3.76	2.80	3.96	2.84	3.99
	1.84	4.47	1.85	4.62	1.89	4.66

[a] Lee *et al.* carried out *ab initio* calculations at the RHF/4-31G/CHARMM22 level (see text).

between our model and the models used in these other QM/MM studies. Relatively small differences in structure can have significant effects on calculated barriers.[14] This is emphasized here also by the effects of Glu78. Selection of an initial structure, and preparation of the model are important considerations. As discussed previously,[14] the TSA complex (as used here) appears to provide a better representation of the structure of the protein at the transition state than the product (prephenate) complex.

Conclusions

Chorismate mutase is an important enzyme in the testing and development of theories of enzyme catalysis. We have modelled the reaction in the enzyme by *ab initio* combined quantum mechanics/molecular mechanics (QM/MM) methods, at reliable levels of QM theory (B3LYP/6-311+G(2d,p)//6-31G(d)/CHARMM22 and MP2/6-31+G(d)//6-31G(d)/CHARMM22). We have studied a large, fully solvated model, with well-tested and appropriate reaction modelling methods, with consistent, combined QM/MM optimization of the whole system. The best estimates of the energy barrier to reaction in the enzyme are 12.7–16.1 kcal mol^{-1} (B3LYP/6-311+G(2d,p)//6-31G(d)/CHARMM22) and 7.4–11.0 kcal mol^{-1} (MP2/6-31+G(d)//6-31G(d)/CHARMM22). These values are comparable to the experimental values of $\Delta H^{\ddagger}$ = 12.7 ± 0.4 kcal mol^{-1}.[1] The reaction has also been modelled in solution (using a variety of continuum solvation models). The best estimates of the barrier to reaction in solution are 24.6 kcal mol^{-1} (B3LYP/6-311+G(2d,p)//RHF/6-31G(d)) and 28.0 kcal mol^{-1} (MP2/6-31+G(d)//RHF/6-31G(d)), similar to the experimental result of $\Delta G^{\ddagger}$ = 24.5 kcal mol^{-1}.[1] Thus the barrier to reaction is lower in the enzyme by 9.4 kcal mol^{-1} and 17.9 kcal mol^{-1} at the same QM theoretical levels (B3LYP/6-311+G(2d,p), MP2/6-31+G(d)). This catalytic effect of the enzyme is achieved primarily by transition state stabilization. These results correspond to a 7.5×10^6 or 1.3×10^{13} rate acceleration over the reaction in solution, respectively, comparable to the experimental estimate of 4.5×10^6. The transition state is stabilized significantly relative to the bound substrate: in accordance with Pauling's hypothesis,[64,65] stabilization of the reacting system by the enzyme reaches a peak at the transition state. The major contributions to transition state stabilization (shown by hydrogen bonding) come from Arg90, Arg7, Arg63, and a crystal water molecule (XSOL124). Glu78 is found to stabilize the product and can also stabilize the transition state, relative to the reactant. The interaction with Glu78 is destabilizing, but destabilizes the product (and potentially the TS) less than the reactant. The results also point to potential variation of the interaction of the reacting system with Glu78. These findings are in agreement with lower-level modelling[2] and recent experimental investigations in demonstrating TS stabilization by *e.g.* Arg90.[20] The product, prephenate, is also calculated to be significantly stabilized (relative to the substrate) by the enzyme. The stabilization of the TS (and product) is primarily due to electrostatic interactions. The structure of chorismate bound to the enzyme is found to be significantly altered from its minimum energy geometry in the gas-phase (and in solution)—in the enzyme, it is further along the reaction coordinate, more similar to the transition state. The amount of distortion/compression is, however, less than predicted by lower-level modelling. This distortion and destabilization of the substrate may also contribute a small amount to the reduction of the activation energy compared to the reaction in solution as suggested previously.[2] The structural change of the substrate can be described as 'strain', in the sense that strain means a physical distortion of an object.[65] However, the energy difference between these conformations in solution does not alone account for the catalytic power of the enzyme, appearing to be relatively small. Selection of the appropriate, reactive, conformation by binding to the enzyme is also likely to make a small contribution to the overall reduction in activation free energy compared to solution.[2,8,13,15] The results presented here, from QM/MM calculations at reliable levels of QM theory, demonstrate unequivocally significant TS stabilization by chorismate mutase. In contrast to some recent suggestions, they demonstrate that TS stabilization is central to catalysis by this important enzyme.

Acknowledgements

We thank Dr J. Harvey for his assistance and helpful discussion throughout the project. Dr L. Ridder thanks the European Union (Quality of Life) Framework Programme 5 for a Marie Curie Fellowship. B. Szefczyk would like to thank the EU SOCRATES program for funding his study visit to the University of Bristol. We also thank the EPSRC, BBSRC, IBM Life Sciences Outreach Programme, and Wroclaw University of Technology for support.

References

1 P. Kast, M. Asif-Ullah and D. Hilvert, *Tetrahedron Lett.*, 1996, **37**, 2691–2694.
2 P. D. Lyne, A. J. Mulholland and W. G. Richards, *J. Am. Chem. Soc.*, 1995, **117**, 11345–11350.
3 O. Wiest and K. N. Houk, *J. Am. Chem. Soc.*, 1995, **117**, 11628.
4 H. A. Carlson and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1996, **118**, 8475.
5 M. Davidson, J. Guest, J. Craw and I. H. Hillier, *J. Chem. Soc., Perkin Trans. 2*, 1997, 1395–1400.
6 R. J. Hall, S. A. Hindle, N. A. Burton and I. H. Hillier, *J. Comput. Chem.*, 2000, **21**, 1433–1441.
7 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *J. Phys. Chem. B*, 2000, **104**, 11308–11315.
8 H. Guo, Q. Cui, W. N. Lipscomb and M. Karplus, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 9032.
9 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Theor. Chem. Acc.*, 2001, **105**, 207–212.
10 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón, J. Bertrán and M. Field, *J. Am. Chem. Soc.*, 2001, **123**, 1709–1712.
11 S. E. Worthington, A. E. Roitberg and M. Krauss, *J. Phys. Chem. B*, 2001, **105**, 7087–7095.
12 Y. S. Lee, S. E. Worthington, M. Krauss and B. R. Brooks, *J. Phys. Chem. B*, 2002, **106**, 12059–12065.
13 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Chem. Eur. J.*, 2003, **9**, 984–991.
14 K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Mol. Phys.*, 2003, **101**, 2695–2714.
15 S. D. Copley and J. R. Knowles, *J. Am. Chem. Soc.*, 1987, **109**, 5008–5013.
16 W. J. Guildford, S. D. Copley and J. R. Knowles, *J. Am. Chem. Soc.*, 1987, **109**, 5013–5019.
17 S. T. Cload, D. R. Liu, R. M. Pastor and P. G. Schultz, *J. Am. Chem. Soc.*, 1996, **118**, 1787–1788.
18 P. Kast, Y. Tewari, O. Wiest, D. Hilvert, K. N. Houk and R. N. Goldberg, *J. Phys. Chem. B*, 1997, **101**, 10976–10982.
19 P. Kast, C. Grisostomi, I. A. Chen, S. Li, U. Krengel, Y. Xue and D. Hilvert, *J. Biol. Chem.*, 2000, **275**, 36832–36838.
20 A. Kienhöfer, P. Kast and D. Hilvert, *J. Am. Chem. Soc.*, 2003, **125**, 3206–3207.
21 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 1472–1473.
22 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 5964–5972.
23 A. J. Mulholland, G. H. Grant and W. G. Richards, *Protein Eng.*, 1993, **6**, 133–147.
24 H. Guo, Q. Cui, W. N. Lipscomb and M. Karplus, *Angew. Chem., Int. Ed.*, 2003, **42**, 1508–1511.
25 M. P. Repasky, C. Ruch Werneck Guimaraes, J. Chandrasekhar, J. Tirado-Rives and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2003, **125**, 6663–6672.
26 A. Warshel, 1991, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, Wiley, New York.
27 N. A. Khanjin, J. P. Snyder and F. M. Menger, *J. Am. Chem. Soc.*, 1999, **121**, 11831–11846.
28 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.
29 P. D. Lyne, M. Hodoscek and M. Karplus, *J. Phys. Chem. A*, 1999, **103**, 3462–3471.

30 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347.
31 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.
32 A. J. Mulholland, P. D. Lyne and M. J. Karplus, *J. Am. Chem. Soc.*, 2000, **122**, 534.
33 L. Ridder, J. Harvey, I. Rietjens, J. Vervoort and A. J. Mulholland, *J. Phys. Chem. B*, 2003, **107**, 2118–2126.
34 P. A. Bartlett and C. R. Johnson, *J. Am. Chem. Soc.*, 1985, **107**, 7792.
35 Y. Chook, H. Ke and W. Lipscomb, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 8600.
36 C. L. Brooks, III, M. Karplus, B. M. Pettitt, 1988, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Wiley, New York.
37 C. L. Brooks, III and M. Karplus, *J. Chem. Phys.*, 1983, **79**, 6312.
38 A. T. Hadfield and A. J. Mulholland, *Int. J. Quantum Chem. (Biophys. Q)*, 1999, **73**, 137–146.
39 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle and J. A. Pople, GAUSSIAN 98, Gaussian, Inc., Pittsburgh, PA, 2001.
40 Jaguar 4.1, 2000, Schrodinger, Inc.: Portland, Oregon.
41 S. Miertus and J. Tomasi, *Chem. Phys.*, 1982, **65**, 239.
42 J. Foresman, T. Keith, K. Wiberg, J. Snoonian and M. Frisch, *J. Phys. Chem.*, 1996, **100**, 16098.
43 B. Marten, K. Kim, C. Cortis, R. A. Fresner, R. Murphy, M. N. Ringnalda, D. Sitkoff and B. Honig, *J. Phys. Chem.*, 1996, **100**, 11775.
44 D. J. Tannor, B. Marten, R. Murphy, R. Friesner, D. Sitkoff, A. Nicholls, M. Ringnalda, W. A. I. Goddard and B. Honig, *J. Am. Chem. Soc.*, 1994, **116**, 11875.
45 P. Mattei, P. Kast and D. Hilvert, *Eur. J. Biochem.*, 1999, **261**, 25–32.
46 B. Szefczyk, W. A. Sokalski, K. E. Ranaghan and A. J. Mulholland, in preparation.
47 B. Szefczyk, A. J. Mulholland, K. E. Ranaghan and W. A. Sokalski, in preparation.
48 L. Ridder, A. J. Mulholland, I. M. C. M. Rietjens and J. Vervoort, *J. Mol. Graphics Mod.*, 1999, **17**, 163–175.
49 L. Ridder, A. J. Mulholland, I. M. C. M. Rietjens and J. Vervoort, *J. Am. Chem. Soc.*, 2000, **122**, 8728.
50 W. A. Sokalski, *J. Mol. Catal.*, 1985, **30**, 395–410.
51 W. A. Sokalski, S. Roszak and K. Pecul, *Chem. Phys. Lett.*, 1988, **153**, 153–159.
52 A. Shurki, M. Štrajbl, J. Villà and A. Warshel, *J. Am. Chem. Soc.*, 2002, **124**, 4097–4107.
53 K. E. Ranaghan and A. J. Mulholland, in preparation.
54 C. Ruch Werneck Guimaraes, M. P. Repasky, J. Chandrasekhar, J. Tirado-Rives and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2003, **125**, 6892–6899.
55 M. Strajbl, A. Shurki, M. Kato and A. Warshel, *J. Am. Chem. Soc.*, 2003, **125**, 10228–10237.
56 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, GAUSSIAN 94, Gaussian, Inc., Pittsburgh, PA, 1994.
57 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Freguson, D. C. Spellmeyer, T. Fox, J. W. M. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.
58 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.
59 L. Addadi, E. K. Jaffe and J. R. Knowles, *Biochemistry*, 1983, **22**, 4494.
60 D. J. Gustin, P. Mattei, P. Kast, O. Wiest, L. Lee, W. W. Cleland and D. Hilvert, *J. Am. Chem. Soc.*, 1999, **121**, 1756–1757.
61 S. Martí, V. Moliner, I. Tuñón and I. H. Williams, *Org. Biomol. Chem.*, 2003, **1**, 483–487.
62 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.
63 J. E. Ladner, P. Reddy, A. Davis, M. Tordova, A. J. Howard and G. L. Gilliand, *Acta Crystallogr., Sect. D*, 2000, **56**, 673.
64 L. Pauling, *Am. Sci.*, 1948, **36**, 51–58.
65 A. Fersht, 1999, *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding*, Freeman, New York.