Molecular Determinants for the Rate Acceleration in the Claisen Rearrangement Reaction

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Molecular Determinants for the Rate Acceleration in the Claisen Rearrangement Reaction

Sebastian Brickel and Markus Meuwly[*]

Department of Chemistry, University of Basel, Klingelbergstrasse 80, Basel, Switzerland

E-mail: m.meuwly@unibas.ch

Phone: +41 (0)61 267 38 21. Fax: +41 (0)61 207 38 55

December 6, 2018

Abstract

The Claisen rearrangement is a carbon-carbon bond-forming, pericyclic reaction of fundamental importance due to its relevance in synthetic and mechanistic investigations of organic and biological chemistry. Despite continued efforts, the molecular origins of the rate acceleration in going from the aqueous phase into the protein is still incompletely understood. In the present work the rearrangement reaction for allyl-vinyl-ether (AVE), its dicarboxylated variant (AVE-$(CO_2)_2$) and the biologically relevant substrate chorismate is investigated in gas phase, water and in chorismate mutase. Only the rearrangement of chorismate in the enzyme shows a negative differential barrier when compared to the reaction in water, which leads to the experimentally observed catalytic effect for the enzyme. The molecular origin of this effect is the positioning of AVE-$(CO_2)_2$ and chorismate in the protein active site compared to AVE. Furthermore, in going from AVE-$(CO_2)_2$ to chorismate entropic effects due to

rigidification and ring formation are operative which lead to changes in the rate. Based on "More O'Ferrall-Jencks" diagrams it is confirmed that C-O bond breaking precedes C-C bond formation in all cases. This effect becomes more pronounced in going from the gas phase to the protein.

1   Introduction

The Claisen rearrangement, discovered at the beginning of the 20th century,[1] is an important [3,3]-sigmatropic rearrangement for high stereoselective[2] C-C bond formation.[3] Especially aliphatic Claisen rearrangement is of key importance in organic synthesis[4] and has been studied extensively by both experiment[5–9] and theory.[10–13]

The text book example of a Claisen rearrangement is the reaction of allyl-vinyl ether (AVE) to pent-4-enal[4] (see Figure 1A). In polar solvent, such as water, the stabilization of the transition state (TS) relative to the reaction in vacuum is the origin of the catalytic effect.[10,11,14] It was noted that the solvent effect is partly due to dipole-dipole interactions, which are more pronounced in polar solvents compared to nonpolar solvents.[15,16] The reaction of chorismate to prephenate (see Figure 1C) in the shikimic acid pathway follows formally a Claisen rearrangement.[17] The shikimic acid pathway is operative in the biosynthesis of aromatic amino acids, e.g. in bacteria, which establishes the biological relevance of the reaction. Therefore, this pathway is an attractive target for developing new antibiotics.[18] This has motivated numerous publications on enzymatic Claisen rearrangements in particular[18–28] and reactions with related substrates.[29–32] Compared to the reaction in aqueous solution the enzymatic catalysis of the Claisen rearrangement reaction in chorismate mutase (CM) leads to a rate acceleration by $\sim 10^6$ due to stabilisation of the TS.[33] As a proxy for studying the interactions of chorismate in CM other, related substrates (e.g. AVE-2,6-dicarboxylate (AVE-$(CO_2)_2$, see Figure 1B) have also been used in the past.[27]
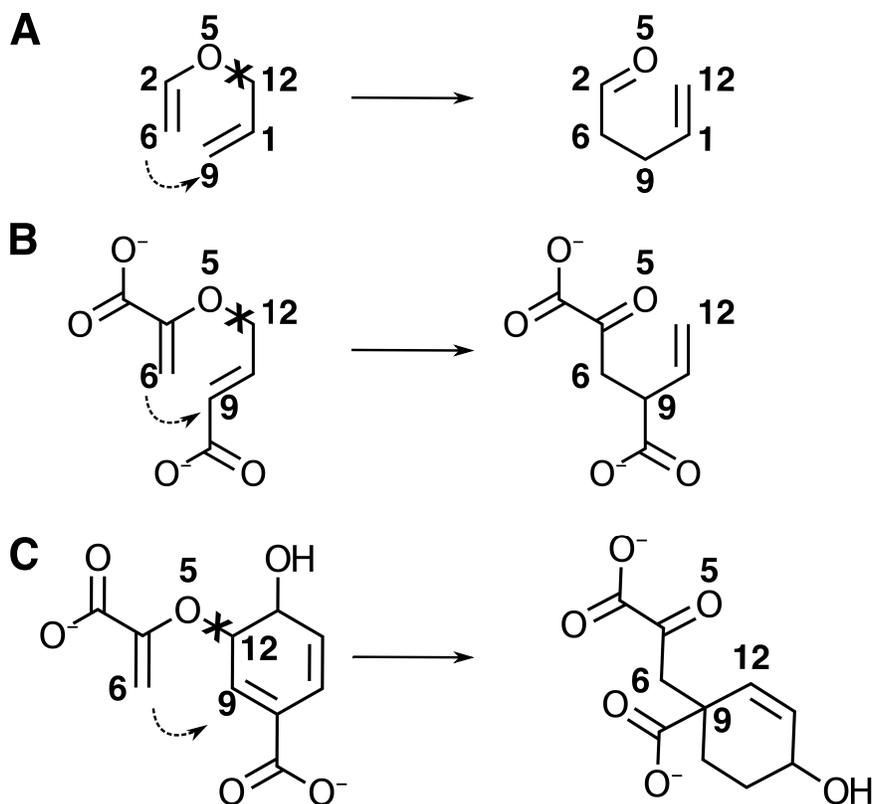
Figure 1: The Claisen rearrangement reaction. A: allyl-vinyl ether → pent-4-enal, including numbering of the heavy atoms. B: AVE-2,6-dicarboxylate → pent-4-enal-dicarboxylate C: Chorismate → prephenate. "X" marks the O5-C12 bond that will be broken, dashed arrow shows the two carbon atoms (C6, C9) that form a bond during rearrangement.

Despite the existing studies on Claisen rearrangement, molecular and certain mechanistic aspects of the reaction remain insufficiently characterized. One of them concerns the structural and molecular origins for the rate enhancement in going from the reaction in solution to that in the protein for chorismate mutase. Another one is the actual pathway between reactant and product for which three possibilities are conceivable: 1) Concerted (via an aromatic TS), 2) asynchronous via 1,4-diyl like TS, or 3) asynchronous via a bis-allyl like TS.[2] The clarification of these aspects - the molecular origins for the rate enhancement in different environments and the pathway between reactant and product - at the molecular level provides the motivation for the present study. Reactive atomistic simulations[34–36] are a meaningful approach for this as they can provide molecular-level insight into reactive processes on sufficiently long time scales in different chemical environments provided that a

meaningful potential energy surface (PES) is available.

2 Computational Details

2.1 MD Simulations

All simulations were performed with CHARMM,[37] including provision for forming and breaking bonds via MS-ARMD.[34] The simulations of AVE-$(CO_2)_2$ and chorismate used the parametrised reactive FF from AVE, with exception of the parameters directly related to the substitution (e.g. bond parameter from C2 to hydrogen in AVE which is replaced by a C-C bond in AVE-$(CO_2)_2$ and chorismate), although further optimization as, e.g., done for the dioxygenation reaction would be possible in principle.[36] This is possible since the substituted groups (e.g. CO_2) are not involved in the rearrangement in AVE-$(CO_2)_2$ and chorismate and remain largely unchanged during the reaction. This was confirmed by *ab initio* optimisation of the reactant and product geometries. Molecular dynamics simulations in the gas phase were performed at 300 K for a total of 100 ps ($\Delta t = 0.25$ fs) using leap-frog Verlet integration scheme.[38] In gas phase the simulations of the three molecules (AVE, AVE-$(CO_2)_2$, and chorismate) started with the respective optimised reactant structure. The simulations in aqueous solution and in the enzyme were started after 50 steps of steepest descent and 50 steps of Adopted Basis Newton-Raphson, followed by 25 ps of NVT dynamics using the velocity Verlet integrator ($\Delta t = 1$ fs) followed by 500 ps ($\Delta t = 2$ fs) of free dynamics using SHAKE.[39] to constrain all bonds involving hydrogen. Periodic Boundary Conditions (PBC) together with the Particle Mesh Ewald (PME[40]) method was used for the long range electrostatic interaction. The cut-off for non-bonded, switching, and smoothing function was 16 Å, 14 Å, and 12 Å, respectively.

For simulations in aqueous solution the system was solvated in a pre-equilibrated TIP3P[41]

water box ($25 \times 25 \times 25$ Å^3) obtained from the CHARMM-GUI.[42] SHAKE[43] was used to constrain all bonds involving hydrogen atoms.

For the simulations in the enzyme the PDB of the 2.2 Å X-ray structure of *Bacillus subtilis* CM (BsCM, PDB: 2CHT[44,45]) was used. BsCM is a homotrimer consisting of six chains with the active sites located at the interface of two chains. The PDB of BsCM contains 1388 residues, 12 transition state analogue (TSA,[46] 8-hydroxy-2-oxa-bicyclo[3.3.1]non-6-ene-3,5-dicarboxylic acid) of the inhibitor molecule, and 530 water molecules which corresponds to four trimers. One such trimer consisting of chains A to C, the crystallographic water and the TSA was used as the starting point to set up the system. This subsystem consists of 6094 atoms, in total.

The simulations in the enzyme environment were set up by replacing the TSA of the inhibitor molecule between the two chains B and C in BsCM by either AVE, AVE-$(CO_2)_2$, or chorismate. The remaining TSA molecules were removed. The enzyme, including the reactant molecule and the crystallographic water molecules, where solvated in a pre-equilibrated water box (including counter ions), of box size 116 Å $\times$ 116 Å $\times$ 116 Å. The overall number of atoms for each of theses systems was 149218, 149207, and 149231 atoms, respectively.

2.2 Umbrella sampling

Umbrella sampling (US)[47] was used to follow the reaction path of the Claisen rearrangement, because the reaction barrier is high and not amenable to direct sampling. The reaction coordinate for the present application was defined as the difference between the breaking O-C bond and the forming C-C bond (see Figure 1), i.e.

$$r_c = d_{O5C12} - d_{C6C9} \tag{1}$$

This corresponds to the previously used reaction coordinate in QM/MM simulations on a chorismate TSA.[30] Simulations were carried out for equidistant windows from $r_c = -2.6$ to 2.2 Å ($\Delta r_c = 0.2$ Å), with $k_{\mathrm{umb}} = 200$ kcal/mol/rad^2 for AVE and AVE-$(CO_2)_2$. For chorismate the sampling was performed in equidistant steps in the range of $r_c = -3.8$ to 2.3 Å, in order to capture the two minimum conformations calculated at the Møller-Plesset MP2/6-311++G(2d,2p) level of theory. Additional umbrellas were added where necessary.

For each window, statistics was accumulated after 5 ps of equilibration simulations. For the calculations in aqueous solution and BsCM US simulations were performed starting each umbrella from a restart file produced in a previous US simulation with the same setting. This procedure prepares the system for the changes in the conformation of the molecules, as well as reduces the difference from initial condition to US conformation along the reaction coordinate. Finally, the analysis of the US simulations was carried out with the Weighted Histogram Analysis Method (WHAM)[48,49] with a tolerance of 0.01.

2.3   Parametrisation of the Force Field

All reactive trajectories were generated with MS-ARMD[34] within the CHARMM[37] software package. MS-ARMD is a powerful, single-valued (by construction), reactive molecular dynamics implementation which combines individual weighted surfaces $w_i(x) * V_i(x)$ (i.e. connectivities), describing different states by parametrized FFs. So called, GAPOs (Gaussian $\times$ POlynomials) for describing the adiabatic barrier are added to form the global reactive Potential Energy Surface (PES). GAPOs are calculated from the energy difference $\Delta V_{ij}(x)$ ($= V_j(x) - V_i(x)$) between two states.

Parametrised FFs for AVE and pent-4-enal were obtained by an iterative procedure, starting with reference parameters from SwissParam.[50] The energy at the equilibrium geometry

of pent-4-enal was chosen as the global zero of energy. Using the initial force field, the equilibrium dynamics of AVE and pent-4-enal in the gas phase was followed for 250 ps at 300 K. A total of 3200 representative structures were extracted and reference energies were determined at the MP2/6-311++G(2d,2p) level of theory using Gaussian09.[51] A downhill simplex algorithm[52] was used for fitting refined FF parameters. The bonds between O5-C12 in the reactant and C6-C9 in the product was described by a Morse oscillator. In a next step, these two FFs were combined to a reactive PES.

In order to parametrise the adiabatic barrier, the IRC of the reaction was calculated at the MP2/6-311++G(2d,2p) level of theory. Such calculations yield a barrier height of 29.7 kcal/mol which agrees favourably with the activation energy of 30.6 kcal/mol determined from experiment[5] and other experiments which find a value of > 28 kcal/mol.[3,9,53] Alternatively, density functional theory calculations at the BLYP/6-31G*[27] and B3LYP/6-31G*[10] level yield somewhat lower barrier heights between 24.1 to 27.7 kcal/mol. Thus, the chosen level of theory was deemed appropriate. For the parametrization 81 structures along the IRC were extracted and their MS-ARMD energy evaluated. A genetic algorithm for parametrising the GAPOs was used to best reproduce the MP2 energies along the IRC path.

3 Results and Discussion

3.1 Quality of the Force Field

All simulations are based on the parametrized force field for the conversion of AVE to pent-4-enal which captures the energetics of the reactant and product to within 0.8 kcal/mol and the reaction path between the end points to within 0.5 kcal/mol, respectively, see Figure 2. The average RMSD between the reference MP2 and the fitted FF energies is 0.7 kcal/mol, which establishes the quality of the FF. For further validation the reactant and the product geometries were minimised at the MP2 level and compared with the force field optimized

structures (see Figure S1) and the normal mode frequencies were found to agree favourably, too (see Tables S1 and S2). The parametrised force field is provided in Tables S3 to S10.
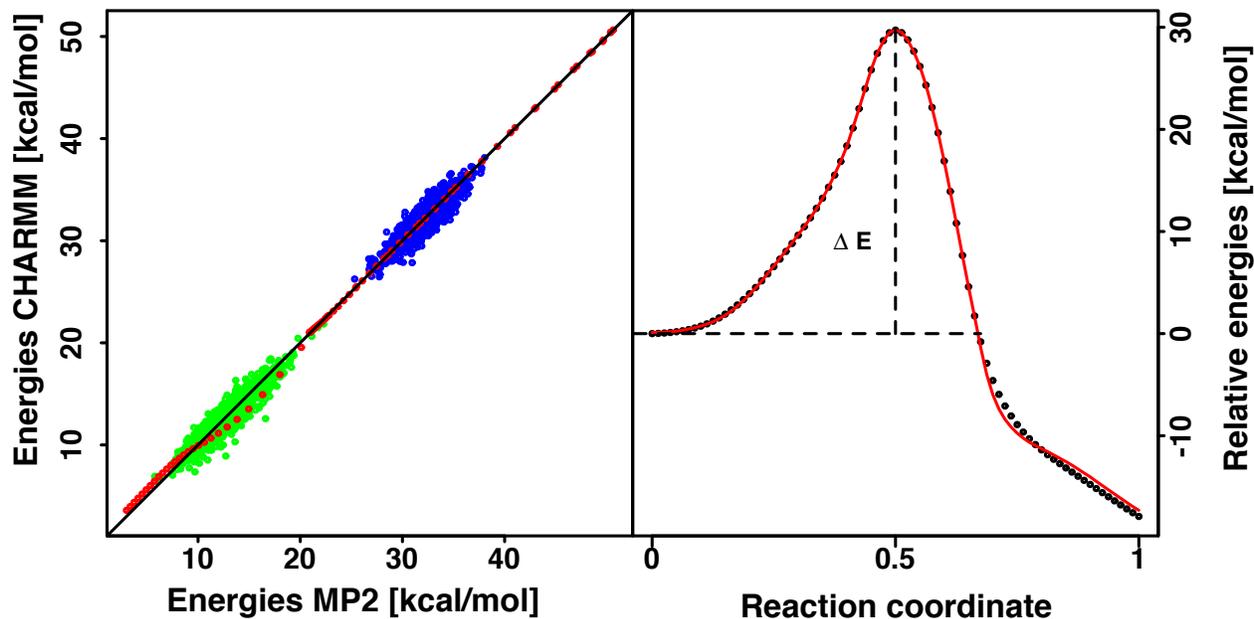

Figure 2: Left panel: Comparison between the MP2/6-311++G(2d,2p) reference energies and the fitted FF. Energies for AVE (blue, RMSD = 0.82 kcal/mol), for pent-4-enal (green, RMSD = 0.8 kcal/mol), and the IRC (red, RMSD = 0.5 kcal/mol). Right panel: *Ab initio* IRC (black points) and GAPO-fitted FF (red line). The right hand side $y-$axis shows the energy relative to the optimised reactant structure. The activation energy is $\Delta E^{\ddagger} = 29.7$ kcal/mol.

3.2 Free Energy Simulations for AVE

Using the parametrised, reactive force field, the barrier free energies $\Delta G^{\ddagger}$ for the rearrangement of AVE in the gas phase, in aqueous solution, and the BsCM enzyme were determined using umbrella sampling (US) simulations. The activation free energies are $\Delta G^{\ddagger} = 23.4$ kcal/mol, 17.7 kcal/mol and 22.3 kcal/mol, respectively (see top panel of Figure 3 and Table 1). Experimentally, barrier free energies have been determined in the gas phase and in apolar solvents to be ~ 28 kcal/mol.[3,5,9] More recent selective integrate-over temperature (SITS) simulations in water using DFTB and the SPC/E water model yield $\Delta G^{\ddagger} = 22.6$ kcal/mol,

~ 5 kcal/mol higher than the present work. This compares with 29.6 kcal/mol obtained here from DFTB3[54] with the TIP3P water model. The activation enthalpy in the gas phase from DFTB calculations is 19.7 kcal/mol which underestimates the experimentally determined $\Delta H^{\ddagger}$ (30.6 kcal/mol). On the other hand, the gas phase free energy barrier height from US is $\Delta G^{\ddagger} = 31.8$ kcal/mol which corresponds to a TS stabilization of $\Delta\Delta G^{\ddagger}_{(gas-solv)} = -2.2$ kcal/mol.

The gas phase $\Delta G^{\ddagger} = 23.4$ kcal/mol underestimates the barrier heights from experiment by ~ 5 kcal/mol. In order to better understand this difference, the distribution functions of all bonds and valence angles involving non-hydrogen atoms were determined from the US simulations around the transition state. Compared with the MP2 structure of the TS (2.32 Å), the O5–C12 bond distribution peaks at 2.63 Å and for the C1-C12-O5 angle the MP2 value is 102° compared with the maximum of the distribution at 122°. All other bonds and angles agree to within 0.05 Å and 5°. Hence, the structure of the TS in the gas phase from US simulations is too open. Furthermore, it was found that despite the quite accurate parametrization of the reactant and product energies, the average harmonic frequencies from the force field differ by 89 and 120 cm^{-1} for the reactant and product, respectively. Therefore, two tests were carried out. First, the US simulations were repeated by increasing all product bond and angle force constants by 10% and second, the US were rerun by placing a constraint (10.0 kcal/mol) on the O5–C12 bond to maintain it near the MP2 TS structure. Both modifications lead to an increase of the barrier, namely by 5.3 kcal/mol when the force constants are changed and by 2.4 kcal/mol when running the simulations with the bias. Hence, for the gas phase simulations the force field could be further improved and mild modifications in the force constants of the product already give near-quantitative agreement with experiment. However, the purpose of the present work is to investigate systematic aspects of the reaction for AVE, AVE-$(CO_2)_2$ and chorismate and to discuss them on a common footing and not to specifically optimize the FFs for individual ligands.

In solution, experiments were carried out in methanol-water mixtures (ranging from 100 % MeOH to 25/75 % MeOH/water) and the reaction barrier reduces by $\Delta\Delta G^{\ddagger}_{\text{gas}-\text{sol}} = -3.5$ to –4.7 kcal/mol,[55] with increasing water content. This compares with a reduction by $\Delta\Delta G^{\ddagger}_{\text{gas}-\text{sol}} = -6.2$ kcal/mol from the present simulations in pure water and is expected due to stabilizing effects of the solvent environment. On the other hand US simulations for this reaction in the protein active site yield $\Delta\Delta G^{\ddagger}_{\text{water}-\text{prot}} = +4.6$ kcal/mol, i.e. a slowdown of the reaction.
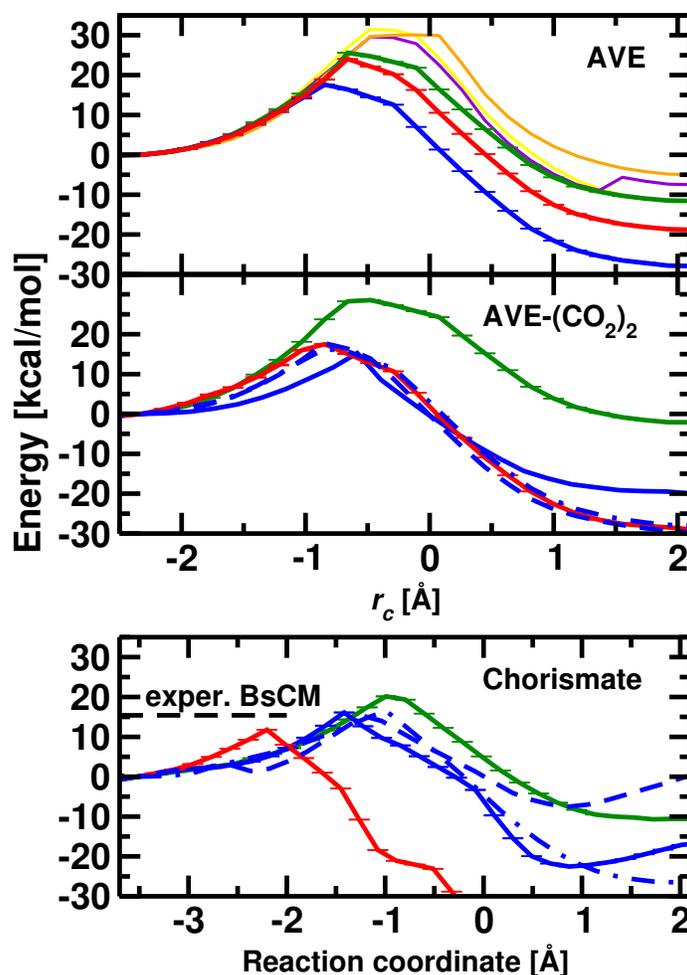
Figure 3: Potentials of mean force for the Claisen rearrangement reaction for AVE (top), AVE-2,6-dicarboxylate (AVE-$(CO_2)_2$, middle), and chorismate (bottom) in gas phase (green), aqueous solution (blue) and in the enzyme BsCM (red). Solid lines for forward simulations and dashed-dotted lines for the reverse reaction (product to reactant). Extended simulations (1 ns per window) are shown as dashed lines. The black dashed line indicates the experimentally measured value.[19] Error bars estimated from a bootstrap error analysis, are overlayed by the line of the respective PMF and are of the size of the line width. DFTB3 forward simulations for AVE in gas phase, water and the enzyme are shown in yellow, violet and orange, respectively.

3.3   Free Energy Simulations for AVE-$(CO_2)_2$

Next, the rearrangement reaction of AVE-2,6-dicarboxylate to pent-4-enal-dicarboxylate was investigated in the three environments. The gas phase PMF (green line in middle panel of Figure 3) for AVE-$(CO_2)_2$ shows an increase in barrier height compared to AVE (32.2

kcal/mol vs. 23.4 kcal/mol). This can be explained by the destabilizing effect of the two negatively charged carboxylate groups which are insufficiently screened in gas phase. In aqueous solution the barrier is considerably reduced (by $\Delta\Delta G^{\ddagger}_{\text{gas-solv}} = 17$ kcal/mol), as expected. The TS stabilisation for $\text{AVE-(CO}_2)_2$ in solution is enhanced compared to AVE due to screening of the repelling charges of the two carboxylate groups during the rearrangement. Extended simulations (1 ns per umbrella, see dashed trace in middle panel of Figure 3) show a change in the free energy of the product, as the surrounding water adapts to the structural changes of the ligand. Such changes on the product side due to slow reorganisation have also been found in ligand-binding simulations for cyano-benzene in Lysozyme.[56] However, for the present work the forward barrier between the reactant and the TS is of primary concern (and can be directly compared with experiment). The forward barrier for the extended sampling increases by 2.5 kcal/mol. This was confirmed by US simulations starting from the solvated product, rather than the solvated reactant, which gives a similar result as the 1 ns forward US simulations (dashed-dotted trace, middle panel Figure 3) and suggests that the simulations are close to converged. No experimental data for the dicarboxylate is available but reactions for the monocarboxylate in $\text{Me}_2\text{SO/H}_2\text{O}$ (ratio 9/1) mixtures have been reported[8] which, however, are not ideal for direct comparison with the computations. For the reaction in the protein the computed $\Delta\Delta G^{\ddagger}$ increases to 18.1 kcal/mol which indicates that the reaction actually slows down in the protein.

3.4 Free Energy Simulations for Chorismate

Finally, the Claisen rearrangement was investigated using the biologically relevant substrate, chorismate (see bottom panel in Figure 3). In addition to the two $-\text{CO}_2^-$ groups, the allyl is completed to form a six-membered ring, see Figure 1C. This makes the substrate in the reactant state less flexible and bulkier compared to the AVE-2,6-dicarboxylate. Contrary to the two other systems, the activation free energy monotonically decreases from $\Delta G^{\ddagger}_{\text{gas}} = 20.5$

kcal/mol in the gas phase to $\Delta G^{\ddagger}_{\text{prot}} = 12.4$ kcal/mol in the protein. In aqueous solution the free energy barrier is intermediate (16.8 kcal/mol) to the two and differs from experiment (24.5 kcal/mol, carried out in 0.02 M Tris-HCl buffer) by 7.7 kcal/mol. However the results are in quite good agreement with QM/MM studies (B3LYP/6-31G(d)/CHARMM27) which yielded 17.4 ± 1.9 kcal/mol.[30] Using an EVB model parametrized to reproduce the barrier in water the free energy barrier height in solvent was found to be 25.1 kcal/mol.[57] As for the results of AVE-$(CO_2)_2$ in solution, extending the sampling in the US simulations to 1 ns for every window increases the free energy of the product by around 15 kcal/mol (dashed blue vs. solid blue line in bottom panel of Figure 3), showing the influence of environmental adaptation on the PMF.[56,58] The barrier height for the forward reaction is not significantly (decrease by 0.8 kcal/mol) influenced.

Table 1: Free energy barrier (in kcal/mol) calculated with MS-ARMD and US in aqueous solution and in the enzyme *Bacillus subtilis* chorismate mutase (BsCM, 116 Å^3 water box). For AVE in the enzyme BsCM the result for the calculation in a 86 Å^3 water box is given in brackets. Experimental (exper.) $\Delta G^{\ddagger}$ from: [a] Ref.[33] (measured in Tris-HCl). [b] values range from 15.0 to 16.2 depending on the protein.[19] All experimental $\Delta G^{\ddagger}$ values based on TST; analysis with an Arrhenius expression yield estimated values ~ 3 kcal/mol lower.[33]

	AVE	AVE-$(CO_2)_2$	Chorismate	Chorismate exper.
Aqueous solution	17.7	15.2	16.8	24.5[a]
Enzyme BsCM	22.3 (24.9)	18.1	12.4	15.4[b]

3.5 Mechanistic Details

In order to clarify the mechanistic details involved in the rearrangement "More O'Ferrall-Jencks" (MOFJ) diagrams[59,60] were generated for all US simulations (50 ps per window). Such diagrams involve two progression coordinates and are useful to illustrate their coupling during the course of a reaction. The corners of the diagram correspond to the reactant, product and the two non-concerted extremes respectively,[7] see left panel in Figure 4. MOFJ diagrams are usually generated from coordinates that quantify the existence of a bond, such

as a distance, bond order or degree of bond making/breaking. In the present case (see Figure 4) the arithmetic average (dots) and the Kernel Density Estimate (KDE, using the R-program)[61,62] minimum and maximum (lines) distribution of the umbrella trajectories are reported. For computing the bond order n' the Pauling relation

$$n' = \exp\left(\frac{D_1 - D(n')}{C}\right) \quad (2)$$

is used where D_1 is the reference interatomic distance from the *ab initio* optimised structures, $D(n')$ is the instantaneous separation during the simulations and $C = 0.6$ Å was used[13,63,64] (for the dependence on the choice of C see Figure S2).
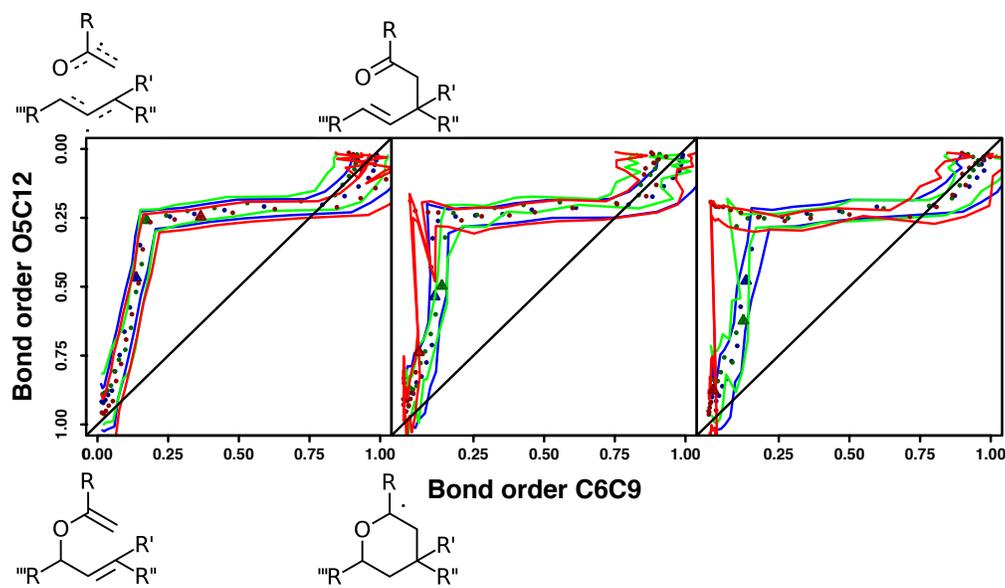

Figure 4: More O'Ferrall-Jencks diagrams for Claisen rearrangement in gas phase (left), aqueous solution (middle) and in the enzyme BsCM (right) for AVE (blue), AVE-2,6-dicarboxylate (green), and chorismate (red). Bond orders are defined using the interatomic distance for the bonds from the respective *ab initio* optimised structure in the Pauling relations (equation 2). From secondary tritium isotope effect in solution the TS of chorismate has been localised at a bond order of ≈ 0.6 for the O5C12 separation with no detectable C6-C9 bond.[65] For MOFJ diagrams by molecule, see Figure S3.

Figure 4 shows that the rearrangement reaction in all three environments follows a step-

wise mechanism, in which first the O5-C12 bond is partially broken and leads towards a bis-allyl-like structure (top left corner), followed by formation of the C6-C9 bond. Running the reaction in polar solvent and in the protein leads to steeper curves during the first phase of the reaction, see Figure 4 (middle panel). In all cases the transition states (indicated by solid triangles) are early, i.e. they occur during the C-O bond-breaking process. This agrees with kinetic isotope experiments according to which C-O cleavage always precedes C-C bond formation.[66] For the nonenzymatic thermal rearrangement of chorismate to prephenate the measured kinetic isotope effects[65,66] indicate that at the TS the C-O bond is about 40 % broken but little or no C-C bond is formed, consistent with the MOFJ diagrams assuming that bond order and C-O distance are linearly related.

Unbiased MD simulations starting at the TS preserve the overall structure of the MOFJ diagrams whereby the TS is shifted towards a more "broken" C-O bond, see Figure 5. These unbiased simulations further establish that C-O bond breaking precedes C-C formation. The average O5-C12 and C6-C9 bond lengths at the transition state from the US simulations are 2.0 and 2.6 Å. At the B3LYP/6-311++G(2d,2p) level the energy of this configuration is ≈ 40 % higher than the minimum energy for the constrained C-C bond at the TS-value which also supports the notion that the C-O bond is appreciably weakened whereas the C-C bond is still not formed. In the enzyme the measured[65] isotope effect of 0.99 suggests an even earlier transition state than in water which is consistent with the free energy simulations (see Figure 3).

For AVE-$(CO_2)_2$ and especially chorismate the effect of TS stabilization is further enhanced in the enzyme BsCM due to spatial constraints. This can be seen by the pronounced lag in C6-C9 bond formation of these two substrates in BsCM. AVE in BsCM shows a distribution intermediate to that in gas phase and aqueous solution. This difference between AVE and the other two substrates is indicative of the importance of the carboxylate groups in
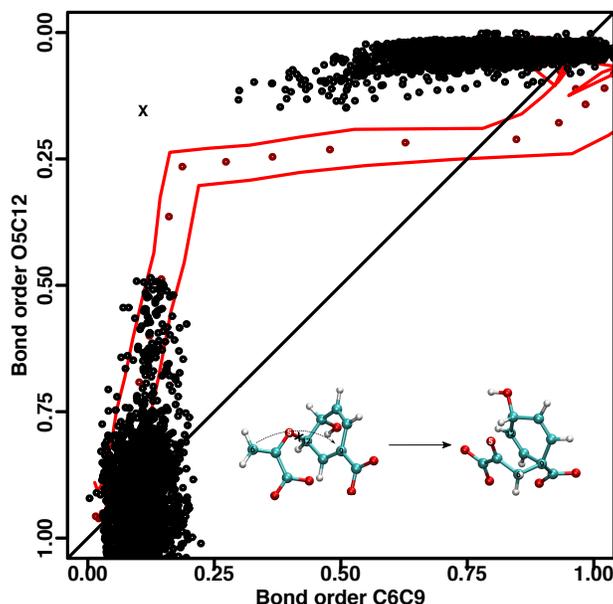
Figure 5: More O'Ferrall-Jencks diagrams for Claisen rearrangement of chorismate in gas phase. The red points show the arithmetic average value of the distribution from the 50 ps US trajectories, lines the maximum and minimum from KDE. Bond orders are defined using the interatomic distance for the bonds from the respective *ab initio* optimised structure in the Pauling relation (equation 2). Black points show the distribution of 32 free dynamics started from TS structures extracted from the US. The black "X" marks the center of gravity of the distribution around the TS.

AVE-$(CO_2)_2$ and chorismate for the catalytic step in BsCM in locking the substrate into place. The importance of the carboxylate groups was also shown by EVB simulations[57] and by measurements of rate constants of different chorismate-like substrates.[67] However, the barrier height for AVE and AVE-$(CO_2)_2$ in the enzyme compared to simulations in aqueous solution increases. This anti-catalytic effect decreases from AVE to AVE-$(CO_2)_2$ due to the presence of the -CO_2 groups and may reduce to close to zero when sampling is extended in the US (see above: increase of the free energy barrier height in water by 2.5 kcal/mol). Additional entropic effects lead to the catalytic effect found for Chorismate.

Starting from 32 chorismate structures extracted from the US simulations at the TS in gas phase, free MD simulations were performed. The resulting MOFJ diagrams from the un-

biased simulations are consistent with those from the US simulation (see Figure 5) with the center of gravity from the unbiased simulations for the TS structure somewhat more advanced along the O5–C12 bond cleavage (see black "X" in Figure 5), compared to that from US.
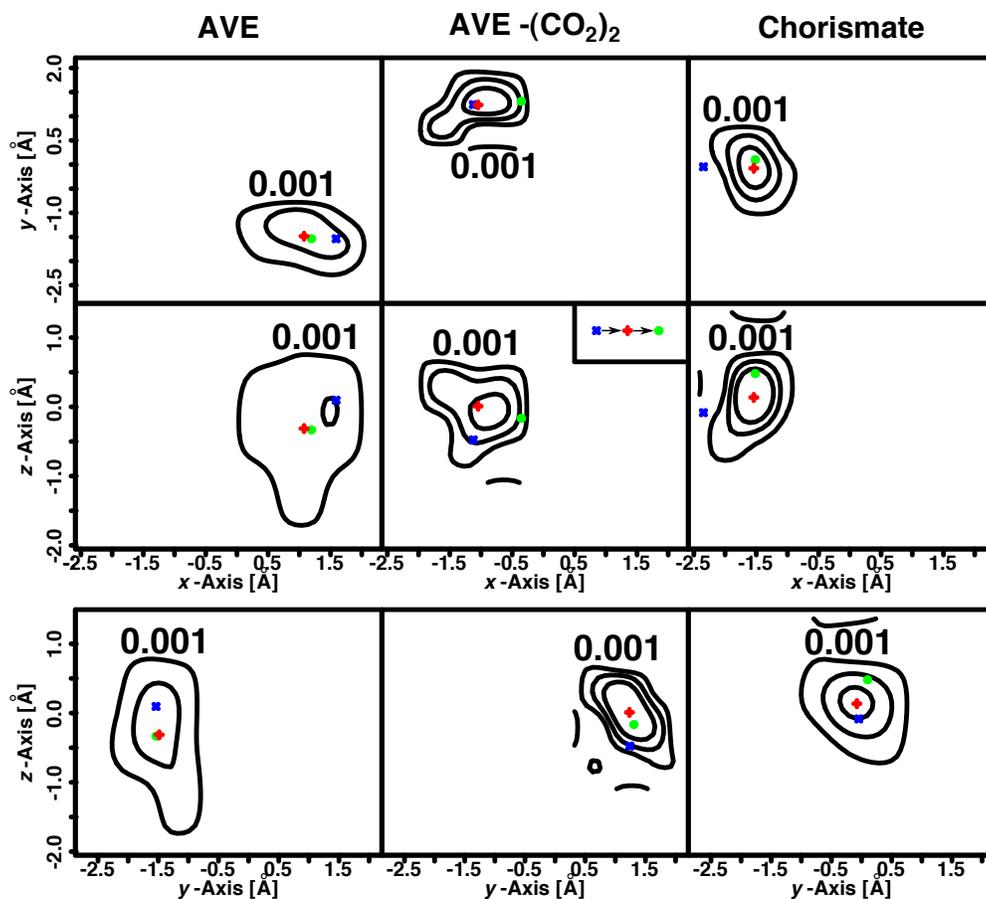
Figure 6: Normalised probability distribution of the $x-$, $y-$, and $z-$coordinate of the center of mass of the heavy atom framework of AVE (left), AVE-$(CO_2)_2$ (middle), and chorismate (right) at the TS in the enzyme BsCM. The center of gravity of the reactant (blue x), TS (red plus), and product (green dot) are displayed in each panel. The origin of the coordinate system was set to the position of the transition state analogue in the X-ray structure (see Figure 7). The enzyme backbone of the US trajectories were aligned to the enzyme backbone of the X-ray structure. For the distribution for the reactant and the product state, see Figures S4 and S5)

The differences in the enzymatic activity for the three substrates can also be related to spa-

tial determinants within the active site. For this, the position of the center of mass (CoM) of the molecules within the active site is analyzed. Figure 6 shows the normalised probability distribution of $x-$, $y-$ and $z-$coordinates of the CoM of the heavy atom framework common to all 3 ligands from the umbrella sampling simulations at the TS of the PMF. A broad or narrow distribution is indicative of a "loose" or "tight" TS, respectively.
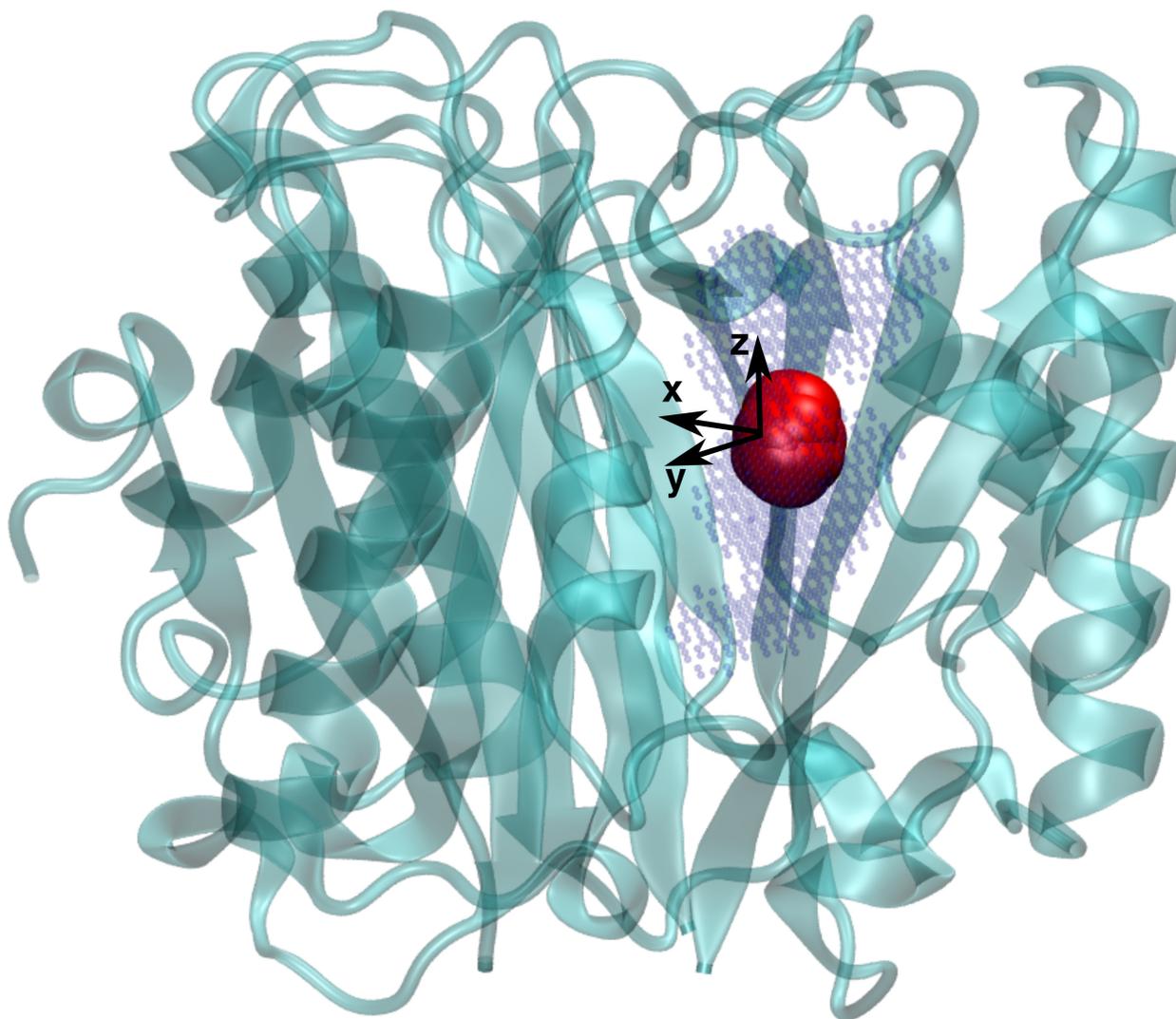
Figure 7: Overview of the simulation system for chorismate mutase and the center of mass of the heavy atom framework of chorismate (red) from the umbrella at the TS in Figure 3. Blue spheres delineate the active site from pocket mapping.[68,69] The origin of the coordinate system for the analysis of the localization for each ligand with the three Cartesian axes is shown in black.

With respect to the localization within the active site, AVE-$(CO_2)_2$ and chorismate occupy comparable regions whereas AVE does not. The TS for AVE is displaced by 2 Å and 3 Å along the $x-$ and $y-$axis, respectively. Furthermore, the widths of the distributions decrease in going from AVE, over AVE-$(CO_2)_2$ to chorismate, i.e. the distributions change from "loose" to "tight", as does the TS itself, see Figure 4. Specifically, the widths along the $x-$ and $z-$directions are different for AVE on one side and AVE-$(CO_2)_2$ and chorismate on the other. Due to the presence of the CO_2^- groups the positioning of the CoM of AVE-2,6-dicarboxylate and chorismate is similar to one another. However, AVE-$(CO_2)_2$ shows a broader distribution in the $x-$direction and a repositioning along y indicating increased flexibility of the CoM compared to chorismate.

Figure 6 also shows the center of gravity of the reactant (blue cross), the TS (red plus), and the product (green dot) of the three ligands in the active site. For AVE the center of gravity of the TS overlays with that of the product which is not the case for the other two molecules. The distribution of the center of gravities of AVE-$(CO_2)_2$ show movement of the ligand during the reaction, except for the $y-$axis position. For chorismate the center of gravity is further displaced between the reactant and the TS, compared to AVE-$(CO_2)_2$, while the movement between TS and product is reduced. This suggests that stabilization of the transition state is coupled to displacement in configurational space and is indicative of TS stabilisation rather than substrate binding.

4   Conclusions

The barrier height for the reaction of chorismate in BsCM is 12.4 kcal/mol, which yields a speedup by 3 orders of magnitude as estimated from transition state theory. Hence, in going from AVE and AVE-$(CO_2)_2$ to chorismate and using the same reactive PES the simulations

find that for smaller substrates the rate slows down when going from water to the protein as the environment whereas for the largest substrate (chorismate) they correctly find that the protein accelerates the reaction. Considering the changes of $+4.6$ (AVE), $+2.9$ (AVE-$(CO_2)_2$) and -4.4 (chorismate) kcal/mol in the activation free energies and correlating them with the actual chemical modifications suggest that both, electrostatic stabilization (AVE$\rightarrow$AVE-$(CO_2)_2$) and entropic contributions (AVE-$(CO_2)_2 \rightarrow$ chorismate, through the rigidification and larger size of chorismate) lead to the rate enhancement observed for chorismate in CM.

The experimentally measured acceleration is 6 orders of magnitude[19] which is underestimated by the present simulations. However, as already mentioned, quantitative agreement for individual systems in specific environments could be further improved by a dedicated parametrization of the force fields compared to experiment.[57] However, the present approach focuses on a systematic assessment of the reaction for the three different systems without specific optimization of the energy function in a particular environment.

In summary, the Claisen rearrangement reaction of AVE, AVE-$(CO_2)_2$, and chorismate in gas phase, aqueous solution and BsCM follows a stepwise mechanism in which the O5-C12 bond breaks prior to C6-C9 bond formation, especially in aqueous solution and BsCM. The analysis of the TS position in the active site of BsCM reveals that the lack of catalytic effect on AVE is due to its loose positioning, insufficient interaction with and TS stabilization by the active site of the enzyme. Major contributions to localizing the substrate in the active site of BsCM originate from the CO_2^- groups. This together with the probability distributions in the reactant, TS and product states suggest that entropic factors must also be considered when interpreting differences between the systems, specifically (but not only) in the protein environment. This requires extended sampling of the configurational space which is currently only possible with reactive force fields and outside the scope of mixed QM/MM MD simulations. The present work provides molecular-level insight into the Claisen rearrangement of

chorismate in BsCM that will serve as a proxy for using computational methods for future development in protein engineering.

Supporting Information Available

Further computational details and validation of the force fields together with all force field parameters for the ligands, additional More O'Ferral-Jencks analysis by system, and center of mass distributions for reactants and products in the active site of BsCM are available in the supporting information. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

Acknowledgment

References

(1) Claisen, L. Über Umlagerung von Phenol-allyläthern in C-Allyl-phenole. *Chem. Ber.* **1912**, *45*, 3157–3166.

(2) Iwakura, I. The Experimental Visualisation of Molecular Structural Changes During Both Photochemical and Thermal Reactions by Real-Time Vibrational Spectroscopy. *Phys. Chem. Chem. Phys.* **2011**, *13*, 5546–5555.

(3) Coates, R. M.; Rogers, B. D.; Hobbs, S. J.; Curran, D. P.; Peck, D. R. Synthesis and

Claisen Rearrangement of Alkoxyallyl Enol Ethers. Evidence for a Dipolar Transition State. *J. Am. Chem. Soc.* **1987**, *35*, 2601–2605.

(4) Ziegler, F. E. The Thermal, Aliphatic Claisen Rearrangement. *Chem Rev* **1988**, *88*, 1423–1452.

(5) Schuler, F. W.; Murphy, G. W. The Kinetics of the Rearrangement of Vinyl Allyl Ether. *J. Am. Chem. Soc.* **1950**, *72*, 3155–3159.

(6) Gajewski, J.; Brichford, N. Secondary Deuterium Kinetic Isotope Effects in the Aqueous Claisen Rearrangement: Evidence Against an Ionic Transition State. *J. Am. Chem. Soc.* **1994**, *117*, 3165–3166.

(7) Gajewski, J.; Conrad, N. Variable Transition State Structure in 3, 3-Sigmatropic Shifts from α-Secondary Deuterium Isotope Effects. *J. Am. Chem. Soc.* **1979**, *101*, 6693–6704.

(8) Gajewski, J. J.; Jurayj, J.; Kimbrough, D. R.; Gande, M. E.; Ganem, B.; Carpenter, B. K. The Mechanism of Rearrangement of Chorismic Acid and Related Compounds. *J. Am. Chem. Soc.* **1987**, *109*, 1170–1186.

(9) Burrows, C.; Carpenter, B. Substituent Effects on the Aliphatic Claisen Rearrangement. 1. Synthesis and Rearrangement of Cyano-Substituted Allyl Vinyl Ethers. *J. Am. Chem. Soc.* **1981**, *103*, 6983–6984.

(10) Guest, J. M.; Craw, J. S.; Vincent, M. A.; Hillier, I. H. The Effect of Water on the Claisen Rearrangement of Allyl Vinyl Ether: Theoretical Methods Including Explicit Solvent and Electron Correlation. *Perkin Trans. 2* **1997**, 71–74.

(11) Severance, D. L.; Jorgensen, W. L. Effects of Hydration on the Claisen Rearrangement of Allyl Vinyl Ether from Computer Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10966–10968.

(12) Davidson, M. M.; Hillier, I. H.; Vincent, M. A. The Claisen Rearrangement of Allyl Vinyl Ether in the Gas Phase and Aqueous Solution. Structures and Energies Predicted by High-Level Ab Initio Calculations. *Chem. Phys. Lett.* **1995**, *246*, 536–540.

(13) Meyer, M. P.; DelMonte, A. J.; Singleton, D. A. Reinvestigation of the Isotope Effects for the Claisen and Aromatic Claisen Rearrangements: The Nature of the Claisen Transition States. *J. Am. Chem. Soc.* **1999**, *121*, 10865–10874.

(14) Cramer, C. J.; Truhlar, D. G. ChemInform Abstract: What Causes Aqueous Acceleration of the Claisen Rearrangement? *J Am Chem Soc* **1992**, *114*, 8794–8799.

(15) White, W. N.; Wolfarth, E. F. The o-Claisen rearrangement. VIII. Solvent effects. *The Journal of Organic Chemistry* **1970**, *35*, 2196–2199.

(16) Acevedo, O.; Armacost, K. Claisen Rearrangements: Insight into Solvent Effects and "on Water" Reactivity from QM/MM Simulations. *J. Am. Chem. Soc.* **2010**, *132*, 1966–1975.

(17) Ganem, B. From Clucose to Aromatics: Recent Developments in Natural Products of the Shikimic Acid Pathway. *Tetrahedron* **1978**, *34*, 3353–3383.

(18) Davidson, M. M.; Guest, J. M.; Craw, J. S.; Hillier, I. H.; Vincent, M. A. Conformational and Solvation Aspects of the Chorismate - Prephenate Rearrangement Studied by Ab Initio Electronic Structure and Simulation Methods. *Perkin Trans. 2* **1997**, 1395–1400.

(19) Kast, P.; Asif-Ullah, M.; Hilvert, D. Is Chorismate Mutase a Prototypic Entropy Trap? - Activation Parameters for the Bacillus Subtilis Enzyme. *Tetrahedron Lett.* **1996**, *37*, 2691–2694.

(20) Ranaghan, K. E.; Ridder, L.; Szefczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J. Insights Into Enzyme Catalysis from QM/MM Modelling: Transition State Stabilization in Chorismate Mutase. *Mol. Phys.* **2003**, *101*, 2695–2714.

(21) Lever, G.; Cole, D. J.; Lonsdale, R.; Ranaghan, K. E.; Wales, D. J.; Mulholland, A. J.; Skylaris, C. K.; Payne, M. C. Large-Scale Density Functional Theory Transition State Searching in Enzymes. *J. Phys. Chem. Lett.* **2014**, *5*, 3614–3619.

(22) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. Theoretical QM/MM Studies of Enzymatic Pericyclic Reactions. *Interdiscip. Sci. Comput. Life Sci.* **2010**, *2*, 115–131.

(23) Ferrer, S.; Martí, S.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J. Molecular Mechanism of Chorismate Mutase Activity of Promiscuos MbtI. *Theor. Chem. Acc.* **2011**, *128*, 601–607.

(24) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J.; Field, M. J. A Hybrid Potential Reaction Path and Free Energy Study of the Chorismate Mutase Reaction. *J. Am. Chem. Soc.* **2001**, *123*, 1709–1712.

(25) Roca, M.; Vardi-Kilshtain, A.; Warshel, A. Toward Accurate Screening in Computer-Aided Enzyme Design. *Biochemistry* **2009**, *48*, 3046–3056.

(26) Madurga, S.; Vilaseca, E. SCRF Study of the Conformational Equilibrium of Chorismate in Water. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3548–3554.

(27) Wiest, O.; Houk, K. N. Stabilization of the Transition State of the Chorismate-Prephenate Rearrangement: An ab Initio Study of Enzyme and Antibody Catalysis. *J. Am. Chem. Soc.* **1995**, *117*, 11628–11639.

(28) Hur, S.; Bruice, T. C. The Mechanism of Catalysis of the Chorismate to Prephenate Reaction by the Escherichia Coli Mutase Enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1176–1181.

(29) Vance, R.; Rondan, N. Transition Structures for the Claisen Rearrangement. *J. Am. Chem. Soc.* **1988**, *110*, 2314–2315.

(30) Claeyssens, F.; Ranaghan, K. E.; Lawan, N.; Macrae, S. J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Analysis of Chorismate Mutase Catalysis by QM/MM Modelling of Enzyme-Catalysed and Uncatalysed Reactions. *Org. Biomol. Chem.* **2011**, *9*, 1578–1590.

(31) Ranaghan, K. E.; Ridder, L.; Szefczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J. Transition State Stabilization and Substrate Strain in Enzyme Catalysis: Ab Initio QM/MM Modelling of the Chorismate Mutase Reaction. *Org. Biomol. Chem.* **2004**, *2*, 968–980.

(32) Carlson, H. A.; Jorgensen, W. L. Monte Carlo Investigations of Solvent Effects on the Chorismate to Prephenate Rearrangement. *J. Am. Chem. Soc.* **1996**, *118*, 8475–8484.

(33) Andrews, P. R.; Smith, G. D.; Young, I. G. Transition-State Stabilization and Enzymic Catalysis. Kinetic and Molecular Orbital Studies of the Rearrangement of Chorismate to Prephenate. *Biochemistry* **1973**, *12*, 3492–3498.

(34) Nagy, T.; Reyes, J. Y.; Meuwly, M. Multisurface Adiabatic Reactive Molecular Dynamics. *J. Chem. Theory. Comput.* **2014**, *10*, 1366–1375.

(35) Soloviov, M.; Das, A. K.; Meuwly, M. Structural Interpretation of Metastable States in Myoglobin-NO. *Angew. Chem. Int. Ed.* **2016**, *55*, 10126–10130.

(36) Das, A. K.; Meuwly, M. Kinetics and Structural Interpretation of Competitive Ligand Binding for NO Dioxygenation in Truncated Hemoglobin N. *Angew. Chemie Int. Ed.* **2018**, 3509–3513.

(37) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. et al. CHARMM: The Biomolecular Simulation Program. *J. Chem. Comp.* **2009**, *30*, 1545–1614.

(38) Verlet, L. Computer Experiemtns on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98–103.

(39) VanGunsteren, W.; Berendsen, H. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327.

(40) Essmann, U.; Perera, L.; Darden, M. L. B.; Lee, H.; Pedersen, L. G.; Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H. et al. A Smooth Particle Mesh Ewald Method A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(42) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Chem. Comp.* **2008**, *29*, 1859–1865.

(43) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(44) Chook, Y.; Ke, H.; Lipscomb, W. Crystal Structures of the Monofunctional Chorismate Mutase from Bacillus Subtilis and its Complex with a Transition State Analog. *Proc. Nat. Acad. Sci. USA* **1993**, *90*, 8600–8603.

(45) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. The Protein Data Bank. `www.rcsb.org`, 2000.

(46) Bartlett, P. A.; Johnson, C. R. An Inhibitor of Chorismate Mutase Resembling the Transition-State Conformation. *J. Am. Chem. Soc.* **1985**, *107*, 7792–7793.

(47) Kottalam, J.; Case, D. A. Dynamics of Ligand Escape from the Heme Pocket of Myoglobin. *J. Am. Chem. Soc.* **1988**, *110*, 7690–7697.

(48) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(49) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40–57.

(50) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: A Fast Force Field Generation Tool for Small Organic Molecules. *J. Chem. Comp.* **2011**, *32*, 2959–2368.

(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

(52) Nelder, J.; Mead, R. A Simplex Method for Function Minimization. *Chem Phys* **1965**, *7*, 308–313.

(53) Zhang, J.; Yang, Y. I.; Yang, L.; Gao, Y. Q. Dynamics and Kinetics Study of In-Water Chemical Reactions by Enhanced Sampling of Reactive Trajectories. *J. Phys. Chem. B* **2015**, *119*, 14505–14514.

(54) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B* **2001**, *105*, 569–585.

(55) Brandes, E.; Grieco, P.; Gajewski, J. Effect of Polar-Solvents on the Rates of Claisen Rearrangements - Assessment of Ionic Character. *J. Org. Chem.* **1989**, *54*, 515–516.

(56) Mondal, P.; Meuwly, M. Vibrational Stark Spectroscopy for Assessing Ligand-Binding Strengths in a Protein. *Phys. Chem. Chem. Phys.* **2017**, *19*, 16131–16143.

(57) Štrajbl, M.; Shurki, A.; Kato, M.; Warshel, A. Apparent NAC Effect in Chorismate Mutase Reflects Electrostatic Transition State Stabilization. *J. Am. Chem. Soc.* **2003**, *125*, 10228–10237.

(58) El Hage, K.; Mondal, P.; Meuwly, M. Free Energy Simulations for Protein Ligand Binding and Stability. *Mol. Simul.* **2018**, *7022*, 1–18.

(59) O'Ferrall, R. A. M. Relationships between E2 and E1cB Mechanisms of β-Elimination. *J Chem Soc* **1970**, *#*, 274–277.

(60) Jencks, W. P. General Acid-Base Catalysis of Complex Reactions in Water. *Chem. Rev.* **1972**, *72*, 705–718.

(61) Sheather, S. J.; Jones, M. C. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **1991**, *53*, 683–690.

(62) Scott, D. W. Multivariate Density Estimation and Visualization. *Pap. / Humboldt-Universität Berlin, Cent. Appl. Stat. Econ. (CASE)* **2004**, *16*, 1–23.

(63) Burton, G. W.; Sims, L. B.; Wilson, J. C.; Fry, A. Calculation of Carbon-14, Chlorine-37, and Deuterium Kinetic Isotope Effects in the Solvolysis of tert-Butyl Chloride. *J. Am. Chem. Soc.* **1977**, *99*, 3371–3379.

(64) Pauling, L. Atomic Radii and Interatomic Distances in Metals. *J. Am. Chem. Soc.* **1947**, *69*, 542–553.

(65) Addadi, L.; Jaffe, E. K.; Knowles, J. R. Secondary Tritium Isotope Effects as Probes of the Enzymic and Nonenzymic Conversion of Chorismate to Prephenate. *Biochemistry* **1983**, *22*, 4494–4501.

(66) Wright, S.; DeClue, M.; Mandal, A.; Lee, L.; Wiest, O.; Cleland, W.; Hilvert, D. Isotope

Effects on the Enzymatic and Nonenzymatic Reactions of Chorismate. *J. Am. Chem. Soc.* **2005**, *127*, 12957–12964.

(67) Pawlak, J. L.; Padykula, R. E.; Kronis, J. D.; Aleksejczyk, R. A.; Berchtold, G. A. Structural Requirements for Catalysis by Chorismate Mutase. *J. Am. Chem. Soc.* **1989**, *111*, 3374–3381.

(68) Kawabata, T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 1195–1211.

(69) Rezácová, P.; Borek, D.; Moy, S. F.; Joachimiak, A.; Otwinowski, Z. Crystal structure and putative function of small Toprim domain-containing protein from Bacillus stearothermophilus. *Proteins* **2008**, *70*, 311–319.