ORIGINAL PAPER

Computational design of a lipase for catalysis of the Diels-Alder reaction

**Mats Linder · Anders Hermansson ·
John Liebeschuetz · Tore Brinck**

Abstract Combined molecular docking, molecular dynamics (MD) and density functional theory (DFT) studies have been employed to study catalysis of the Diels-Alder reaction by a modified lipase. Six variants of the versatile enzyme Candida Antarctica lipase B (CALB) have been rationally engineered in silico based on the specific characteristics of the pericyclic addition. A kinetic analysis reveals that hydrogen bond stabilization of the transition state and substrate binding are key components of the catalytic process. In the case of substrate binding, which has the greater potential for optimization, both binding strength and positioning of the substrates are important for catalytic efficiency. The binding strength is determined by hydrophobic interactions and can be tuned by careful selection of solvent and substrates. The MD simulations show that substrate positioning is sensitive to cavity shape and size, and can be controlled by a few rational mutations. The well-documented S105A mutation is essential to enable sufficient space in the vicinity of the oxyanion hole. Moreover, bulky residues on the edge of the active site hinders the formation of a sandwich-like nearattack conformer (NAC), and the I189A mutation is needed to obtain enough space above the face of the α,β-double bond on the dienophile. The double mutant S105A/I189A performs quite well for two of three dienophiles. Based on binding constants and NAC energies obtained from MD simulations combined with activation energies from DFT computations, relative catalytic rates (v_{cat}/v_{uncat}) of up to 103 are predicted.

M. Linder · A. Hermansson · T. Brinck (✉)
Physical Chemistry, Royal Institute of Technology (KTH),
Stockholm, Sweden
e-mail: tore@physchem.kth.se

M. Linder
e-mail: linder@physchem.kth.se

J. Liebeschuetz
Cambridge Crystallographic Data Centre,
Cambridge, UK
e-mail: john@ccdc.cam.ac.uk

Introduction

The Diels-Alder reaction is remarkably rare in nature [1], which may be attributed to a large activation energy due to the highly ordered transition state. To date, only a few naturally occurring enzymes coined "Diels-Alderases" have been characterized [2], and these are very substrate-specific. The mechanistic studies performed on macrophomate synthase (MPS), the only enzyme performing bimolecular Diels-Alder [3], also show that the catalytic mechanism is likely to be a two-step, Michael-Aldol pathway, rather than a concerted [4 + 2] addition [4]. Attempts to employ artificial 'enzyme mimics', for example cyclodextrin, catalytic antibodies and other macrocycles [5] that utilize substrate recognition and transition state (TS) binding, have yielded only modest results due to product inhibition [6].

To find a versatile Diels-Alderase is highly desirable, owing to the synthetic use of creating up to four stereocenters in one concerted step. The concept of noncovalent, hydrogen bond (H-bond) catalysis of Diels-Alder using carbonyl-activated dienophiles has undergone rapid development in the last decade [7–9], and since activation of

carbonyl-containing substrates to nucleophilic attack is very common in enzymes, in particular the large hydrolase family (EC 3.x), the idea to utilize a promiscuous member of such an enzyme family for Diels-Alderase activity is not far-fetched, and has been suggested in the literature [10].

Among the hydrolase family, lipases are by far the most commonly utilized for in vitro catalysis of organic reactions, such as (dynamic) kinetic resolution of secondary alcohols, transesterifications and acylations [11–14]. Many lipases have been found to be stable under a broad range of temperature and pH, and active in organic solvents [15, 16], which greatly expands their utility in synthetic applications. More recently, evidence has emerged of catalytic promiscuity [17, 18] (i.e., the ability of an enzyme to exhibit catalysis of other than its native reaction) toward, e.g., aldol condensations [19, 20], epoxidations [21] and Michael additions [22–26], not seldom enhanced by point mutations [27]. An often-seen feature in promiscuous catalysis is that only parts of the catalytic machinery are used, so that removal of one or more of these residues might actually enhance the desired activity [28]. Following these concepts, the hypothetical reaction mechanism for hydrolase-catalyzed Diels-Alder is shown in Fig. 1a. Note that the only part of the catalytic machinery thought to influence Diels-Alder catalysis is the oxyanion hole, rendering the catalytic triad obsolete. This effectively removes the dependence on dynamic binding in the terminology of Kirby [6], and means that there is a large degree of freedom in screening and (re)design of candidate enzymes.

Building on previous studies of Candida Antarctica Lipase B (CALB), performed by us and collaborators,[19–22] we set out to test this versatile enzyme for Diels-Alder activity. CALB was chosen because of its highly documented catalytic promiscuity and tolerance for a broad range of solvents and reaction conditions. Figure 1b shows the structure of CALB, which belongs to the large α/β hydrolase superfamily [29].

Initial experimental efforts with CALB wild type (wt) and S105A mutant (shown to have enhanced catalytic activity for carbon-carbon bond formation [20, 22–24]) did not indicate any Diels-Alder activity [30]. Therefore, we turned our attention to modeling the enzyme-substrate interactions, in order to i) identify the obstacles for CALB Diels-Alder activity, and ii) explore how rational design might improve the system. We have employed both molecular dynamics (MD) and quantum chemical (QC) methods, as well as molecular docking, in order to obtain all facets of the complex enzyme-substrate(s) system. To our knowledge, in silico studies of enzymes with non-natural substrates are scarce, and we therefore seek to find a theoretical framework for prediction of reactivity, catalytic power, and productive combinations of enzyme species and substrates. Unlike most other computational design studies which focus on optimum transition state stabilization, including the groundbreaking de novo designed enzymes by Baker and colleagues [31, 32], we attempt to monitor the full reaction pathway, including substrate binding and substrate-substrate interactions. This approach is supported

Fig. 1 (a) The hypothetic CALB-catalyzed Diels-Alder mechanism, (b) CALB crystal structure (PDB entry 1LBT), with residues important for the present study explicitly shown. Residues subject to mutations are shown in green
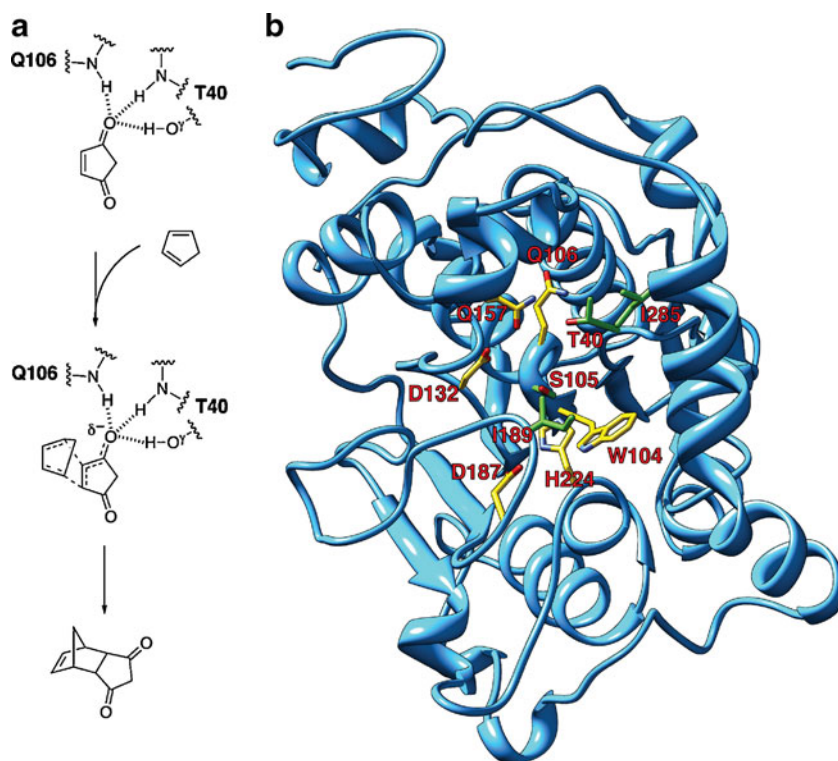
by a very recent study by Simón and Goodman [33], who show that oxyanion holes are not optimized for maximum TS stabilization, but rather for minimizing the reaction barrier.

As seen in our previous work, H-bond catalysts can provide an enthalpic TS stabilization of ~1–5 kcal·mol^{-1}, depending on reactants and polarizability of the medium [34]. These values are expected to be similar using lipase oxyanion holes, so the main concern for the viability of lipases such as CALB therefore lies in the ligand-receptor dynamics. From data on molecular catalysis reported in the literature, we can conclude that rather high catalyst concentrations are required for a reasonable activation, typically around 1–20 mol-% [35–38], which is not possible with enzymes. Thus the free energy of binding between the enzyme and both substrates must be sufficiently low.

Six CALB variants were simultaneously studied in this work. The S105A mutant was explored alongside the wild-type (wt), due to its previously reported utility. I189 was identified as a steric constraint near the oxyanion hole. Thus, the I189A and S105A/I189A mutants were probed. I189 and I285 constitute the 'channel' dividing the active site in the small 'stereoselectivity pocket' and larger hydrophobic (or acyl) pocket [39]. This channel is the most sterically crowded region near the oxyanion hole, yet the region where the sandwichlike DA TS must be located in order to exploit the catalytic machinery of CALB. Consequently, we also modeled the I285A and S105A/I285A mutants.

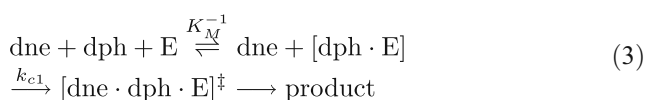## Theory & methods

Kinetics and binding dynamics

In order to estimate the catalytic potential of the Diels-Alder reaction by a given enzyme, we need expressions relating the different states along the catalytic pathway, as well as a means to calculate their relative energies. The reaction formula for the uncatalyzed reaction is given in Eq. 1.

$$\text{dne} + \text{dph} \xrightarrow{k_u} [\text{dne} \cdot \text{dph}]^{\ddagger} \rightarrow \text{product}, \tag{1}$$
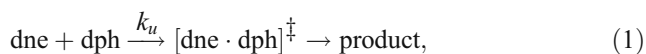
for which the rate expression is

$$\upsilon_u = k_u[\text{dne}][\text{dph}]. \tag{2}$$

As a first approximation, we can consider the case where we have no specific interaction between the diene and the enzyme active site, i.e., there is no productive binding of the diene (note that the dienophile is assumed to bind to the oxyanion hole with the binding constant K_M, as in Fig. 1a).

$$\text{dne} + \text{dph} + \text{E} \overset{K_M^{-1}}{\rightleftharpoons} \text{dne} + [\text{dph} \cdot \text{E}]$$
$$\xrightarrow{k_{c1}} [\text{dne} \cdot \text{dph} \cdot \text{E}]^{\ddagger} \longrightarrow \text{product} \tag{3}$$

This reaction can be described by the well-known Michaelis-Menten kinetics [40].

$$K_M = \frac{[\text{E}][\text{dph}]}{[\text{dph} \cdot \text{E}]} = \frac{([\text{E}^0] - [\text{dph} \cdot \text{E}])[\text{dph}]}{[\text{dph} \cdot \text{E}]} \tag{4}$$

The rate expression is then given by Eq. 5

$$\upsilon_{\text{cat}} = k_{c1}[\text{dne}][\text{dph} \cdot \text{E}] = \frac{k_{c1}[\text{E}^0][\text{dne}][\text{dph}]}{K_M + [\text{dph}]} \tag{5}$$
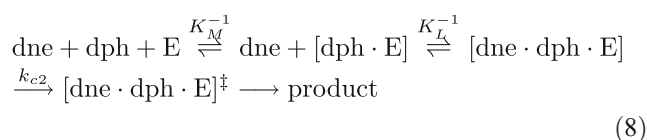
Dividing Eq. 5 by Eq. 2, we obtain the ratio

$$\frac{\upsilon_{cat}}{\upsilon_u} = \frac{k_{c1}}{k_u} \frac{[\text{E}^0]}{K_M + [\text{dph}]} \tag{6}$$

We see that according to Eq. 6, the catalytic effect is maximized by using high enzyme and low dienophile concentrations. Moreover, the relative rate is independent of diene concentration, and as long as $K_M \ll [\text{dph}]$, the value of K_M has little influence on the relative rate. It should be noted that this model simplifies the comparison between the catalyzed and the uncatalyzed reaction. The entropic cost of bringing the diene to the dienophile is similar for the two reactions, and the solvation/desolvation effects are also similar. Thus, the relative rate can easily be estimated from two sets of quantum chemical calculations; one on the uncatalyzed reaction and one on the catalyzed reaction with a simple model of the active site.
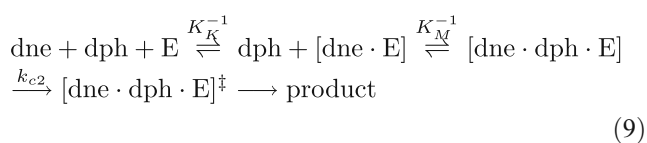
$$\frac{k_{c1}}{k_u} = e^{-\left(\Delta G_{c1}^{\ddagger} - \Delta G_u^{\ddagger}\right)/RT} \approx e^{-\left(\Delta E_{c1}^{\ddagger} - \Delta E_u^{\ddagger}\right)/RT} \tag{7}$$

Equation 6 describes well the kinetics we have for a reaction catalyzed by an organic catalyst without a binding pocket. It is also expected to work well for enzymes when we have a nonpolar diene and a nonpolar solvent, i.e., when there is no binding to the active site due to hydrophobic or electrostatic effects. Clearly for this model, small organic catalysts are more favored since they allow for higher catalyst concentration.

In an enzyme it is also possible to have a specific binding of the diene to the active site. This can lead to increased reaction rates, since the entropic cost of bringing the reactants together is compensated for by the attractive binding of diene. In this case the model have to be refined according to Eq. 8.

$$\text{dne} + \text{dph} + \text{E} \overset{K_M^{-1}}{\rightleftharpoons} \text{dne} + [\text{dph} \cdot \text{E}] \overset{K_L^{-1}}{\rightleftharpoons} [\text{dne} \cdot \text{dph} \cdot \text{E}]$$
$$\xrightarrow{k_{c2}} [\text{dne} \cdot \text{dph} \cdot \text{E}]^{\ddagger} \longrightarrow \text{product} \tag{8}$$

By neglecting any bias on the order of substrate binding, one can equvivalently write

$$\text{dne} + \text{dph} + \text{E} \overset{K_K^{-1}}{\rightleftharpoons} \text{dph} + [\text{dne} \cdot \text{E}] \overset{K_M^{-1}}{\rightleftharpoons} [\text{dne} \cdot \text{dph} \cdot \text{E}]$$
$$\overset{k_{c2}}{\longrightarrow} [\text{dne} \cdot \text{dph} \cdot \text{E}]^{\ddagger} \longrightarrow \text{product} \tag{9}$$

The rate expression is in this case given by Eq. 10 [41, 42].

$$\upsilon_{cat} = \frac{k_{c2}[E^0][\text{dne}][\text{dph}]}{K_M K_L + K_M[\text{dne}] + K_L[\text{dph}] + [\text{dne}][\text{dph}]} \tag{10}$$

The rate relative the uncatalyzed reaction is given by:

$$\frac{\upsilon_{cat}}{\upsilon_u} = \frac{k_{c2}}{k_u} \frac{[E^0]}{K_M K_L + K_M[\text{dne}] + K_L[\text{dph}] + [\text{dne}][\text{dph}]} \tag{11}$$

The difference between k_{c1} and k_{c2} is the reference state. By comparing Eqs. 3, 8 and 9, one realizes that

$$k_{c2} = k_{c1} K_L. \tag{12}$$

Inserting this expression into Eq. 11 yields Eq. 13.

$$\frac{\upsilon_{cat}}{\upsilon_u} = \frac{k_{c1}}{k_u} \frac{[E^0]}{K_M + [\text{dph}] + \frac{K_M[\text{dne}]}{K_L} + \frac{[\text{dne}][\text{dph}]}{K_L}} \tag{13}$$

We note that for large K_L, i.e., no specific binding of the diene by the enzyme, Eq. 13 reduces to Eq. 6. However, it is important to note that the value of k_{c1} is not independent of K_L. When K_L is low there are interactions in the active site that stabilize the diene relative its free solvated state. These interactions are also likely to be present at the transition state, and lower its energy relative the free state. Thus, the approach to calculate $\frac{k_{c1}}{k_u}$ from quantum chemical calculations using a simple model of the active site is no longer applicable. Figure 2 illustrates the relation between the various states discussed in the text, as well as the difference between k_{c1} and k_{c2}.

The challenge now lies in accurately calculating the (relative) free energies of the relevant states. The free energy of activation for the uncatalyzed reaction can be calculated directly with quantum chemical methods by correcting the energies for solvation effects and thermal effects, using implicit solvent models and independent particle statistical mechanics, respectively. Binding constants between substrate and enzyme can be calculated from MD trajectories, using classical methods based on statistical sampling, e.g., MM/PB(GB)SA [43] or linear interaction energy (LIE) [44, 45]. The main problem is computing the relative free energy of the enzyme-catalyzed TS. In principle this can be accomplished by free energy integration methods and a QM/MM description of the enzyme system. However, these methods are very computationally demanding due to the need for extensive sampling of the configurational space, and there are still technical problems associated with the representation of the quantum-classical interface. Instead, we prefer to use a simple model of the oxyanion hole as a QC representation of the enzyme, and seek a common reference state along the reaction coordinate for the quantum and classical regimes. One such reference state is the 'near-attack conformer' (NAC) [46, 47].

The NAC can be viewed as a point, or an area, on the potential energy surface, located between the bound enzyme-substrate complex and the TS, where the geometry of the system is preorganized for forming the TS and the atoms of the bonds that are formed in the TS are within van der Waals distance. Since the geometry of the NAC is similar to the TS, the interactions between the active site and the substrate(s) are similar in the NAC and the TS, except for those that directly influence the bond breaking/forming. Thus it can be anticipated that quantum chemical calculations with even a small model of the active site can reproduce the free energy difference between the two states. Justification for the division of the potential energy surface also comes from the reaction force concept of Toro-Labbe and coworkers [48–50]. They have shown for a number of different reaction types that there is a narrow transition

Fig. 2 Reaction pathway showing the difference between the catalyzed and uncatalyzed reactions. ΔG_M^0 and ΔG_L^0 refer to the indexing in Eqs. 3, 4, 5, 6, 7, 8, 9, 10 and 11
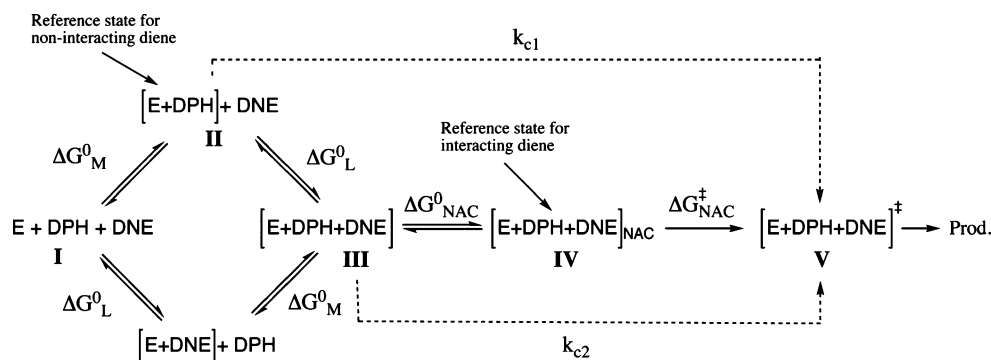
region on the potential energy surface (PES) where the main changes in electronic properties take place, and where the system goes from activated reactants to activated products. Structural changes not associated with bond breaking/formation and environmental effects, such as solvation, mainly inuence the PES before and after the transition region.

The free energy of activation going from the enzyme substrate complex to the TS is then given by Eq. 14.

$$\Delta G_{c2}^{\ddagger} = \Delta G_{NAC} + \Delta G_{NAC}^{\ddagger},\qquad(14)$$

where ΔG_{NAC} is the free energy of forming a NAC, and $\Delta G_{NAC}^{\ddagger}$ is the free energy of activation from the NAC. MD simulations provide information of the NAC frequency, which is obtained by statistical analysis of the geometries in the trajectory. Analogous to transition state theory, one can regard the relationship between NACs and non-NACs as an equilibrium. Thus, the relative population of NACs from a set of trajectory points gives an estimate of ΔG_{NAC}.

If the diene binds tightly to the active site and in a geometry that is close to the NAC geometry, ΔG_{NAC} will be low; typically for a natural enzyme binding a natural substrate, ΔG_{NAC} is in the order of 0.5–2 kcal·mol^{-1} [46, 47]. However, if the substrate binds loosely to the active site and retains substantial freedom of moment, ΔG_{NAC} will be higher. Even higher values of ΔG_{NAC} are expected if the substrate binds tightly but in a nonproductive fashion, i.e., in a geometry that is very different from the NAC. In the case that ΔG_{NAC} is high, its determination by unconstrained MD simulations also becomes less accurate since only a very small number of configurations will be NACs and the sampling will be insuficient. This is particularly true if stringent criteria for NAC geometries are enforced.

On the basis of ΔG_{NAC} obtained from the MD simulation and $\Delta G_{NAC}^{\ddagger}$ from the quantum chemical calculation it is possible to estimate $\Delta G_{c2}^{\ddagger}$. Combined with the free energy estimate of the uncatalyzed reaction, k_{c2}/k_u can be calculated (c.f. Fig. 2):

$$\frac{k_{c2}}{k_u} = e^{-\left(\Delta G_{c2}^{\ddagger} - \Delta G_{u}^{\ddagger}\right)/RT}\qquad(15)$$

Thus, by a combination of MD simulation and quantum chemical calculations it is possible to estimate the catalytic performance for different substrates and enzyme combinations using Eq. 11. There is, unfortunately, substantial uncertainty involved in the determination of the different energy terms, which makes it hard to obtain accurate global energy estimates. However, relative rates are likely to be of better precision. In particular, for ranking the performance of different enzyme variants, e.g., mutations, the ΔG_{NAC}

can be of great importance. For this purpose less stringent NAC criteria can be employed, which improves the statistical significance of the results. A first estimate of the quality of a given enzyme variant or substrate is however most easily obtained by visual and statistical analysis. The ability to form NACs has for this purpose been used as an important indicator of the viability of a given enzyme variant.

Enzyme variants

The CALB model was obtained from the PDB structure 1LBT [39]. The entry was chosen due to the presence of a large lipid analog bound to the oxyanion hole, which putatively induces a fully active conformation of the enzyme. The all-atom RMSD to two other common structures in modeling is small; 1TCA [51]=0.29, 1LBS [39]=0.44. The 1LBT structure was stripped of all water molecules and protonated (default protonation state) in the viewer Hermes, a part of the GOLD Suite [52]. This structure was subsequently used as the basis for mutant models, which were generated in SwissPDB viewer [53].

Molecular docking

Selection of dienophiles and generation of initial studies were accomplished by docking simulations, using the GOLD Suite [54–57]. The docking procedure was conducted according to the following protocol: The built-in genetic algorithm (GA) was used to generate different poses, which were evaluated with the ChemScore scoring function [58, 59]. The space of possible poses was limited to relevant ones by inducing hydrogen bond constraints between the dienophile carbonyl oxygen and T40-NH, T40 − O$^{\gamma}$H and Q106-NH. To somewhat counteract the limitations created by these constraints, the 'diverse solution' ag was switched on, using default parameters. 10^5 GA runs were used for each docking pose, and 50 poses were generated for each molecule. Statistics were calculated on the top 10 of these solutions, based on total score. The poses were further analyzed according to a composite scoring function, which is explained in detail in the Electronic Supplementary Material (ESM). After numerical and visual analysis, the best poses of each dienophile were saved for docking of the diene. The same set of parameters were used as for docking the dienophiles, except now, no constraints were used to limit the possible poses.

Docking in the MD-relaxed variants was performed analogously. Since the average structures (Enzyme variants) resulted in unphysical positions of a large portion of hydrogens, all hydrogens were stripped of and the enzymes were reprotonated in GOLD. The H-bond constraints were limited to T40-NH and Q106-NH since the T40 − O$^{\gamma}$H was

seen to contribute less to binding of the substrates, and turned away somewhat during relaxation.

Molecular dynamics simulations

All MD simulation were performed with the AMBER 9 & 10 packages [60], and the AMBER 99 force field [61] was employed throughout. All initial structures were taken from molecular docking. Force field parameters for the nonprotein atoms (dienophiles and diene) were generated from the general Amber force field (gaff) [62], using the Antechamber program. The generation of gaff-parameters was based on AM1 geometries, and the AM1-BCC approach was used for partial charges [63, 64]. All amino acid residues were assumed to be in their natural protonation state.

The LEaP program was used to create topology files, neutralize the enzyme complexes with sodium ions, and add an octahedral box of solvent in a minimum layer of 8 Å. Acetonitrile (ACN) was treated by the available united-atom model [65], while chloroform (CHL) [66] and water (WAT, described by TIP3P) [67] parameters were taken from all-atom models.

Simulations were performed under periodic boundary conditions and the long-range electrostatics was handled by particle Mesh Ewald summation. Initial minimization was performed in two steps of 1000 and 2500 iterations, respectively, where the protein was held fixed during the first step. Next, the systems were heated to 300 K over 20 ps with mild restraints on the protein atoms (10 kcal·mol^{-1}·Å). Unconstrained simulations followed for 2–4 ns at 300 K and 1.0 bar. Two fs time steps were used, both during warmup and the following simulations, and hydrogen movements were constrained by the SHAKE algorithm [68].

Trajectories were based on snapshots taken every 0.2 ps. Statistics were obtained using the PTRAJ tool within AMBER. For structural analysis, an average structure was generated based on the last 1000 snapshots (200 ps). In these structures, hydrogen positions were poorly defined, so when performing subsequent docking, all hydrogens were stripped of and the relaxed enzymes were reprotonated in GOLD.

For simulations of substrates 1 and 3 in water (see Further optimization and Binding constants), the binding free energies of both diene and dienophile were estimated using the LIE method [44, 45]. LIE approximates ΔG_{bind} by an electrostatic contribution V^{elec} (Coulomb interactions) and a non-polar (van der Waals) contribution V^{vdw}. These contributions are weighted by coefficents α and β, calibrated to free energy perturbation (FEP) data and empirical data, respectively.

$$\Delta G_{bind} = \alpha \left[\left\langle V^{elec}_{l-s} \right\rangle_{bound} - \left\langle V^{elec}_{l-s} \right\rangle_{free} \right]$$
$$+ \beta \left[\left\langle V^{vdw}_{l-s} \right\rangle_{bound} - \left\langle V^{vdw}_{l-s} \right\rangle_{free} \right], \quad (16)$$

where α has been determined to 0.43 and β to 0.18 for the neutral molecules used here.[69] 300 samples were taken from the last half of the 2 ns trajectories. The LIE method requires explicit simulations of the free solvated substrates; they were thus simulated for 500 ps, from which 200 samples were used. Both average values and standard deviation were stable with respect to a 5-fold increase in sampling rate.

It should be noted that we also attempted to calculate binding free energies using the MM/PBSA and MM/GBSA approaches, but they lead to more divergent results and irrational magnitudes.

DFT calculations

A model of the oxyanion hole was created from the parent crystal structure 1LBT, using the docked structures as starting guesses for the dienophile and TS. The central peptide atoms of the backbone Q106 and T40 nitrogens were included together with adjacent α-carbons, in order to correctly model the electrostatics of the hydrogen bonds. The T40 side chain was included as well a methyl group on residue 105, representing both serine and alanine. Finally, the entire Q106 side chain was included because it is seen to H-bond to the back of theT40 OH-group in CALB crystal structures. A graphical representation of the oxyanion hole model is provided in ESM Fig. 5.

Geometry optimizations were performed at the B3-LYP/6-31G(d) level [70–73], for the [E·dph], NAC-[E·dph·dne] (the $C^4_{dne} - C^\alpha_{dph}$ and $C^1_{dne} - C^\beta_{dph}$ distances being fixed at 4.0 Å, see the discussion in MD simulations of substrates) and [E·dph·dne] states. Gaussian 03 was used for all calculations [74]. An endo conformation of the TS was chosen, since it is generally accepted to be the kinetically favored geometry [75]. Because of the truncated oxyanion hole model, several atoms were kept frozen throughout the optimization, as illustrated in ESM Fig. 5. As this generally leads to more than one negative force constant in the calculated Hessian, no thermodynamic corrections were made on the oxyanion hole structures. Where stated, estimates of the corrections were instead made by using the Hessian from the corresponding uncatalyzed reactions. This is a reasonable estimate since eigenstates of the enzyme are expected to change little during the course of reaction. The transition states were detected by identifying the most negative imaginary frequency as being the formation of the two incipient C-C bonds.

Single point calculations on the optimized structures were performed using a 6–31G++(d,p) basis set and the C-PCM solvent model [76, 77]. We used a dielectric constant of $\varepsilon = 4$ to simulate the polarity of a typical enzymatic environment, and the shape of the cavity was determined by the united atom topological model [78].

Graphical representation

All molecular representations were created using the UCSF Chimera package [79, 80]. Diagrams were created using XMGrace [81] and Plot [82].

Results and discussion

Enzyme variants

To test for relative stability and look for unexpected distortions of the enzyme variants, 4 ns MD simulations were carried out on each variant in the three solvents. In all simulations, the backbone RMS trajectories level out after 1 ns, indicating that 4 ns is enough simulation time for this study.

For structural comparison, average structures of the last 200 ps of simulations were abstracted. Their backbone RMS deviations (RMSDs) with respect to the crystal structure are shown in ESM Table 1. No major structural change occurs in any solvent, compared to that of the wild type, which suggests that all mutants are potentially stable. As expected, the water simulations display the lowest RMSDs with respect to the crystal structure, while the CHL-relaxed variants show the highest RMSD values.

An extensive analysis of solvent and mutant influences on the structure is given in the ESM. Here, we summarize by noting that the essential catalytic machinery is well-preserved, but that the T40 side-chain moves away from its crystal position where the OH-group stabilizes the oxy-anion. A question thus arises concerning its role in catalysis, which will be addressed below. The S105 residue, where present, tends to move into the designated oxyanion hole region, an observation that advocates mutation. Finally, the I189A and I285A mutations did not generate any major structural distortions, and should thus be viable.

ACN was chosen for further MD studies of enzyme-substrate interactions, since it does not significantly distort the cavity, while it is a moderately polar solvent, which is believed to improve the affinity between the substrates and the hydrophobic active site. Moreover, using an organic solvent is of interest in synthesis, and (immobilized) CALB is known to operate well in dry environments.

Molecular docking

Among a larger set of molecules, benzoquinone **1**, maleic anhydride **2** and cyclohexenone **3** (Fig. 3), all three common reactants for Diels-Alder, were selected for further study. One strong motivation for this choice is that all dienophiles are cyclic, thus locked in an s-*trans* conformation. Although s-*cis* is widely considered to be a more activated conformation [75], a trans dienophile has the β-carbon further away from
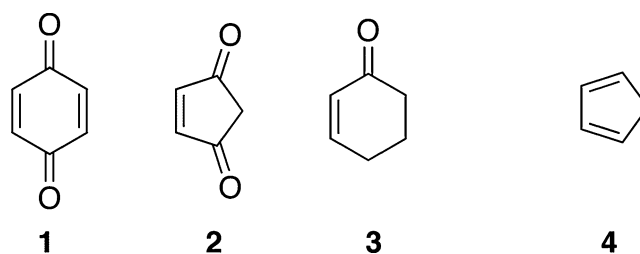
Fig. 3 Structure of Diels-Alder substrates used in this study

the oxyanion hole when bonded, which might be necessary in the crowded active site. In addition, rings are more facile to model since the conformational space is limited.

From each dienophile docking, one pose was selected onto which the diene (cyclopentadiene **4**) was subsequently docked, using the same set of parameters as for the dienophile runs. **4** was chosen due to it's frequent use in Diels Alder in the literature, and again the fact that the cyclic structure limits its mobility and facilitates simulation. The best [E·dph·dne] complex for each variant and dienophile was then subjected to MD simulations, together with the best [E·dph] complexes.

MD simulations of substrates

Both the [E·dph] and [E·dph·dne] complexes were simulated for 2 ns each, since the relaxation of the uncomplexed variants indicated that the RMSD leveled off after 1 ns. From the simulations, trajectories of backbone RMSDs, H-bond distances between the oxyanion hole and carbonyl oxygen, and B-factors were calculated. We note in passing that the atomic fluctuations are similar to those of the uncomplexed variants in ACN, but that there is no clear trend regarding how the exibility is affected by different point mutations, similar to the uncomplexed simulations (ESM Fig. 7). Nor is there any significant difference between the exibility pattern for complexes of different substrates. RMS deviations over the last 0.2 ns of simulation are summarized in Table 1. Distance trajectories for oxyanion hole hydrogens and dienophile oxygen are presented in Fig. 4.

The occurrence of H-bonding between dienophile and oxyanion hole was evaluated using a threshold distance of 2.75 Å between the hydrogen atom and the acceptor, and calculating the fraction of saved geometries fulfilling this threshold for each H-bond donor (Q106-NH, T40-NH and T40 − O^γH). To avoid bias originating from the unrelaxed starting guess, only the last 1 ns (5000 frames) of each simulation was used. The fractions for each variant and dienophile are presented in Fig. 5. One can see, both from Fig. 4 and Fig. 5, that in some cases, H-bonding of the dienophile is slightly induced by the introduction of a diene, which effectively blocks the escape route. But overall, the pattern from the [E·dph] simulations is consistent.

Table 1 Backbone RMS deviations (Å) for substrate complexes compared to corresponding uncomplexed variant in ACN[a]

Variant	1	1+4	2	2+4	3	3+4
WT	0.83 (0.93)	0.94 (0.75)	0.76 (0.85)	0.80 (1.13)	0.90 (0.80)	0.99 (0.75)
S105A	1.00 (0.87)	1.07 (0.88)	0.94 (1.01)	1.06 (1.01)	0.98 (0.86)	1.06 (1.04)
I189A	0.85 (1.04)	0.75 (1.03)	0.92 (0.97)	1.01 (0.92)	0.92 (1.06)	1.07 (0.97)
I285A	0.89 (0.92)	0.76 (0.76)	1.34 (1.41)	0.71 (0.99)	0.84 (1.06)	1.13 (1.31)
S105A/I189A	1.05 (1.15)	0.98 (1.09)	1.16 (1.09)	1.04 (0.86)	1.41 (1.39)	1.05 (1.04)
S105A/I285A	1.18 (1.04)	1.16 (0.92)	1.24 (0.90)	1.16 (1.19)	1.24 (0.77)	1.13 (0.88)

a Based on average structures from the last 0.2 ns of simulation (1000 frames). Numbers in parentheses refer to RMS deviations with the 1LBT crystal structure.

From the H-bond trajectories in Fig. 4, one clearly sees that the two most important donors are the backbone nitrogens of T40 and Q106. $T40 - O^\gamma H$ participates only sparsely. The available crystal structures of CAL-B [39, 51] reveal that T40 accepts H-bonds from the amide side chains of Q106 and Q157 (Fig. 1), and at least one is maintained in the simulations, accompanied by a slight distortion of the backbone, also observed in the uncomplexed variants (ESM Fig. 2). Thus, it is questionable that T40-OH participates in the oxyanion hole

during substrate recognition. This supports the previous observation that only two donors are important in the oxyanion hole at any one time, even in enzymes having donors in the sidechains [20, 21]. However, as seen in the 1LBS crystal structure [39], the covalent intermediate (and possibly the TS) binds to all three H-bond donors, perhaps by some geometric distortion induced by the tetrahedral structure.

To evaluate the frequency of NACs between diene and dienophile, we used a similar procedure as given above for
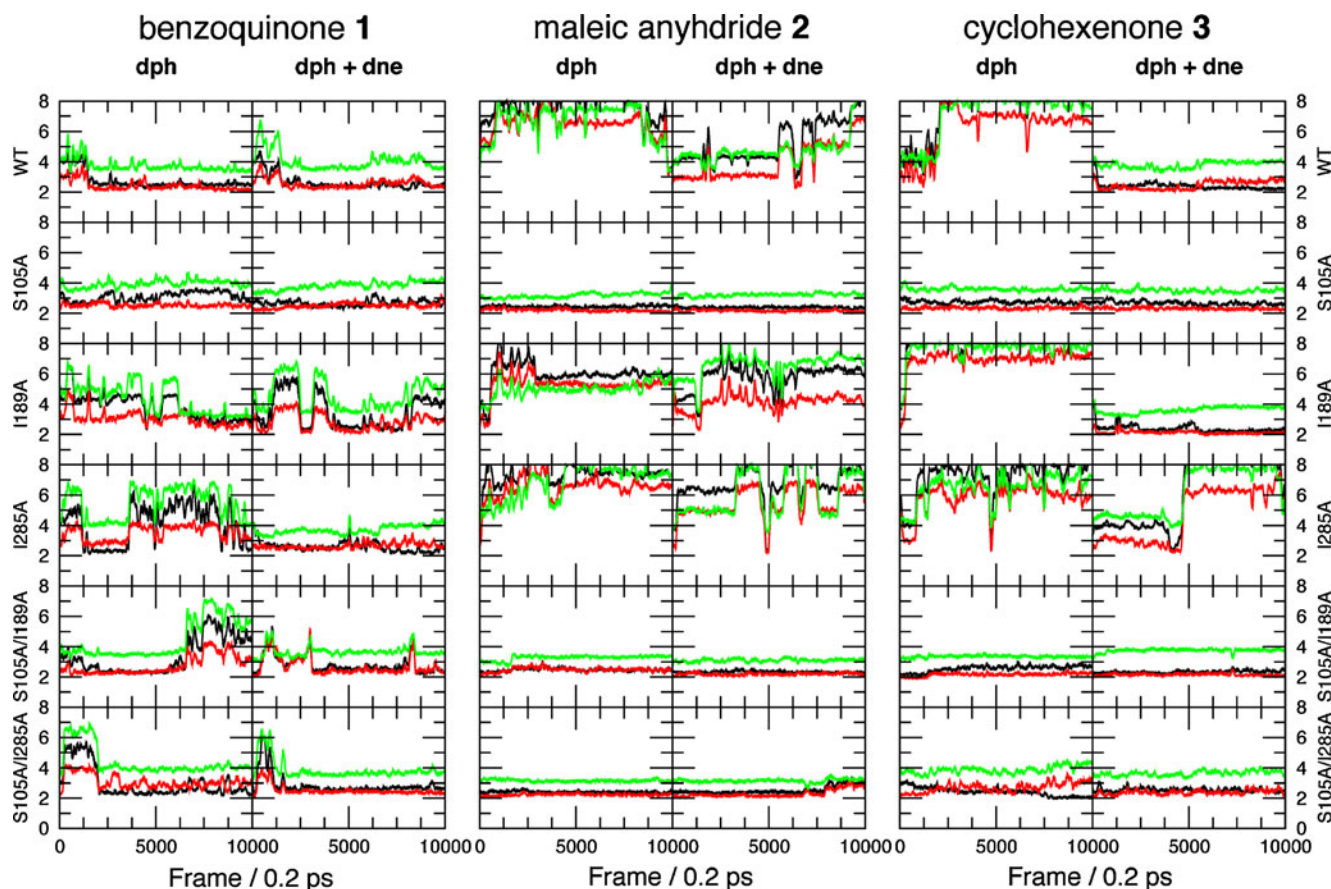
Fig. 4 H-bond trajectories from the 2 ns simulations of all substrates in the six variants. Distances are between Q106-NH (black), T40-NH (red) and $T40 - O^\gamma H$ (green), and acceptor oxygen without and with cyclopentadiene 4 present. Distances are in Angstroms and each trajectory is presented as a 100-frame running average, for clarity, meaning that extreme values are cut off
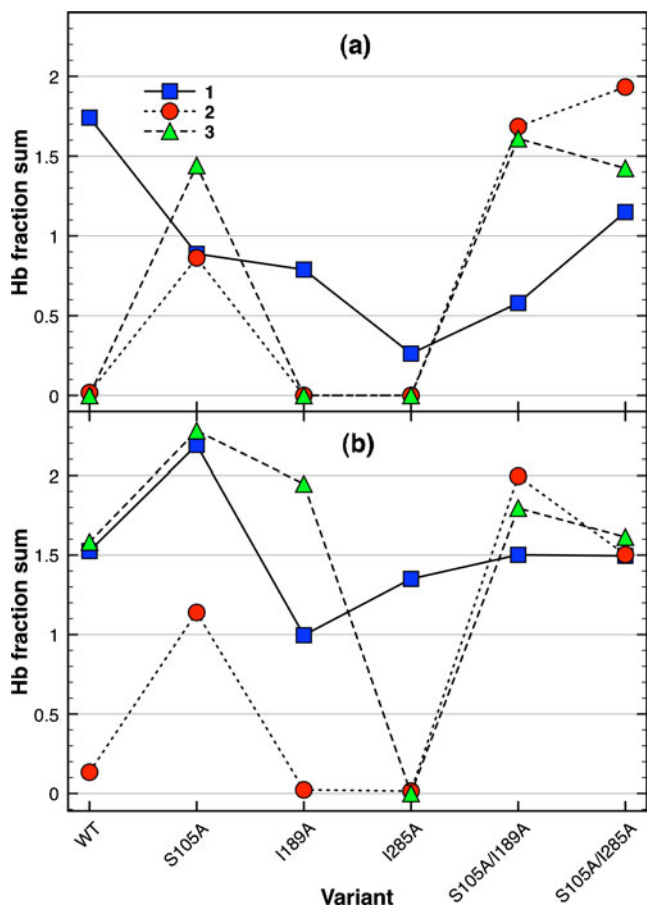
Fig. 5 Sum of fraction of the three putative H-bonds between oxyanion hole and substrate acceptor (Q106-NH, T40-NH and T40 − O$^{\gamma}$H, threshold 2.75 A), calculated on the last half of each simulation. (**a**) Dienophile complex, (**b**) dienophile+diene complex

H-bonds. We have used a slightly less rigorous deffinition than the original one stated by Bruice and colleagues [46, 47], as this allows us to reduce the sampling of data. Most important for our model to be valid is that the NAC deffinition is identical in both the MD and QC regime (see Kinetics and binding dynamics). The 'loose NAC' is thus defined as a structure where both the dne−C^1−dph−C$^{\alpha}$ and dne−C^1−dph−C$^{\beta}$ distances are ≤4.0 Å simultaneously. As with the H-bond evaluation, only the last 5000 frames (1 ns) are taken into account. Figure 6 shows the fractions of structures fulfilling these criteria, accompanied by the corresponding fractions with only one incipient bond being within 4.0 Å. Note that for 1, which has two identical α,β-double bonds, data is presented for the one with the highest number of frames fulfilling the NAC criterion. It should be noted that angular parameters are only evaluated visually, therefore the calculated NAC frequencies are overestimations and should be taken as upper limits. With these loose NAC criteria, we obtain up to 1–4 % NAC-structures with **1**, corresponding to a ΔG_{NAC} of ~2.8–1.9 kcal/mol. Even though seldom in close association with the dienophile, the diene is clearly attracted by the hydrophobic cavity, which is expected to compensate for the entropic penalty of bringing the two reactants together. Following the distance trajectories (not shown), it is apparent that the diene occasionally diffuses away before termination. This happens in the [**1**·S105A], [**1**·I189A], [**1**·S105A/I189A], [**2**·WT], [**2**·S105A], [**2**·I285A], [**2**·S105A/I189A], [**3**·I285A] and [**3**·S105A/I285A] complexes. It is difficult to tell whether this is a stochastic event or a direct effect of the particular substrate-enzyme combination. The relatively high fraction of NACs in **1** is partly due to the molecular

Fig. 6 Simultaneous incipient bond NAC fractions (threshold 4.0 Å). The inset shows corresponding fraction with only one of two distances fulfilled. Results from additional simulations of **1** and **3** in S105A/I189A (see Furter optimization) are shown in gray
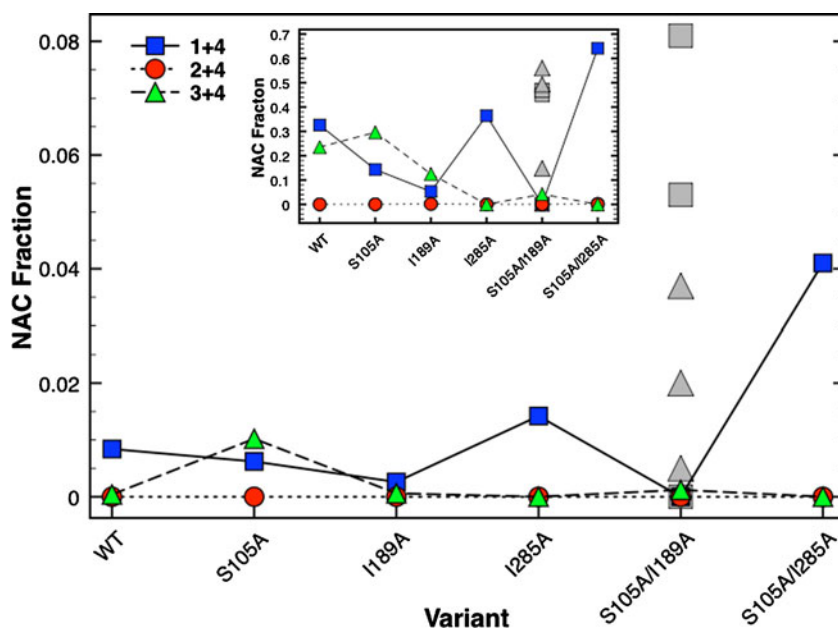
symmetry; benzoquinone has two identical double bonds that could be subject to attack. A more detailed discussion for each dienophile follows below.

Benzoquinone Statistically, benzoquinone **1** is the dienophile that appears to be best suited for Diels-Alder in a CALB variant. However, no simulation of **1** + **4** display any real 'sandwich'-like conformation between the substrates, which is essential for the appropriate orbital overlap. One reason is that **1** is a long molecule that clashes with residues surrounding the catalytic triad, especially I189, hence it cannot assume a productive conformation, which exposes the at face to the solvent, as seen in Fig. 7a. We therefore expected the situation to improve when introducing the I189A mutation, but were surprised to see that the the NAC frequencies in I189A and S105A/I189A were in fact lower than in all other variants. From visual analysis, it appears that 1 tends to fall out of the active site in absence of I189, possibly due to lack of confinement. On the other hand, it is clear that I189 hinders a good sandwichlike NAC conformation in the other variants, including S105A/I285A, which at first glance seems to fulfill the loose NAC criteria as defined (see Fig. 7b). But because I285 is on average further away from the oxyanion hole, we deem it less inuential on forming activated complexes.

Maleic anhydride Maleic anhydride **2** clearly needs the S105A mutation to bind to the oxyanion hole (Fig. 4). Interactions with the diene is poor in all variants, as illustrated by virtually no detected NACs (Fig. 6). The diene remains reasonably close to **2** in the two double mutants, but the position of both diene and dienophile is less consistent

than for the other substrates; in S105A/I285A, the diene resides in the small subpocket, close to W104 (Fig. 8a), while it prefers the large subpocket in S105A/I189A (which is the common region in most simulations), as displayed in Fig. 8b. **2** also tends to tilt more deeply into the cavity, which effectively blocks the α,β-double bond; see Fig. 8a and b. This could be due to undesired interaction with the second, equivalent carbonyl group.

Cyclohexenone Like **2**, cyclohexenone **3** is more tightly bound to the oxyanion hole with the S105A mutation present (Fig. 4). The double bond generally faces the large subpocket, where the diene resides in all simulations except the S105A/I285A variant, where it escapes quickly. From Fig. 6, we see that the NAC frequencies are not very impressive, and that the 1-bond NAC frequency tends to decrease with mutations of the bulky I189 and I285 residues. The latter observation can be a confinement effect; visual analysis reveals that S105A/I189A produces the desired sandwich-type conformation of substrates, albeit with a low NAC frequency. The S105A mutation alone generates T-shaped diene-dienophile interactions (Fig. 9a), similar to the few interactions produced with **2**. It is evident that with both I189 and I285 present, little space is available for a Diels-Alder TS. [3·4·S105A/I189A] is the one simulation where at least one conformation fulfills the more rigorous NAC criterion that the two incipient bonds should be within 3.5 Å, while also maintaining a reasonable NAC geometry (Fig. 9b).

Even when conditions are seemingly optimized for Diels-Alder, such as in S105A/I189A, the C^α of the dienophiles is obscured by the loop containing T40, in particular the T40 side chain. This is a further reason why
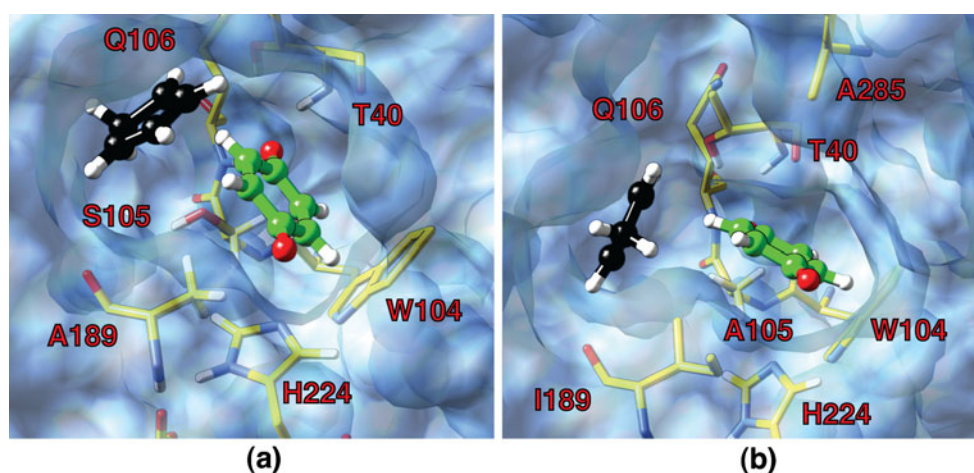
Fig. 7 Selected snapshots from MD simulations of 1 in (**a**) I189A and (**b**) S105A/I285A, showing selected residues and hydrogens. Both illustrations show typical T-shaped interactions. The reactants come close, but not typically in the sandwich-like conformation required for TS formation. In (**a**) it is shown that S105 has as much space in the immediate vicinity of the oxyanion hole, forcing 1 to tilt away and bind less strongly. (**b**) shows that the l285A mutation does not create as much space in the immediate vicinity of the oxyanion hole. There is no increase in sandwich-like conformations relative to the S105A mutant

Fig. 8 Selected snapshots from MD simulations of **2** in (**a**) S105A/I285A and (**b**) S105A/I189A, showing selected residues and hydrogens. In both illustrations, it is shown how **2** tilts to the left, effectively making the α,β-double bond unavailable for attack. (**a**) shows a unusual diene position, made available because of the dienophile tilt. A more frequent position is shown in (**b**), but the buriedness of the dienophile hinders attack by the diene
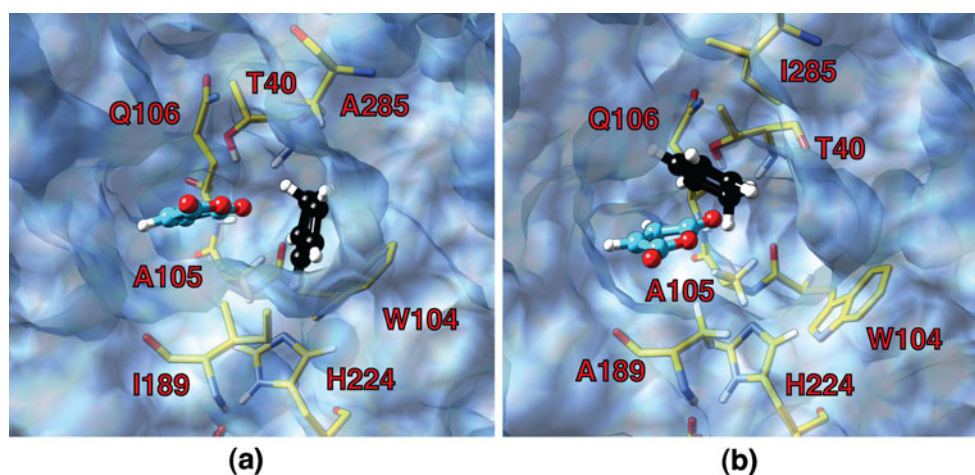
so few 2-bond NAC conformations are detected, although 1-bond proximitiy is quite common in many instances.

Further optimization

The two main purposes with each point mutation was to create more space for the dienophile to fit into the oxyanion hole without having to be covalently bound to the enzyme, and make possible the NAC geometry neccessary for Diels-Alder. S105 tends to block the oxyanion hole during a large part of all simulations; consequently, as Fig. 5a shows, the best H-bonding is obtained with the variants containing the S105A mutation. This is further evidenced by the H-bond trajectories. The fact that **1** breaks the pattern somewhat could be due to favorable interactions of the conjugated molecule with, e.g., W104 and some of the hydrophobic residues.

The I189A and I285A mutants were investigated since the large isobutyl side chains tend to clash with either the dienophile or the diene during molecular docking. In terms of H-bonding to the oxyanion hole, however, netiher mutant provided any major improvement. If anything, the dienophile recognition became worse, which

may have to do with unrealistic starting guesses. Since the molecules start within the active site, clashes with I189 or I285 may actually lock the substrate in position, while the smaller alanine replacement allows for escape. In the two double mutants, the H-bond frequency increases, but now without any steric clashes locking the substrates in the active site.

From the above discussion, we draw the conclusions that *i*) the S105A mutation is necessary to obtain acceptable binding between the oxyanion hole and the cyclic enones **1–3**; *ii*) the I189A mutation makes possible an adequate positioning of the diene for a pericyclic reaction; *iii*) T40 obstructs the diene from approaching the dienophile C^α, while the alcohol side chain is obsolete in the oxyanion hole; *iv*) **2**, although the most reactive of the dienophiles, fails to assume favorable geometries in the active site; *v*) **1** associates weakly with the oxyanion hole, possibly due to more delocalized charge, but has general good affinity to the diene, although it is to bulky to assume productive NACs in most variants; *vi*) **3** binds strongly to all S105A mutants and produces good NAC-like geometries in some variant, but at low frequencies. It is also a less reactive substrate.

Fig. 9 Selected snapshots from MD simulations of **3** in (**a**) S105A and (**b**) S105A/I189A, showing selected residues and hydrogens. In (**a**) it clearly shows that with both isoleucins present, there is insufficient space above and below the dienophile for TS formation. (**b**) shows an example of a good NAC conformation. The indicated incipient bond distances are: $C^4_{dne} - C^\alpha_{dph} = 3.79$ Å, $C^1_{dne} - C^\beta_{dph} = 4.15$ Å. Closer contact is hindered by T40
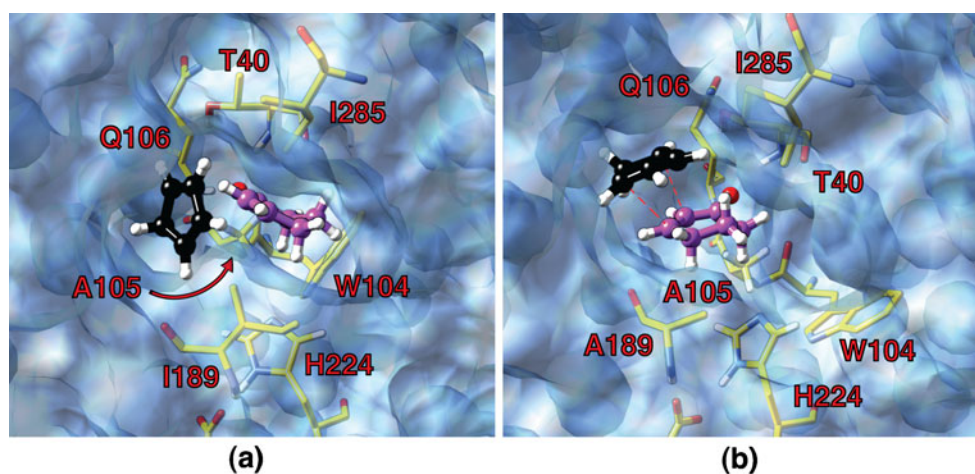
Additional support for the S105A and I189A mutants are found when comparing results of molecular docking in 4 ns relaxed variants (last 1000 frame average structures) with those in the unrelaxed crystal structure (as was used originally). When docking dienophiles in the unrelaxed variants, the results are quite invariant to mutations. In contrast, in ACN and WATrelaxed structures, S105A mutants give rise to higher scores, while the remaining three scored poorly. This is mainly due to clashes; the steric constraints in the relaxed variants are much higher, unless S105 is substituted (see Figure S1). Furthermore, a 'consensus' approach, utilizing dienophile, diene and TS composite scores,[1] shows that when considering the complete reaction pathway in the relaxed structures, S105A/I189A gives the highest score for both 1 and 3 (see Fig. 10).

Based on the above discussion, we therefore conducted the following additional investigations:

(A) *Multiple simulations of 1 and 3 with 4 in S105A/I189A.* One problem with our approach is that it is difficult to sample the conformational space of the diene in the cavity. We wanted to see if more NACs could be found in S105A/I189A by additional MD simulations, since post-relaxation docking so strongly indicated this variant to be favorable. Therefore, using three different starting geometries and velocities, three additional short (1 ns) simulations were performed on these two systems, using the same protocol as previously. The geometries were selected from the previous simulation, selecting ones where NAC-like interactions were present.

(B) *1 and 3 with 4 in S105A/I189A, using water as solvent.* To investigate if water, which is more polar than ACN, increases the affinity between the substrates and active site, one 2 ns simulation was executed for each of the [S105A/I189A·1], [S105/I189A·1·4], [S105A/I189A·3] and [S105A/I189A·3·4] complexes. The generated trajectories were also used to calculate LIE energies.

(C) *Simulation of a T40A/S105A/I189A variant in ACN.* Since a general obstacle for forming good NACs is the T40 sidechain, a triple mutant T40A/S105A/I189A was generated *in silico*, and relaxed in ACN for 4 ns, while 2 ns simulations of [T40A/S105A/I189A·3] and [T40A/S105A/I189A·3·4] were performed in parallel.

Results from (A) are presented in Table 2, giving H-bond data and NAC statistics. Protein RMSDs and exibilities were similar to the previous 2 ns simulations. The high NAC fractions, displayed in gray in Fig. 6, give further support to the S105A/I189A variant. Once again, we see evidence that the T40 alcohol is no participant in the oxyanion hole in our
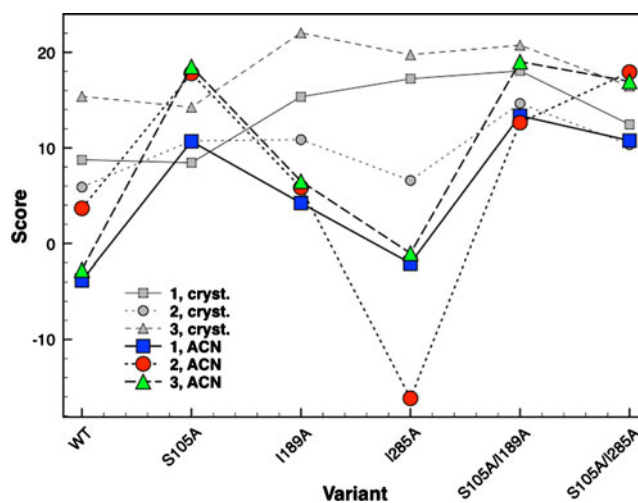
Fig. 10 Docking consensus score compared between the six variants and solvent-relaxed structures versus the crystal form. The consensus score is calculated according to the protocol given in the supplementary information, and includes contributions from individual dienophile, diene and TS poses

system. Intriguingly, the loose NAC geometries are now quite frequent in two of three simulations with each substrate. We also detect some 'tight' NAC poses, although they are still very scarce. With this set of simulations, we have shown that 1 can bind to the S105A/I189A and form NACs, and that the null result from the first setup was not generalizable, but that multiple simulations are preferable for sufficient sampling. However, 1 seems to be bonded more weakly to the oxyanion hole than 3, as implied by the average H-bond distances in Table 2. Visual analysis of the trajectories also reveals that 1 tumbles around in the active site more than 3, and occasionally falls out of the oxyanion hole for a short time. The best single simulation in this set is entry 2 in Table 2, in which a sandwich-like NAC is maintained almost constantly (see the supplementary information for an animation). From these data further multiple trajectories are indeed important for efficient sampling of the conformational space, which is a standard approach of, e.g., Pleiss and coworkers [83, 84].

The saturated framework of 3 makes it more difficult to form sandwich-like interactions with 4 than for 1. There is a smaller volume available for the diene to move in to form NACs with 3, and this volume is often occupied by surrounding residues, most dominantly I285, T40 and Val140. The prevalent interaction is a T-shaped geometry, with 4 on the double bond side of 3. Still, 4 displays good affinity with the active site and dienophile.

Results from the water simulations (B) are given in Table 3. We note that while 3 sticks tightly to the oxyanion hole, 1 tumbles around slightly, as observed in ACN. Hence the high average H-bond distances. We see a 14-fold increase in NAC fractions for 3, compared to the initial

---

[1] See the ESM for a detailed description of scoring and consensus

Table 2 H-bond and NAC statistics for the three 1 ns simulations of **1** and **3** with **4** in S105A/I189A[a]

Entry		dph H-bond $\bar{d}$ [Å][b]			NAC fractions[c]	
		Q88-NH	T40-NH	T40 − O$^\gamma$H	Loose[d]	Tight[e]
1	1, I	2.87 (0.61)	3.01 (0.61)	3.47 (0.07)	0.053	0.0016
2	1, II	2.37 (0.91)	2.30 (0.91)	3.62 (<0.01)	0.081	0.0022
3	1, III[f]	3.05 (0.35)	2.47 (0.81)	4.21 (0.01)	0	0
4	3, I	2.75 (0.52)	2.31 (0.91)	3.76 (<0.01)	0.037	0.0004
5	3, II	2.36 (0.87)	2.14 (0.98)	3.83 (<0.01)	0.020	0.0002
6	3, III	2.58 (0.75)	2.14 (0.98)	3.43 (<0.01)	0.005	0.0002

[a] 1 ns simulations, 5000 saved geometries. [b] Distances are averaged over all 5000 frames. Numbers in parentheses show the fraction of distances below a threshold value of 2.75 A. [c] Based on the two-bond criterium described in the text. [d] 4.0 Å threshold. e 3.5 Å threshold. [f] The diene left the cavity shortly after the beginning of the simulation, similar to the original 2 ns run.

simulation in ACN, and the NAC fraction for **1** is even higher. The relationship is more or less in concert with the findings in (A). More importantly, in both simulations, there is an event where the diene temporarily falls out of the active site but readily moves back in. In ACN, this does not occur; if the diene starts to diffuse out, it does not stop. Thus, a highly polar solvent is preferable when working with substrates such as those in the present study. The issue of Diels-Alder catalysis by water[75] is not explored further here, but clearly, there will be competition.

An extensive analysis of the T40A/S105A/I189A triple mutant (C) is given in the ESM. Here it suffices to say that while H-bonding of the dienophile was surprisingly improved, the formation of NACs with the diene was poor. Both phenomena are believed to arise from a relief of steric crowding around the oxyanion hole. This allows the dienophile to bind harder and become more buried, which in turn makes it less accessible to the diene. This finding highlights the importance of delicate tuning around the catalytic site in order to align both reactants.

Binding constants

We have estimated the binding constants (K_M and K_L) of substrates **1**, **3** and **4** to S105A/I189A in water using the LIE method, see Table 4. The results for the dienophiles are in agremeent with earlier experimental and theoretical studies on the binding of small ketones and aldehydes to CALB and other lipases; the binding constants in water are typically in the range of 0.1–1 M, and generally decrease with increasing hydrophobicity of the substrates [21, 85, 86]. Our lowest binding constant is obtained for cyclopentadiene **4**, which has no specific binding to the active site. This supports the interpretation that the main driving force for binding is hydrophobic interactions. In earlier study on the epoxidation of butenal in CALB wt and S105A we obtained experimentally a K_M value of 0.2 M in aqueous buffer, whereas no saturation was found in ACN up to concentrations of 3 M [21]. Thus, to obtain sufficiently low binding constants we conclude that nonpolar substrates and water as solvent should be used.

DFT calculations

Solvent-corrected B3LYP/6-31++G(d,p) relative energies are reported in Table 5, together with corresponding values for the uncatalyzed reaction. Energies for [E·dph] are also presented for reference. Figure 11 shows the [E·dph] and TS optimized geometries, including oxyanion hole H-bond distances.

Table 3 RMS, H-bond and NAC statistics for simulations of **1** and **3** with and without **4** in S105A/I189A, with WAT as solvent[a]

Entry	dph H-bond $\bar{d}$ [Å][b]				NAC fractions[c]	
	RMSD [Å][d]	Q88-NH	T40-NH	T40 − O$^\gamma$H	Loose[e]	Tight[f]
1	1.01 (0.91)	2.85 (0.39)	2.15 (0.98)	3.55 (0.00)		
1+4	0.95 (0.80)	3.11 (0.32)	3.13 (0.33)	3.19 (0.12)	0.085	0.0020
3	1.05 (1.03)	3.64 (0.01)	2.35 (0.88)	3.52 (0.03)		
3+4	0.83 (0.76)	2.35 (0.94)	2.16 (0.98)	3.40 (0.01)	0.017	0.0006

[a] 2 ns simulations, 10,000 saved geometries. [b] Distances are averaged over last 5000 frames. Numbers in parentheses show the fraction of distances below a threshold value of 2.75 Å. [c] Based on the two-bond criterium described in the text. [d] Backbone atoms, vs. the 4 ns relaxed structure (vs. crystal structure) [e] 4.0 Å threshold. [e] 3.5 Å threshold.

Table 4 Calculated estimates of binding parameters[a]

	I→II	II→III	III→IVc
	Benzoquinone 1		
$\Delta\Delta E_{el}$	2.6	0.6	-
$\Delta\Delta E_{vdW}$	-1.4	-1.2	-
ΔG	1.2	-0.7	1.8
const	$K_M=7.3$	$K_L=0.33$	$K_{N\,AC}=0.047$
	Cyclohexenone 3		
$\Delta\Delta E_{el}$	1.3	-0.5	-
$\Delta\Delta E_{vdW}$	-1.6	-1.0	-
ΔG	-0.3	-0.4	2.4
const.	$K_M=0.61$	$K_L=0.48$	$K_{N\,AC}=0.017$

[a] Roman numbers refer to Fig. 2. All energies are in kcal·mol⁻¹. [b] Electrostatic and van der Waals energies calculated by the LIE method, and scaled by 0.43 and 0.18, respectively. [c] Calculated from NAC averaged over all five [S105A/I189A·dph·dne] simulations in ACN and WAT.

It should be noted that because of the observation from MD that the T40-OH frequently points away from the oxyanion hole, we were led to investigate if it has any inuence on the stabilization of the TS. For this comparison, suitable snapshots were extracted from the S105A/I189A trajectories with **4** and **1**, **2** and **3**, respectively, and oxyanion hole models were generated as described in the methods section. It was found that the crystal structure version of the model consistently stabilized the TS more than the relaxed structure models, while the dienophile stabilization was more similar in both (slightly better in the crystal model for **2** and **3**, and slightly better in the relaxed model for **1**. This finding supports the above proposal that all three H-bond donors present in the CALB oxyanion hole contribute to TS stabilization, while the association with a carbonyl substrate does not necessarily involve T40-OH. We concluded that the crystal structure geometry was a better representation of the enzyme-TS complex, and therefore only present results obtained from this model.

We note that in the activated TS, $d_\beta < d_\alpha$ which is a general finding for this type of reaction. The d_β bond forms earlier along the reaction coordinate [87], which could somewhat mitigate the steric hindrance from T40. H-bond distances are generally smaller in the TS than in the [E·dph] complex, as seen in Fig. 11, which reflects the oxyanion hole preference for (partial) negative charge. It was somewhat surprising to see that the T40-OH distances are constantly among the shortest in the model, but this illustrates the sensitivity of enzyme-substrate bond dynamics. With the loop in the right position, T40-OH is a strong H-bond donor, while it is often excluded from the oxyanion hole by the enzyme motion. Moreover, neither dienophile nor TS assume the otherwise optimum planar geometry between carbonyl group and proton donors. This is in concert with the finding that oxyanion holes are optimized to give a minimum barrier between the TS and substrate complex, which occurs at a geometry which is not a minimum in either case [33].

With the [E·dph] complex taken as reference state (simple model), **1** is the dienophile that undergoes the largest TS stabilization $\left(\Delta G_{c1}^{\ddagger} - \Delta G_{u}^{\ddagger} = -3.9\,\text{kcal}\cdot\text{mol}^{-1}\right)$. The large $\Delta G_{u}^{\ddagger}$ of **3** is surprising considering this is a known Diels-Alder reactant. The relative rate constants (k_{c1}/k_u) of all three dienophiles (~10^1–10^2) correlate well to experimental rates for H-bonding molecular catalysts [35]. Moreover, they confirm the limitation of H-bond stabilization as the only means of rate enhancement; one cannot obtain more than a few kcal·mol⁻¹ in stabilization, and thus $\leq 10^3$ in k_{c1}/k_u.

Kinetics

Using the LIE computed K_M values and the k_{c1}/k_u values from the quantum chemical calculations we can estimate relative catalytic rate using the simplified kinetic model that assumes no binding of the diene (K_L). For substrates **1** and **3** the relative rate expressions are (inserting relative rate constants from Table 5):

$$\frac{v_{c1}}{v_u} = \frac{660\left[\text{E}^0\right]}{7.3 + [\text{dph}]}$$

$$\frac{v_{c1}}{v_u} = \frac{46\left[\text{E}^0\right]}{0.61 + [\text{dph}]}.$$

Table 5 B3LYP/6-31++G(d,p) energies in oxyanion hole model[a]

Dph	$d_\beta - d_a\,[\text{Å}]$[b]	$\Delta G_u^{\ddagger}$	$\Delta G_{c1}^{\ddagger}$[c]	$\frac{k_{c1}}{k_u}$	ΔG_{NAC}^d	$\Delta G_{NAC}^{\ddagger}$[e]	$\Delta G_{c2}^{\ddagger}$[f]	$\frac{k_{c2}}{k_u}$
1	0.52 (0.00)	28.1	24.2	660	1.8	13.2	15.1	$3.1\cdot10^9$
2	0.20 (0.00)	24.8	23.5	8.6	-	11.4	-	-
3	0.21 (0.28)	40.0	36.7	46	2.4	26.3	28.7	$3.0\cdot10^7$

[a] All energies are in kcal·mol⁻¹. Gibbs free energies are estimated using thermodynamic corrections added from the corresponding uncacatalyzed reaction. [b] Incipient bond distances in TS: $d_a = C_{dne}^4 - C_{dph}^a$, $d_\beta = C_{dne}^\beta - C_{dph}^\beta$. Values in parentheses refer to the corresponding uncatalyzed reaction. [c] Relative to the model [E·dph] complexes depicted n Fig. 11a–c. [d] NAC defined by $d_a = d_\beta = 4.0$ Å. ΔG_{NAC} obtained in Table 4. [e] Relative to QC NAC model. [f] From Eq. 14.
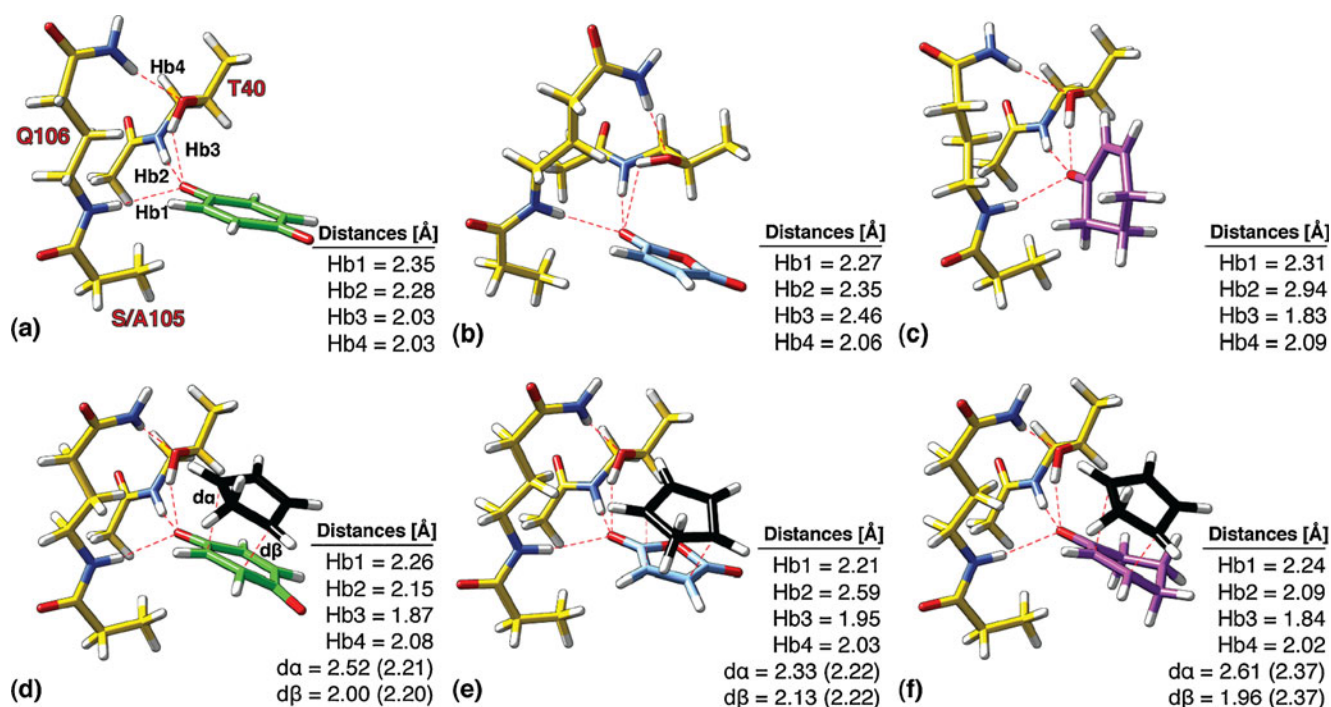
Fig. 11 B3LYP/6-31G(d) optimized geometries of oxyanion hole-bound dienophiles (**a-c**) and TSs (**d-f**) model. Distances within parentheses refer to the uncatalyzed TS

Inserting standard enzyme and substrate concentrations (10^{-5} M and 0.001-1 M, respectively) in the above expressions, it is clear that without binding of the diene, $v_{c1}/v_u < 0.001$, i.e., no catalytic effect can be expected. It is not realistic to raise the k_{c1}/k_u much above 660 by utilizing hydrogen bonding for catalysis. The corresponding v_{c1}/v_u values for the most efficient organic catalyst can be estimated to approximately 10–100 [35]. The dependence on the catalyst concentration clearly explains the superiority of small organic catalyst over enzymes for catalyzing the Diels-Alder reaction; organic catalysts can be, and typically are, used in concentrations >1 M.

The picture changes if we also consider productive binding of the diene. Using Eq. 11, and our calculated K_M, K_L and k_{c2}/k_u values, we obtain the following rate expressions for substrates 1 and 3, respectively.

$$\frac{v_{c1}}{v_u} = \frac{3.1 \cdot 10^9 \left[\mathrm{E}^0\right]}{2.4 + 7.3[\mathrm{dne}] + 0.33[\mathrm{dph}] + [\mathrm{dne}][\mathrm{dph}]}$$

$$\frac{v_{c2}}{v_u} = \frac{3.0 \cdot 10^7 \left[\mathrm{E}^0\right]}{0.29 + 0.61[\mathrm{dne}] + 0.48[\mathrm{dph}] + [\mathrm{dne}][\mathrm{dph}]}$$

The k_{c2}/k_u values are of similar magnitude to those we previously have obtained for Michael-type additions and direct epoxidation in wt and S105A CALB [21, 22]. Inserting the standard enzyme and substrate concentrations

we obtain rate enhancements of the order 10^3 for **1** and 10^2 for **3**, respectively. These values are competitive with conventional molecular catalysts, such as Schreiner-type thioureas [35], which give modest rate enhancements of about 10, despite that they are used at much higher concentrations. However, it should be emphasized that the rate enhancements are likely to be overestimated. In particular, the description of the NAC is generous and the ΔG_{NAC} values could be considered as lower limits. Moreover, the estimates come with a standard deviation of the same order as the average value. Note that to sample instances of $K_{NAC} < 10^{-3}$, we need an order of magnitude longer simulation time, which is beyond the scope of this study. There is also an uncertainty in the rate constants for the uncatalyzed reaction, where we have not considered hydrogen-bonding catalysis by the solvent explicitly. Still, the results are encouraging, and clearly demonstrate that with productive binding of both dienophile and diene to the active site it is possible to obtain rate enhancements of similar magnitude or better to those of molecular catalysts.

Summary and conclusion

Using standard steady-state kinetics and Eyring rate theory, we have derived expressions for calculating the rate-enhancements for catalysis of bimolecular reactions by enzymes or molecular catalysts. The input for the calcu-

lations is a combination of data from molecular dynamics simulations and quantum chemical calculations. MD simulations on solvated substrates and solvated enzyme-substrate systems provide means for obtaining substrate binding constants and free energies for forming NAC complexes. Quantum chemical calculations on model systems allows estimates of activation energies for the enzyme and uncatalyzed reactions.

Our analysis shows that the amount of transition state stabilization that can be obtained from hydrogen bonding is limited to a few kcal/mol, and it is difficult to optimize enzymes or substrates to enhance this effect further. It is not realistic to significantly surpass the hydrogen bonding effect that is obtained with organic catalysts, which have the great advantage of allowing for much higher catalyst concentrations due to their smaller size. Instead the key component to obtain effective catalysis in enzyme catalysts lies in substrate binding. Favorable interactions between the substrates and the active site can compensate for the entropic penalty of bringing the two substrates together. The binding constants for small organic molecules toward lipases are typically in the range of 0.1–1 M in aqueous buffer. Since this is similar to the typical substrate concentrations, even small improvements in binding can have significant effects on reaction rates. Hydrophobic interactions are the main driving force for binding, thus decreasing substrate polarity and increasing substrate size will improve the catalytic performance. The catalytic effect increases with solvent polarity, and the best effect is expected in aqueous solution. Another key parameter, related to substrate binding, is the free energy required to form near-attack conformers, ΔG_{NAC}. The relatively narrow cavity of CALB hinders the formation of the proper sandwich-like NAC required for the pericyclic reaction pathway. This is also the parameter that is most productive to optimize by point mutations. Extensive MD simulations were performed on a number of rationally-designed enzyme mutants, and it was found that the S105A/I189A variant gives up to 5% 'loose' NAC geometries in ACN for one given simulation, and even higher in WAT (8.5%). The corresponding value for the wild-type in ACN was below 1%.

On the basis of our computed parameters we have analyzed the catalytic performance of the S105A/I189A mutant in water. The computed rate enhancements with standard enzyme and substrate concentrations are in the order of 10^2–10^3. This is comparable to the catalytic effects that we have seen in earlier studies of catalysis of promiscuous reactions in CALB[21, 22]. Whereas we expect the computed rate enhancements to be overestimated, the results clearly point toward the potential for effective Diels-Alder catalysis in redesigned enzymes.

We intend to continue optimization of enzyme-catalysts for this important reaction based on the current findings. In addition, we have provided important information for designing experiments to enhance the potential for detecting Diels-Alder catalysis in enzymes.

References

1. Kelly WL (2008) Org Biomol Chem 6:4483–4493
2. Pohnert G (2001) ChemBioChem 2:873–875
3. Ose T, Watanabe K, Mie T, Honma M, Watanabe H, Yao M, Oikawa H, Tanaka I (2003) Nature 422:185–189
4. Guimaraes C, Udier-Blagovic M, Jorgensen WJ (2005) Am Chem Soc 127:3577–3588
5. Kim SP, Leach AG, Houk KN (2002) J Org Chem 67:4250–4260
6. Kirby AJ (1996) Angew Chem Int Ed 35:706–724
7. Pihko PM (2004) Angew Chem Int Ed 43:2062–2064
8. Schreiner PR (2003) Chem Soc Rev 32:289–296
9. Zhang Z, Schreiner PR (2009) Chem Soc Rev 38:1187–1198
10. Zhang X, DeChancie J, Gunaydin H, Chowdry AB, Clemente FR, Smith AJT, Handel TM, Houk KNJ (2007) Org Chem 73:889–899
11. Kazlauskas RJ, Weissoch ANE, Rappaport AT, Cuccia LAJ (1991) Org Chem 56:2656–2665
12. Stecher H, Faber K (1997) Synthesis pp. 1–16
13. Kazlauskas R, Weber H (1998) Curr Opin Chem Biol 2:121–126
14. Pamies O, Bäckvall J-E (2003) Chem Rev 103: 3247-3262
15. Zaks A, Klibanov AM (1985) Proc Natl Acad Sci 82:3192–3196
16. Klibanov AM (2001) Nature 409:241–246
17. O'Brien PJ, Herschlag D (1999) Chem Biol 6:R91–R105
18. Hult K, Berglund P (2007) Trends Biotech 25:231–238
19. Branneby C, Carlqvist P, Magnusson A, Hult K, Brinck T, Berglund PJ (2003) Am Chem Soc 125:874–875
20. Branneby C, Carlqvist P, Hult K, Brinck T, Berglund PJ (2004) Mol Catal B 31:123–128
21. Svedendahl M, Carlqvist P, Branneby C, Allnr O, Frise A, Hult K, Berglund P, Brinck T (2008) ChemBioChem 3:2443–2451
22. Carlqvist P, Svedendahl M, Branneby C, Hult K, Brinck T, Berglund P (2004) ChemBioChem 6:331–336
23. Svedendahl M, Hult K, Berglund PJ (2005) Am Chem Soc 127:17988–17989
24. Svedendahl M, Jovanovic B, Berglund P (2009) ChemCatChem 1:252–258
25. Strohmeier GA, Sovic T, Steinkellner G, Hartner FS, Andryushkova A, Purkarthofer T, Glieder A, Gruber K, Griengl H (2009) Tetrahedron 65:5663–5668
26. Torre O, Alfonso I, Gotor V (2004) Chem Commun, pp. 1724–1725
27. Kazlauskas RJ (2005) Curr Opin Chem Biol 9:195–201
28. Peracchi A (2001) Trends Biochem Sci 26:497–503
29. Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, Sussman JL, Verschueren KH, Goldman A (1992) Protein Eng 5:197–211
30. Svedendahl M, Berglund P (2009) Personal communication
31. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althof EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Nature 453:190–195
32. Jiang L, Althof EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas ICF, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) Science 319:1387–1381

33. Simón L, Goodman JM (2010) J Org Chem 75:1831–1840
34. Linder M, Brinck T (2009) Org Biomol Chem 7:1304–1311
35. Wittkopp A, Schreiner PR (2003) Chem Eur J 9:407–414
36. Thadani AN, Stankovic AR, Rawal VH (2004) Proc Natl Acad Sci 101:5846–5850
37. Ishihara K, Nakano K, Akakura M (2008) Org Lett 10:2893–2896
38. Kano T, Tanaka Y, Osawa K, Yurino T, Maruoka K (2009) Chem Comm 15:1956–1958
39. Uppenberg J, Oehrner N, Norin M, Hult K, Kleywegt GJ, Patkar S, Waagen V, Anthonsen T, Jones TA (1995) Biochemistry 34:16838–16851
40. Michaelis L, Menten ML (1913) Biochem Z 49:333
41. Cleland WW (1970) Enzymes 2:1–65
42. Dalziel K (1970) Enzymes 11:2–60
43. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Acc Chem Res 33:889–897
44. Åqvist J, Medina C, Samuelsson J (1994) Protein Eng 7:385–391
45. Marelius J, Hansson T, Åqvist J (1998) Int J Quantum Chem 69:77–88
46. Lightstone FC, Bruice TCJ (1996) Am Chem Soc 118:2595–2605
47. Bruice TC, Lightstone FC (1999) Acc Chem Res 32:127–136
48. Toro-Labbe A, Gutierrerez-Oliva S, Murray J, Politzer P (2007) Mol Phys 105:2619–2625
49. Labet V, Morell C, Grand A, Toro-Labbe AJ (2008) Phys Chem A 112:11487–11494
50. Toro-Labbe A, Gutierrez-Oliva S, Murray J, Politzer PJ (2009) Mol Model 15:707–710
51. Uppenberg J, Hansen MT, Patkar S, Jones TA (1994) Structure 2:453–454
52. GOLD 4.1, 2008, http://www.ccdc.cam.ac.uk/-products/life_sciences/gold/.
53. Guex N, Peitsch M (1997) Electrophoresis 18:2714–2723, http://www.expasy.org/spdbv/
54. Jones G, Willet P, Glen RJ (1995) Mol Biol 245:43–53
55. Jones G, Willett P, Glen RC, Leach AR, Taylor RJ (1997) Mol Biol 267:727–748
56. Nissink JWM, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R (2002) Proteins 49:457–471
57. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Proteins 52:609–623
58. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) J Comput-Aided Mol Des 11:425–445
59. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD (1998) Proteins 33:367–382
60. Case D et al. AMBER 10, 2008.
61. Wang JM, Cieplak P, Kollman PAJ (2000) Comput Chem 21:1049–1074
62. Wang JM, Wolf RM, Caldwell JW, Kollman PAJ (2004) Comput Chem 25:1157–1174
63. Jakalian A, Bush BL, Jack DB, Bayly CI (2000) J Comput Chem 21:132–146
64. Jakalian A, Jack DB, Bayly CI (2002) J Comput Chem 23:1623–1641
65. Guardia E, Pinzon R, Casulleras J, Orozoco M, Luque F (2001) J Mol Simulat 26:287–306
66. Cieplak P, Caldwell JW, Kollman PA (2001) J Comput Chem 22:1048–1057
67. Jorgensen W, Chandrasekhar J, Madura J, Klein M (1983) J Chem Phys 79:926–935
68. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341
69. Almlöf M, Brandsdal BO, Åqvist J (2004) J Comput Chem 25:1242–1254
70. Becke AD (1988) Phys Rev A 38:3098–3100
71. Becke ADJ (1993) Chem Phys 98:1372–1377
72. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
73. Ditchfield R, Hehre WJ, Pople JA (1971) J Chem Phys 54:724–728
74. Frisch MJ et al. (2003) Gaussian 03, Revision C.02.
75. Kong S, Evanseck JJ (2000) Am Chem Soc 122:10418–10427
76. Barone V, Cossi MJ (1998) Phys Chem A 102:1995–2001
77. Cossi M, Rega N, Scalmani G, Barone V (2003) J Comput Chem 24:669–681
78. Barone V, Cossi M, Tomasi J (1997) J Chem Phys 107:3210–3221
79. UCSF Chimera (2004) http://www.cgl.ucsf.edu/chimera.
80. Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin T (2004) J Comput Chem 25:1605–1612
81. Turner PJ. Grace v.5.1.20, 2007, http://plasmagate.weizmann.ac.il/Grace.
82. Wesemann M (2007) Plot, v. 0.997, http://plot.micw.eu/.
83. Trodler P, Pleiss J (2008) BMC Struct Biol 8:9
84. Trodler P, Schmid R, Pleiss J (2009) BMC Struct Biol 9:38
85. Nini L, Sarda L, Comeau L-C, Boitard E, Dubes J-P, Chahinian H (2001) Biochim Biophys Acta 1534:34–44
86. Chahinian H, Nini L, Boitard E, Dubes J-P, Comeau L-C, Sarda L (2002) Lipids 37:653–662
87. Garcia J, Mayoral J, Salvatella L (1996) J Am Chem Soc 118:11680–11681