DiVA
http://www.diva-portal.org

Postprint

This is the accepted version of a paper published in *ACS Catalysis*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Amrein, B A., Bauer, P., Duarte, F., Janfalk Carlsson, Å., Naworyta, A. et al. (2015)
Expanding the catalytic triad in epoxide hydrolases and related enzymes.
ACS Catalysis, 5(10): 5702-5713
http://dx.doi.org/10.1021/acscatal.5b01639

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-260232

Expanding the Catalytic Triad in Epoxide Hydrolases and Related Enzymes

Beat A. Amrein[1], Paul Bauer[1], Fernanda Duarte[1], Åsa Janfalk Carlsson[2], Agata Naworyta[1], Sherry L. Mowbray[,1], Mikael Widersten[*,2] and Shina C. L. Kamerlin[*,1].*

[1] Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, SE-751 24, Uppsala, Sweden. [2] Department of Chemistry-BMC, Uppsala University, BMC Box 576, SE-751 23, Uppsala, Sweden.

*sherry.mowbray@icm.uu.se, mikael.widersten@kemi.uu.se, kamerlin@icm.uu.se.

Abstract

Potato epoxide hydrolase 1 exhibits rich enantio- and regioselectivity in the hydrolysis of a broad range of substrates. The enzyme can be engineered to increase the yield of optically pure products, as a result of changes in both enantio- and regioselectivity. It is thus highly attractive in biocatalysis, particularly for the generation of enantiopure fine chemicals and pharmaceuticals. The present work aims to establish the principles underlying the activity and selectivity of the enzyme through a combined computational, structural, and kinetic study, using the substrate *trans*-stilbene oxide as a model system. Extensive empirical valence bond simulations have been performed on the wild-type enzyme together with several experimentally characterized mutants. We are able to computationally reproduce the differences in activities between different stereoisomers of the substrate, and the effects of mutations in several active-site residues. In addition, our results indicate the involvement of a previously neglected residue, H104, which is electrostatically linked to the general base, H300. We find that this residue, which is highly conserved in epoxide hydrolases and related hydrolytic enzymes, needs to be in its protonated form in order to provide charge balance in an otherwise negatively-charged active site. Our data show that unless the active-site charge balance is correctly treated in simulations, it is not possible to generate a physically meaningful model for the enzyme that can accurately reproduce activity and selectivity trends. We also expand our understanding of other catalytic residues, demonstrating in particular the role of a non-canonical residue, E35, as a "backup-base" in the absence of H300. Our results provide a detailed view of the main factors driving catalysis and regioselectivity in this enzyme, and identify targets for subsequent enzyme design efforts.

KEYWORDS: Potato epoxide hydrolase - StEH1 - enzyme selectivity - empirical valence bond - trans-stilbene oxide - biocatalysis - X-ray crystallography.

1. Introduction

Epoxide hydrolases catalyze the hydrolysis of epoxides to the corresponding vicinal diols. They are widely distributed, with biological functions depending on the organism of origin. Known functions can be largely grouped into detoxification through the breakdown of toxic epoxides, catabolism, and cellular signaling[1]. Furthermore, the enzymes' ability to catalyze the resolution of racemic epoxides makes them attractive tools in biocatalysis, as they can help fulfill the ever increasing need for enantiopure pharmaceuticals and fine chemicals[2]. In this context, potato (*Solanum tuberosum*) epoxide hydrolase 1 (StEH1) has been a system of particular interest (*e.g.* refs. [2-9]). We have recently demonstrated that it is possible to obtain engineered variants of this enzyme that produce optically pure products as a result of changes in both enantio- and regioselectivity[8]. Clearly, a detailed understanding of the catalytic mechanism of StEH1 would facilitate progress in this direction, as would the availability of a computational approach that can reliably and efficiently predict mutational effects on function.

A generalized mechanism for the reaction catalyzed by StEH1 is shown in **Figure 1**, based on proposals put forward in the literature[4,6]. Following substrate binding (I), the reaction occurs in three sequential steps: (II) nucleophilic attack of D105 on one of the two epoxide carbons (C1 or C2) of the bound substrate, leading to a covalent alkyl-enzyme adduct, (III) nucleophilic attack by a conserved active-site water molecule (usually believed to be activated by H300 as a general base) to form a tetrahedral intermediate, followed by (IV) decay into product, which is subsequently released from the enzyme (V). It should be noted that there has been some controversy about the details of the reaction[5,6,10,11], in particular with respect to the putative general base. The authors of a recent computational study of murine soluble EH (sEH)[11] suggested that the histidine of the active-site triad (H523 in this enzyme) must be protonated.

The conclusion was based on the calculated activation barrier (which was lower for the protonated form of the histidine), correct prediction of the regioselectivity, and the presence of improved hydrogen-bonding interactions to stabilize the relevant transition states, as well as empirical pK_a calculations that suggested a strongly shifted pK_a value (up to 10-11) for this residue[11]. However, protonating this histidine would require unorthodox approaches to activate the hydrolytic water molecule. Additionally, mutation of the equivalent residue in StEH1, H300, to asparagine almost completely abolished activity during the hydrolysis step[4], suggesting that this histidine is indeed the most likely general base.

In the present work, we have performed long-timescale empirical valence bond (EVB)[12,13] simulations of the enantio- and regioselective hydrolysis of *trans*-stilbene oxide (TSO, **Figure 1**). TSO was chosen as our model substrate because of its size; it almost completely fills the StEH1 active site, thus avoiding the complications that would arise from the multiple potential binding modes expected with smaller substrates. In addition, extensive experimental data (e.g. refs. [4-8]) are available for this particular substrate that can be used for the benchmarking and validation of our theoretical models. It should be noted that computational studies of enzyme selectivity are made particularly challenging by the very small energy differences between various substrates and between mutant forms of the enzyme, as well as the long simulation times required to obtain convergent free energies (see *e.g.* [14,15]). As discussed in ref. [11], while it can be very hard to obtain quantitatively reliable results, an increasing number of elegant studies (*e.g.* ref. [11,15,16]) show that there is potential for real progress using current techniques.

The computational work herein has been coupled with a crystallographic study of the H300N mutant form of the enzyme, in which the conventional general-base functionality of StEH1 has been removed. We have also performed biochemical characterization of an E35Q/H300N

double-mutant form of the enzyme, which has no viable active-site base for the hydrolytic step shown in **Figure 1**. We are able to reproduce both the catalytic activity of the wild-type enzyme and the experimentally observed effects upon mutation of the known active-site residues with quantitative accuracy. Additionally, we identify a key active-site residue, H104, that has been overlooked in both previous experimental[4,6,17-19] and computational works[6,10,11,19,20]. This histidine, which plays an indirect electrostatic role in modulating StEH1 activity, is highly conserved among epoxide hydrolases and related hydrolytic enzymes. Taken together, our observations strongly suggest that the canonical "catalytic triad" in epoxide hydrolases (and related enzymes) needs to be expanded to include E35 and H104.

2. Materials and Methods

2.1. System Setup. The crystal structure of StEH1 without substrate was obtained from the Protein Data Bank[21] (PDB entry 2CJP, 1.95 Å resolution[17]). Potential flips of histidine, asparagine and glutamine side chains were evaluated using the MolProbity server[22] and those suggested for H269, H308 and H313 were applied to the structure. Protonation states of active-site side chains were investigated using PROPKA 3.1[23] and by visual inspection of their hydrogen-bonding patterns. The C_γ of the active-site nucleophile D105 was defined as the simulation center, and all crystallographic waters within 16 Å of this center were retained. The substrate TSO was manually placed in the active site and covalently connected to D105 to form the alkyl-enzyme intermediate. This structure was used as starting point for simulations in both backward and forward direction, leading to the Michaelis complex and the tetrahedral intermediate, respectively.

The E35Q, Y154F and Y235F mutant forms of the enzyme were manually generated by deletion and/or replacement of atoms in the wild-type structure. For Y149F, the published crystal structure (PDB entry 3CXU[18]) was used; the RMSD to the C_α atoms of the wild type structure was 0.2 Å, with only minor local changes observed. The structure of the H300N mutant reported here (see below) was used in all calculations for this variant. While this structure is overall very similar to wild type (RMSD of 0.3 Å for 320 matched C_α atoms), there are significant differences in the active sites, particularly at residues 35 and 300, and in the positioning of the nearby waters (**Figures 2** and **S1**).

2.2. Molecular dynamics simulations. Molecular dynamics (MD) simulations were conducted using the OPLS-AA forcefield[24] implemented in the Q simulation package (Version 5.06)[25]. For the Michaelis complex with TSO, the alkyl-enzyme, and the tetrahedral intermediates (see **Figure 1**), OPLS-AA compatible force field parameters were generated using the Macromodel 9.1 force field version 2001 (Schrödinger LCC)[26]. Partial charges were generated using the HF/6-31G* RESP procedure described in ref. [27]. All simulations were performed by solvating the system in a spherical water droplet of 20-Å radius subject to Surface Constrained All Atom Solvent (SCAAS) boundary conditions[28]. The water droplet was comprised of TIP3P[29] water molecules centered on the C_γ atom of the active-site nucleophile, D105. All atoms within 17 Å of the simulation center were fully mobile, atoms between 17 Å to 20 Å were restrained with a 10 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic restraint, and atoms outside the boundary were restrained to their initial positions with a harmonic force constant of 200 $kcal \cdot mol^{-1} \cdot Å^{-2}$, following the standard procedure used in our previous work (see *e.g.* ref. [30]). Therefore, the water droplet complements the 16 Å sphere of crystallographic waters that were retained for the simulations, and ensures that these waters fall within the mobile region of the simulation. A cutoff of 10 Å was used for calculation

of non-bonded interactions between all atoms except for those in the EVB-region, for which all interactions were explicitly calculated up to 99 Å from each atom (*i.e.* no cutoff was applied). Long-range electrostatic interactions beyond this cutoff were treated with the Local Reaction Field (LRF)[31] method. The SHAKE[32] algorithm was applied to all solvent molecules except the hydrolytic water molecule, which is directly involved in the chemical reaction and needs to be able to break and form O-H and C-O bonds, respectively.

The systems were initially heated from 1 K to 300 K over a short 25-ps simulation, applying a 200 kcal·mol^{-1}·Å^{-2} harmonic force constant on the solute atoms to restrain them to their crystallographic positions. This allowed for solvent equilibration around the solute molecule. The system was cooled to 5 K for another 10 ps and then reheated to 300 K for 70 ps of simulation time while the force constants of the harmonic restraint were gradually decreased from 200 to 0.5 kcal·mol^{-1}·Å^{-2}. Finally, a 10 ns equilibration was performed at 300 K, using a 0.5 kcal·mol^{-1}·Å^{-2} force constant on atoms treated as EVB atoms. The endpoint of this equilibration run was then used as a starting point for subsequent EVB simulations. The temperature of the system was regulated using the Berendsen thermostat[33] (with a 100 fs bath relaxation time). The time-step used for all simulations was 1 fs.

For the corresponding reaction in water, the positions of the atoms in the reactive region (the atoms treated as EVB atoms) were extracted directly from the equilibrated protein structure. The nucleophilic aspartate and general-base histidine were truncated to propionate and ethyl imidazole, respectively. The resulting system was solvated in a 20 Å water sphere centered at the carboxylic carbon of the propionate ion, following the protocol described above. Finally, the system was equilibrated for 1 ns before starting EVB calculations.

2.3. Empirical valence bond calculations. The computational method of choice in this work was the empirical valence bond (EVB) approach[12,13], which is a semi-empirical multiscale approach to describe chemical reactivity based on valence bond theory. The EVB approach is sufficiently fast to allow easy access to up to microseconds of conformational sampling and provides a physically meaningful description of bond-breaking and bond-making processes[34]. Thus, quantitatively reliable modeling of enzyme selectivity is possible[14,15]. Additionally, the success of the EVB approach, as applied to numerous biological systems, has been extensively documented[13,34] and its ability to predict and rationalize mutational effects has made it a powerful tool for computer-guided enzyme re-design[35-37] (which is the ultimate goal of this work).

EVB relies on a well-defined reference state[38]. For enzymatic reactions, the most natural choice is the uncatalyzed reaction in solution. In cases where the enzyme-catalyzed reaction is fully understood, the wild-type enzyme can be used as a reference state for a series of mutant forms of the enzyme. In the present work, we have extensive experimental data available for several mutant forms[4,5,18] and therefore, in principle, the wild-type enzyme should be an adequate reference state. However, as discussed below, it is not immediately obvious which protonation states of the coupled active-site residues H104 and H300 are likely to be relevant during the catalytic step (see discussion in ref.[11,20]). Therefore, we use the reaction in solution as a reference, to allow us to fit our EVB simulations to a common reference point, and compare different combinations of histidine protonation states. Once the reference state has been clearly and rigorously defined and parameterized, the same parameter set can then be used unchanged for the wild-type and mutant forms of the enzyme (see discussion in ref. [34]), allowing us to directly examine the catalytic effect of the enzyme and compare the effect of mutations without the need for any adjustable parameters. When calibrating the reference state, one important aspect to take

into account is that the preferred mechanism in solution, without the catalyst present, can be different from the one in the enzyme active site, as one feature of the enzyme may be to change the mechanism of the uncatalyzed reaction (making a usually unfavorable mechanism favorable). Therefore, the effect of having different mechanisms and different environments should be considered separately[39]. In the first case, it is normally straightforward to estimate the difference between these mechanisms in aqueous solution, and therefore the only relevant state to consider is the "chemically filtered" reference reaction, defined as the reaction that follows exactly the same mechanism as the one in the enzyme, but in absence of it. The energetics of this reaction can be obtained either from actual or extrapolated experimental data, or, when such data is not available, by means of high-level (DFT or *ab initio*) quantum mechanical (QM) calculations. In such a way, it is possible to obtain a reasonable estimate of the energetics of the reference reaction to within a few kcal/mol, which can seem quite large in terms of absolute energy, but is still an acceptable error margin compared to the huge rate-accelerations afforded by the majority of enzymes[40]. This has been clarified in much greater detail in ref. [38].

As outlined in **Figure 1**, the StEH1 reaction starts with nucleophilic attack on the bound epoxide to form an alkyl-enzyme intermediate, followed by general-base ester hydrolysis of the alkyl-intermediate proceeding *via* a short-lived tetrahedral intermediate, which decomposes to the final product. As the last step will be extremely fast and certainly not rate-limiting[41], for computational simplicity, we have focused here on the first two steps, which were described in terms of two valence bond structures, as illustrated in **Figure 3**. Additionally, mutagenesis studies[4] provide evidence that D105 (**Figure 2**) acts as the nucleophile to open the epoxide ring, and that Y154 and Y235 are also crucial for the catalytic activity, forming an oxyanion hole to stabilize the alkyl-enzyme intermediate. Even more importantly, rapid kinetics suggests that the

opened epoxide ring is only protonated later in the reaction[5]. Taking into account this information, we have calibrated the two first steps of the process as described in **Sections S1.1** and **S1.2**.

All EVB calculations were performed using the standard EVB free-energy perturbation/umbrella sampling (EVB-FEP/US) procedure outlined in ref. [13,42]. For the enzymatic reaction, the EVB region always included the substrate, the side chains of D105 and H300 (including $C_\beta H_2$), and the hydrolytic water molecule. During formation of the alkyl-enzyme intermediate, the H300 side chain and the hydrolytic water molecule are present as spectator atoms in order to keep the same EVB region throughout. In the hydrolysis step, the H300 side chain acts as a potential general base to activate the hydrolytic water molecule for nucleophilic attack on the alkyl-enzyme intermediate. As all atoms in all valence bond states are treated *using the same force field*, the only differences between the reacting (EVB) and non-reacting atoms are the use of Morse rather than harmonic potentials to describe bonds that are changing during the reaction (see the **Supporting Information**), and the fact that the EVB atoms do not have a cutoff for calculation of the non-bonded interactions. Hence, the distinction between non-reacting EVB atoms and the remainder of the system is primarily for chemical bookkeeping, and does not affect the outcome of the simulation. Finally, for the corresponding background reaction, D105 and H300 were truncated to propionate and ethyl imidazole, respectively (ensuring the same reacting atoms in both enzymatic and background reactions), and ethyl imidazole and the hydrolytic water molecule were only included in our model of the hydrolytic step of the reaction.

The starting point for our EVB simulations was the endpoint of the initial 10 ns equilibration run. As a first step, we performed an additional 250 ps of equilibration, taking structures every 25 ps to provide ten different starting structures for subsequent EVB simulations. All EVB

simulations were performed in 51 windows of 200 ps length each, applying either a 0.5 (protein) or 1 kcal·mol^{-1}·Å^{-2} (water) harmonic restraint to the reactive region. The larger restraint in the water simulation was necessary to focus the sampling, since greater conformational space is available in the absence of the protein. For similar reasons, in the second reaction step (hydrolysis of the alkyl-enzyme intermediate, **Figure 1**), it was necessary to apply a weak (3 kcal·mol^{-1}·Å^{-2}) distance restraint between the oxygen atom of the hydrolytic water molecule and the C_γ atom of D105, and between the water hydrogen and the proton acceptor ($N_{\varepsilon 2}$ or $O_{\delta 1}$ for H300 and E35, respectively). The calibration of the reference reaction in aqueous solution and EVB parameters are provided in the **Supporting Information**. All other parameters were identical to those used for the equilibration runs. A total protein simulation time of 11.8 μs was conducted, including all equilibration runs, and EVB simulations for all setups and variants considered in this work.

2.4. Empirical pK_a estimates. Estimates of pK_a values for key active-site residues were obtained by rapid empirical screening using PROPKA 3.1[23]. This program provides reasonable estimates of protein pK_a values[23], and can also identify non-covalently coupled groups and model their titration behavior[43]. The latter is particularly relevant here, as the protonation states of H104 and H300 are likely to be coupled due to their close proximity. We used a local installation of PROPKA 3.1[23] to obtain estimates of pK_as from snapshots of the equilibrated system at different stages of the reaction cycle. For the alkyl-enzyme intermediate, the pK_a calculations were performed directly on the equilibrated structure after 10 ns of molecular dynamics equilibration. For the Michaelis complex and tetrahedral intermediate, we used structures from the last window of the relevant EVB calculations. The pK_a values for the substrate-free enzyme were calculated

directly from the crystal structures of wild-type StEH1 (PDB entry 2CJP[17]) and the H300N mutant reported here (PDB entry 4Y9S).

2.5. Sequence alignment and conservation of H104. To identify other residues that could be of importance in catalysis, sequences from the epoxide hydrolase superfamily were obtained from the ESTHER sequence database[44]. The sequences were grouped into three different pools: (1) the whole epoxide hydrolase superfamily, including putative epoxide hydrolase-like enzymes, dehalogenases and Mesoderm Specific Transcript (MEST)-like enzymes, (2) only epoxide hydrolase-like enzymes, and (3) only dehalogenase enzymes. All of these are thought to use an aspartate nucleophile within their catalytic triad, introducing an additional negative charge in the active site and making it possible to form a salt bridge with the general-base histidine. The sequences were aligned using Clustal Omega[45], with default alignment parameters. Sequence logo plots were created using a local version of WebLogo 3[46]. Plots were generated according to the probability of finding the residue at a given position without the need for additional weighting. After a first alignment, sequences not containing an aspartate at the expected nucleophile position were removed, and the alignments were repeated. The StEH1 sequence was also used as a query in BLAST[47] searches for related sequences among mammals, plants and fungi. Several annotated sequences were then used for an additional alignment to check for conservation of H104.

2.6. Construction and characterization of the E35Q/H300N double mutant. The E35Q/H300N double mutant was constructed by combining pGTacStEH1-5H-based mutants E35Q[6] and H300N[4]. The plasmid carrying the H300N mutation was digested with restriction enzymes *Mun*I and *Spe*I (Thermo Scientific) and the resulting fragment was ligated into the

corresponding plasmid encoding the E35Q mutant, digested with the same pair of enzymes, to yield the final variant. The plasmid was transformed into *E. coli* XL1-Blue cells (Stratagene). The recombinant protein was produced, then purified by Ni(II)-IMAC and size exclusion chromatography, as previously described[4]. For steady-state kinetic measurements, the purified enzyme (1.5 μM) was mixed with a range of concentrations of (*R*,*R*)-TSO (up to 500 μM) in 0.1 M sodium phosphate (pH 6.8) with 1% (v/v) acetonitrile. Hydrolysis was measured at 30 °C by following the absorbance of the substrate at 225 nm[4]. No decrease in epoxide concentration was observed after 5 min, and so the mutant was considered to be inactive. The formation of the alkyl-enzyme intermediate was detected as a decrease in tryptophan fluorescence using a sequential stopped-flow spectrophotometer (Applied Photophysics SX.20MV)[4]. The excitation wavelength was 290 nm and the detected light was filtered through a 320 nm-cutoff filter. Multiple-turnover experiments were performed at 30 °C in 0.1 M sodium phosphate buffer (pH 6.8) with 1% (v/v) acetonitrile. Enzyme concentration was 0.5 μM and substrate concentration varied between 1.6 and 50 μM. Apparent rate constants were determined by fitting a single exponential function (**Eq. 1**) to the data points, where A is the amplitude of fluorescence change, k_{obs} the observed rate constant and C the floating end point.

$$F(t) = A\exp(-k_{obs}t) + C \qquad (1)$$

Curve fitting was performed using SIMFIT's (www.simfit.org.uk) QNFIT from which the microscopic rate constants (**Eq. 2**) were determined. In this model, k_2 is the rate of alkyl-enzyme formation, and k_{-2} and k_3 the rates of decay of the alkyl-enzyme, either back to the Michaelis complex, or forward to the product, respectively. The quenched tryptophan fluorescence signal was not recovered after 5 min incubation, confirming the lack of product formation; thus k_3 in **Eq. 2** was set to zero and the intercept on the y-axis (*i.e.* at [S] = 0) will correspond to k_{-2}.

$$k_{obs} = \frac{k_2[S]}{K_s + [S]} + (k_{-2} + k_3) \quad (2)$$

2.7. Crystal structure of H300N mutant. Mutagenesis and protein production of the H300N mutant have been described previously[4]. The StEH1 crystals were obtained under the same conditions as the wild type[17], then soaked for 5 min at pH 3.5 in cryoprotectant solution prior to flash-cooling. X-ray data were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble. The structure was solved and refined using the programs of the CCP4 interface[48]. After processing the data with MOSFLM[49] and SCALA[50], the two molecules in the asymmetric unit of the wild-type enzyme (PDB entry 2CJP) were placed using rigid-body refinement with REFMAC5[51]. H300 was mutated to asparagine and the structure was rebuilt as needed using O[52]. Waters were added with COOT[53], in conjunction with restrained refinement in REFMAC5. Statistics for the data and final refined model are summarized in **Table S1**. The H300N structure and X-ray data have been deposited at the Protein Data Bank[21] with entry code 4Y9S.

3. Results and Discussion

3.1. Probing likely protonation states of the active-site histidines. A generalized mechanism for the reaction catalyzed by StEH1 is shown in **Figure 1**. In our study, we considered both (*R*,*R*) and (*S*,*S*) enantiomers of TSO, as well as nucleophilic attack at either the C1 or C2 carbon of the epoxide ring. We first re-assessed the protonation state of H300, following discussion arising from previous computational studies[6,11,20]. While we obtained a similarly high estimate for the pK_a of H300 in the substrate-free enzyme to that cited in ref. [11], this value drops substantially once the substrate is bound, and stays comparably low at each step of the catalytic cycle (**Table 1**). When combined with experiments strongly supporting the role of this histidine as a general

base[4,6,54], it seems unlikely that it is protonated during the catalytic cycle. Additionally, the pK_a of the nucleophilic D105 is elevated to ~6 upon substrate binding, suggesting that both the catalytic base and the nucleophile could contribute to the basic limb of the experimental pH-rate profile[4] (**Figure 4**).

We also propose the involvement of a second, previously neglected but strongly conserved histidine in the active site, H104 (**Figures 2** and **5)**. Using sequence alignments, the canonical conserved residues could be identified in the wider epoxide hydrolase-like superfamily (Figure **S2A**), the epoxide hydrolase family specifically (**Figure S2B**), dehalogenases (**Figure S2C**), and the annotated sequences from mammals, plants, fungi and bacteria obtained after aligning BLAST results to the StEH1 sequence (**Figure 5**). From these alignments and the accompanying residue-occurrence plots, it is clear that the residue corresponding to H104 in StEH1 is highly conserved throughout the superfamily. The side chain of this residue is near that of H300 (the closest atoms are separated by only 3.3 Å in the wild-type structure) and, although the geometry is not consistent with a hydrogen bond, the protonation states of these two residues are probably linked. For instance, having both residues simultaneously protonated would create large charge-charge repulsion. Our pK_a calculations (**Table 1**) suggest that, in contrast to H300, H104 is likely to be protonated during the catalytic cycle. Protonation of either H104 or H300 would provide charge balance in the highly negatively-charged environment of the active site, which contains three anionic residues (**Figure 2**). When examining the patterns seen for nearby residues (**Figure 2 and Table S2**), it can be seen that even though it is possible to replace E35 with Q without losing H104, whenever H104 is replaced by another residue, in particular in dehalogenases, the adjacent E35 is replaced by either a Ser or Thr, compensating for the loss of positive charge in

the active site. The combined evidence strongly suggests that H104 is preferentially protonated in the StEH1 active site.

To further test this hypothesis, all wild-type EVB simulations were performed with H104 in either its neutral or ionized form, after separate molecular dynamics equilibration runs for each possible protonation state. The calculated activation and reaction free energies are summarized in **Table S3** of the **Supporting Information**, along with the observed values. Modeling H104 in its neutral form results in unrealistically low activation barriers, and extreme exothermicity, with a predicted stabilization in excess of 6 kcal·mol^{-1} for the alkyl-enzyme intermediate. These results are in poor agreement with pre-steady-state kinetic data, which suggest a stabilization of 0.2 to 1.0 kcal·mol^{-1} for this intermediate (see refs. [4,6]). Similar extreme exothermicities were described in earlier QM cluster model studies[10], which considered the equivalent of H300 (H523 in sEH) as being neutral and did not include the counterpart of H104 (H332 in sEH) in the model[10,20]. In contrast, by protonating H104 and thus introducing a positive charge into the active site, much better agreement is obtained with the experimental results in terms of both activation free energies and intermediate stability, as shown in **Figure 6** and **Tables 2** and **S3**. It now becomes possible to reproduce the energetics of the wild-type enzyme as well as the trends of the series of tyrosine mutants tested (**Figure 6** and **Table 2**). Note, however, that in the exceptional case of the E35Q variant, the pK_a of H104 is predicted to be dramatically lowered on removal of the adjacent negative charge (**Table 1**); therefore, in this instance, H104 was modeled as being deprotonated, as discussed below.

3.2. Modeling the selectivity of StEH1. While one would not expect either regio- or enantioselectivity in the uncatalyzed epoxide ring opening of TSO in aqueous solution (as the two carbons are presumably chemically indistinguishable in a symmetrical substrate), StEH1

preferentially hydrolyzes (*R*,*R*)-TSO at pH 6.8[4]. Our calculations suggest a nucleophilic attack primarily at carbon C1 for (*S*,*S*)-TSO ($\Delta G_1^{\ddagger}$ is 1.7 kcal·mol^{-1} lower compared to C2), and at carbon C2 for (*R*,*R*)-TSO (preferred by 3.6 kcal · mol^{-1}), see **Table S3** and **Figure 6**. However, in the hydrolysis step, the barrier is substantially higher for both enantiomers when the epoxide has been opened at C1 compared to C2. This means that trajectories following attack at C1 are likely to be non-productive, as return to substrate will be faster than the rate of hydrolysis. Therefore, we suggest that the regioselectivity is determined in the second step of the reaction (hydrolysis), with both enantiomers reacting through the intermediate formed after nucleophilic attack at C2. As seen in **Figure 6**, this leads to good quantitative agreement with the experimental values. Note that while we get clear regioselectivity for attack at C1 *vs*. C2 in both steps, we are only able to reproduce enantioselectivity in the first (alkylation) step, due to the small energy differences and the dependence of the energetics of the hydrolysis step on the calculated free energy of the previous alkylation step.

To assess the role of individual residues in determining the observed regioselectivity in the hydrolysis step, we performed a quantitative analysis of the electrostatic contributions of key residues to the calculated activation barriers for hydrolysis using the linear response approximation, as outlined in previous studies[30,55] (see **Figure 7A)**. The main contributions come from residues H104, W106 and D265, with the latter two residues compensating for the positive (destabilizing) contribution from H104 to the hydrolysis step. This creates an electrostatically balanced active site, as modeling H104 as being deprotonated in the simulation system results in very large artificial stabilization of the alkyl-enzyme intermediate (see **Table S3)**. Even though the individual contributions are very similar for different pathways (**Figure 7A**), the sum of these contributions results in a larger stabilizing effect when hydrolysis takes place in the alkyl-

enzyme intermediate formed after nucleophilic attack at C2**.** This arises predominantly from the interplay between the effects of H104 and W106.

To probe this further, we considered the electrostatic interactions between the protein and different parts of the substrate (**Figures 7B** and **C**). The interactions have been separated into three groups: those coming from each of the two benzene rings and those arising from interactions of the core of the substrate. As expected (**Figure 7B**), the most important stabilizing interactions are those between the oxyanion formed upon epoxide ring opening and its surrounding electrostatic environment. These stabilizing interactions are stronger, by 5 to 10 kcal·mol^{-1}, when nucleophilic attack has taken place at C2 for both (*R,R*)-TSO and *(S,S*)-TSO (**Figure 7B**). This difference in stabilization energy is similar to that obtained for the absolute activation barrier for hydrolysis following attack at C1 *vs* C2, *i.e*. the values obtained before correction for the calculated free energies of the alkyl-enzyme intermediate (see **Table S4**). Additionally, an overlay of the TS for C1 *vs* C2 attack shows a substantial shift in one of the benzyl rings for the disfavored TS (**Figure S3**), suggesting a further steric/strain contribution.

3.3. Exploring the effect of key active-site mutations. To further validate our calculations, we performed EVB simulations on a series of key active-site mutants (**Tables 2**, **S5 - S7**). **Figure 8** shows that we were able to reproduce the effect of each of these mutants with an accuracy of 1-2 kcal·mol^{-1} (where experimental data are available). Our key observations are summarized below.

E35Q: In the wild-type enzyme (**Figure 2**), the γ-carboxylate of E35 is within hydrogen-bonding distance of the hydrolytic water molecule and H104 (although the geometry for the latter is not optimal). E35 is also part of a hydrogen-bond network that connects the active site with the solvent interface, *i.e*. E35 → S39 → Y219 → R41 → E215. Previous studies have demonstrated

strong conservation of E35 as well as the hydrogen-bond network it is involved in[6]. Kinetic studies have shown that while the reaction of the (*R*,*R*) enantiomer is affected at both alkylation and hydrolysis steps by this mutation, its impact is smaller for the (*S*,*S*) enantiomer (and mainly at the hydrolysis step)[6]. Furthermore, there is a complete inversion in the pH-rate profile for the (*R*,*R*) enantiomer, whereas that for the (*S*,*S*) enantiomer is only slightly shifted compared to the wild-type **(Figure 4)**. However, what appears to be an inverted pH profile is in fact a drastic shift of the acidic limb to lower pH values. The cause of this shift is unclear but it seems reasonable to link the effect to the extremely low predicted pK_a of the H104 imidazolium, even if this value is clearly exaggerated (**Table 1**). The finding that the modeled conformations of the side chain of Q35 are different in the alkyl-enzyme complexes of the different enantiomers (and also from the wild-type enzyme complexes) (**Figure S4**) is also consistent with the idea that the E35Q substitution affects the hydrolysis of the TSO enantiomers differently.

In a previous simulation study[6], it was proposed that two key roles of E35 are to (1) participate in activation of the aspartate nucleophile (presumably by facilitating channeling of protons out of the active site), and (2) orient the catalytic water for optimal hydrogen bonding during the hydrolysis half-reaction. However, these conclusions were based on the premise that H300 was protonated and H104 deprotonated[6]. As discussed above, our EVB calculations (**Tables 2** and **S3**) suggest that, for both enantiomers, H104 is preferentially protonated in the wild-type enzyme, creating a strong interaction between H104 and the adjacent E35 (see **Figure 2).** We do not observe an interaction between E35 and the hydrolytic water molecule in the wild-type enzyme. Note however that our simulations start with equilibration of the alkyl-enzyme intermediate rather than the substrate-free enzyme, and there is a significant charge shift in the active site upon formation of this intermediate, due to the displacement of the negative charge on

the nucleophile towards the newly-formed oxyanion. This could, in turn, break the interaction between E35 and the hydrolytic water molecule.

In the E35Q variant, however, the orientation of Q35 seems to be heavily dependent on which substrate enantiomer is bound to the active site. In the wild-type, the E35-H104 interaction blocks solvent access to the active site, and in the case of the complex of the E35Q variant with (*S*,*S*)-TSO, solvent accessibility is similarly prevented by an interaction between Q35 and the hydroxyl group of the tetrahedral intermediate formed after attack of the hydrolytic water molecule (**Figure S4**). In the case of (*R*,*R*)-TSO, the substrate pushes Q35 out of the active site, preventing any interaction with either H104 or the tetrahedral intermediate, and allowing solvent access to the active site. These critical structural differences, which in turn affect the solvation of the reacting atoms, can provide a starting point for understanding the different behavior of the E35Q variant towards the two enantiomers.

Y154/Y235: As shown in **Figure 2**, Y154 and Y235 together form an oxyanion hole that stabilizes the oxyanion of the alkyl-enzyme intermediate formed by epoxide ring opening, as suggested in ref. [5]. To explore their role in intermediate stabilization in more detail, we modeled two single-mutant forms of StEH1 based on the wild-type crystal structure, and studied the full reaction profiles for the hydrolysis of the two TSO enantiomers. The calculated activation barriers are presented in **Tables 2** and **S5-S7.** For both enantiomers, the alkylation barriers are higher for the mutants than for the wild-type enzyme, and the energy difference between initial nucleophilic attack at either carbon atom on the substrate is very small. It is important to note that the experimental rates for these mutants are based on steady-state data, and k_{cat} may represent different properties in these mutants compared to the wild-type enzyme or other mutants (as the rate-limiting step may have changed from hydrolysis to alkylation). Additionally,

in the case of (*R*,*R*)-TSO, we overestimate the activation barrier for the hydrolytic step (Step II) by *ca*. 2 kcal·mol^{-1}. However, it is fully plausible that in the absence of a tyrosine residue to form the oxyanion hole shown in **Figure 2** (which in turn stabilizes charge on the oxyanion intermediate), this oxygen atom could become protonated earlier in the reaction cycle, thus reducing the barrier to the hydrolytic step.

H300N and E35Q/H300N: The H300N mutant involves a substantial perturbation of the active site, so it was fortunate that we could solve this crystal structure. Importantly, the structure shows that N300 takes a different side-chain conformation than H300 (**Figure 2B**). Additionally, no electron density was observed for the hydrolytic water (**Figure S1**), indicating that its usual site is occupied rarely in this mutant, if ever. The structure further suggests a strong interaction between E35 and H104 that is not present in the wild-type protein, and only appeared in the mutant simulations after the initial equilibration run (see **Figure 2**). Our calculations suggest that the pK_a of D265 is significantly elevated in the H300N mutant compared to the wild type, probably due to the missing hydrogen bond between D265 and H300 (**Table 1**). Experimental chemical shift data have suggested that the corresponding histidine-aspartate pairs in enzymes with related mechanisms, such as serine proteases, also share a proton[56]. Therefore, with H300 removed from the system, D265 could plausibly be protonated, and we have modeled it as such in our EVB calculations on this variant. The mutation also seems to affect the protonation state of H104; a protonated H104 would unfavorably perturb the pK_a of the nearby E35.

Experimentally, this variant shows drastically reduced kinetic efficiencies (**Table S8)**. Due to the low signal-to-noise ratio in the measurements of the steady-state kinetics, it was not possible to determine k_{cat} and K_M values for this mutant[4] and therefore the experimental values for k_{cat} have to be taken as lower limits. The biggest challenge with the H300N mutant is modeling the second

step of the reaction, as the conventional general base of the enzyme has been removed and the hydrolytic water is lacking. This complicates modeling of the second step, since it is not clear where the needed water molecule actually lies. Although there is no detectable activity for (*S,S*)-TSO, the enzyme still maintains some level of activity for the (*R,R)*-TSO enantiomer, as shown in **Table S8**. Reasonable agreement with experimental results for both (*R,R*)-TSO and (*S,S*)-TSO are only obtained when H104 is deprotonated.

In the absence of H300, we posit that E35 could act as a general base with some efficiency. As the hydrolytic water seems to be lost, we are not convinced a computational model of the subsequent hydrolytic step would be very meaningful. However, we have experimentally explored this possibility by preparing and biochemically characterizing an E35Q/H300N double mutant form of the enzyme. As seen in **Table S8**, both individual mutants have some activity for (*R,R*)-TSO. The double mutant retains the activity of the H300N single mutant for the first step of the reaction, forming the alkyl-enzyme intermediate, but all activity is lost in the hydrolytic step of the reaction. This provides a clear indication that both residues are involved in the catalytic step, and when combined with our computationally study of the role of H104, strongly suggests that the canonical catalytic triad should in fact be expanded to include residues E35 and H104.

4. Overview and Conclusions

The present work describes a combined computational, biochemical and structural study of the hydrolysis of *trans*-stilbene oxide by StEH1. The result is a concrete and unified molecular model that for the first time explains the pH-dependence of catalysis by StEH1[6], its regiopreference, and changes in activity and selectivity upon mutation. Among many new insights, we find that the regioselectivity of the reaction is determined not at the alkylation step,

as one would intuitively assume, but rather at the subsequent hydrolytic step shown in **Figure 1**. This suggests that the latter reaction step needs to be carefully examined in any computational design studies that aim at controlling the regioselectivity of this enzyme.

In addition to the well-established contributions of the active-site triad D105/H300/D265 and the two lid residues, Y154 and Y235, we uncovered the vital roles of two previously neglected residues, E35 and H104, which form an ion-pair and are highly conserved among epoxide hydrolases and related hydrolytic enzymes. Our EVB calculations show that unless H104 is included in the simulation system in its protonated form, it is impossible to reproduce experimental trends in activities and selectivities for the wild-type and mutant enzymes. Previous quantum chemical cluster calculations that did not include the equivalent of H104 (H332 in sEH)[20] resulted in extreme exothermicities, as indeed we obtain with a neutral H104.

Our new model also resolves the ongoing controversy with regard to the protonation state of H300[11]. Protonating either histidine would provide charge balance in the active site, however, protonation of H104 (which is linked to H300 through the hydrolytic water molecule) rather than H300 allows for retention of the general-base capability of the latter. All known sequences of related hydrolytic enzymes in which H104 is replaced by a different residue also replace the adjacent negatively-charged E35 with either Ser or Thr (**Table S2**), allowing charge balance to be maintained in the active site. This strongly suggests that our conclusions are not limited to *St*EH1, but rather, can be extended to other epoxide hydrolases and related hydrolytic enzymes (such as dehalogenases) that use an Asp nucleophile. Considering different protonation states of H104 also allows us to provide a rationale for the unusual pH-rate profile observed upon mutating E35 to Q[6], as commented on in other recent work[11]. Removal of this negative charge

most likely results in a neutral H104 and allows Q35 to be a modulator of water access to the StEH1 active site.

The distinction between residues that are directly involved in catalysis through, for instance, nucleophilic attack or proton transfer (such as D105 and H300), and those that act by electrostatically modulating the active site, is proving to be subtle. In StEH1, the changing protonation state of H104 in the E35Q and H300N mutations appears to play a direct role in modulating the selectivity of the enzyme. Thus, we show that electrostatic interactions from residues not directly involved in bond breaking/making processes can still play an important role in even reactions such as enantioselective ring opening, which are usually considered to be dominated by steric considerations[57]. This emphasizes the need for complete and accurate enzyme models for rational computational design of enzymes of biocatalytic interest, such as the one we present for epoxide hydrolases here.

Supporting Information. A detailed description of the calibration of the background reaction in aqueous solution, comparison of the energetics of all reactions considered, extended sequence analysis, and all EVB parameters used in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgments

The European Research Council has provided financial support under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement No. 306474. The authors also acknowledge support from the Swedish Research Council (MW, grant 621-2011-6055 and SLM, grant 621-2010-5145), including an assistant professorship to LK, as well as a doctoral studies fellowship from the Sven and Lilly Lawski Foundation to PB. Finally,

MW acknowledges support from COST action 1303 (Systems Biocatalysis). The authors would like to thank Anthony Kirby and John Wilkie for valuable discussion.

Abbreviations

EVB: Empirical Valence Bond. **Calc**.: Calculated values of the enzyme-catalyzed reaction in this work. **Exp**.: experimental values of the enzyme-catalyzed reaction derived from kinetic data. **MD**: Molecular Dynamics. **RMSD**: Root-Mean-Square Deviation. StEH1: *Solanum tuberosum* Epoxide Hydrolase 1. **SCAAS**: Surface Constrained All Atom Solvent. **sEH**: murine soluble EH. **TSO**: *trans*-stilbene oxide.

References

(1) Morisseau, C.; Hammock, B. D. *Annu. Rev. Pharmacol. Toxicol.* **2004**, *45*, 311-333.

(2) Archelas, A.; Furstoss, R. *Curr. Opin. Chem. Biol.* **2001**, *5*, 112-119.

(3) Monterde, M. I.; Lombard, H.; Archelas, A.; Cronin, A.; Arand, A.; Furstoss, R. *Tetrahedron Asymm.* **2004**, *15*, 2801-2805.

(4) Elfström, L. T.; Widersten, M. *Biochem. J.* **2005**, *390*, 633-640.

(5) Elfström, L. T.; Widersten, M. *Biochemistry* **2006**, *45*, 205-212.

(6) Thomaeus, A.; Carlsson, J.; Åqvist, J.; Widersten, M. *Biochemistry* **2007**, *46*, 2466-2479.

(7) Lindberg, D.; de la Fuente Revenga, M.; Widersten, M. *Biochemistry* **2010**, *49*, 2297-2304.

(8) Carlsson, Å.; Bauer, P.; Ma, H.; Widersten, M. *Biochemistry* **2012**, *51*, 7627-7637.

(9) Wu, S.; Chen, Y.; Xu, Y.; Li, A.; Xu, Q.; Glieder, A.; Li, Z. *ACS Catal.* **2014**, *4*, 409-420.

(10) Hopmann, K. H.; Himo, F. *Chemistry* **2006**, *6*, 6898-6909.

(11) Lonsdale, R.; Hoyle, S.; Grey, D. T.; Ridder, L.; Mulholland, A. J. *Biochemistry* **2012**, *51*, 1774-1786.

(12) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218-6226.

(13) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions, 1st ed.*; Wiley: New York, 1991.

(14) Frushicheva, M. P.; Warshel, A. *ChemBioChem* **2012**, *13*, 215-223.

(15) Schopf, P.; Warshel, A. *Proteins* **2014**, *82*, 1387-1399.

(16) Lind, M. E. S.; Himo, F. *Angew. Chem. Int. Ed.* **2013**, *52*, 4563-4567.

(17) Mowbray, S. L.; Elfström, T. L.; Ahlgren, K. M.; Andersson, C. E.; Widersten, M. *Protein Sci.* **2006**, *15*, 1628-1637.

(18) Thomaeus, A.; Naworyta, A.; Mowbray, S. L.; Widersten, M. *Protein Sci.* **2008**, *17*, 1275-1284.

(19) Reetz, M. T.; Bocola, M.; Wang, L.-W.; Sanchis, J.; Cronin, A.; Arand, M.; Zou, J.; Archelas, A.; Bottalla, A.-L.; Naworyta, A.; Mowbray, S. L. *J. Am. Chem. Soc.* **2009**, *131*, 7734-7343.

(20) Hopmann, K. H.; Himo, F. *J. Phys. Chem. B* **2006**, *110*, 21299-21310.

(21) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235-242.

(22) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2010**, *66*, 12-21.

(23) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525-537.

(24) Jorgensen, W. L.; Maxwell, D. S.; Tiado-Rives, J. J. *J. Am. Chem. Soc.* **1996**, *118*, 1125-11236.

(25) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graph. Modell.* **1998**, *16*, 213-225.

(26) Schrödinger Release 2013-3: MacroModel version 10.2. Schrödinger LLC. New York 2013.

(27) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357-1377.

(28) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647-3661.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(30) Aqvist, J.; Kamerlin, S. C. L. *Biochemistry* **2015**, *54*, 546-556.

(31) Lee, F. S.; Warshel, A. *J. Chem.Phys.* **1992**, *97*, 3100-3107.

(32) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

(33) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(34) Kamerlin, S. C. L.; Warshel, A. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 30-45.

(35) Roca, M.; Vardi-Kilshtain, A.; Warshel, A. *Biochemistry* **2009**, *48*, 3046-3056.

(36) Frushicheva, M. P.; Cao, J.; Chu, Z. T.; Warshel, A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16869-16874.

(37) Frushicheva, M. P.; Cao, J.; Warshel, A. *Biochemistry* **2011**, *50*, 3849-3858.

(38) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210-3235.

(39) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425-443.

(40) Wolfenden, R.; Snider, M. *Acc. Chem. Res.* **2001**, *34*, 938-945.

(41) McClelland, R. A.; Kanagasabapathy, V. M.; Steenken, S. *Can. J. Chem.* **1990**, *68*, 375-382.

(42) Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297-5311.

(43) Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R. *BMC Biochem.* **2006**, *7*, 18 (1-12).

(44) Lenfant, N.; Hotelier, T.; Velluet, E.; Bourne, Y.; Marchot, P.; Chatonnet, A. *Nucleic Acids Res.* **2013**, *41*, D423-D429.

(45) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W. Z.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G. *Mol. Sys. Biol.* **2011**, *7*, 539 (1-6).

(46) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res.* **2004**, *14*, 1188-1190.

(47) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 430-410.

(48) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2011**, *D67*, 235-242.

(49) Leslie, A. G. W.; Powell, H. R. In *Evolving Methods for Macromolecular Crystallography*; Read, R. J., Sussman, J. L., Eds.; Springer: Dordrecht, 2007; Vol. 245, p 41-51.

(50) Evans, P. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2006**, *62*, 72-82.

(51) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2011**, *67*, 355-367.

(52) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. *Acta Crystallogr. Sect. A* **1991**, *47*, 110-119.

(53) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2010**, *66*, 486-501.

(54) Rink, R.; Fennema, M.; Smids, M.; Dehmel, U.; Janssen, D. B. *J. Biol. Chem.* **1997**, *272*, 14650-16457.

(55) Muegge, I.; Tao, H.; Warshel, A. *Protein Eng.* **1997**, *10*, 1363-1372.

(56) Robillard, G.; Shulman, R. G. *J. Mol. Biol.* **1972**, *71*, 507-511.

(57) Wijma, H. J.; Floor, R. J.; Bjelic, S.; Marrink, S. J.; Baker, D.; Janssen, D. B. *Angew. Chem. Int. Ed.* **2015**, *54*, 3726-3730.

(58) The sequences were found by BLAST with the StEH1 sequence as the query, selecting annotated sequences from the results. The residue corresponding to H104 in StEH1 is strongly conserved (with the exceptions discussed in the main text), and, when it is changed, it is most commonly replaced by the structurally-similar glutamine.

(59) Kinetic information for the reverse reaction of the second step was not experimentally available, and therefore the calculated and experimental energetics of the tetrahedral intermediate cannot be compared. All values were normalized to the reactant state and are shown in kcal · mol^{-1}. Note that the calculated free-energy barrier for the second step of the reaction is dependent on the corresponding calculated reaction free energy of the first reaction step, so any uncertainties in this value will also shift the calculated barrier to the second part of the reaction.

(60) These correspond to k_{cat}, which provides an upper boundary for the rate-limiting step and allows for direct comparison to the calculated values. The only exception is the active-site mutant H300N, where the microscopic rate of alkyl-enzyme hydrolysis, k_3, is used, as k_{cat} is not available for both substrates. Additionally, there are no experimental data available for the reactions of the Y235F and H300N variants of StEH1 with (*S*,*S*)-TSO.

Figures & Tables

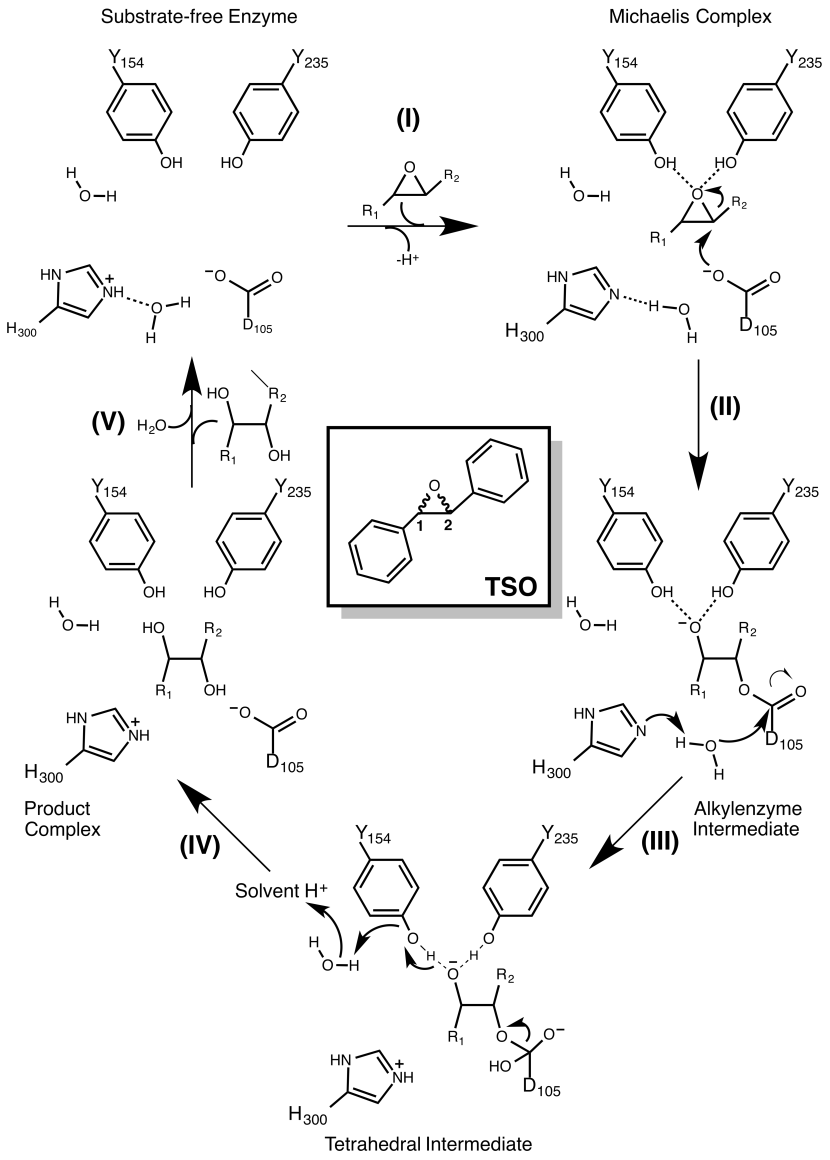


Figure 1: Generalized mechanism for the reaction catalyzed by StEH1. Also included is the structure of the substrate studied, trans-stilbene oxide (TSO).

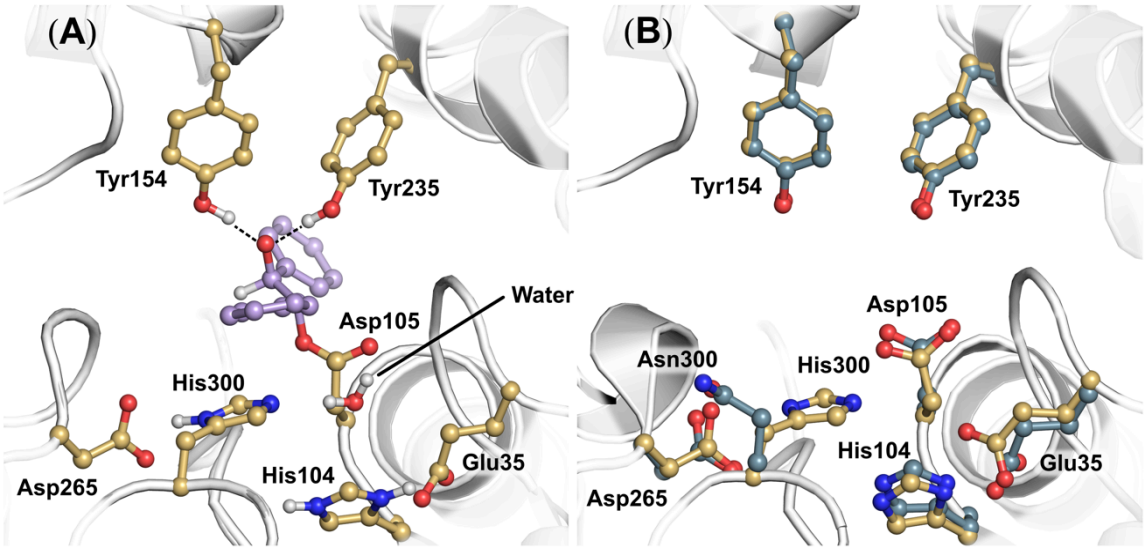


Figure 2: (A) Model of the StEH1 active site, showing the alkyl-enzyme intermediate in which the ring-opened substrate, (*S*,*S*)-TSO (purple sticks), and the nucleophile D105 are covalently bonded. Also shown are other key catalytic residues, as discussed in the main text. The model is based on the StEH1 crystal structure, PDB ID 2CJP[17], after initial equilibration of the covalently-bound intermediate. (B) Overlay of the active sites of wild-type StEH1 (gold carbons) and the H300N mutant (gray carbons), taken directly from the crystal structures (PDB IDs: 1CJP and 4Y9S respectively). For the corresponding electron density maps, see **Figure S1**.

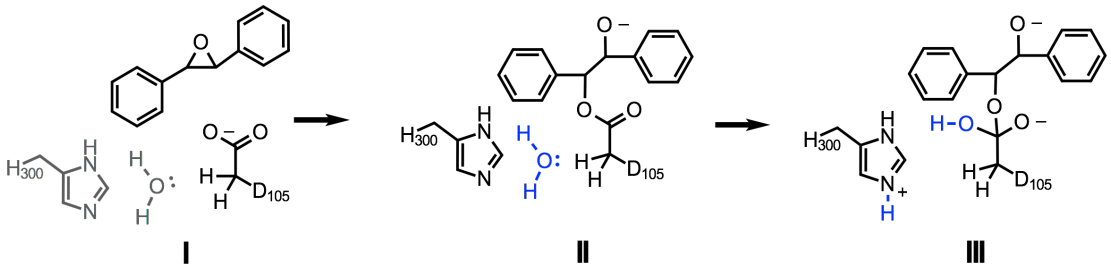


Figure 3: Overview of the valence bond states used to describe (I→II) nucleophilic attack on the epoxide oxygen to form an alkyl-enzyme intermediate, and (II→III) hydrolysis of the alkyl-enzyme intermediate by the hydrolytic water molecule (blue), using H300 as a general base. The same valence bond structures were used to describe the corresponding background reaction in aqueous solution, except that D105 and H300 were modeled as propionate and ethyl imidazole, respectively, for simplicity. The hydrolytic water molecule and the side chain of the general base were only present as spectators in the first step of the enzymatic reaction (shaded in gray), and were not included in the corresponding background reaction (see the discussion in the main text).

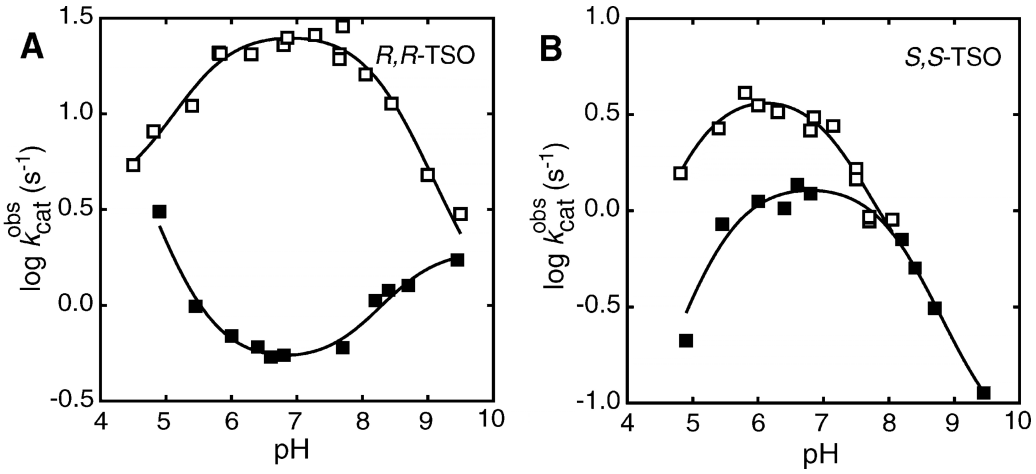


Figure 4: Effect of pH on the turnover number, k_{cat}, for wild-type (□) and E35Q mutant (▪) catalyzed reactions of (**A**) (*R*,*R*)- and (**B**) (*S*,*S*)-TSO. The lines represent fits of equation 5 in ref. [17,21] to experimental data, and describe the titration of a doubly-ionizing system. Adapted with permission from ref. [6]. Copyright © 2007, American Chemical Society.

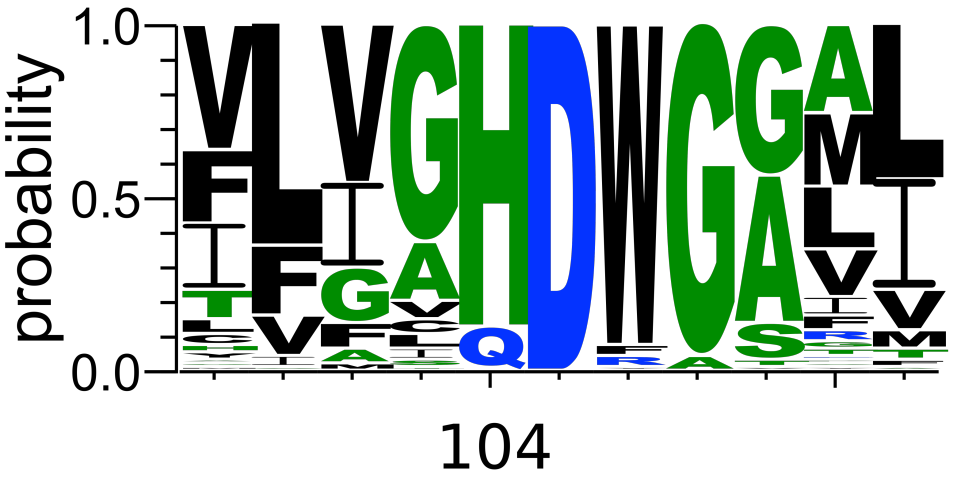


Figure 5: Residue conservation analysis in the epoxide hydrolase-like superfamily. For details see note [58]. Image created with WebLogo 3.4[46].

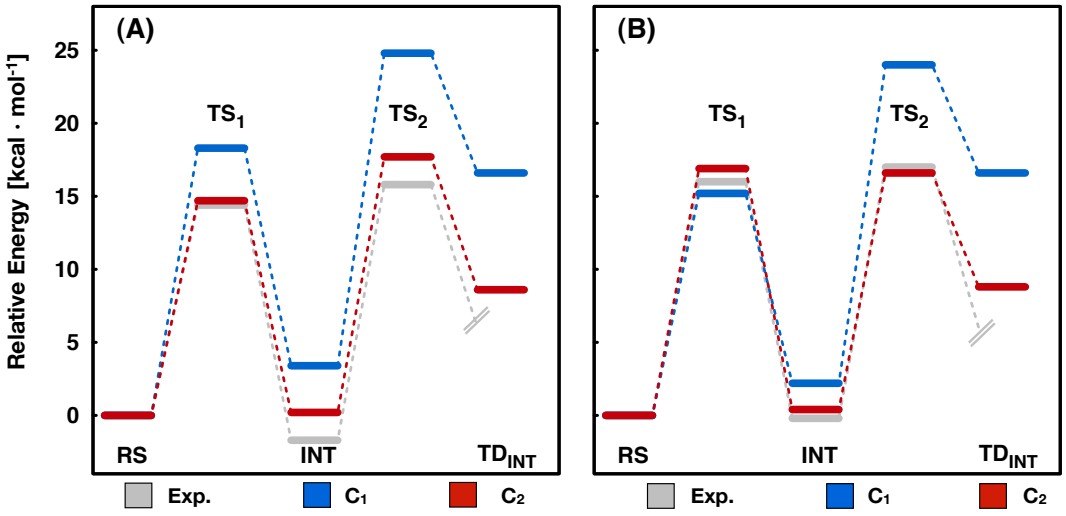


Figure 6: Calculated (C1 attack, blue and C2 attack, red) and experimentally-derived (grey) free-energy profiles for the StEH1-catalyzed reactions of **(A)** (*R*,*R*)-TSO and **(B)** (*S*,*S*)-TSO for wild-type StEH1 (with H104 protonated). This figure is based on data presented in **Table 2**. The experimental values (Exp.) were derived from the kinetic data presented in refs. [4-6,18]. For more details see note [59].

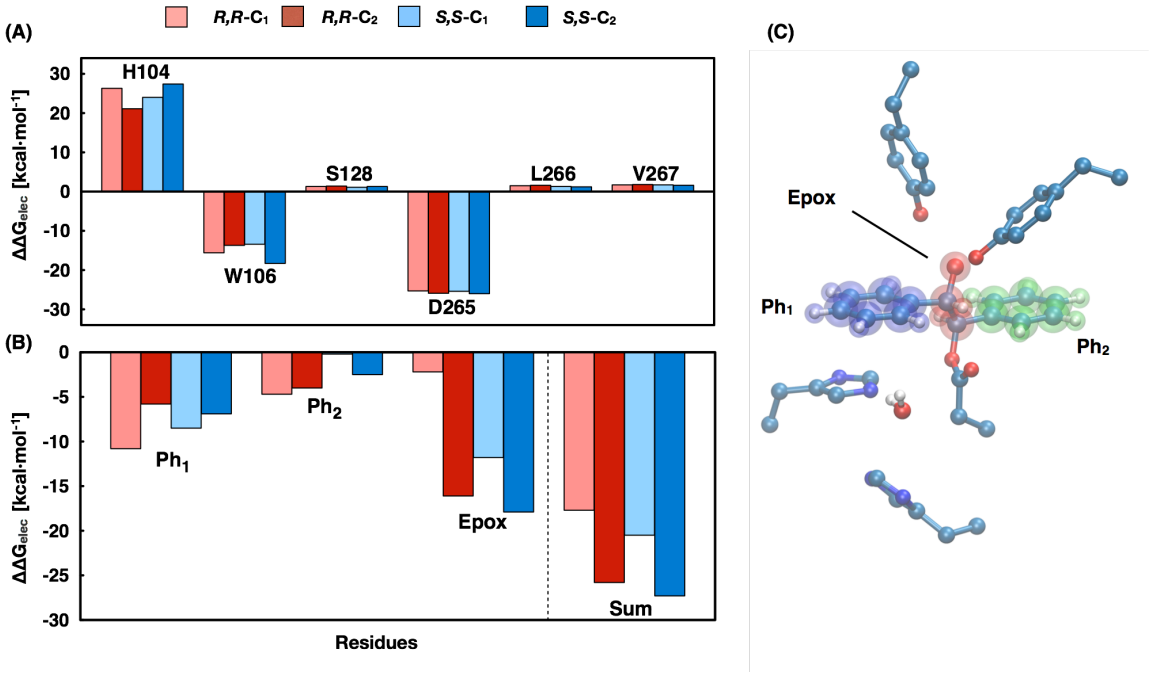


Figure 7: (**A**) Electrostatic contribution of key active-site residues to the calculated activation barrier. (**B**) Electrostatic contribution from different parts of the substrate. (**C**) Orientation of the substrate in the active site, illustrating the annotation used for (B). Positive values correspond to destabilizing contributions and negative values to stabilizing contributions. For clarity, only the residues that contribute 1 kcal·mol^{-1} or more to the calculated activation barrier are shown.

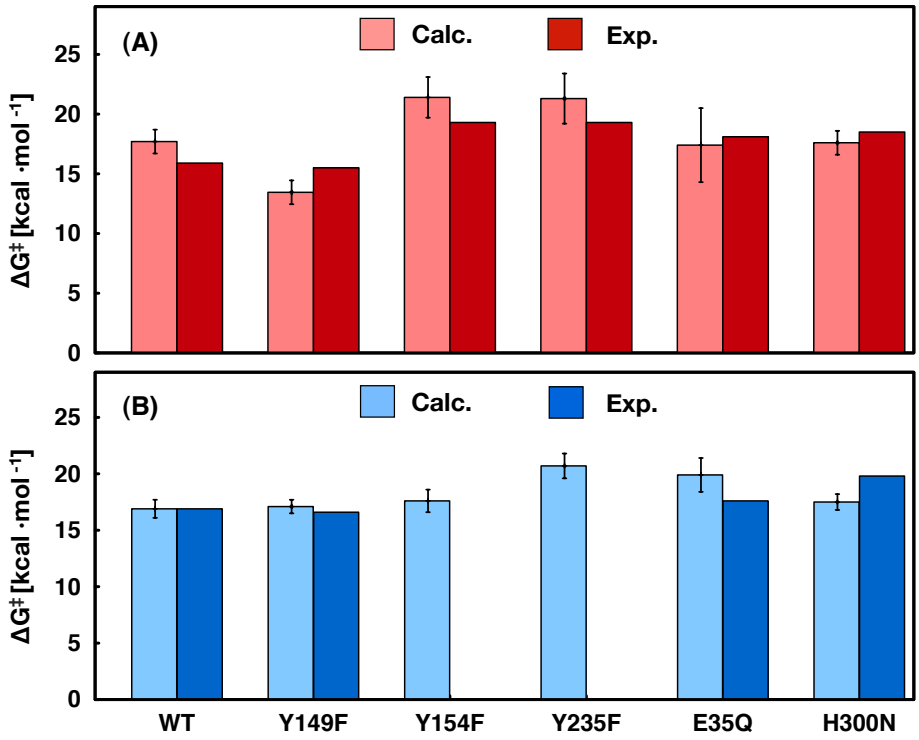


Figure 8: Calculated (light color) and experimentally derived (dark color) activation energies for the StEH1-catalyzed reactions of **(A)** (*R*,*R*)-TSO and **(B)** (*S*,*S*)-TSO. The calculated activation free energies ($\Delta G^{\ddagger}_{calc}$) correspond to the rate-limiting step for each system, with the appropriate protonation state of H104 (see **Tables 2 and S3-S6**). The experimental activation free energies ($\Delta G^{\ddagger}_{expp}$), were derived from the kinetic data presented in ref. [4-6,18]; for further details see note [60].

Table 1: Predicted pK_a values of the relevant residues of StEH1 at different stages of the reaction, *i.e.* substrate-free enzyme (SF), reactant state (RS), alkyl-enzyme intermediate (INT), and tetrahedral intermediate (TD)[(a)]. All pK_a values were estimated with PROPKA 3.1 as described in the Methodology section.

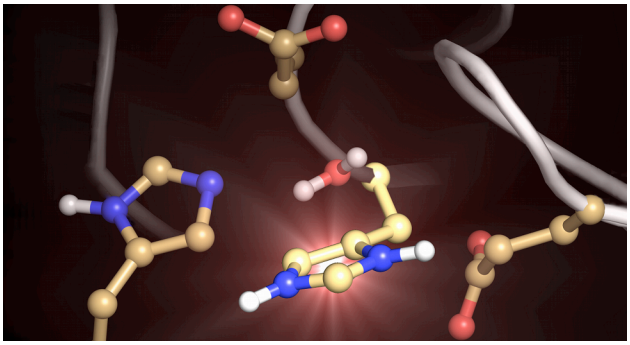
	SF	RS	INT	TD
Wild-type				
E35	3.6	2.3±0.3	2.4±0.4	2.6±0.4
H104	7.3	8.2±0.7	6.7±1.1	8.0±0.5
D105[(b)]	1.8	6.1±1.0	n.d.	n.d.
D265	2.7	2.6±0.3	2.5±0.4	2.6±0.4
H300	11.0	7.1±1.5	7.3±0.7	6.5±0.5
E35Q Mutant				
H104	--	2.1±0.7	2.5±0.5	2.8±0.5
D105(b)	--	5.8±1.0	n.d.	n.d.
D265	--	3.0±0.7	2.7±0.6	3.0±0.6
H300	--	8.4±1.1	7.2± 0.1	7.3±0.1
H300N Mutant[(c)]				
E35	3.4	3.0±0.4	3.3±0.4	n.d
H104	9.0	8.7±0.2	7.8±0.2	n.d.
D105[(b)]	5.3	7.3±0.4	n.d.	n.d.
D265	6.7	6.8±0.6	6.6±0.5	n.d.

[(a)] For RS, INT and TD, the pK_a values were calculated as averages and standard-deviations over the 10 starting snapshots for each of the different enantiomers, as well as structures obtained *via* attack at carbons C1 and C2 at each stage of the reaction pathway. For SF, values were obtained from chain A of the respective crystal structure. [(b)] In the intermediate (INT) and tetrahedral state (TD), this residue is covalently bound to the substrate and therefore its pK_a value was not calculated (n.d.). [(c)] The pK_a for the H300N mutant for RS and INT was obtained after MD simulations with H104 protonated. Due to the missing catalytic water in this crystal structure, the second step was not modeled and pK_a estimates are not presented.

Table 2: Calculated[(a)] and observed[(b)] activation energies ($\Delta G^{\ddagger}$) and reaction free energies ($\Delta G°$) for the alkylation ($\Delta G_1^{\ddagger}$) and hydrolysis ($\Delta G_2^{\ddagger}$) of (*R*,*R*)- and (*S*,*S*)-TSO by wild-type (WT) and mutant StEH1.

		Step I				Step II[(d)]			
	Epoxide Carbon	$\Delta G_1^{\ddagger}$ Calc.	Exp.	$\Delta G_1°$ Calc.	Exp.	$\Delta G_2^{\ddagger}$ Calc.	Exp.	$\Delta G_2°$ Calc.	$\Delta G^{\ddagger}_{kcat}$ Exp.
(*R*,*R*)-TSO									
WT	C2	14.7±1.0	14.4	0.2±1.4	-1.7	17.7±1.6	15.8	8.6±1.7	15.9
E35Q	C2	15.5±0.4	16.1	0.6±1.3	-0.5	17.4±3.1	17.8	6.2±3.3	18.1
Y149F	C2	13.0±0.5	14.3	-3.1±0.9	-1.4	13.5±1.0	15.3	5.6±0.9	15.5
Y154F	C1	18.7±0.8	n.d.	4.1±1.0	n.d.	21.4±1.7	n.d.	12.4±2.1	19.3
Y235F	C2	18.4±1.3	n.d.	5.7±1.5	n.d.	21.3±2.1	n.d.	12.0±2.3	19.3
H300N[(c)]	C1	17.6±1.0	18.5	-0.1±1.1	-0.2	n.d.	n.d.	n.d.	n.d.
(*S*,*S*)-TSO									
WT	C2	16.9±0.8	16.0	0.4±1.1	-0.2	16.6±0.7	17.0	8.8±1.8	16.9
E35Q	C1	17.0±0.8	15.6	2.2±1.2	-1.2	19.9±1.5	17.6	8.9±1.6	17.6
Y149F	C2	17.1±0.6	15.5	-0.3±0.8	0.0	15.5±1.0	16.2	7.4±0.9	16.6
Y154F	C2	17.6±1.0	n.d.	1.1±1.0	n.d.	16.8±1.4	n.d.	9.3±1.5	n.d.
Y235F	C2	17.6±0.5	n.d.	3.3±0.6	n.d.	20.7±1.1	n.d.	12.1±1.2	n.d.
H300N[(c)]	C1	17.5±0.7	19.8	2.5±0.8	-1.2	n.d.	n.d.	n.d.	n.d.

[(a)] All values are averages and standard deviations based on 10 individual EVB simulations generated from different starting structures, in kcal·mol^{-1}. Shown here are the lowest free-energy pathways when considering attack at both carbon atoms. The energetics for all relevant pathways is presented in **Tables S3 – S5**. In almost all cases H104 is protonated and D265 is deprotonated; the exceptions to this are the reactions involving the E35Q and H300N mutants, where H104 is most likely neutral. Additionally, in the H300N variant, D265 is protonated, as discussed in the main text. [(b)] **Exp**. refers to experimental values of the enzyme-catalyzed reaction for the respective step, derived from the kinetic data presented in ref [4-6,18]. [(c)] Due to the low signal-to-noise ratio in the measurements of the steady-state kinetics, it was not possible to determine k_{cat} and K_M values for this mutant. [(d)] $\Delta G^{\ddagger}_2$ corresponds to activation barriers for the hydrolysis steps corrected by adding the calculated activation barrier of the hydrolysis step to the free energy of the intermediate. For the corresponding uncorrected (absolute) activation barriers see **Table S7.** n.d.: Not determined.



For Table of Contents Only