

Catalytic Cycle of Haloalkane Dehalogenases towards Unnatural Substrates Explored by Computational Modeling

Sérgio M. Marques, Zuzana Dunajova, Zbynek Prokop,
Radka Chaloupkova, Jan Brezovsky, and Jiri Damborsky

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.7b00070 • Publication Date (Web): 11 Jul 2017

Downloaded from <http://pubs.acs.org> on July 12, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



Catalytic Cycle of Haloalkane Dehalogenases Towards Unnatural Substrates Explored by Computational Modeling

Sérgio M. Marques,^{1,2} Zuzana Dunajova,¹ Zbynek Prokop,^{1,2} Radka Chaloupkova,^{1,2}

Jan Brezovsky,^{1,2} Jiri Damborsky^{1,2}*

¹Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment RECETOX, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic

²International Clinical Research Center, St. Anne's University Hospital Brno, Pekarska 53, 656 91 Brno, Czech Republic

***Corresponding author.** E-mail: jiri@chemi.muni.cz

KEYWORDS: haloalkane dehalogenases; enzyme engineering; molecular dynamics; catalytic cycle modeling; substrate binding; product release

ABSTRACT

The anthropogenic toxic compound 1,2,3-trichloropropane is poorly degradable by natural enzymes. We have previously constructed the haloalkane dehalogenase DhaA31 by focused directed evolution (Pavlova, M. et al., *Nat. Chem. Biol.* **2009**, 5, 727–73), which is 32 times more active than the wild-type enzyme and is currently the most active variant known against that substrate. Recent evidence has shown that the structural basis responsible for the higher activity of DhaA31 was poorly understood. Here we have undertaken a comprehensive computational study of the main steps involved in the biocatalytic hydrolysis of 1,2,3-trichloropropane to decipher the structural basis for such enhancements. Using molecular dynamics and quantum mechanics approaches we have surveyed: (i) the substrate binding, (ii) the formation of the reactive complex, (iii) the chemical step, and (iv) the release of the products. We showed that the binding of substrate and its transport through the molecular tunnel to the active site is a fast process. The cleavage of the carbon-halogen bond was previously identified as the rate-limiting step in the wild-type. Here we demonstrate that this step was enhanced in DhaA31 due to a significantly higher number of reactive configurations of the substrate, and a decrease of the energy barrier to the S_N2 reaction. C176Y and V245F were identified as the key mutations responsible for most of those improvements. The release of the alcohol product was found to be the rate-limiting step in DhaA31 primarily due to the C176Y mutation. Mutational dissection of DhaA31 and kinetic analysis of the intermediate mutants confirmed the theoretical observations. Overall, our comprehensive computational approach has unveiled mechanistic details of the catalytic cycle which will enable a balanced design of more efficient enzymes. This approach is applicable to deepen the biochemical knowledge of a large number of other systems and may contribute to robust strategies in the development of new biocatalysts.

INTRODUCTION

The utilization of enzyme biocatalysts in the industry has seen a boom in recent years and is still growing. Due to the increasing demand for safer and environmentally friendly industrial processes and their proficiency as catalysts, the enzymes have become highly competitive options as compared to the traditional chemical catalysts^{1,2}. The haloalkane dehalogenases (HLDs, E.C.3.8.1.5) are bacterial enzymes that convert a wide range of halogenated organic compounds into the respective alcohols. This fact has made them very interesting targets for a number of biotechnological applications, such as industrial biocatalysis, detoxification, bioremediation, biosensing and molecular imaging³⁻⁵. However, one main limitation of HLDs is their low-to-moderate activity towards many substrates. Hence, there is high interest in engineering these enzymes to make them more active catalysts. The catalytic activity of the HLDs consists of the hydrolytic cleavage of the carbon-halogen bond with aid of the residues forming the catalytic pentad⁶⁻⁸. Due to the simplicity of this three-step catalytic reaction (Fig. S1), the HLDs have frequently been used as model systems for the study of molecular evolution, development and validation of novel approaches in protein design, as well as several computational tools to assist protein engineering⁹⁻¹³.

Like many other enzymes, the HLDs have their catalytic site buried within the protein core, which is connected to the bulk solvent through transport pathways known as molecular tunnels. These are used for the transport and exchange of substrate, products, water and ions, and may have important functional roles^{14,15}. In the HLDs, the main access tunnel (p1) and auxiliary tunnels (p2, p3) have been studied in detail, allowing finding several key residues that have been subjected to mutagenesis studies^{12,16,17}. Computer-assisted protein design combined with directed

evolution techniques have succeeded in finding new mutants with (i) improved activity towards degradation of toxic pollutants^{16,18}, (ii) higher operational stability and resistance to organic solvents⁹, (iii) increased enantioselectivity¹⁹, (iv) and modified reaction mechanisms²⁰.

1,2,3-trichloropropane (TCP) is a persistent anthropogenic toxic pollutant that can be converted by the HLDs to the less toxic 2,3-dichloropropan-1-ol (DCP). The kinetic parameters towards the hydrolysis of TCP by DhaA (HLD from *Rhodococcus rhodochrous*, NCIMB 13064) are only $k_{\text{cat}} = 0.04 \text{ s}^{-1}$ and $k_{\text{cat}}/K_m = 40 \text{ s}^{-1}\text{M}^{-1}$. Hence, the utility of this enzyme for the bioremediation of this toxic compound would greatly benefit from further improvements in catalytic efficiency^{16,21}. A combination of rational protein design and directed evolution has resulted in variants with higher activity and efficiency towards the hydrolysis of TCP. DhaA31 is the best HLD available to date for hydrolyzing TCP and contains five mutations: I135F, C176Y, V245F, L246I, and Y273F (Fig. 1).

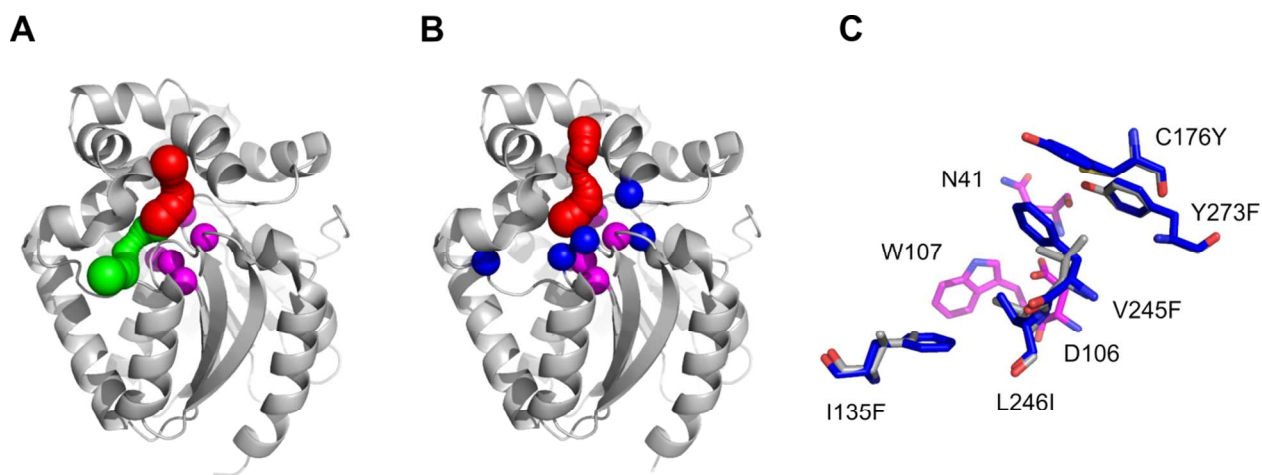


Figure 1. Structures of the DhaA variants studied. A) The crystal structures of the wild-type DhaA from *R. rhodochrous* (DhaAwt; PDB-ID: 4E46) and (B) DhaA31 variant (PDB-ID: 3RK4). The tunnels were calculated using CAVER software¹²; the main tunnel (p1) is shown as the red spheres and the slot tunnel (p2) as green; the catalytic residues are represented as

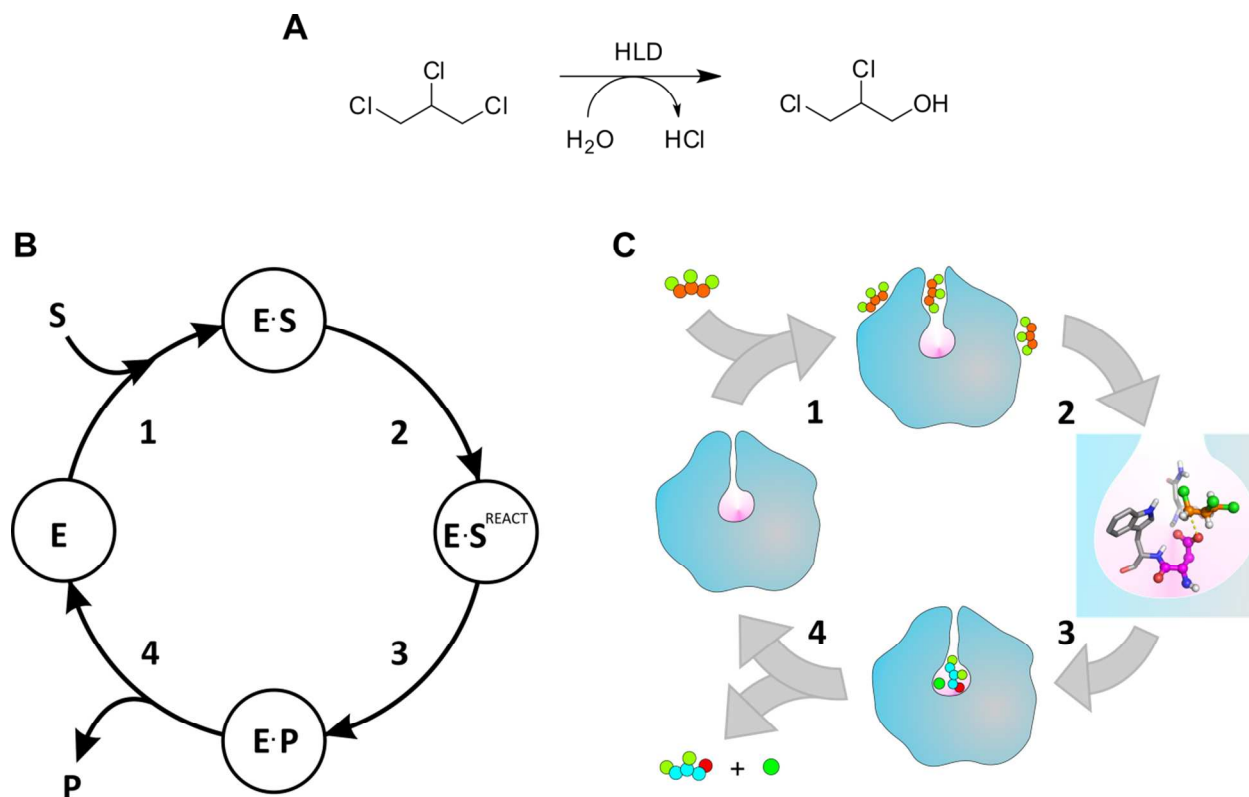
magenta spheres and the mutations as blue spheres (C_{α} atoms)¹⁶. C) The residues I135F, C176Y, V245F, L246I, and Y273F, mutated from DhaA (grey sticks) to DhaA31 (blue sticks), are superimposed with three catalytic residues (magenta sticks) – the nucleophile (D106) and the halide-stabilizing residues (N41 and W107).

Four out of five mutations in DhaA31 are located in the tunnels and introduced bulkier residues, which narrowed the p1 and p2 tunnels leading to its active site (Fig. 1). Nonetheless, DhaA31 was found to have an activity towards TCP higher than the wild-type (DhaAwt), with enhancements of 32-fold on k_{cat} and 26-fold on the catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$). Moreover, while the rate-limiting step in DhaAwt is the $S_{\text{N}}2$ chemical reaction, i.e. the cleavage of the carbon-chlorine bond, in DhaA31 the rate-limiting step was found to be the release of the products. The working hypothesis for the increased $S_{\text{N}}2$ rate was that the narrower tunnels blocked the penetration of water molecules into the catalytic site, which would reduce the hindrance for the $S_{\text{N}}2$ reaction due to the solvent around the nucleophile D106¹⁶. However, recent computational studies at longer time scales did not support this view. This urged us to investigate more deeply the molecular basis for the catalytic enhancements observed in DhaA31.

We have decided to perform a comprehensive computational study of the catalytic cycle with both DhaAwt and DhaA31 enzymes. The catalytic cycle of the HLDs is presented in Scheme 1A and the detailed reaction mechanism is shown in Fig. S1^{22,23}. A simplification of this cycle that includes the relevant steps studied here is presented in Scheme 1^{24,25}. It starts with the binding of the substrate to the enzyme to form the Michaelis complex E·S. This complex can have multiple configurations, some of which are productive and are represented henceforth as the reactive complex, E·S^{REACT}. This corresponds to the population of the ground state complex that contains

the substrate in such a configuration that may result in the subsequent chemical step, i.e., is susceptible to the S_N2 attack by the carboxylate group of D106 to form the alkyl-enzyme intermediate and the halide ion. The intermediate is then attacked by a water molecule, activated by the catalytic base H272, which results in the alcohol product. This sequence of reactions converts the $E \cdot S^{\text{REACT}}$ into the enzyme complexed with the alcohol and halide products, $E \cdot P$. Finally, the $E \cdot P$ complex dissociates and results in the release of the products and regeneration of the free enzyme. A computational dissection of the catalytic mechanism of LinB from *Sphingobium japonicum* UT26 with several halogenated derivatives has previously been performed by Negri *et al.* to provide insight on the differences between different substrates²⁶.

Here we simulated the physical processes using molecular mechanics: the binding of the substrate TCP, its positioning in the reactive conformation and the release of the products DCP and chloride ion. The first reaction step – nucleophilic attack S_N2 on the carbon atom of TCP and release of chloride – was studied using quantum mechanics (Scheme 1b). The second chemical step – the hydrolytic reaction – was neglected for simplification, since it was found not to be a limiting step in any of the studied variants. The S_N2 reaction is the rate-limiting step in DhaAwt, which was shown to have been improved 145-fold in DhaA31, and therefore deserves careful attention¹⁶. From this study, we aim at understanding each one of the steps of the enzymatic cycle at the molecular level. This will allow us to explain the difference in the kinetic profiles of the two enzymes and the increased activity and efficiency of DhaA31 as compared to DhaAwt. Ultimately, the knowledge attained here may help us develop more efficient HLD variants for the hydrolysis of TCP or other substrates.



Scheme 1. Catalytic cycle of the HLD enzymes. A) Schematic representation of the hydrolytic dehalogenation reaction catalyzed by the HLDs converting TCP to DCP. B) Canonical representation of the catalytic cycle: (1) binding of the substrate S to the enzyme E, resulting in Michaelis complex E·S; (2) reorganization of the E·S complex to generate a reactive configuration E·S^{REACT}; (3) chemical steps that convert the halogenated substrate into the alcohol and halide products complexed with the enzyme E·P; (4) release of the products P from the E·P complex. C) The same cycle showing an illustration of the structures under study here: the HLD enzyme (blue surface) with a tunnel connecting the buried active site (pink cavity) to the bulk solvent; TCP (orange and green balls) binding to the HLD; TCP in reactive configuration within the active site, the nucleophile D106 (magenta) and two halide-stabilizing residues N41-W107 (grey); DCP (cyan, green and red balls) and chloride ion (dark green ball) products in the active site, that will be released.

MATERIALS AND METHODS

Molecular dynamics with unbound enzymes

Preparation of the protein structures

The crystal structures of DhaA31 and DhaAwt proteins (UniProtKB accession number P0A3G2) were obtained from the RCSB Protein Data Bank²⁷ (PDB entries 3RK4 and 4E46, respectively). The duplicated side chains of some residues were removed, keeping only the conformation closest to those observed in other crystal structures (1CQW and 1BN7). The water molecules and ions were removed and all hydrogen atoms were added using the H++ server, calculated in implicit solvent at pH 7.5, 0.1 M salinity, with an internal dielectric constant of 10 and external of 80²⁸. The protonation state of the catalytic H272 was preserved as doubly protonated, as predicted by H++. The water molecules from the original crystal structure were then added to DhaA31. In the case of DhaAwt, the more hydrated structure 1CQW was aligned with the protein structure prepared as previously described, and its water molecules were incorporated into the later structure. The tLEaP program of AmberTools 14²⁹ was used to prepare the topology and trajectory input files for performing the dynamics simulation. The ff10, ff12SB, or ff14SB force fields^{30–33} were specified for testing the ideal conditions, and ff12SB was used in all subsequent simulations. The Na⁺ and Cl[−] ions required for ionic strength of 0.1 M were added, and a TIP3P³⁴ truncated octagonal box of water molecules with edges 10 Å away from the original system was added.

Molecular dynamics simulations

The PMEMD.CUDA^{35,36} module of AMBER 12³⁷ was used for running the MD simulations. Four minimization steps, composed of 500 cycles of steepest descent and 500 cycles of

conjugate gradient, were performed as follows: (i) in the first step, all the heavy atoms of the protein were restrained with 500 kcal/mol·Å² harmonic force constant; (ii) in the following three steps, only the backbone atoms of the protein were restrained, respectively, with 500, 125, and 25 kcal/mol·Å² force constant, respectively. A fifth minimization step, composed of 100 cycles of steepest descent and 400 cycles of conjugate gradient, was performed without any constraints. The subsequent MD simulations employed periodic boundary conditions, the particle mesh Ewald method for treatment of the long-range interactions beyond a 10 Å cut-off³⁸, the SHAKE algorithm³⁹ to constrain bonds involving the hydrogen atoms, the Langevin thermostat with collision frequency 1.0 ps⁻¹, and a time step of 2 fs. Equilibration dynamics were performed in two steps: (i) 20 ps of gradual heating from 0 to 310 K, under constant volume, restraining the protein atoms with 5 kcal/mol·Å² harmonic force constant; (ii) 2 ns of unrestrained MD at constant pressure (1 bar) and constant temperature (310 K). Finally, the production MD simulations were run with no restraints using the settings employed in the second equilibration dynamics, saving the energy and coordinates every 10 ps with total simulation lengths of 500 ns. Independent minimization-equilibration-production sequences were performed in quadruplicate for each protein.

Accelerated molecular dynamics simulations

The accelerated MD (aMD) method consists of adding a boost to the energy of the system to decrease the depth of the energy basins and increase the transitions between conformational states. The new energy profile, $V^*(r)$, is given by Eq. 1, and the boost of energy, $\Delta V(r)$, is defined by Eq. 2, where $V(r)$ is the real energy function, E is the threshold energy above which there is no energy boost, and α (> 0) is the acceleration factor. The energy boost, increases with the increase of the difference between $V(r)$ and the energy threshold E ; on the other hand, it

decreases with the increase in α , and for $V(r) < E$, the energy profile resembles more the constant potential with value of E as α approaches 0^{40,41}. Since the energy barriers are reduced, up to the value of the energy threshold, E , the conformational sampling is enhanced. This method has been used to explore greater conformational spaces of biomolecules and sample more than would be possible with common classical MD, being equivalent to timescales more than three orders of magnitude larger. Hence, it is a complementary method to the classical MDs and can be very useful for studying systems with slow conformational dynamics (μs – ms time scales)^{42–48}.

Several authors recommend an empirical approach for assessing the acceleration parameters as defined by Eqs. 3 and 4^{42,29,43}. These equations allowed to set: the torsional (or dihedral) energy threshold, E^{dih} , calculated from the average dihedral energy in a classical MD simulation (V_0^{dih}) and the number of protein residues (N^{resid}); the dihedral acceleration factor, α^{dih} , which depends on N^{resid} only; the total potential energy threshold, E^{tot} , calculated from the average potential energy of the system in a classical MD simulation (V_0^{tot}) and the total number of atoms in the system (N^{atom}); and the total potential acceleration factor, α^{tot} , which depends on N^{atom} .

$$V^*(r) = \begin{cases} V(r), & V(r) \geq E \\ V(r) + \Delta V(r), & V(r) < E \end{cases} \quad \text{Eq. 1}$$

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))} \quad \text{Eq. 2}$$

$$E^{dih} = V_0^{dih} + 3.5 N^{resid}; \quad \alpha^{dih} = 0.2 (3.5 N^{resid}) \quad \text{Eq. 3}$$

$$E^{tot} = V_0^{tot} + 0.2 N^{atom}; \quad \alpha^{tot} = 0.2 N^{atom} \quad \text{Eq. 4}$$

Dual torsional and total potential energy boosts were applied to our systems. V_0^{dih} and V_0^{tot} average energies were extracted for each system from a 20 ns production simulation of classical

MD at the same initial conditions and were used to calculate the respective energy thresholds (E) and acceleration factors (α) according to Eqs. 3 and 4. The aMD simulations were carried out with the PMEMD.CUDA^{35,36} module of AMBER 12³⁷. They were performed with DhaA31 and DhaAwt without any constraints after the systems were fully equilibrated, as previously described for the MD simulations. The calculation step was 2 fs and the energy and coordinates were saved every 2 ps, in a total of 200 ns aMD simulation. Four aMD replicates were run for each protein. The root mean square deviation (RMSD) of the catalytic residues, calculated as described below, was used to assess the proteins' catalytic stability. As we were only interested in the protein's catalytically active conformations, any aMD simulation presenting irreversible disruption of the catalytic site was discarded.

Tunnel calculation

CAVER version 3.02¹² was used to calculate and cluster the tunnels in the MD and aMD simulations with DhaA31 and DhaAwt. The tunnels were calculated for every snapshot of each simulation, using a probe radius of 1.0 Å, a shell radius of 3 Å and shell depth 4 Å. The starting point for the tunnel search was a point in the center of the active site cavity, defined by the center of mass of four atoms from surrounding residues: Y176-C_β, F205-C_α, L209-C_α and H272-C_α for DhaA31, and C176-C_β, F205-C_α, L209-C_α and H272-C_α for DhaAwt. This point was regarded as the center of the active site throughout this paper. The clustering was performed by the average-link hierarchical Murtagh algorithm, with a weighting coefficient of 1 and clustering threshold of 5.0. Approximate clustering was allowed only when the total number of tunnels was higher than 100,000, and it was performed using 15 training clusters. A tunnel was considered open when the radius of its narrowest point, the bottleneck, was ≥ 1.4 Å.

Molecular dynamics with ligands

Ligand construction and parametrization

The TCP and *R*-DCP structures were constructed and minimized using Avogadro 1.1.1⁴⁹. The initial minimization was performed by the Auto Optimization Tool of Avogadro, using the UFF force field⁵⁰ with steepest descent algorithm. The resulting structure was then submitted to further optimization and calculation of partial atomic charges using the R.E.D. server⁵¹. The optimization was performed by the Gaussian 2009 D.01 program, interfaced with the R.E.D. server, with the Hartree-Fock method and 6-31G* basis set. The multi-orientation RESP fit was performed with the RESP-A1A charge model, to obtain the MOL2 file with the partial atomic charges of the ligand, to be used in the dynamics simulations. The structure and charges of the *S*-DCP enantiomer were obtained from the *R* enantiomer by manually converting it to its mirror image. PREPI parameter files of TCP and DCP were prepared from the MOL2 structure, using the Antechamber module of AmberTools 14²⁹, and compiled using AMBER force field atom types.

Molecular docking

To obtain the initial structures for studying the TCP reactivity, or the release of the *R*-DCP and *S*-DCP products, the optimized structures of those molecules were docked into the protein structures of DhaA31 and DhaAwt using AutoDock 4⁵². The region of interest for molecular docking was defined by a $20 \times 20 \times 20$ Å box centered at the carboxylic OD1-atom of the nucleophilic D106 residue. AutoDock atom types were assigned to the alcohol product and protein structures by the AutoDockTools4 module of MGLTools 1.5.6⁵² and the input files were converted to AutoDock-compliant format. Electrostatic and van der Waals energies, H-bonds, and desolvation energy were calculated for the protein with AutoGrid 4.0⁵² prior to docking. The docking was performed with 250 runs of Lamarckian Genetic algorithm and initial population

size of 300. The final orientations from every docking run were clustered with RMSD tolerance of 2 Å. Two main binding conformations of TCP were found, the reactive *R* and *S* orientations (i.e., upon reaction they would yield the *R*- and *S*-enantiomer of DCP, respectively). These two configurations were used as different starting points for the subsequent simulations to have unbiased tracking of the preferred reactive orientation in each protein.

System preparation

To prepare the initial structures for studying TCP's binding to DhaA31 and DhaAwt, three molecules of the substrate were added to each protein structure. This proved to be the suitable number to mimic the concentration of substrate similar to the experimentally used (0.01 M), according to the volume of the systems after equilibration. The same protein structures were used as mentioned above. The three molecules of TCP were manually added and placed $\geq$ ca. 5.0 Å away from any atom of the protein with PyMOL 1.7.4⁵³, each along a different Cartesian axis. The substrate molecules were then rotated with respect to the protein, to place the TCP in different regions and generate randomized starting points. In total, four initial structures were created, in which the TCP molecules were globally covering most of the protein surface. For the systems containing the docked alcohol (*R*- or *S*-DCP), the chloride ion present at the halide-binding site of the respective crystal structures was also extracted and incorporated into those structures. The structures of DhaA31 and DhaAwt proteins complexed with TCP or DCP and chloride were hydrated with the waters from the respective crystal structures, as described for the unbound enzymes, and any water overlapping with the protein or ligand atoms was removed. The tLEaP program was used as described above to prepare the topology and trajectory input files for the dynamics simulation, using the ff12SB force field. In the simulations with three TCP molecules in the bulk solvent, a truncated octagonal box of water molecules with edges 12

1
2
3 Å away from the protein system was added, while with TCP or DCP docked in the active site the
4
5 octagonal box of water was defined with the edges 10 Å away from the protein.
6
7

8 *MD and aMD simulations of the complexes*
9

10 The MD simulations were carried out under the same conditions as described for the unbound
11 proteins, while applying a smoother and extended equilibration cycle. The first four
12 minimization steps were composed of 2,500 cycles of steepest descent followed by 7,500 cycles
13 of conjugate gradient, with the same restrains sequence as before. The fifth minimization step,
14 performed without any constraints, was composed of 5,000 cycles of steepest descent and 15,000
15 cycles of conjugate gradient. The subsequent equilibration dynamics were performed in twelve
16 steps: (i) 20 ps of gradual heating from 0 to 310 K, under constant volume, restraining the
17 protein atoms and ligand with 200 kcal/mol·Å² harmonic force constant; (ii) ten MDs, of 400 ps
18 each, at constant pressure (1 bar) and constant temperature (310 K), with gradually decreasing
19 restraints on the backbone atoms of the protein and heavy atoms of the ligand with harmonic
20 force constants of 150, 100, 75, 50, 25, 15, 10, 5, 1, 0.5, and 0 kcal/mol·Å². The production MD
21 simulations were run for 200 ns without any restraints. Each starting structure containing TCP in
22 the bulk solvent was simulated once (four independent MDs were performed), each structure
23 with docked TCP in *R* and *S*-orientation was simulated in octuplicate, and with docked *R* and *S*-
24 DCP was simulated in quadruplicate.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46 The aMD simulations were performed on each system after an entire equilibration cycle,
47 without any constraints, for a total of 100 ns simulation time. Dual torsional and total potential
48 energy boosts were applied to the systems as described before. Each starting structure containing
49 TCP in the bulk solvent was simulated twice (eight independent aMDs were performed), and
50
51
52
53
54
55
56
57
58
59
60

each structure with docked TCP in *R* and *S*-orientation, and with *R*- and *S*-DCP, was simulated in octuplicate.

Simulation treatment and analysis

Post-simulation treatment and analysis were performed using the cpptraj⁵⁴ module of AmberTools 14²⁹, such as centering and alignment of trajectories, removal of water molecules and ions, measurement of distances and angles, calculation of RMSD values, B-factors, calculation of secondary structure elements, solvation spheres and radial distribution function of water (RDF). Unless stated otherwise, the RMSD of the protein atoms are reported with respect to the crystal structures and were calculated with fitting excluding the terminal residues with highest fluctuation (1-3 residues). To determine the stability of the catalytic pentad during the MD and aMD simulations we calculated the RMSD of the heavy atoms of residues N41, D106, W107, E130 and H272, excluding the equivalent oxygen atoms of the respective carboxylic groups. For a simulation to be considered to have a stable catalytic site we required such RMSD values to remain below $\leq$ ca. 1 Å, or to be reversible to such values within the simulation time. For calculating the hydration of the nucleophile residue and the active site, we monitored the number of waters within 3.4 Å and 8 Å from the C_γ-atom of D106, respectively, and calculated the RDF of waters around the same atoms over each simulation. The trajectories were visualized using PyMOL 1.7.4⁵³ and VMD 1.9.1⁵⁵. Grubbs' test⁵⁶ was used for detection of outliers in the estimation of different mean properties. Any outliers found were excluded from the calculation of the respective mean values, standard error of the mean (SEM), standard deviation (SD), and *P*-values. The SEM was calculated based on the respective SD and number of independent simulations or experiments. The *P*-values were calculated with the unpaired *t*-test for quantifying

the statistical significance of the difference between two data sets based on the respective averages, SDs and sample size.

The position of the ligands (TCP, DCP and Cl⁻) were evaluated by the distance of their center of mass to the active site cavity, defined by the same point used as the origin of the calculated tunnels. To discard the fluctuation noise, the running average of the distance (averaging over 0.5 ns, equivalent to 250 snapshots) was used to define the location of those ligands according to the boundary values: inside the protein tunnel for average distances ≤ 9 Å, at the tunnel mouth for 9 Å $<$ distances < 13 Å, and outside for distances ≥ 13 Å. These threshold values were chosen according to the length calculated for the p1 tunnel and proved reasonably accurate by visual inspection. The residence times were accounted accordingly. The transport rates of substrates or products aim at accounting the full binding or full unbinding with respect to the active site, and therefore more demanding thresholds were required. A binding event was accounted when the ligand entered from the exterior to average distances < 6 Å from the active site, and a release event when it was released from the interior to average distances > 20 Å. The occupancy density of the ligands was calculated and displayed with VMD 1.9.1⁵⁵. For that, the 20 ps-spaced snapshots from all the simulations of each type were loaded and aligned by minimizing the displacement of the backbone atoms. The linear interaction energies (LIE) of TCP and DCP with the proteins and the respective hydrogen bonds were calculated for the MD simulations with ccpptraj⁵⁴ module of AmberTools 14²⁹. The reactive configurations of TCP in the active site were evaluated in order to find the positions for the S_N2 reaction during the MD simulations with TCP docked in the active site. This was done by measuring, for each snapshot, the distances and angles between the nucleophile and substrate atoms, according to Hur *et al.*⁵⁷: the distance between one of the nucleophile's carboxyl oxygen atoms (D106-OD atoms) and the carbon atom

of TCP to be attacked had to be ≤ 3.41 Å, and the angle formed by that oxygen, carbon and the halide atoms must be $\geq 157^\circ$. As reported before⁵⁸, we also required at least a weak H-bonding between the reactive chlorine atom and the halide stabilizing residues, defined by the filter of the distance between the halide and the indole hydrogen of W107, or the side chain NH hydrogen of N41, to be ≤ 3.0 Å. The fraction of reactive complexes in each MD was obtained by dividing the number of snapshots in reactive configuration found in each MD by the total number of snapshots per individual simulation (10^5).

Adiabatic mapping

To evaluate and compare the energetics of the first reaction of TCP in the two enzymes and calculate the respective energy barrier, $\Delta G^\ddagger$, we studied the potential energy surface (PES) along the reaction coordinate using a QM/MM approach^{59,60}. For that the 100 best reactive complexes of each protein with TCP in each orientation (*R* and *S*) were chosen according to the closest distance between the reacting atoms (TCP C-atom and D106 OD-atom), and were subjected to the QM/MM calculations. In order to avoid biasing unevenly the quality of the structures tested, the reactive complexes from the outlier MDs found previously were excluded from this selection. The topology of each structure was prepared by tLEaP module using the ff12SB force field for the proteins and the PREPI parameters for the TCP molecule. The complexes were minimized in vacuum (*igb*=6). Five rounds of optimization, each one consisting of 500 cycles of steepest descent followed by 500 conjugate gradient cycles, were performed as: (i) one step with all heavy atoms restrained with 500 kcal/mol·Å² harmonic force constant, and (ii) four steps with decreasing restraints on the protein backbone atoms with 500, 125, 25 and 1 kcal/mol·Å² force constant. An adiabatic mapping along the reaction coordinate was performed by the sander module of AMBER 14²⁹. The QM part of the system contained TCP and the side-chains of the

halide stabilizing residues (N41 and W107) and the catalytic aspartate (D106) and had charge -1. The semi-empirical PM6 Hamiltonian was used to treat the QM part of the system⁶¹ and the ff12SB force field to treat the MM part. The QM/MM boundary was treated through explicit link atoms and the cutoff for the QM/MM charge interactions was set to 999 Å. The backbone was constrained with a force constant of 1.0 kcal/mol·Å². The reaction coordinate was defined as the distance between the nearest OD atom of D106 and the C-atom of TCP under attack. The tracking along the reaction coordinate was performed in decrements of 0.05 Å, each involving 1,000 minimization steps of the limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm⁶². $\Delta G^\ddagger$ was calculated as the difference between the lowest energy of the ground state and the energy of the transition state (respectively, GS and TS). The ratio between the kinetic rates of the S_N2 reaction for two different systems was calculated based on the difference between their energy barriers, $\Delta\Delta G^\ddagger$, according to the transition state theory (Eq. 5).

$$k/k' = e^{\frac{\Delta\Delta G^\ddagger}{RT}} \quad \text{Eq. 5}$$

Binding free energy

The molecular mechanics/generalized Born surface area (MM/GBSA)^{63,64} method was applied to calculate the binding free energy of TCP or DCP with the proteins, ΔG_{bind} , and the respective residue-by-residue interactions. This was performed on every structure of the reactive complexes, or on the entire MD simulations. The tLEaP was used to prepare the coordinate files from each PDB structure of the reactive complex, and the *ante-MMPBSA.py* module of AmberTools 14²⁹ was used to remove the solvent and ions from the original topology files of each system, define the Born radii as *mbondi2*, and generate the corresponding topology files for the *complex*, *receptor* and *ligand*, to be used in the MM/GBSA calculations. The ΔG_{bind} of the

ligand in each structure was calculated with *MMPBSA.py* program⁶³. The generalized Born method was used (*&gb* namelist) with implicit generalized Born solvent model (*igb*=8) and salt concentration as before (*saltcon*=0.1). The solvent accessible surface area was computed with an LCPO algorithm⁶⁵. Decomposition of the pairwise interactions was generated (*&decomp* namelist), with discrimination of all types of energy contributions (*idecomp*=4) for the whole residue (*dec_verbose*=0).

For the MD simulations, *cpptraj*⁵⁴ was used to remove all the ions and water molecules from the full MD trajectories. The *ante-MMPBSA.py* module of AmberTools 14²⁹ was used to remove the solvent and ions from the original topology files of each system, define the Born radii as *mbondi2*, and generate the corresponding topology files for the *complex*, *receptor* and *ligand*, to be used in the MM/GBSA calculations. The ΔG_{bind} of the ligand was calculated in parallel with the *MMPBSA.py.MPI* module⁶³ for all 10 ps-spaced snapshots of each simulation. The same settings were used as described for the individual reactive complex structures.

Alanine mutants

The alanine mutants of DhaA31 were prepared using the mutagenesis tool of PyMOL 1.7.4⁵³, from the unbound protein structure prepared and protonated as described above, to generate the single point mutations Y176A, F245A and I246A. TCP was docked into the active site of each mutant and the two binding modes obtained, the *R* and *S*-oriented, were used to build the starting structures for the MD simulations, prepared as described for DhaA31-TCP complexes. Each complex containing the *R*- and *S*-orientated TCP docked to the alanine mutant was simulated for 200 ns in quadruplicate.

Experimental assays

Construction of DhaA variants

The mutant recombinant genes of *dhaA04* (C176Y)⁶⁶, *dhaA31* (I135F, C176Y, V245F, L246I and Y273F)¹⁶, *dhaA137* (L246I) and *dhaA138* (V245F)⁶⁷ were constructed using techniques of directed evolution and site-directed mutagenesis as described previously. The recombinant gene of *dhaA133* (C176Y, V245F) was synthesized artificially (GeneArt, LifeTechnologies, Germany) according to the wild-type sequence and the gene sequence (Fig. S2) was optimized for an expression in *Escherichia coli*. All genes were cloned into an expression vector pET21b (Novagen, Madison, USA) using restriction endonucleases *NdeI* and *XhoI/HindIII* (Fermentas, Burlington, Canada) and T4 DNA ligase (Promega, Madison, USA).

Enzyme expression and purification

To overproduce all DhaA variants in *E. coli* BL21(DE3) cells, the corresponding genes were transcribed by T7 RNA polymerase, which is expressed by the isopropyl β -D-1-thiogalactopyranoside (IPTG)-inducible lac UV5 promoter. Cells containing these plasmids were cultured in Luria broth medium at 37 °C. When the culture reached an optical density of 0.6 at a wavelength of 600 nm, enzyme expression (at 20 °C) was induced by the addition of IPTG to a final concentration of 0.5 mM. The cells were subsequently harvested and disrupted by sonication using a Soniprep 150 (Sanyo Gallenkamp PLC, Loughborough, England). The supernatant was collected after centrifugation at 100 000g for 1 h. The crude extract was further purified on a HiTrap Chelating HP 5 ml column charged with Ni²⁺ ions (GE Healthcare, Uppsala, Sweden). The His-tagged enzyme was bound to the resin in the presence of equilibration buffer (20 mM potassium phosphate buffer pH 7.5, 0.5 M sodium chloride, 10 mM imidazole). Unbound and non-specifically bound proteins were washed out with buffer containing 37.5 mM imidazole. The target enzyme was eluted with buffer containing 300 mM imidazole. The active fractions were pooled and dialyzed against 50 mM potassium phosphate buffer pH 7.5 overnight.

Stopped-flow fluorescence analysis

The binding experiments were performed in an SFM-300 stopped-flow instrument (BioLogic, France) combined with MOS-500 spectropolarimeter. The fluorescence emission from tryptophan residues was observed through a 320 nm cut-off filter upon excitation at 295 nm. The reactions were performed at 30 °C in a pH 8.6 glycine buffer. The dissociation constants were calculated from the amplitudes of fluorescence quench for rapid-equilibrium phase. Amplitudes and observed rate constants of the slow kinetic phase were evaluated from single exponentials. Data fitting was performed by using software Origin 6.1 (OriginLab Corp., USA).

Dehalogenase activity test

Enzymatic activity towards TCP was assayed using the colorimetric method developed by Iwasaki et al.⁶⁸ The release of halide ions was analyzed spectrophotometrically at 460 nm using a SUNRISE microplate reader (Tecan, Grödig/Salzburg, Austria) after reaction with mercuric thiocyanate and ferric ammonium sulfate. The dehalogenation reaction was performed at 37 °C in 25 ml Reacti-flasks with Mininert valves. The reaction mixture contained 10 ml glycine buffer (100 mM, pH 8.6) and 10 µl of TCP. The reaction was initiated by the addition of enzyme in concentration 1.5-5.0 mg/ml. The reaction was monitored by withdrawing 1 ml samples at periodic intervals from the reaction mixture and immediate mixing of the samples with 0.1 ml 35% nitric acid to terminate the reaction. The dehalogenation activity was quantified as the rate of product formation with time.

Steady-state kinetics

Steady-state kinetics of DhaA variants with TCP were determined by the isothermal titration calorimetry method (ITC)⁶⁹⁻⁷¹, using a VP-ITC microcalorimeter (MicroCal, USA) at 37 °C. For this purpose all enzymes were dialyzed against 100 mM glycine buffer, pH 8.6 overnight. The

same buffer was used to dissolve the substrate to a final concentration of 13.1 mM. Substrate concentration was verified by gas chromatography (Finnigen, USA). The reaction mixture vessel of the microcalorimeter was filled with 1.4 ml of enzyme solution at a concentration 0.028- 0.547 mg/ml. The enzyme was titrated in 150-s intervals in the reaction mixture vessel with increasing amounts of a substrate, while pseudo-first-order conditions were maintained. After every injection, the peak of heat due to dilution was observed, which was followed by a relaxation of the signal to a level corresponding to the heat produced by the enzymatic reaction (Fig. S3). The rate of the heat generated by the enzymatic reaction is equivalent to the decrease in instrumental thermal power. Every injection increased the substrate concentration, leading to a further increase in enzyme reaction rate. The reactions proceeded at a steady-state since only negligible substrate depletion occurred after every injection. A total of 28 injections were carried out during every titration cycle. The reaction rates reached after every injection (in units of thermal power) were converted to enzyme turnover by using apparent molar enthalpy (ΔH_{app}), as shown in Eq. 6, where [P] is the molar concentration of product generated and Q is the enzymatically generated thermal power.

$$\text{Rate} = \frac{d[P]}{dt} = \frac{1}{V \cdot \Delta H_{app}} \cdot \frac{dQ}{dt} \quad \text{Eq. 6}$$

The ΔH_{app} was determined in a separate experiment by allowing the reaction to proceed to the completion, and then integrating the signal to obtain the total heat evolved (Eq. 7), where [S] is the molar concentration of substrate converted.

$$\Delta H_{\text{app}} = \frac{1}{[S]_{\text{Total}} \cdot V} \int_{t=0}^{t=\infty} \frac{dQ(t)}{dt} \cdot dt \quad \text{Eq. 7}$$

The calculated enzyme turnover plotted against the actual concentration of the substrate after every injection (Fig. S4) was then fitted by nonlinear regression to kinetic models using the software Origin 8.0 (OriginLab, USA). The steady-state kinetic parameters were calculated by using the equation describing cooperativity (Eq. 8), substrate inhibition (Eq. 9), and cooperativity with substrate inhibition (Eq. 10).

$$\frac{v}{V_{\text{lim}}} = \frac{[S]^n}{K_{0.5}^n + [S]^n} \quad \text{Eq. 8}$$

$$\frac{v}{V_{\text{lim}}} = \frac{[S]}{K_m + [S] \left(1 + \frac{[S]}{K_{\text{si}}} \right)} \quad \text{Eq. 9}$$

$$\frac{v}{V_{\text{lim}}} = \frac{[S]^n}{K_{0.5}^n + [S]^n \left(1 + \frac{[S]}{K_{\text{si}}} \right)} \quad \text{Eq. 10}$$

In the previous equations, k_{cat} is the catalytic constant ($k_{\text{cat}} = V_{\text{lim}}/[E]$), K_m is the Michaelis constant, $K_{0.5}$ is the concentration of substrate at half-maximal velocity ($K_{0.5} = K_m$ when $n = 1$), K_{si} is the substrate inhibition constant and n is Hill coefficient.

RESULTS

Unbound enzymes

DhaA31 and DhaAwt enzymes were studied by classical molecular dynamics (MD) and accelerated molecular dynamics (aMD) to track the conformational changes naturally occurring in the native ligand-free structures. These were performed in quadruplicate simulations of 500 and 200 ns, respectively. The aMD simulations have proven very useful to study rare or slow events in biomolecules, by sampling the equivalent to much longer time scales, in some cases beyond three orders of magnitude larger than the actual simulation time^{42–45,47,72}. In this work, we are dealing with potentially slow processes of ligand transport through the tunnels of DhaAwt and DhaA31, and therefore we decided to use this method.

A preliminary study was carried out to evaluate the performance of three force fields available with AMBER with our systems, the ff10, ff12SB and ff14SB force fields. This study showed that ff10 produced quite unstable MD simulations, with average root mean square deviation (RMSD) values ca. 1.5 Å for the protein backbone atoms and frequently unstable catalytic residues, evaluated by the irreversible drifting of their RMSD to values above 1.0–1.5 Å. Despite the ff14SB produced slightly more stable MD simulations than ff12SB, it also generated more frequently aMD simulations with unstable catalytic site. This led us to opt for ff12SB over ff14SB in all the subsequent simulations with DhaAwt and DhaA31.

The simulations with the unbound enzymes provided a benchmark for the simulations in the presence of substrates or products. They revealed very stable structures for all four replicate 500 ns MD simulations with both enzymes (average RMSD of 0.95 ± 0.08 Å for the backbone atoms and 1.41 ± 0.08 Å for the protein heavy atoms; see Fig. S5), and reasonably stable for the four 200 ns replicate aMD simulations (average RMSD of 1.71 ± 0.27 Å for the backbone atoms and 2.25 ± 0.29 Å for the heavy atoms). This clearly showed that the conformational space sampled by the aMDs was considerably larger than in the MDs.

One important parameter to be evaluated in these simulations is the time-dependence the enzyme tunnels, or tunnel dynamics. DhaA31 variant is more active than DhaAwt towards the hydrolysis of TCP, in spite of all the existing crystal structures of DhaA31 (PDB IDs 3RK4, 4FWB, 4HZG) showing its access pathways leading to the catalytic site totally occluded. This suggested that its structure must be flexible enough to open a pathway that can accommodate and transport the substrates and products. The tunnels present in DhaA31 and DhaAwt were calculated for every simulation by CAVER 3.02¹², using a specific point in the middle of the catalytic site to define the tunnel origin. Based on the fraction of snapshots with detected tunnels and the respective opening rates (a tunnel is considered open when has radius ≥ 1.4 Å) (Table S1), the only relevant tunnel observed in DhaA31 was the one commonly known as p1 tunnel, which was also the main tunnel in DhaAwt (Fig. 1). The relevant parameters obtained for this tunnel in the MD and aMD simulations of DhaA31 and DhaAwt are presented in Table 1.

It was found that in both enzymes the p1 tunnel adopted two main conformations, a closed and an open one (Fig. 2). DhaAwt had about twice as much of tunnel openings than DhaA31, both in the MD and aMD simulations. However, the maximum bottleneck radii observed in each enzyme were similar (ca. 3 Å in aMDs). Comparing MDs and aMDs, it was obvious that each enzyme showed larger fluctuations on the tunnel dynamics and wider maximum radii. Moreover, higher tunnel opening rates were observed in the aMDs. This was expected, since the aMDs sample a larger conformational space equivalent to long time scales.

Table 1. Tunnel dynamics and hydration data. Opening rates and maximum bottleneck radius observed for the p1 tunnel and hydration spheres, calculated for the MD and aMD simulations with DhaA31 and DhaAwt.*

		% tunnel opening	Max. radius (Å)	Solvation shells	
				Nucleophile ($r \leq 1.4$ Å)	Active site ($r \leq 8.0$ Å)
MD	DhaA31	12.6 ± 1.0	2.32 ± 0.02	2.23 ± 0.28	7.10 ± 1.05
	DhaAwt	23.9 ± 1.7	2.36 ± 0.10	2.01 ± 0.19	5.36 ± 0.92
aMD	DhaA31	17.4 ± 5.4	3.03 ± 0.16	2.34 ± 0.34	6.64 ± 1.05
	DhaAwt	42.6 ± 13.4	2.96 ± 0.27	2.29 ± 0.31	6.69 ± 1.15

*Average values $\pm$ SEM calculated over the respective 4×500 ns MD and 4×200 ns aMD simulations; tunnel opening refers to the ratio of snapshots containing tunnel with radius ≥ 1.4 Å; max. radius refers to the largest tunnel radius observed in each simulation; solvation shells are the number of water molecules found within the given distance from the nucleophile D106 C γ -atom.

The solvation of the nucleophile D106 and the active site was calculated for the MD and aMD simulations (Table 1 and Figs. S6-S7). No significant differences were observed between the two proteins in terms of the hydration of the active site. This evidence does not agree with the previous working hypothesis that the structural basis for the enhanced activity of DhaA31 towards TCP was the lower accessibility of the water molecules to the respective active site¹⁶. This hypothesis was based on 2 ns long MD simulations, which clearly were not sufficiently long to have a properly equilibrated system in terms of solvation. Moreover, due to the unavailability of a crystal structure at that time, those simulations were performed on a modeled structure of DhaA31 with hydration modeled by homology with DhaAwt.

In our study, we did not test different explicit water models since the TIP3P model is routinely used in detailed studies of water interactions with biomolecules^{73,74}. The literature suggests that the TIP4P and SPC/E models reproduce well the water's physical properties⁷⁵. The recently proposed OPC3 model brings a different approach to the common rigid models and has also been proven very accurate^{76,77}. The selection of water model and force field do influence the

interactions observed, although in published cases the differences were minimal⁷⁸. In a comparative study such as ours any inaccuracy of the water model will be similar in both enzymes, which differ in only five residues, making our conclusions regarding hydration valid independently of the water model chosen.

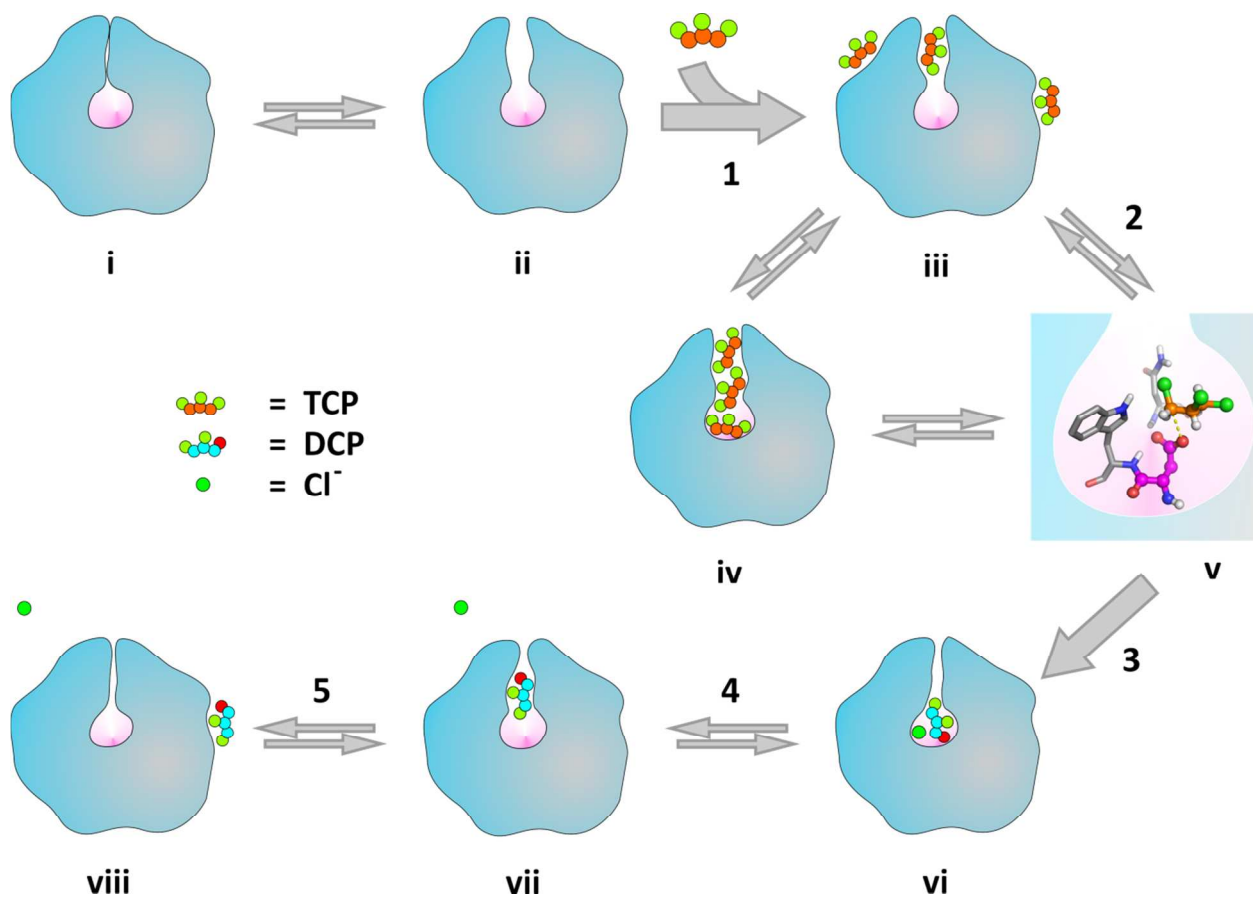


Figure 2. Molecular processes observed in the simulations with DhaA31 and DhaAwt. Catalytic steps: 1) binding of TCP substrate to the protein; 2) formation of the reactive complex; 3) chemical steps; 4 and 5) release of the products. The structures i) and ii) represent the closed and open conformations of p1 tunnel of the unbound enzyme, respectively; iii) protein bound with substrate molecules; iv) multiple binding of substrate molecules to the p1 tunnel; v) reactive complex, $\text{E} \cdot \text{S}^{\text{REACT}}$; vi) protein complexed with the products in the active site; vii) protein with

the alcohol in the bottleneck region of p1 tunnel; viii) protein after both products have been released. The nucleophile D106 is represented in magenta ball-and-sticks, the active site as the pink shaded pocket, and TCP, DCP and chloride ion are colored as indicated in the legend.

Substrate binding

To study the binding of TCP substrate to the proteins, we carried out quadruplicate 200 ns long MD simulations and octuplicate 100 ns aMD simulations, starting with the substrate in the bulk solvent at the experimental concentration which corresponded to three TCP molecules. The stability of these simulations was equivalent to that of the unbound proteins (Fig. S8). Similarly, the aMDs also showed higher RMSDs due to the expected enhancement of the conformational sampling. The distance of each molecule to the active site was calculated as the distance from the center of mass of TCP to the origin of the tunnels calculated by CAVER.

Table 2 summarizes the transport events and average location of the TCP molecules during the MD and aMD simulations, according to those criteria. The time variation of the TCP's distance to the active site (Fig. S9) shows that its binding was observed in the MD and aMD simulations with both enzymes (Fig. 2). As expected, the aMD simulations showed a higher number of binding events than the MDs. The binding of TCP was more frequent with DhaAwt than with DhaA31. However, DhaAwt also showed some releases of TCP after binding, which was not observed with DhaA31. This may explain why the overall affinity of the enzymes for TCP was very similar, with reported K_m values of 1.0 and 1.2 mM for DhaAwt and 31, respectively¹⁶.

The p1 tunnel was the only pathway used for the substrate binding or release. Fig. 3 shows the density of TCP occupancy, which can be associated with the respective histogram distribution in Fig. 4. It can be observed that TCP was more frequently localized in the main tunnel for DhaAwt

than DhaA31. These observations correspond to faster transport rates of TCP through the p1 tunnel of DhaAwt than DhaA31 (Table 2).

Table 2. Location of TCP with respect to the protein tunnels. Transport events per simulation, residence times and opening rates of p1 tunnel, calculated for the MD and aMD simulations of DhaA31 and DhaAwt in the presence of 3 molecules of TCP substrate.*

		TCP transport rates (per simulation) [§]		% TCP residence time			% tunnel opening	
		binding	release	inside	mouth	Outside		
MD	DhaA31-3TCP	0.5 ± 0.6	0 ± 0	2.6 ± 1.9%	6.7 ± 2.1%	90.7 ± 3.1%	31.4 ± 8.3%	
	DhaAwt-3TCP	0.8 ± 0.5	0.8 ± 0.5	1.3 ± 0.6% [#]	19.7 ± 3.5%	75.5 ± 4.3%	38.9 ± 3.0%	
aMD	DhaA31-3TCP	0.8 ± 0.5	0 ± 0	8.6 ± 2.6%	3.7 ± 0.5% [#]	86.1 ± 2.9%	44.2 ± 5.1%	
	DhaAwt-3TCP	1.9 ± 1.0	0.4 ± 0.7	44.1 ± 2.3%	5.3 ± 1.1%	50.6 ± 3.0%	80.8 ± 2.4%	

*Average values ± SEM calculated over the respective 4 × 200 ns MD and 8 × 100 ns aMD simulations; transport rates refer to full binding (< 6 Å from the active site) of TCP or full release (> 20 Å); residence times defined as “inside” for distances from the active site ≤ 9 Å, at the tunnel “mouth” for 9 Å < distances < 13 Å, and “outside” for distances ≥ 13 Å; tunnel opening refers to the ratio of snapshots containing tunnel with radius ≥ 1.4 Å. [§]Rates are in average number of events ± SD. [#]An outlier was excluded from these statistics.

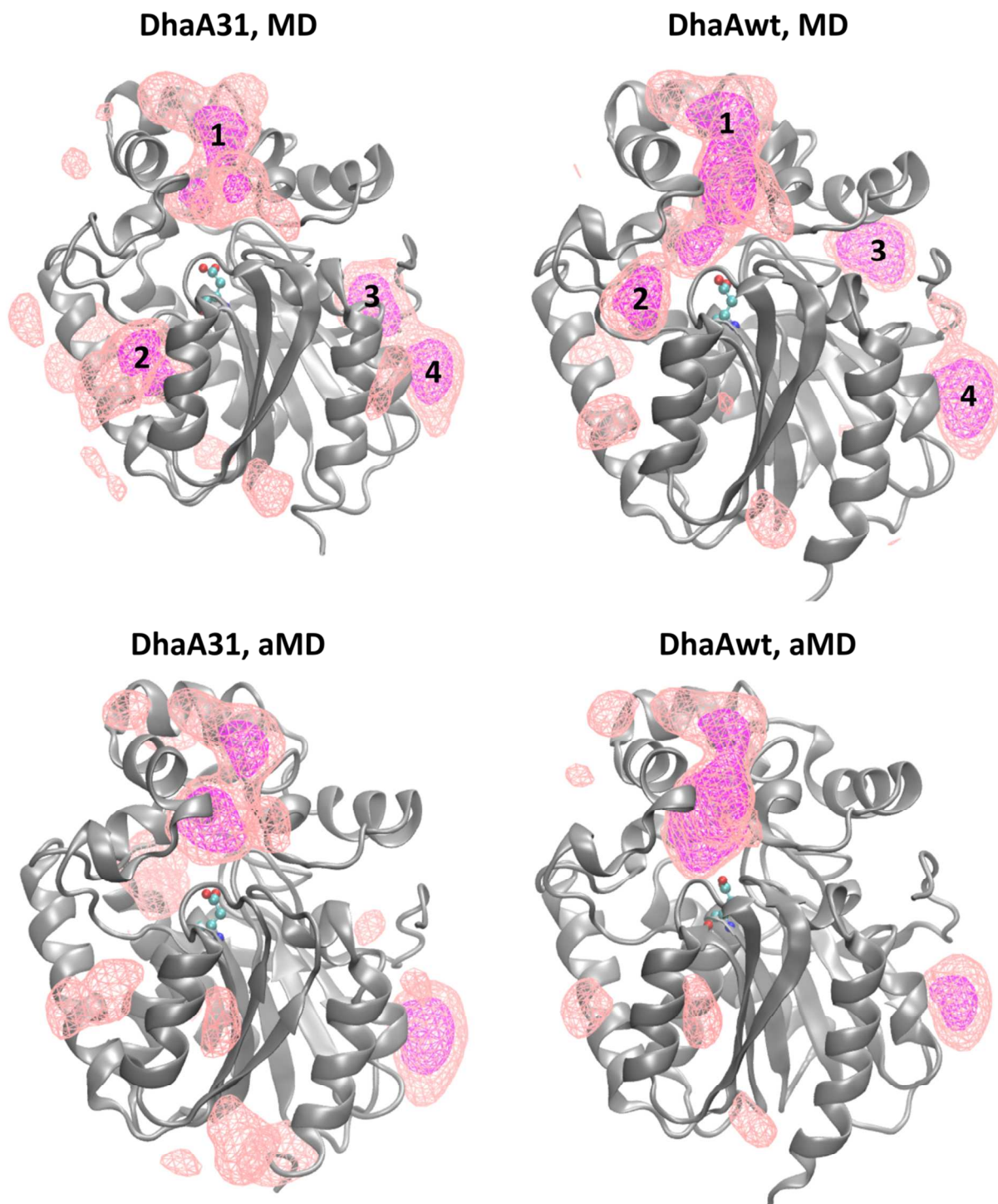


Figure 3. Distribution of TCP over the proteins' structures. Occupancy density of TCP over the MD and aMD simulations with DhaA31 and DhaAwt at the isovalues of 0.05 (magenta) and 0.01

(pink). The nucleophile D106 is represented as green ball-and-sticks. The most occupied regions at the protein surface are numbered by 1-4.

The aMD simulations repeatedly showed multiple binding of TCP molecules to the tunnel of DhaAwt, which was not observed with DhaA31 (Figs. 2 and S9). The TCP molecules also tended to bind for long periods to specific regions of the protein surface. These regions can easily be associated with specific peaks in the histograms of the ligand distances (Figs. 3 and 4): region 1 or entrance to p1 tunnel (maximum ca. 10-12 Å), region 2 or entrance to p2 tunnel (maximum ca. 15 Å), region 3 (maximum ca. 19-22 Å) and region 4 (maximum ca. 29 Å). The protein interior corresponds to distances $\leq$ ca. 9 Å.

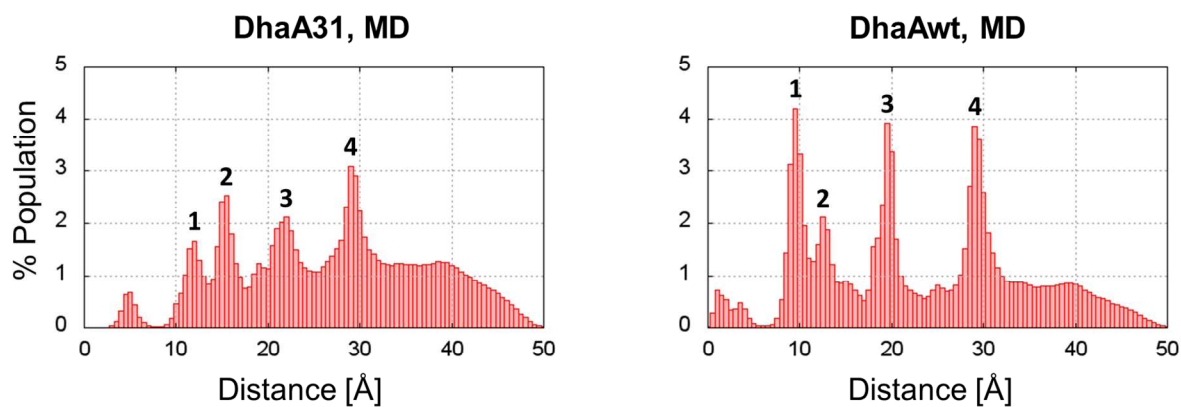


Figure 4. Histogram distribution of TCP during the simulations. Histogram distribution of the TCP distances to the active site for the MD simulations with DhaA31 and DhaAwt. The most occupied regions of the protein surface are numbered by 1-4.

Higher rates of tunnel opening were observed in the presence of TCP than with the unbound proteins (Table 2). The variation of the tunnel bottleneck radius with the distance of TCP to the active site (Fig. S10) shows a clear dependency when the substrate is located in the bottleneck region, ca. 5-8 Å from the tunnel origin. This is consistent with the view that TCP may induce the tunnel opening during its binding (Fig. 2) to allow its transport to the active site. The bottleneck region corresponds to the natural barrier encountered by the ligands that pass through the tunnel, which is why this region is the least populated by TCP while traveling along the pathway. The exception to this trend were the aMDs with DhaAwt, where the tunnel radius was not influenced by a single but several molecules due to the multiple binding events.

The linear interaction energies (LIE) between TCP and the proteins were calculated from the classical MD simulations. Fig. 5 shows the variation of the respective electrostatic and van der Waals components with the distance of TCP to the active site. In general, the van der Waals contribution was remarkably dominating the interaction energy, which is consistent with the hydrophobic character of the TCP substrate. As expected, the lowest energy minima corresponded to the TCP molecules binding to the hydrophobic main tunnel or active site. The interactions with DhaA31 for smaller distances of TCP are not as low as with DhaAwt, which is probably due to undersampling of the lower distances in the MD simulations with the more occluded DhaA31. At the protein surface there are always minima of the interaction energy, which correspond to the highest populated regions previously identified (Figs. 3 and 4). The TCP molecules formed quite strong interactions with the protein at some of these regions, such as the tunnel mouth (ca. 10-12 Å), region 2 for DhaA31 (15 Å) and region 3 for DhaAwt (19.5 Å) (Fig.

5). TCP also displayed significant electrostatic interactions with the protein at the tunnel mouth and region 3, arising mostly from the positively charged residues K175 and K71.

The solvation of the active sites of DhaA31 and DhaAwt was calculated for all simulations performed with ligands (Table S2 and Fig. S11-S14). As found for the unbound proteins, DhaA31 and DhaAwt showed comparable levels of hydration, confirming again that the enhanced rates of the S_N2 step in DhaA31¹⁶ are not due to lower hydration in this mutant.

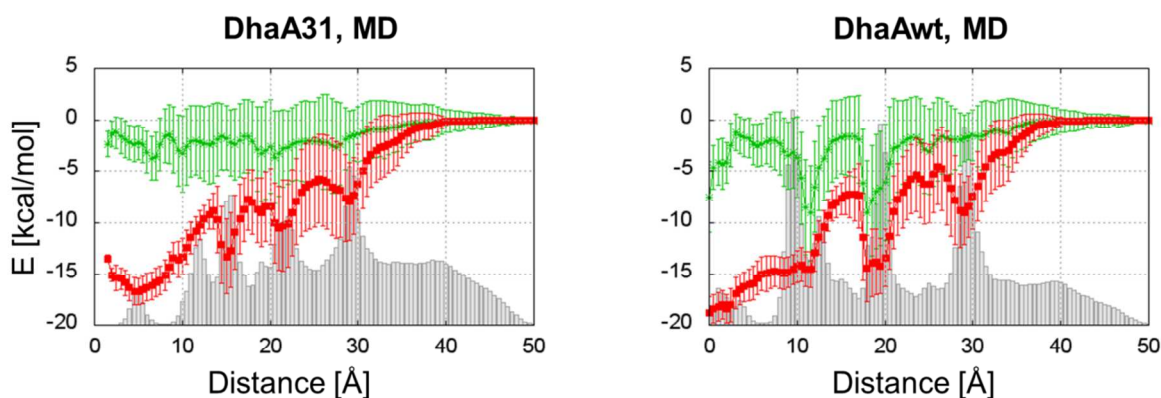


Figure 5. Linear interaction energy of TCP during the simulations. Linear interaction energy contributions binned according to the distance of TCP to the active site cavity for MD simulations of DhaA31 and DhaAwt. Simulations carried out in the presence of 3 molecules of TCP substrate. Average electrostatic (green) and van der Waals (red) are represented with the respective SD (vertical bars). The respective histogram distributions of the TCP distances to the active site are represented by the gray bars.

Catalytic step

Reactivity of the Michaelis complex

We studied here the predisposition of the two enzyme variants to generate reactive complexes with the substrate and the respective energy barriers to the first reaction, the S_N2 step, which is known to be rate-limiting in DhaAwt. We started by making a conformational analysis of the Michaelis complex, E·S (Scheme 1), during a set of MD simulations to determine the ratio of reactive configurations of TCP (here termed as “reactive complex”, ES^{REACT}). To study and compare the ratios of reactive configurations observed with DhaA31 and DhaAwt, TCP was docked into the active site of the two enzymes in two binding modes, one oriented to form the *R*-enantiomer of DCP, and the other one to form the *S*-enantiomer. These two initial conformations were subsequently studied with MD and aMDs simulations (both run in octuplicate, for 200 ns and 100 ns, respectively).

All simulations were stable, with all the catalytic residues fluctuating around their reactive conformation (Fig. S15). From the 16 MDs performed with each enzyme, only 3 exhibited the full release of TCP from DhaAwt, whereas none was observed with DhaA31. The aMD simulations showed that the substrate was rarely released from either of the DhaA variants (3 in 16 aMDs for DhaA31 and 1 for DhaAwt), confirming that TCP binds tightly to the hydrophobic interior of those proteins. The percentage of p1 tunnel opening was significantly higher in the presence of TCP than with the unbound proteins (Table 3). A strong dependence of the tunnel bottleneck radius with the position of TCP was observed as previously (Fig. S16). Likewise, the tunnel displayed larger average bottleneck radius when the substrate occupied the bottleneck regions, ca. 6.5-8 Å from the active site. Because the energy boost applied during the aMDs strongly disfavors the formation of the reactive conformations, these simulations were not considered further in the study of the interactions at the active site.

The evolution of the reactive complexes (ES^{REACT}) of TCP with DhaA31 and DhaAwt was determined according to the positions of TCP with respect to the nucleophile, D106, and the halide stabilizing residues, N41 or W107. The aim was to evaluate the propensity of TCP to adopt a reactive orientation once located in the active site and how each enzyme will preserve such reactive configuration. It is known that the productive encounters are extremely rare in biological systems, which means that there is a large number of enzyme-substrate conformations that are not reactive⁷⁹. For this reason we started this study with TCP in the reactive orientation. Despite the limited sampling, such simulations will provide valuable information regarding the relative probability of each system to proceed with the S_N2 reaction. This analysis resulted in a number of snapshots containing TCP in reactive configurations (Fig. 2). For both enzymes there were MDs presenting fractions of reactive complexes that were outlying from the normal distribution among the respective data set, and hence were excluded from subsequent analysis (Fig. S17). As expected, the highest numbers of reactive complexes were observed at the beginning of each MD, but often TCP regenerated the reactive configurations after long periods (Fig. S17B). This was observed more frequently with DhaA31 than with DhaAwt. It was found that the MD simulations with DhaA31 retained significantly higher rates of reactive complexes than with DhaAwt by 5.5-fold (P -value 0.0029). Moreover, the distribution of the O-C distances and O-C-Cl angles for the different ES^{REACT} complexes (Fig. S18) showed better quality of reactive complexes in DhaA31 than in DhaAwt, meaning that the ratios of ES^{REACT} mean fractions for DhaA31 vs. DhaAwt was always higher with stricter thresholds, up to 11.6-fold for distances ≤ 3.0 Å and angles $\geq 164^\circ$. These results thus suggest that DhaA31 may be more reactive than DhaAwt due to both the number and quality of reactive conformations adopted by TCP in the respective active sites.

Table 3. Location of TCP with respect to the protein tunnels. Transport events per simulation, residence times and opening rates of p1 tunnel, calculated for the MD simulations of DhaA31 and DhaAwt, containing one molecule of TCP in the active site.*

		TCP transport rates(per simulation) [§]		% TCP residence time			% tunnel opening
		release	rebinding	inside	mouth	Outside	
MD	DhaA31-TCP	0 ± 0	0 ± 0	100.0 ± 0.0% [#]	± 0.0 ± 0.0% [#]	± 0.0 ± 0.0% [#]	46.6 ± 5.3%
	DhaAwt-TCP	0.2 ± 0.4	0 ± 0	83.8 ± 7.7%	6.0 ± 3.0% [#]	± 10.2 ± 5.4%	± 53.9 ± 5.1%
aMD	DhaA31-TCP	0.2 ± 0.4	0 ± 0	94.5 ± 2.9%	0.2 ± 0.1% [#]	± 1.3 ± 1.2% [#]	32.1 ± 2.7%
	DhaAwt-TCP	0.1 ± 0.3	0.1 ± 0.3	98.1 ± 0.7% [#]	1.5 ± 0.5%	0.2 ± 0.2% [#]	72.0 ± 1.4%

*Average values ± SEM calculated over the respective 8 × 200 ns MD and 16 × 100 ns aMD simulations; transport rates refer to full release (> 20 Å) of TCP or full binding (< 6 Å from the active site); residence times defined as “inside” for distances from the active site ≤ 9 Å, at the tunnel “mouth” for 9 Å < distances < 13 Å, and “outside” for distances ≥ 13 Å; tunnel opening refers to the ratio of snapshots containing tunnel with radius ≥ 1.4 Å. [§]Rates are in average number of events ± SD. [#]Outliers were excluded from these statistics.

Table 4. Reactivity of TCP with DhaA31, DhaAwt and the alanine mutants. Fraction of reactive complexes found in the MD simulations with TCP in the active site, in total and according to the *R* and *S*-orientation of TCP. The respective activation energy barriers calculated for the S_N2 reaction (ΔG[‡]) are reported.*

Variant	Fraction of total ES ^{REACT} × 10 ⁵	Fraction of ES ^{REACT} reactive orientation × 10 ⁵	by ΔG [‡] (kcal/mol)
	Average	Orientation	Average

DhaA31	$285 \pm 67^{\#}$	<i>R</i>	53 ± 13	18.36 ± 0.14
		<i>S</i>	$229 \pm 63^{\#}$	17.08 ± 0.12
DhaAwt	$52 \pm 15^{\#}$	<i>R</i>	16 ± 6	19.09 ± 0.07
		<i>S</i>	$44 \pm 16^{\#}$	18.69 ± 0.14
DhaA31+Y176A	173 ± 50	<i>R</i>	83 ± 20	
		<i>S</i>	91 ± 34	
DhaA31+F245A	12 ± 4	<i>R</i>	5 ± 2	
		<i>S</i>	7 ± 4	
DhaA31+I246A	188 ± 55	<i>R</i>	115 ± 51	
		<i>S</i>	$43 \pm 18^{\#}$	

* Average values $\pm$ SEM; the fractions of ES^{REACT} refer to the reactive complexes found over the 16×200 ns MD simulations, in total and by orientation. [#]Outliers were excluded from these statistics.

We found that both initial orientations of TCP were able to yield *R*- and *S*-oriented ES^{REACT} with both enzymes, which means that the substrate was able to rotate in the active sites and adopt different binding modes. The total enantiomeric distribution of the ES^{REACT} was determined and showed that the *R*-oriented conformations of TCP were less populated than the *S*, both in DhaA31 and DhaAwt (Table 4). This suggests the preference of both enzymes to generate the *S*-isomer of the alkyl-enzyme complex rather than the *R*.

Considering the reactive configurations of the E-S complex, DhaA31 showed a significantly higher predisposition to the reaction than DhaAwt. To further confirm this observation, we studied the energy of the systems towards the S_N2 reaction, by simulating the transition state. For that, we determined the potential energy surface (PES) along the reaction coordinate using hybrid quantum mechanics/molecular mechanics (QM/MM) calculations. The activation energy,

$\Delta G^\ddagger$, was calculated as the energy difference between the ground state and the transition state (Fig. S19A). This calculation was performed on the reactive complexes of DhaA31 and DhaAwt with TCP in *R*- and *S*-orientation, using the 100 best structures of each type. The results showed that the activation barrier was lowest for DhaA31 with *S*-oriented TCP, 1.61 kcal/mol below what was observed for DhaAwt with TCP in *S*-orientation (Table 4). This difference was statistically very significant (P -value < 0.0001), and, according to the transition state theory, it may represent an increase in the reaction rates of DhaA31 about 14-fold with respect to DhaAwt. When analyzing the composition of the active site of these enzymes, we found that during the S_N2 reaction the residue F245 is in close contact with TCP in DhaA31, while V245 in DhaAwt is farther and has limited contact with this substrate (Fig. S19B). Hence we propose that V245F mutation may be the major responsible for lowering the energy barrier of this reaction. In spite of the different reactivity of TCP oriented towards the production of *R*- and *S*-enantiomer, it is known that neither DhaAwt nor 31 are significantly enantioselective⁸⁰. This can only be explained by the subsequent steps of the catalytic cycle, such as the hydrolysis of the alkyl-enzyme intermediate or the product release. These processes might give preference to the formation and release of *R*-DCP and level out the enantiomeric yield, as was previously described for other systems⁸¹.

Dissection of interactions in the reactive complex

The binding free energy of TCP (ΔG_{bind}) was calculated for all snapshots containing the reactive complex using the molecular mechanics/generalized born surface area (MM/GBSA) method^{63,64} to further dissect the reactivity improvements in DhaA31. The residue-wise decomposition of ΔG_{bind} was performed for all ES^{REACT} structures, and Fig. 6 shows the residues

contributing the most to the reactive binding energy of TCP. As expected, all residues that form the catalytic pentad (N41, D106, W107, E130 and H272) showed strong interactions with TCP. H272 contributed the most for the stability of the complex, while E130 interacted unfavorably. A few other residues in the active site contributed favorably to the stability of TCP in the ES^{REACT} , such as F149, F168 and L209. Regarding the residues mutated in DhaA31, those at positions 245 and 246 presented strong interactions with TCP in the reactive complexes with both enzymes, whereas residue 176 had only marginal interactions due to its considerable distance to TCP. The calculated differences between the averaged interactions (Fig. 6) revealed that H272 stabilized remarkably less TCP in DhaA31 than in DhaAwt, whereas D106 appeared to stabilize more TCP in DhaA31. Besides the catalytic residues, only the residues at positions 176, 245 and 246 presented significantly different interactions with TCP in the ES^{REACT} structures, although the difference was small for residue 176. V245F mutation showed a significant increase in the stabilization of TCP in the complex with DhaA31, which explains its positive role in improving the reactivity of DhaA31 towards TCP. Conversely, L246I mutation showed a decrease of the stabilization in DhaA31, suggesting a negative contribution to the reactivity of DhaA31 with respect to DhaAwt.

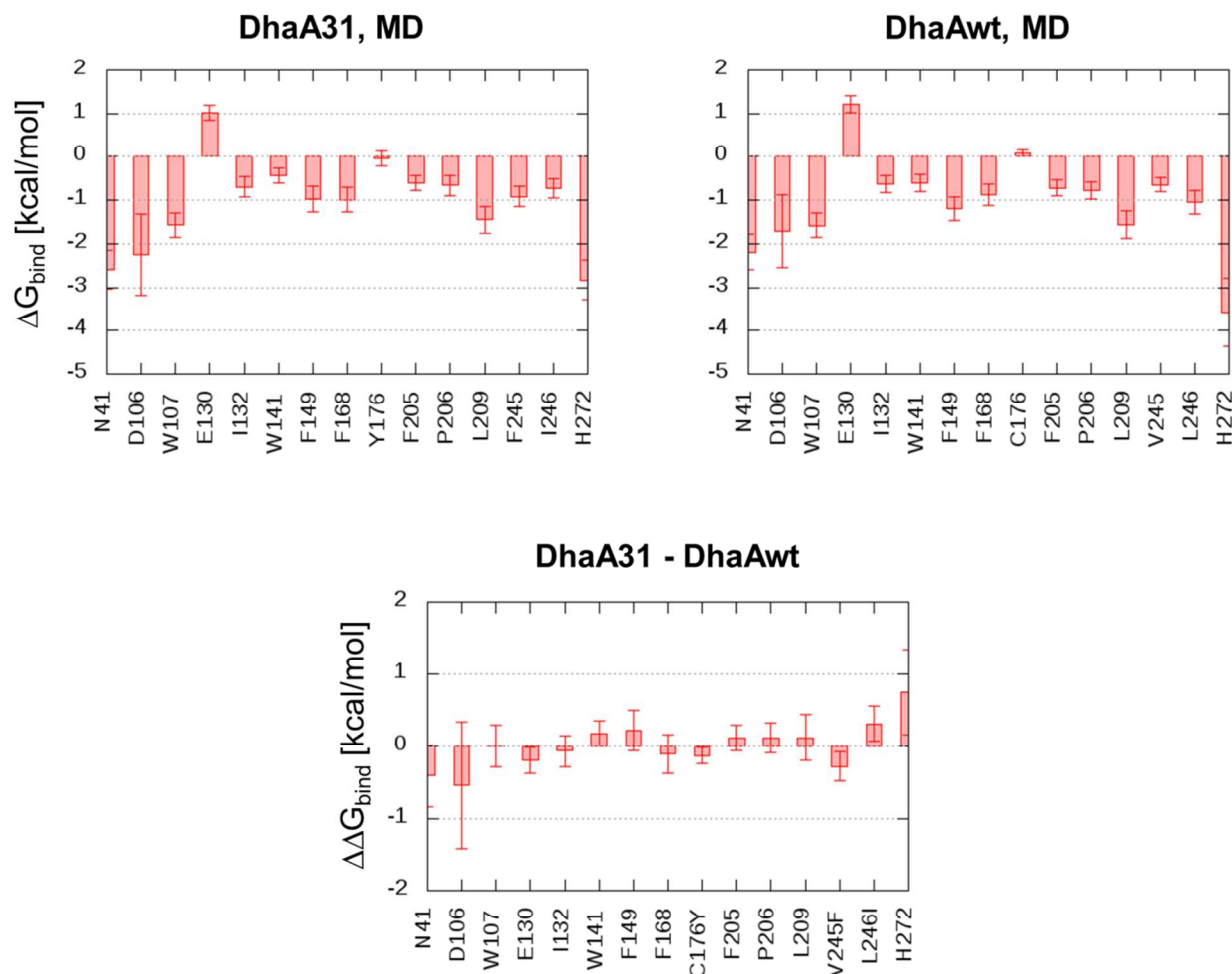


Figure 6. Interactions of TCP in the reactive configuration. Residue-wise decomposition of the binding interactions of TCP (ΔG_{bind}) in the reactive complexes with DhaA31 and DhaAwt, and respective differences ($\Delta\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^{31} - \Delta G_{\text{bind}}^{\text{wt}}$). The error bars represent the SD from the mean values.

To verify the importance of these three residues – Y176, F245 and I246 – in DhaA31, they were mutated *in silico* to alanine. These mutants were simulated with TCP in their active site in the same way as DhaA31, and the reactivity in each case was accounted by the number of ES^{REACT} complexes observed (Table 4). All of these mutants presented lower fractions of

1
2
3 reactive configurations of the substrate than DhaA31, but F245A showed the most significant
4 decrease (P -value 0.0078 when compared to DhaA31). This demonstrates that F245 is
5
6 fundamental for the correct positioning of TCP in the active site of DhaA31 to have a reasonable
7
8 rate of the S_N2 reaction. The residues at positions 176 and 246 also seem to influence the
9
10 reactivity, but the differences were less significant (P -values 0.27 and 0.35, respectively) than for
11
12 residue at position 245. With the Y176A mutant TCP was released to the bulk solvent in 6 out of
13
14 8 MD simulations, whereas for the other alanine mutants only one release was observed with
15
16 I246A (data not shown). This suggests the importance of the bottleneck residue Y176 in keeping
17
18 the substrate in the active site, which leads to increasing the chances of the chemical reaction to
19
20 take place when compared to DhaAwt.
21
22
23
24
25
26
27
28

29 To disclose the global role of the different residues in keeping TCP within the active site,
30
31 TCP's ΔG_{bind} was also calculated in the MD simulations in which TCP was not released from the
32
33 protein interior. The respective residue-wise decomposition (Fig. S20) revealed that TCP had
34
35 some of the strongest interactions with residues in the catalytic site of DhaA31, while most of
36
37 these interactions were lowered in DhaAwt. These data reflect the fact that TCP spent less time
38
39 in the catalytic site with DhaAwt than with DhaA31. Of all residues, the one at position 176 had
40
41 the largest drop of interaction with TCP when comparing DhaAwt with DhaA31. Since this was
42
43 one of the most frequent residues lining the bottleneck of the p1 tunnel in both systems, such a
44
45 difference in the interactions with the substrate also translates a strong change in the gatekeeping
46
47 roles of tyrosine and cysteine present in DhaA31 and DhaAwt, respectively. The side chain of
48
49 F245, which interacted the most with TCP in the reactive configuration with DhaA31 (Fig. 6),
50
51 was found to have significantly different orientation (P -value < 0.0001) in Y176A mutant as
52
53
54
55
56
57
58
59
60

1
2
3 compared to DhaA31 (Fig. S21). The improper orientation of this residue could contribute to the
4
5 less frequent reactive configurations of TCP in the active site of the Y176A mutant. This also
6
7 suggests the long-range effect of Y176 residue in shaping the active site of DhaA31 in order to
8
9 enhance the positioning of the substrate in the reactive orientation.
10
11
12
13
14

15 **Product release**

16
17 The release of DCP and Cl⁻ products from DhaA31 and DhaAwt was studied by MD and aMD
18
19 simulations. The initial binding conformations of the *R*- and *S*-DCP in the active sites were
20
21 obtained from docking, while the chloride ion was taken directly from the crystal structures (Fig.
22
23 2). Each system was simulated with quadruplicate 200 ns long MD simulations and octuplicate
24
25 100 ns long aMD simulations, which revealed similarly stable as the previously reported
26
27 simulations (Fig. S22). The transport rates and residence times of both products in these
28
29 simulations are summarized in Table 5. It was found that the Cl⁻ ion was released from DhaA31
30
31 and DhaAwt in most of the simulations, and, in general, much faster than DCP (Fig. S23). Cl⁻
32
33 was also released significantly faster from DhaAwt than from DhaA31 (*P*-values < 0.070 for the
34
35 MDs and 0.0049 for the aMDs). Regarding the DCP, no significant differences were observed
36
37 between the *R*- and *S*-enantiomers (data not shown). DCP was not released from DhaA31 in any
38
39 MD simulation, and it remained inside the protein 100% of the time. Conversely, DCP was very
40
41 frequently released from DhaAwt (8 times in 8 MDs), and even some rebinding events were
42
43 observed (3 in 8 MDs). DhaA31 showed a few releases of DCP in the aMD simulations (7 in 16
44
45 aMDs), but notably less than DhaAwt (21 in 16 aMDs). The distribution of the residence times
46
47 of DCP was also very different with DhaA31 and DhaAwt, reflecting the slower traffic of DCP
48
49 along the protein tunnel with DhaA31. Altogether, these results showed that the release of both
50
51
52
53
54
55
56
57
58
59
60

Cl⁻ and DCP products occurred significantly faster in DhaAwt than in DhaA31. In general, no significant differences were found in the releases of the *R*- and *S*-enantiomers of DCP. Although *R*-DCP was released twice as much from DhaA31 than *S*-DCP in the aMDs (Fig. S23), a larger number of simulations would be required to achieve the statistical significance that allowed us draw reliable conclusions. The p1 tunnel was the only pathway used in the release or rebinding of the products. As for TCP, the influence of DCP on the bottleneck radius and opening rates of the tunnel was remarkable (Table 5 and Fig. S24). This effect was more pronounced for distances between ca. 5-8 Å from the active site, corresponding to the bottleneck region.

Table 5. Location of the products with respect to the protein tunnels. Chloride release times, transport events per simulation, residence times and opening rates of p1 tunnel, calculated for the MD and aMD simulations of DhaA31 and DhaAwt containing one molecule of *R/S*-DCP and Cl⁻ in the active site.*

	Cl ⁻ release time (ns)	DCP transport rates (per simulation) [§]		% DCP residence time			% tunnel opening
		release	rebinding	inside	mouth	outside	
A31-DCP	$> 75.0 \pm 27.7$	0 ± 0	0 ± 0	$100.0 \pm 0.0\%^{\#}$	$0.0 \pm 0.0\%^{\#}$	$0.0 \pm 0.0\%$	$58.5 \pm 9.6\%$
Awt-DCP	20.2 ± 4.5	1.0 ± 0.5	0.4 ± 0.7	$41.5 \pm 13.3\%$	$8.4 \pm 2.5\%$	$50.1 \pm 13.1\%$	$52.5 \pm 5.9\%$
A31-DCP	9.6 ± 1.6	0.4 ± 0.6	0.1 ± 0.3	$82.1 \pm 6.7\%$	$1.4 \pm 0.6\%^{\#}$	$15.9 \pm 6.1\%$	$33.5 \pm 2.8\%$
Awt-DCP	$2.9 \pm 0.6^{\#}$	1.3 ± 0.8	0.8 ± 0.8	$60.0 \pm 6.6\%$	$4.1 \pm 0.8\%$	$36.0 \pm 6.7\%$	$52.6 \pm 3.5\%$

*Average values ± SEM calculated over the respective 8 × 200 ns MD and 16 × 100 ns aMD simulations; transport times or rates refer to full release (> 20 Å) of Cl⁻ or DCP or full binding (< 6 Å from the active site); residence times defined as “inside” for distances from the active site ≤ 9 Å, at the tunnel “mouth” for 9 Å < distances < 13 Å, and “outside” for distances ≥ 13 Å; tunnel opening refers to the ratio of snapshots containing tunnel with radius ≥ 1.4 Å. [§]Rates are in average number of events ± SD. [#]An outlier was excluded from these statistics.

The LIE analysis of DCP with the proteins was carried out over the MD simulations (Fig. 7). The interactions were highest when DCP was in the active site, with both electrostatic and van der Waals playing equally important roles. This is in agreement with the H-bond calculated in these simulations (Table S3), which showed that the nucleophile D106 prevailed in terms of H-bonding with DCP in DhaA31. All starting structures in these simulations contained DCP forming an H-bond with D106, and this was the first interaction that needed to be broken for achieving the release of DCP. Other important electrostatic interactions in the active site corresponded to H-bonds with the catalytic N41 and W107. When moving along the p1 tunnel DCP also formed H-bonds with bottleneck residues (F168, Y173 and W141) and the electrostatic component of LIE lost relevance in favor of the van der Waals interactions with the hydrophobic tunnel-lining residues (Fig. 8). The difference in interactions may hold the key to explaining the different mobility of DCP through the main tunnels of DhaA31 and DhaAwt. The only residues in the p1 tunnel that differ in those two variants are V245F and C176F. Residue 245 is also part of the active site and V245F mutation is not expected to hinder much the release of DCP. On the other hand, residue 176 is in the bottleneck region, making strong interactions with DCP in DhaA31. Y176 has been identified as a key residue for holding TCP in the active site, and therefore its role in hampering the release of DCP from DhaA31 seems clear. At the tunnel mouth (ca. 9-12 Å from the active site), DCP showed strong electrostatic interaction with several residues in this region – T148, A172, F144 and K175 (Table S3). After being fully released from DhaAwt, DCP frequently interacted at specific regions of the protein surface (Fig. 9). These regions roughly overlap with the regions that interacted with TCP (compare with Fig. 3). The

strong electrostatic interactions found in region 3 of DhaAwt (maximum ca. 19.5 Å, Fig. 7) were mostly due to H-bonds with H188, S44, K71 and E191 (Table S3).

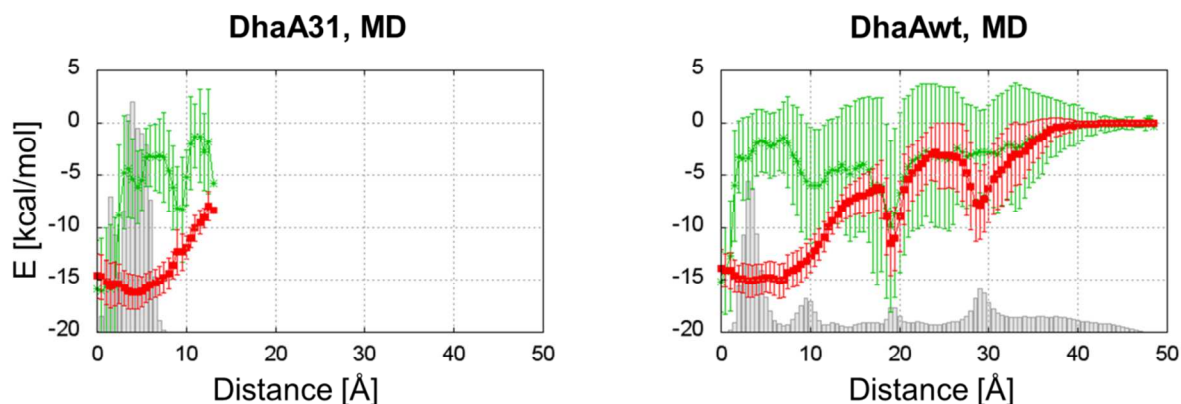


Figure 7. Linear interaction energy of DCP during the simulations. The linear interaction energy contributions binned according to the distance of DCP to the active site cavity, for all MD simulations of DhaA31 and DhaAwt containing DCP and Cl⁻ in the active site. Average electrostatic (green) and van der Waals (red) are represented with the respective SD (vertical bars). The respective histogram distributions of the TCP distances to the active site are represented by the gray bars.

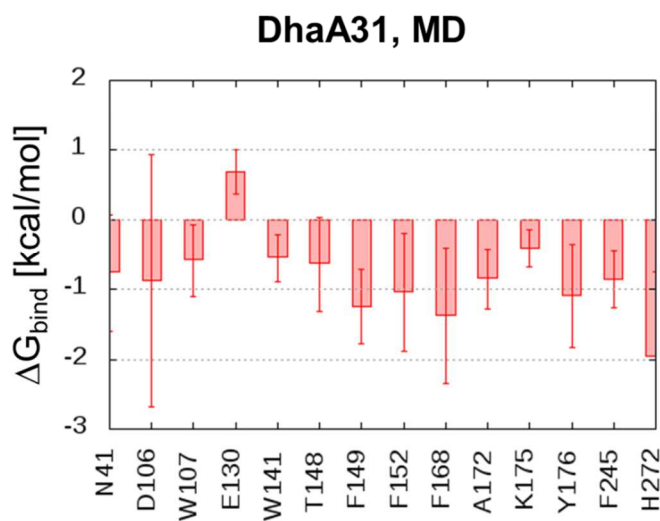


Figure 8. Interactions of DCP with DhaA31. Residue-wise decomposition of the binding interactions of DCP (ΔG_{bind}) in the MD simulations with DhaA31. The error bars represent the SD from the mean values.

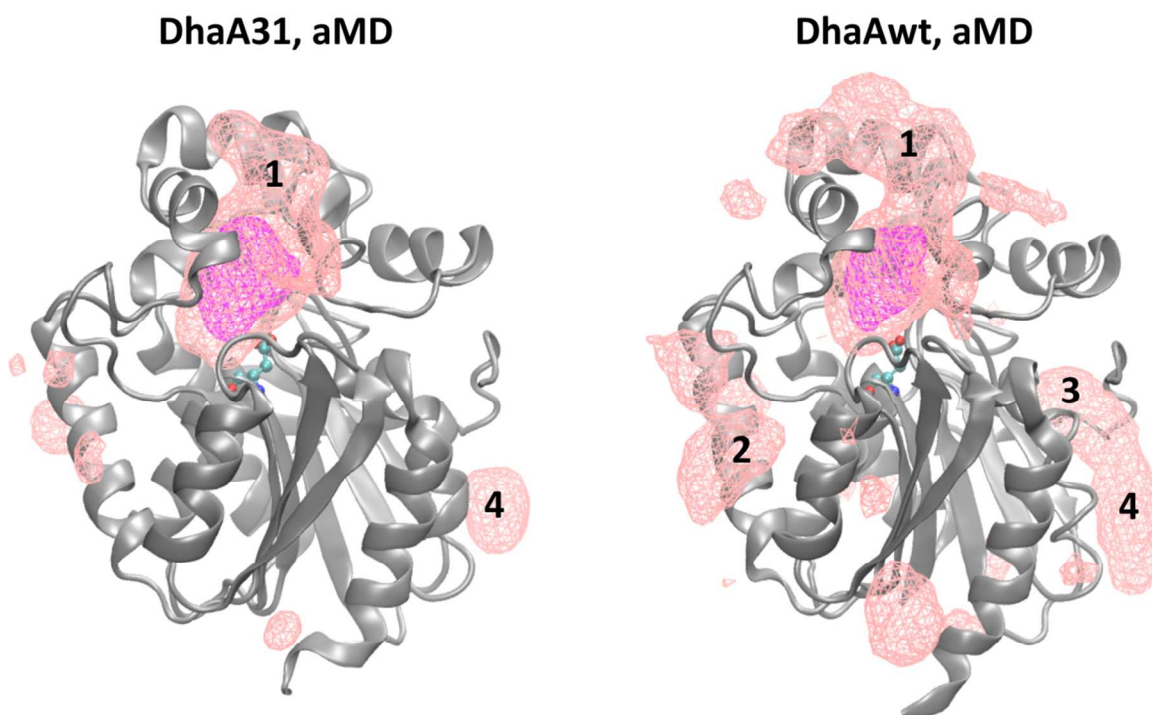


Figure 9. Distribution of DCP over the proteins' structures. Occupancy density of DCP over the aMD simulations with DhaA31 and DhaAwt, at the isovalues of 0.05 (magenta) and 0.01 (pink). The nucleophile D106 is represented as green ball-and-sticks. The most occupied regions at the protein surface are numbered by 1-4.

Experimental validation

Transient state kinetics of halide and alcohol release

We have conducted a series of kinetic studies with DhaA31, DhaAwt and several intermediate mutants to experimentally validate the structural basis for the improved activity of DhaA31 with TCP inferred from molecular modelling. Pre-steady state kinetic analyses were performed to disclose the mechanism and kinetics of halide and alcohol product release in DhaA31 and DhaAwt. Detailed binding experiments were performed with the DCP and chloride products, and, for comparison, also with bromide, which provides strong quenching of the intrinsic fluorescence from the halide-stabilizing residue W107.

The binding of DCP to DhaA31 and DhaAwt was investigated by using stopped-flow fluorescence. The rapid mixing of DCP with DhaAwt was associated with slow quenching of the fluorescence signal (Fig. 10A). The fluorescence traces could be fitted to a single exponential. The observed rate constant decreases with increasing the concentration of ligand (Fig. 10B), which is indicative of the presence of a slow conformational step preceding the fast binding⁸². This is described by Scheme 2 (with k_{obs} given by the relation as in Eq. 11), where the transition from E to E' represents a conformational change of the enzyme, which is required prior to the binding of a ligand L (here the alcohol). The results indicate that DhaAwt exists in two conformational isomers at the equilibrium ratio of 2:1, favoring the nonbinding form (Table 6).

This is in agreement with the previous theoretical results, which predicted that the enzymes have the open and closed states of the p1 tunnel. However, these data provide no clue regarding the equilibrium ratios in the absence of a ligand, in order to test the hypothesis of ligand-induced gating of the p1 tunnel.

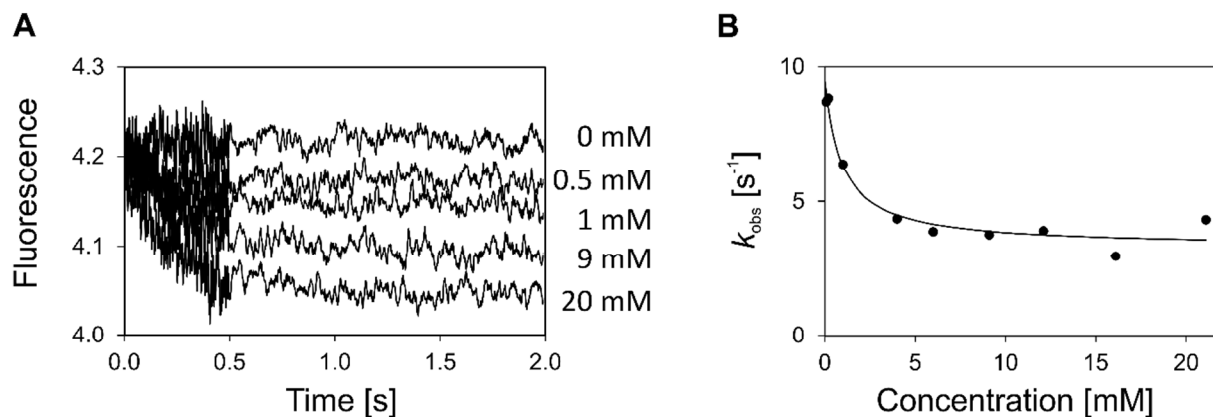
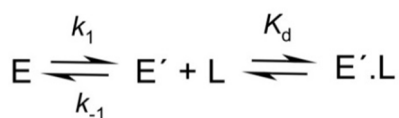


Figure 10. Kinetics of DCP binding to DhaAwt. Fluorescence traces obtained upon mixing of DhaAwt (70 μM) with DCP to a final concentration of 0-20 mM (A), and dependence of the observed rate constants of the observed fluorescence quench with DCP concentration (B). The solid lines represent the best fits to the data.



Scheme 2

$$k_{\text{obs}} = k_1 + \frac{k_{-1} \cdot K_d}{[\text{L}] + K_d}$$

Eq. 11

Scheme 2. Conformational change of an enzyme and binding of a ligand. Transition from E to E' represents a conformational change of an enzyme required for the binding of a ligand L. k_1 and k_{-1} are the kinetic constants for protein opening and closing, respectively, K_d is the dissociation constant.

Table 6. Kinetic constants describing the binding of DCP, Cl⁻ and Br⁻ to DhaAwt and DhaA31.*

	DCP			Br ⁻	Cl ⁻
	k_1	k_{-1}	K_d	K_d	K_d
	(s ⁻¹)	(s ⁻¹)	(mM)	(mM)	(mM)
DhaAwt	3.31 ± 0.27	6.16 ± 0.42	0.95 ± 0.34	730 ± 280	> 2000
DhaA31	-	-	> 20	1520 ± 670	> 2000

*Average values ± SEM; k_1 and k_{-1} are the rate constants and K_d is the dissociation constant; see Scheme 2 and Eq. 11

The binding of DCP to DhaA31 did not result in observable changes in the fluorescence signal even at concentrations reaching the solubility limit. This clearly indicates a significant drop in the affinity of the DhaA31 variant for the binding of DCP with respect to DhaAwt, which is also in agreement with the theoretical observations. The binding of bromide and chloride to DhaAwt and DhaA31 was investigated by using stopped-flow fluorescence/anisotropy measurements. Fluorescence anisotropy provided reasonable signal quality for the binding of both halide ions, while the fluorescence intensity could be analyzed only for bromide, but not for chloride. Fast binding kinetics for both chloride and bromide occurred in the dead time of the instrument (0.3 ms) with both enzymes. This suggests that the halide release is a fast process. The fluorescence and its anisotropy signal reached rapid equilibrium and only the difference in initial signal could

be observed. The dissociation constant (K_d) calculated from the equilibrium levels of fluorescence intensity and fluorescence anisotropy obtained upon binding of bromide are presented (Table 6). The results show that the overall binding affinity of DhaAwt towards bromide is low in comparison to other haloalkane dehalogenases^{24,25}, and it was further decreased by the mutations present in DhaA31. In the case of chloride, the dissociation constant was higher than the solubility limit (2 M) for both enzyme variants, which represents even lower binding affinity. We conclude that the halide release is a fast process which does not limit the overall steady-state catalytic performance of both DhaA31 and DhaAwt.

Steady-state kinetics

To understand the effect of key mutations introduced in DhaAwt towards DhaA31 and support theoretical predictions with new experimental data, a set of single and multiple-point mutants were constructed and evaluated in terms of activity and kinetics. These studies complement the simulations performed with the alanine mutants of DhaA31 at positions 176, 245 and 246, reported above, and assess the importance of those mutations alone in the wild-type. It was found that the single-point mutations C176Y and V245F increased significantly the activity of DhaAwt towards TCP by 2.5- and 1.9-fold, respectively. Conversely, mutation L246I impaired the activity as compared to the wild-type (Table 7). The double-point mutant C176Y+V245F exhibited 6.1-fold activity increase. This is close to the activity achieved from all five mutations in DhaA31 (8.1-fold). Interestingly, the effects of C176Y and V245F mutations are not additive but synergistic. This is in good agreement with our theoretical findings that residue Y176 in DhaA31 influences the positioning of F245, which subsequently pushes TCP towards its reactive configuration.

Table 7. Activity and steady-state kinetic parameters of the DhaA variants towards the hydrolysis of TCP.*

Enzyme variant	Mutation	Specific activity ($\mu\text{mol.s}^{-1}.\text{mg}^{-1}$)	Relative activity (mut/wt)	k_{cat} (s^{-1})	$K_{0.5}$ (mM)	K_{si} (mM)	n	$k_{\text{cat}}/K_{0.5}$ ($\text{s}^{-1}.\text{mM}^{-1}$)	Rel. $k_{\text{cat}}/K_{0.5}$ (mut/wt)
DhaAwt	n.a.	0.0020	1.0	0.046 ± 0.009	0.61 ± 0.03	n.a.	1.63 ± 0.12	0.083 ± 0.005	1.0
DhaA04	C176Y	0.0050	2.5	0.066 ± 0.000	0.35 ± 0.01	23.15 ± 0.21	1.62 ± 0.03	0.189 ± 0.002	2.3
DhaA137	L246I	0.0012	0.6	0.025 ± 0.003	0.57 ± 0.03	n.a.	1.48 ± 0.01	0.043 ± 0.006	0.52
DhaA138	V245F	0.0037	1.9	0.042 ± 0.001	0.54 ± 0.02	n.a.	1.38 ± 0.02	0.079 ± 0.001	0.95
DhaA133	C176Y, V245F	0.0122	6.1	0.29 ± 0.01	0.38 ± 0.01	n.a.	1.56 ± 0.11	0.76 ± 0.04	9.2
DhaA31	I135F, C176Y, V245F, L246I, Y273F	0.0161	8.1	1.09 ± 0.02	0.65 ± 0.01	5.47 ± 0.16	n.a.	1.68 ± 0.02	20.3

*Average values $\pm$ SEM; k_{cat} , catalytic constant; $K_{0.5}$, concentration of substrate at half-maximal velocity, $K_{0.5} = K_{\text{m}}$ when $n = 1$; K_{m} , Michaelis constant; K_{si} , substrate inhibition constant; n , Hill coefficient; $k_{\text{cat}}/K_{0.5}$, catalytic efficiency in; n.a., not applicable. The water solubility of TCP is 12.2 mM (1.8 g/L) at 25 °C⁸³.

It was found that k_{cat} increased with the C176F single mutation, while it remained almost the same in the case of V245F mutant, not following the same trend observed for the activity. However, when those two mutations were combined in DhaA133 variant, the increase in k_{cat} was much enhanced as compared to the single mutations alone, in a similar way as observed for the activity, revealing again the strong synergistic effect of those mutations. L246I mutation reduced

1
2
3 k_{cat} by half as compared to DhaAwt, which is in accordance with the drop observed in the
4
5 enzyme activity. When considering the catalytic efficiencies ($k_{\text{cat}}/K_{0.5}$), the trend was the same as
6
7 for k_{cat} but with higher relative improvements for the mutants bearing C176Y or C176Y+V245F
8
9 mutations (Table 7).
10
11

12
13 It was found that most of the variants studied here exhibited positive substrate cooperativity
14
15 with TCP ($n > 1$), except for DhaA31, indicating that the binding of the first TCP molecule to
16
17 those proteins facilitates the binding of other molecules. The cooperative effect of TCP in
18
19 DhaAwt translates into the fact that the binding of one TCP molecule to the catalytic cavity will
20
21 facilitate the binding of other TCP molecules, probably due to the widening of the tunnel
22
23 connecting the enzyme active site with the surrounding solvent. Of all tested DhaA variants, only
24
25 DhaA04 (C176Y mutation) and DhaA31 exhibited substrate inhibition towards TCP. In the case
26
27 of DhaA04, only weak substrate inhibition was observed, with K_{si} more than 60 times higher
28
29 than $K_{0.5}$ for the reaction with TCP. DhaA31 with TCP exhibited stronger substrate inhibition
30
31 that DhaA04. Such substrate inhibition found in DhaA31 may explain the lower activity
32
33 improvement from DhaAwt (8.1 fold) when compared to the improvement in the catalytic
34
35 efficiency of the same enzyme (20.3 fold).
36
37
38
39
40
41
42

43 DISCUSSION

44
45 DhaA31 is currently the best HLD enzyme in hydrolyzing the non-natural recalcitrant toxic
46
47 pollutant TCP, and it has been included in a multienzyme degradation pathway which converts
48
49 TCP to glycerol that can be used as the sole carbon and energy source^{21,84}. Therefore, this variant
50
51 is a remarkable achievement of protein engineering which deserves a proper mechanistic
52
53 understanding, as it can inspire the redesign of other biocatalysts. Most of the mutations present
54
55
56
57
58
59
60

1
2
3 in DhaA31 introduced bulky residues which considerably narrowed the transportation pathways
4
5 to the active site. The explanation generally accepted for the better performance of DhaA31 as
6
7 compared to DhaAwt was that the reduced accessibility of the solvent molecules to the active
8
9 site decreased their inhibitory effect on the nucleophilic attack to the carbon atom of the
10
11 substrate¹⁶. This hypothesis was disproved by all sets of simulations included here. These new
12
13 simulations were much longer and in larger number than the previous ones, were performed
14
15 under more varied conditions, and used a recent and more reliable force field¹⁶. They showed
16
17 that the hydration of the active sites of DhaA31 and DhaAwt do not differ significantly, and
18
19 hence the molecular basis for the higher catalytic performance of DhaA31 must lie elsewhere.

20
21
22 The diffusion of ligands through the protein structures is often a slow process and the
23
24 transitions between the bound and unbound states are usually rare events⁸⁵, particularly for large
25
26 molecules. In this work we studied both of these processes using classical and enhanced methods
27
28 of MD. The simulations with the free enzymes showed that p1 tunnel was the only relevant
29
30 tunnel in DhaA31 and that its opening rate was higher in DhaAwt than in DhaA31, although it
31
32 was able to open to maximum values in the same range (ca. 3 Å of bottleneck radius). This
33
34 finding elucidates why both enzymes can be active against a wide range of substrates¹⁶. Our
35
36 simulations with TCP were in agreement with those data and showed several events where the
37
38 TCP was bound to the enzymes in the time scales of hundreds of nanosecond. This evidence
39
40 showed that the binding of TCP to the tunnels of both enzymes is a relatively fast process, which
41
42 is in agreement with the experimental kinetic data. In the long time scales surveyed by the aMDs,
43
44 DhaAwt showed approximately twice more binding than DhaA31, but it also showed the release
45
46 of TCP after its binding, whereas in DhaA31 the substrate release was never observed. This may
47
48 explain the similar overall substrate affinities presented by the two enzymes. DhaAwt also
49
50
51
52
53
54
55
56
57
58
59
60

showed the simultaneous binding of several TCP molecules, not observed in DhaA31. This was in good agreement with the cooperative binding of TCP observed in the experimental assays with DhaAwt and not with DhaA31.

The nucleophilic attack to the substrate (S_N2) is the rate-limiting step in DhaAwt which was improved in the DhaA31 mutant¹⁶. This step was studied by evaluating the propensity of the substrate to preserve and possibly re-generate the productive binding during MD simulations. We observed considerably higher fractions of reactive configurations of TCP sampled in the simulations with DhaA31 than with DhaAwt by 5.5-fold. Moreover, TCP attained more favorable distributions in terms of the quality of the reactive configurations in DhaA31 than in DhaAwt. This suggests that DhaA31 favors more the reactive positioning of TCP in the active site than DhaAwt. The free energy of binding (ΔG_{bind}) of TCP calculated for these reactive complexes showed no significant difference between DhaA31 and DhaAwt. This fact suggests that the main factor leading to the adoption of reactive configurations is not a higher enthalpy stabilization, but rather a lower entropic penalty – TCP is restricted to a lower number of possible conformations in the whole active site of DhaA31 than in DhaAwt. This is supported by the fractions of snapshots with TCP in closer distances to the nucleophile, which was increased in DhaA31 as compared to DhaAwt. Hence, the higher confinement of TCP within the active site of DhaA31 seems to be the major promoter of the larger number of reactive configurations found with this variant. We have determined the reaction potential energy surface using hybrid QM/MM calculations to estimate the energy barrier for the nucleophilic attack between the D106 and the substrate to produce the chloride ion and the alkyl-enzyme intermediate. It is known that the subsequent hydrolytic step is fast in both of these enzymes, and therefore the overall chemical step of the catalytic cycle is limited by the S_N2 reaction, which allowed us to omit the

simulation of this step. The activation energy barrier was lowest for DhaA31 with TCP in *S*-orientation, with $\Delta G^\ddagger$ of 17.08 ± 0.12 kcal/mol, which is 1.61 kcal/mol below the lowest value found for DhaAwt. This difference is statistically very significant and represents 14-fold higher probability of the reaction occurrence in DhaA31 at 37 °C. These results, in combination with the 5.5 to 11.6 times higher rates of reactive complexes observed with DhaA31, are in very good agreement with the experimental evidence that the rate of the S_N2 step in DhaA31 increased 145-fold with respect to DhaAwt¹⁶. The differences between the calculated interactions of TCP with the enzymes suggested the important roles of V245F and C176Y mutations in increasing the productive binding of TCP in DhaA31, as compared to DhaAwt. It also suggested the unfavorable role of the L246I mutation. The kinetic experiments performed on the dissected mutants of DhaAwt confirmed these predictions and showed C176Y and V245F as the individual mutations that increased the activity the most towards TCP (2.5 and 1.9-fold, respectively), and L246I decreasing the activity with respect to DhaAwt. They also confirmed the synergistic effects of mutations C176Y and V245F combined, as theoretically predicted.

The calculated binding interactions showed that TCP strongly interacts with the catalytic residues, but also with several residues in the tunnel bottleneck region. The residues of DhaA31 with the most significant interaction differences as compared to DhaAwt – Y176, F245 and I246 – were subjected to alanine scanning and simulated with TCP, in order to determine their importance to the number of reactive complexes observed in DhaA31. All three positions were found important to the formation of reactive configurations of TCP within the active site. The residue at position 245 was the most significant of all, with a dramatic decrease of the reactive complexes in the F245A mutant, suggesting that the presence of a bulky residue at position 245 is fundamental for the positioning of TCP in the reactive conformation. The Y176A mutant of

DhaA31 showed high rates of TCP release as compared to any other mutant. This highlights the role of the bottleneck residue Y176 in keeping TCP in the active site. This finding, together with the fact that Y176 had very low interaction with TCP in the reactive complex, demonstrate the main role of this residue as a gatekeeper to increase the residence time of TCP within the active site, thus increasing the chances of the first chemical step to occur. It also seemed to induce the proper orientation of F245 that favors the reactive configuration of TCP.

The product release is the rate-limiting step in DhaA31. The previous studies were uncertain whether it was the chloride or the alcohol release limiting this step, but the transient kinetic studies reported here clearly showed that DCP was the slowest product to be released from both DhaA31 and DhaAwt. Our simulations with the DCP and chloride products are in excellent agreement with these findings. Cl^- tended to be released quite quickly, in average within tens of nanoseconds, although such time scale was influenced by the lack of polarizable force fields in these simulations. DCP proved to be released much more slowly from DhaA31 than from DhaAwt. Whereas in the MD simulations several unbinding events were observed with DhaAwt, none was found with DhaA31, and DCP was released from this enzyme only in the long time scales surveyed by the aMDs. Based on the interactions with DCP, the C176Y mutation present in DhaA31 was the main responsible for keeping DCP within the active site and inner part of the tunnel for significantly longer time than for DhaAwt. After DCP was released, it often rebound into the tunnels of DhaAwt, both in the MDs and aMDs, while with DhaA31 it was observed only in 1 out of 16 aMD trajectories. This is indicative of product inhibition in the case of DhaAwt, also confirmed experimentally. Both TCP and DCP, while outside of the enzymes, tended to interact for long periods at specific regions of the protein surface. These regions, in the shape of cavities or grooves, were identified and clearly corresponded to local minima of energy.

The residues in these regions formed mostly van der Waals interactions with TCP and DCP, but some also formed hydrogen bonds.

CONCLUSIONS

The comprehensive investigation of the catalytic cycle carried out here has provided a global understanding of the different processes involved in the enzymatic hydrolysis of TCP by the haloalkane dehalogenases. The comparative study of two enzyme variants (the DhaA wild type and the mutant DhaA31) has shed new light on the molecular basis for their different catalytic activities and kinetics. It has also allowed identifying the key elements responsible for the improvements observed in the mutant DhaA31 and the new kinetic limitations. The combination of such knowledge will allow us to improve the catalytic efficiency even further. This can be achieved through a balanced optimization of the different processes, aiming at facilitating the rate-limiting steps without slowing other ones. Such computational approach is applicable to a large number of other systems, and may contribute to a deep understanding of biomolecular systems of very diverse interests and more robust strategies for rational design of biocatalysts.

ASSOCIATED CONTENT

Supporting information

Figure S1. Reaction mechanism of the haloalkane dehalogenase DhaA.

Figure S2. Sequence of *dhaA133* gene.

Figure S3. Raw data from isothermal titration calorimetry experiments of DhaA variants with TCP.

Figure S4. Steady-state kinetics of DhaA variants towards TCP.

Figure S5. Time variation of the RMSD values in the MD and aMD simulations of DhaA31 and DhaAwt.

Table S1. Main parameters obtained for the top-ranked tunnels calculated in the MD simulations of DhaA31 and DhaAwt.

Figure S6. Radial distribution function (RDF) of water molecules around the D106-Cg atom in the MD and aMD simulations with DhaA31 and DhaAwt.

Figure S7. Time variation of the solvation shells in the MD and aMD simulations of DhaA31 and DhaAwt.

Figure S8. Time variation of the RMSD values in the MD and aMD simulations of DhaA31 and DhaAwt in the presence of three molecules of TCP substrate.

Figure S9. Time variation of the TCP distances to the active site in the MD and aMD simulations of DhaA31 and DhaAwt in the presence of three molecules of TCP substrate.

Figure S10. Average bottleneck radii of p1 tunnel binned according to the distance of TCP to the active site cavity for all MD and aMD simulations of DhaA31 and DhaAwt.

Table S2. Hydration spheres of the nucleophile and the active site in all sets of MD and aMD simulations.

Figure S11. Radial distribution function (RDF) of water molecules around the D106-Cg atom in the MD and aMD simulations with DhaA31 and DhaAwt.

Figure S12. Time variation of the solvation shells in the MD and aMD simulations of DhaA31 and DhaAwt in the presence of three molecules of TCP substrate.

Figure S13. Time variation of the solvation shells in the MD and aMD simulations of DhaA31 and DhaAwt containing one molecule of TCP in the active site.

Figure S14. Time variation of the solvation shells in the MD and aMD simulations of DhaA31 and DhaAwt containing one molecule of DCP and Cl⁻ in the active site.

Figure S15. Time variation of the RMSD values in the MD and aMD simulations of DhaA31 and DhaAwt in the presence of one molecules of TCP substrate in the active site.

Figure S16. Average bottleneck radii of p1 tunnel binned according to the TCP distance to the active site cavity for all MD and aMD simulations of DhaA31 and DhaAwt containing one molecule of TCP in the active site.

Figure S17. Time evolution of the total number of reactive complexes sampled in the MD simulation of DhaA31 and DhaAwt containing one molecule of TCP in the active site.

Figure S18. Quality of the reactive conformation of TCP in the reactive complexes.

Figure S19. S_N2 reaction of TCP with DhaA31 and DhaAwt.

Figure S20. Residue-wise decomposition of the binding interactions of TCP (ΔG_{bind}) during the MD simulations with DhaA31 and DhaAwt.

Figure S21. Orientation of the side chain of F245 residue in representative structures of DhaA31 and Y176A mutant.

Figure S22. Time variation of the RMSD values in the MD and aMD simulations of DhaA31 and DhaAwt in the presence of one molecule of DCP and Cl⁻ in the active site.

Figure S23. Time variation of DCP and Cl⁻ distances to the active site in the MD and aMD simulations of DhaA31 and DhaAwt.

Figure S24. Average bottleneck radii of p1 tunnel binned according to the distance of DCP to the active site cavity, for all MD and aMD simulations of DhaA31 and DhaAwt containing DCP and Cl⁻ in the active site.

Table S3. Hydrogen bonds formed by DCP with the protein residues, in the MD simulations with DhaA31 and DhaAwt.

ACKNOWLEDGEMENTS

The authors would like to thank the Grant Agency of the Czech Republic (P503/12/0572) and the Czech Ministry of Education (LO1214, LQ1605, LH14027), the Masaryk University (MUNI/M/1888/2014) for financial support. SM is supported by the SoMoPro II Programme (project BIOGATE, Nr. 4SGA8519), which is co-financed by the People Programme (Marie Curie action) of the EU's Seventh Framework Programme according to the REA Grant Agreement No. 291782, and the South Moravian Region. The computational resources were provided by CESNET (LM2015042) and the CERIT Scientific Cloud (LM2015085), as part of the "Projects of Large Research, Development, and Innovations Infrastructures" programme

(LM2015051, LM2015047, LM2015055). The authors are also grateful to Veronika Liskova, MSc, for her valuable help with the construction of the DhaA variants. This publication only reflects the authors' views and the Union is not liable for any use that may be made of the information contained herein.

ABBREVIATIONS: HLD, haloalkane dehalogenase; TCP, 1,2,3-trichloropropane; DCP, 2,3-dichloropropan-1-ol; aMD, accelerated molecular dynamics; $E \cdot S^{\text{REACT}}$ or ES^{REACT} , reactive complex; RMSD, root mean square deviation; RDF, radial distribution function; LIE, linear interaction energy; PES, potential energy surface; QM/MM, quantum mechanics/molecular mechanics; MM/GBSA, molecular mechanics/generalized Born surface area; ΔG_{bind} , binding free energy; $\Delta G^{\ddagger}$, Gibbs energy of activation.

REFERENCES

- (1) Adrio, J. L.; Demain, A. L. Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules* **2014**, *4*, 117–139.
- (2) Choi, J.-M.; Han, S.-S.; Kim, H.-S. Industrial Applications of Enzyme Biocatalysis: Current Status and Future Aspects. *Biotechnol. Adv.* **2015**, *33*, 1443–1454.
- (3) Swanson, P. E. Dehalogenases Applied to Industrial-Scale Biocatalysis. *Curr. Opin. Biotechnol.* **1999**, *10*, 365–369.
- (4) Bidmanova, S.; Chaloupkova, R.; Damborsky, J.; Prokop, Z. Development of an Enzymatic Fiber-Optic Biosensor for Detection of Halogenated Hydrocarbons. *Anal. Bioanal. Chem.* **2010**, *398*, 1891–1898.
- (5) Koudelakova, T.; Bidmanova, S.; Dvorak, P.; Pavelka, A.; Chaloupkova, R.; Prokop, Z.; Damborsky, J. Haloalkane Dehalogenases: Biotechnological Applications. *Biotechnol. J.* **2013**, *8*, 32–45.
- (6) Janssen, D. B. Evolving Haloalkane Dehalogenases. *Curr. Opin. Chem. Biol.* **2004**, *8*, 150–159.
- (7) Chovancová, E.; Kosinski, J.; Bujnicki, J. M.; Damborský, J. Phylogenetic Analysis of Haloalkane Dehalogenases. *Proteins* **2007**, *67*, 305–316.

- (8) Koudelakova, T.; Chovancova, E.; Brezovsky, J.; Monincova, M.; Fortova, A.; Jarkovsky, J.; Damborsky, J. Substrate Specificity of Haloalkane Dehalogenases. *Biochem. J.* **2011**, *435*, 345–354.
- (9) Koudelakova, T.; Chaloupkova, R.; Brezovsky, J.; Prokop, Z.; Sebestova, E.; Hesseler, M.; Khabiri, M.; Plevaka, M.; Kulik, D.; Kuta Smatanova, I.; Rezacova, P.; Ettrich, R.; Bornscheuer, U. T.; Damborsky, J. Engineering Enzyme Stability and Resistance to an Organic Cosolvent by Modification of Residues in the Access Tunnel. *Angew. Chem. Int. Ed Engl.* **2013**, *52*, 1959–1963.
- (10) Bednar, D.; Beerens, K.; Sebestova, E.; Bendl, J.; Khare, S.; Chaloupkova, R.; Prokop, Z.; Brezovsky, J.; Baker, D.; Damborsky, J. FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLoS Comput. Biol.* **2015**, *11*, e1004556.
- (11) Pavelka, A.; Chovancova, E.; Damborsky, J. HotSpot Wizard: A Web Server for Identification of Hot Spots in Protein Engineering. *Nucleic Acids Res.* **2009**, *37*, W376–383.
- (12) Chovancová, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8*, e1002708.
- (13) Bendl, J.; Stourac, J.; Salanda, O.; Pavelka, A.; Wieben, E. D.; Zendulka, J.; Brezovsky, J.; Damborsky, J. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Comput. Biol.* **2014**, *10*, e1003440.
- (14) Prokop, Z.; Gora, A.; Brezovsky, J.; Chaloupkova, R.; Stepankova, V.; Damborsky, J. Engineering of Protein Tunnels: Keyhole-Lock-Key Model for Catalysis by the Enzymes with Buried Active Sites. In *Protein Engineering Handbook*; Lutz, S., Bornscheuer, U. T., Eds.; Wiley-VCH: Weinheim, 2012; Vol. 3, pp 421–464.
- (15) Marques, S. M.; Brezovsky, J.; Damborsky, J. Role of Tunnels and Gates in Enzymatic Catalysis. In *Understanding Enzymes: Function, Design, Engineering, and Analysis*; Svendsen, A., Ed.; Pan Stanford Publishing: Singapore, 2016; pp 421–463.
- (16) Pavlova, M.; Klvana, M.; Prokop, Z.; Chaloupkova, R.; Banas, P.; Otyepka, M.; Wade, R. C.; Tsuda, M.; Nagata, Y.; Damborsky, J. Redesigning Dehalogenase Access Tunnels as a Strategy for Degrading an Anthropogenic Substrate. *Nat. Chem. Biol.* **2009**, *5*, 727–733.
- (17) Klvana, M.; Pavlova, M.; Koudelakova, T.; Chaloupkova, R.; Dvorak, P.; Prokop, Z.; Stsiapanava, A.; Kutý, M.; Kuta-Smatanova, I.; Dohnalek, J.; Kulhanek, P.; Wade, R. C.; Damborsky, J. Pathways and Mechanisms for Product Release in the Engineered Haloalkane Dehalogenases Explored Using Classical and Random Acceleration Molecular Dynamics Simulations. *J. Mol. Biol.* **2009**, *392*, 1339–1356.
- (18) Moriuchi, R.; Tanaka, H.; Nikawadori, Y.; Ishitsuka, M.; Ito, M.; Ohtsubo, Y.; Tsuda, M.; Damborsky, J.; Prokop, Z.; Nagata, Y. Stepwise Enhancement of Catalytic Performance of Haloalkane Dehalogenase LinB towards β -Hexachlorocyclohexane. *AMB Express* **2014**, *4*, 72.
- (19) Prokop, Z.; Sato, Y.; Brezovsky, J.; Mozga, T.; Chaloupkova, R.; Koudelakova, T.; Jerabek, P.; Stepankova, V.; Natsume, R.; van Leeuwen, J. G. E.; Janssen, D. B.; Florian, J.; Nagata, Y.; Senda, T.; Damborsky, J. Enantioselectivity of Haloalkane Dehalogenases and Its Modulation by Surface Loop Engineering. *Angew. Chem. Int. Ed Engl.* **2010**, *49*, 6111–6115.

- (20) Biedermannová, L.; Prokop, Z.; Gora, A.; Chovancová, E.; Kovács, M.; Damborsky, J.; Wade, R. C. A Single Mutation in a Tunnel to the Active Site Changes the Mechanism and Kinetics of Product Release in Haloalkane Dehalogenase LinB. *J. Biol. Chem.* **2012**, *287*, 29062–29074.
- (21) Dvorak, P.; Bidmanova, S.; Damborsky, J.; Prokop, Z. Immobilized Synthetic Pathway for Biodegradation of Toxic Recalcitrant Pollutant 1,2,3-Trichloropropane. *Environ. Sci. Technol.* **2014**, *48*, 6859–6866.
- (22) Verschueren, K. H.; Seljée, F.; Rozeboom, H. J.; Kalk, K. H.; Dijkstra, B. W. Crystallographic Analysis of the Catalytic Mechanism of Haloalkane Dehalogenase. *Nature* **1993**, *363*, 693–698.
- (23) Kutý, M.; Damborský, J.; Prokop, M.; Koča, J. A Molecular Modeling Study of the Catalytic Mechanism of Haloalkane Dehalogenase. 2. Quantum Chemical Study of Complete Reaction Mechanism. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 736–741.
- (24) Schanstra, J. P.; Kingma, J.; Janssen, D. B. Specificity and Kinetics of Haloalkane Dehalogenase. *J. Biol. Chem.* **1996**, *271*, 14747–14753.
- (25) Prokop, Z.; Monincová, M.; Chaloupková, R.; Klvaňa, M.; Nagata, Y.; Janssen, D. B.; Damborský, J. Catalytic Mechanism of the Haloalkane Dehalogenase LinB from *Sphingomonas Paucimobilis* UT26. *J. Biol. Chem.* **2003**, *278*, 45094–45100.
- (26) Negri, A.; Marco, E.; Damborsky, J.; Gago, F. Stepwise Dissection and Visualization of the Catalytic Mechanism of Haloalkane Dehalogenase LinB Using Molecular Dynamics Simulations and Computer Graphics. *J. Mol. Graph. Model.* **2007**, *26*, 643–651.
- (27) Rose, P. W.; Bi, C.; Bluhm, W. F.; Christie, C. H.; Dimitropoulos, D.; Dutta, S.; Green, R. K.; Goodsell, D. S.; Prlić, A.; Quesada, M.; Quinn, G. B.; Ramos, A. G.; Westbrook, J. D.; Young, J.; Zardecki, C.; Berman, H. M.; Bourne, P. E. The RCSB Protein Data Bank: New Resources for Research and Education. *Nucleic Acids Res.* **2013**, *41*, D475–D482.
- (28) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating PKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33* (suppl 2), W368–W371.
- (29) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*; University of California: San Francisco, 2014.
- (30) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712–725.
- (31) Joung, I. S.; Cheatham, Thomas E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (32) Joung, I. S.; Cheatham, T. E. Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, *113*, 13279–13290.

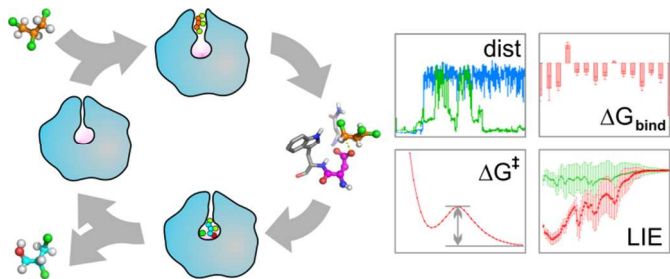
- (33) Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *J. Am. Chem. Soc.* **2014**, *136*, 13959–13962.
- (34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (35) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.
- (36) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.
- (37) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California: San Francisco, 2012.
- (38) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (39) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (40) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
- (41) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. Sampling of Slow Diffusive Conformational Transitions with Accelerated Molecular Dynamics. *J. Chem. Phys.* **2007**, *127*, 155102.
- (42) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F de Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
- (43) Markwick, P. R. L.; McCammon, J. A. Studying Functional Dynamics in Bio-Molecules Using Accelerated Molecular Dynamics. *Phys. Chem. Chem. Phys. PCCP* **2011**, *13*, 20053–20065.
- (44) Bucher, D.; Grant, B. J.; Markwick, P. R.; McCammon, J. A. Accessing a Hidden Conformation of the Maltose Binding Protein Using Accelerated Molecular Dynamics. *PLoS Comput. Biol.* **2011**, *7*, e1002034.
- (45) de Oliveira, C. A. F.; Grant, B. J.; Zhou, M.; McCammon, J. A. Large-Scale Conformational Changes of Trypanosoma Cruzi Proline Racemase Predicted by Accelerated Molecular Dynamics Simulation. *PLoS Comput. Biol.* **2011**, *7*, e1002178.
- (46) Wereszczynski, J.; McCammon, J. A. Nucleotide-Dependent Mechanism of Get3 as Elucidated from Free Energy Calculations. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7759–7764.
- (47) Miao, Y.; Nichols, S. E.; Gasper, P. M.; Metzger, V. T.; McCammon, J. A. Activation and Dynamic Network of the M2 Muscarinic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10982–10987.

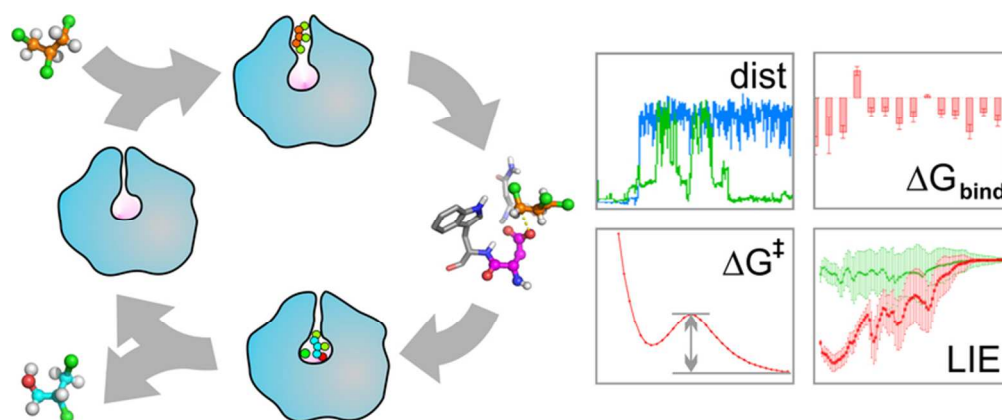
- (48) Wiesler, S. C.; Weinzierl, R. O. J.; Buck, M. An Aromatic Residue Switch in Enhancer-Dependent Bacterial RNA Polymerase Controls Transcription Intermediate Complex Activity. *Nucleic Acids Res.* **2013**, *41*, 5874–5886.
- (49) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4*, 17.
- (50) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (51) Vanqualef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39*, W511–517.
- (52) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- (53) *The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC.*
- (54) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- (55) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph.* **1996**, *14*, 33–38.
- (56) Grubbs, F. E. Sample Criteria for Testing Outlying Observations. *Ann. Math. Stat.* **1950**, *21*, 27–58.
- (57) Hur, S.; Kahn, K.; Bruce, T. C. Comparison of Formation of Reactive Conformers for the SN2 Displacements by CH₃CO₂⁻ in Water and by Asp124-CO₂⁻ in a Haloalkane Dehalogenase. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2215–2219.
- (58) Daniel, L.; Burycka, T.; Prokop, Z.; Damborsky, J.; Brezovsky, J. Mechanism-Based Discovery of Novel Substrates of Haloalkane Dehalogenases Using in Silico Screening. *J. Chem. Inf. Model.* **2015**, *55*, 54–62.
- (59) Ranaghan, K. E.; Mulholland, A. J. Investigations of Enzyme-Catalysed Reactions with Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods. *Int. Rev. Phys. Chem.* **2010**, *29*, 65–133.
- (60) Lonsdale, R.; N. Harvey, J.; J. Mulholland, A. A Practical Guide to Modelling Enzyme-Catalysed Reactions. *Chem. Soc. Rev.* **2012**, *41*, 3025–3038.
- (61) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.
- (62) Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Trans Math Softw* **1997**, *23*, 550–560.
- (63) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

- (64) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461.
- (65) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230.
- (66) Bosma, T.; Damborský, J.; Stucki, G.; Janssen, D. B. Biodegradation of 1,2,3-Trichloropropane through Directed Evolution and Heterologous Expression of a Haloalkane Dehalogenase Gene. *Appl. Environ. Microbiol.* **2002**, *68*, 3582–3587.
- (67) Liskova, V.; Stepankova, V.; Bednar, D.; Brezovsky, J.; Prokop, Z.; Chaloupkova, R.; Damborsky, J. Different Structural Origins of the Enantioselectivity of Haloalkane Dehalogenases toward Linear β -Haloalkanes: Open–Solvated versus Occluded–Desolvated Active Sites. *Angew Chem Int Ed Engl* **2017**, *56*, 4719–4723.
- (68) Iwasaki, I.; Utsumi, S.; Ozawa, T. New Colorimetric Determination of Chloride Using Mercuric Thiocyanate and Ferric Ion. *Bull. Chem. Soc. Jpn.* **1952**, *25*, 226–226.
- (69) Todd, M. J.; Gomez, J. Enzyme Kinetics Determined Using Calorimetry: A General Assay for Enzyme Activity? *Anal. Biochem.* **2001**, *296*, 179–187.
- (70) Freyer, M. W.; Lewis, E. A. Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. *Methods Cell Biol.* **2008**, *84*, 79–113.
- (71) Frasca, V. Using Isothermal Titration Calorimetry Techniques to Quantify Enzyme Kinetics. *Ind. Biotechnol.* **2016**, *12*, 207–211.
- (72) Kamenik, A. S.; Kahler, U.; Fuchs, J. E.; Liedl, K. R. Localization of Millisecond Dynamics: Dihedral Entropy from Accelerated MD. *J. Chem. Theory Comput.* **2016**, *12*, 3449–3455.
- (73) Janosi, L.; Ceccarelli, M. The Gating Mechanism of the Human Aquaporin 5 Revealed by Molecular Dynamics Simulations. *PLoS ONE* **2013**, *8*, e59897.
- (74) Alberga, D.; Nicolotti, O.; Lattanzi, G.; Nicchia, G. P.; Frigeri, A.; Pisani, F.; Benfenati, V.; Mangiatordi, G. F. A New Gating Site in Human Aquaporin-4: Insights from Molecular Dynamics Simulations. *Biochim. Biophys. Acta* **2014**, *1838*, 3052–3060.
- (75) Zielkiewicz, J. Structural Properties of Water: Comparison of the SPC, SPCE, TIP4P, and TIP5P Models of Water. *J. Chem. Phys.* **2005**, *123*, 104501.
- (76) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.
- (77) Izadi, S.; Onufriev, A. V. Accuracy Limit of Rigid 3-Point Water Models. *J. Chem. Phys.* **2016**, *145*, 074501.
- (78) Ozu, M.; Alvarez, H. A.; McCarthy, A. N.; Grigera, J. R.; Chara, O. Molecular Dynamics of Water in the Neighborhood of Aquaporins. *Eur. Biophys. J.* **2013**, *42*, 223–239.
- (79) Bar-Even, A.; Milo, R.; Noor, E.; Tawfik, D. S. The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. *Biochemistry (Mosc.)* **2015**, *54*, 4969–4977.
- (80) van Leeuwen, J. G. E.; Wijma, H. J.; Floor, R. J.; van der Laan, J.-M.; Janssen, D. B. Directed Evolution Strategies for Enantiocomplementary Haloalkane Dehalogenases: From Chemical Waste to Enantiopure Building Blocks. *Chembiochem Eur. J. Chem. Biol.* **2012**, *13*, 137–148.

- (81) Lafaquière, V.; Barbe, S.; Puech-Guenot, S.; Guieysse, D.; Cortés, J.; Monsan, P.; Siméon, T.; André, I.; Remaud-Siméon, M. Control of Lipase Enantioselectivity by Engineering the Substrate Binding Site and Access Channel. *ChemBioChem* **2009**, *10*, 2760–2771.
- (82) Schanstra, J. P.; Janssen, D. B. Kinetics of Halide Release of Haloalkane Dehalogenase: Evidence for a Slow Conformational Change. *Biochemistry (Mosc.)* **1996**, *35*, 5624–5632.
- (83) NTP (National Toxicology Program). 2016. Report on Carcinogens, 14th Ed.; Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service. <https://ntp.niehs.nih.gov/ntp/roc/content/profiles/trichloropropane.pdf>.
- (84) Kurumbang, N. P.; Dvorak, P.; Bendl, J.; Brezovsky, J.; Prokop, Z.; Damborsky, J. Computer-Assisted Engineering of the Synthetic Pathway for Biodegradation of a Toxic Persistent Pollutant. *ACS Synth. Biol.* **2014**, *3*, 172–181.
- (85) Rydzewski, J.; Nowak, W. Ligand Diffusion in Proteins via Enhanced Sampling in Molecular Dynamics. *Phys. Life Rev.* in press; doi:10.1016/j.plrev.2017.03.003.

TABLE OF CONTENTS GRAPHIC





TOC

35x14mm (600 x 600 DPI)

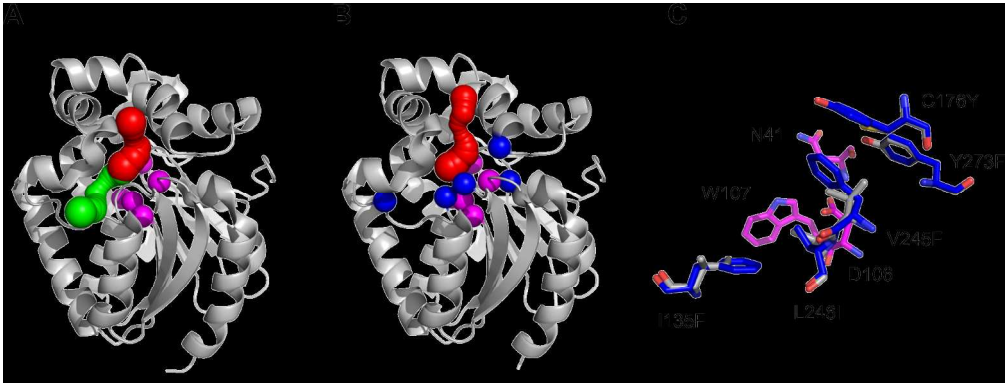


Figure 1

448x170mm (300 x 300 DPI)

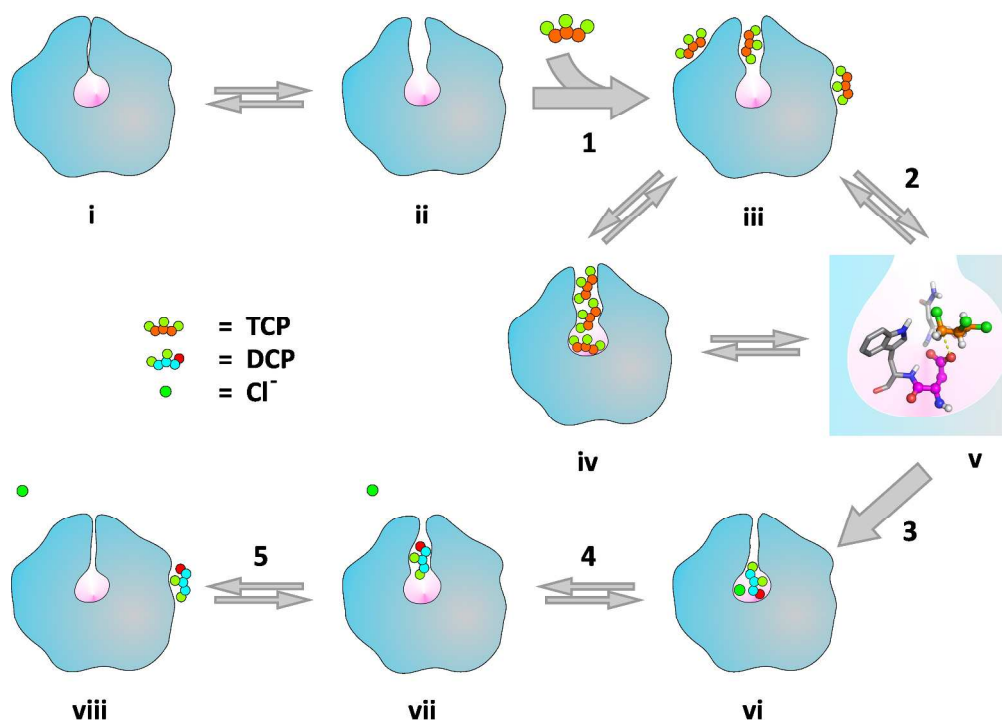


Figure 2

302x213mm (300 x 300 DPI)

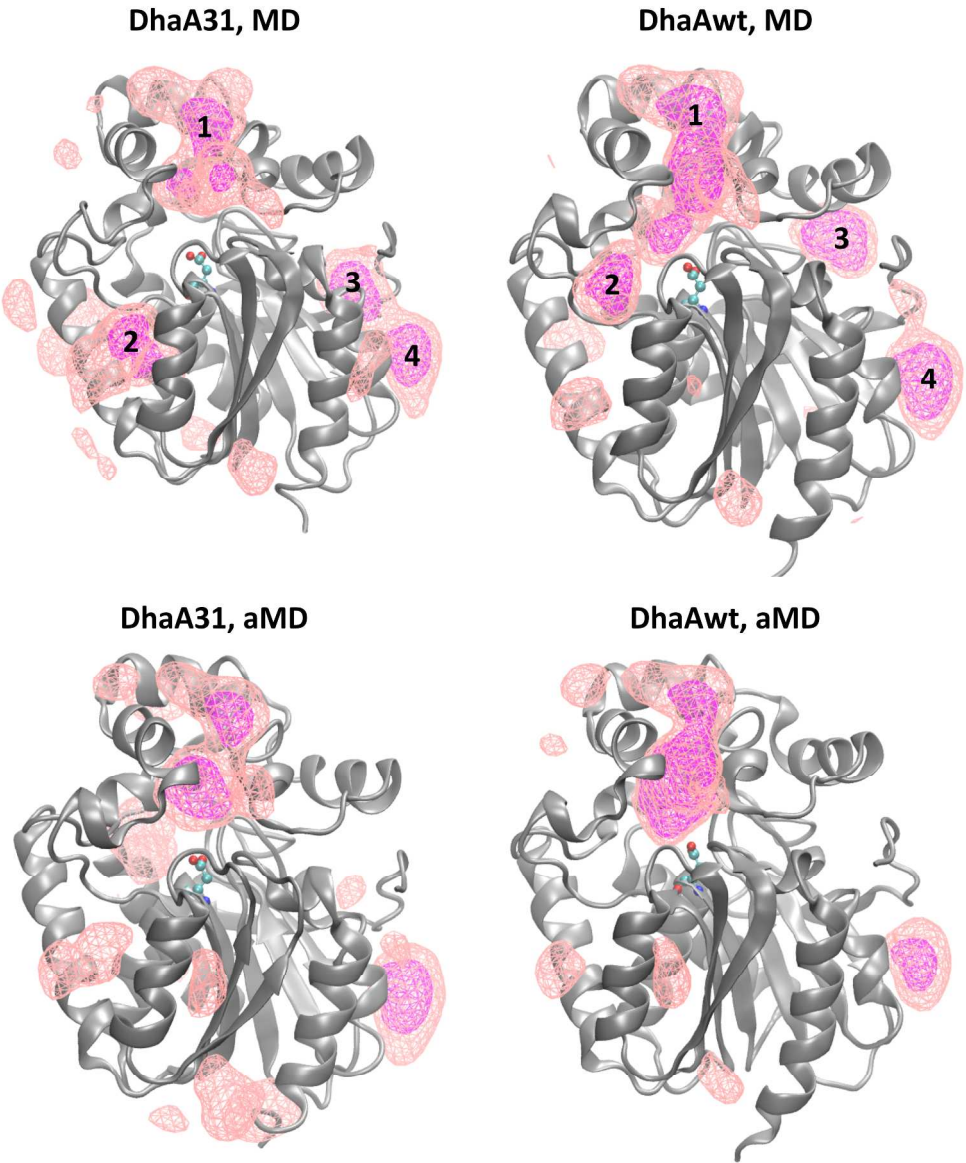


Figure 3

237x283mm (300 x 300 DPI)

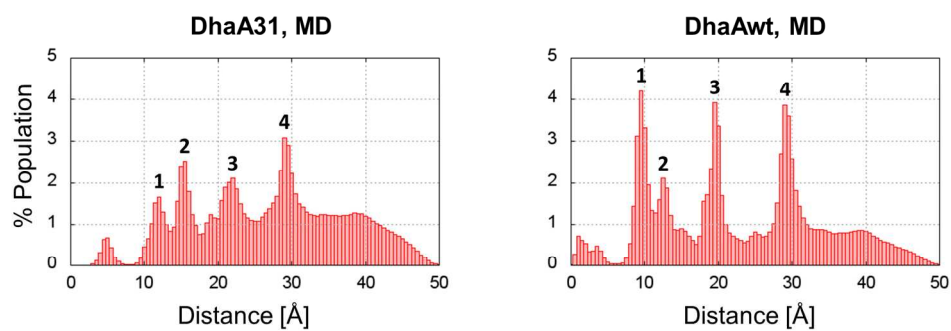


Figure 4

662x230mm (72 x 72 DPI)

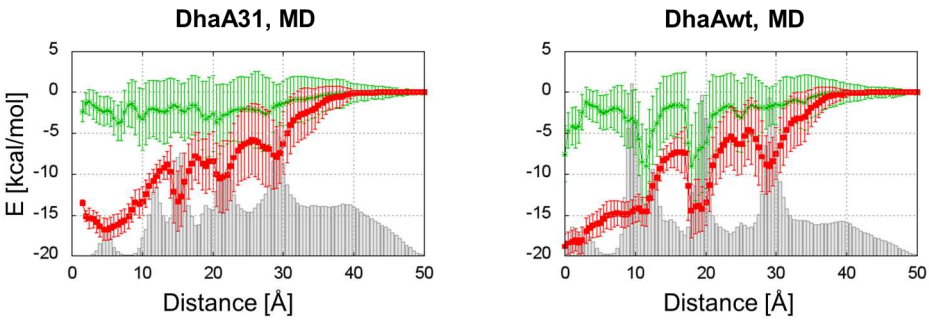


Figure 5

668x236mm (72 x 72 DPI)

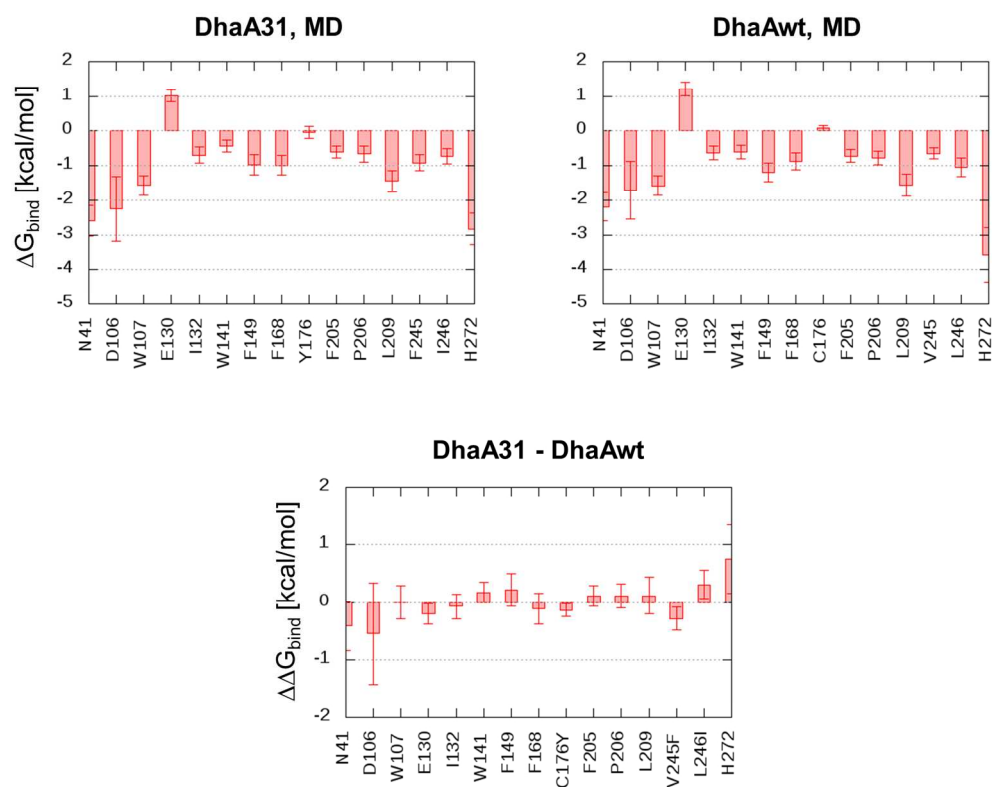


Figure 6

631x517mm (72 x 72 DPI)

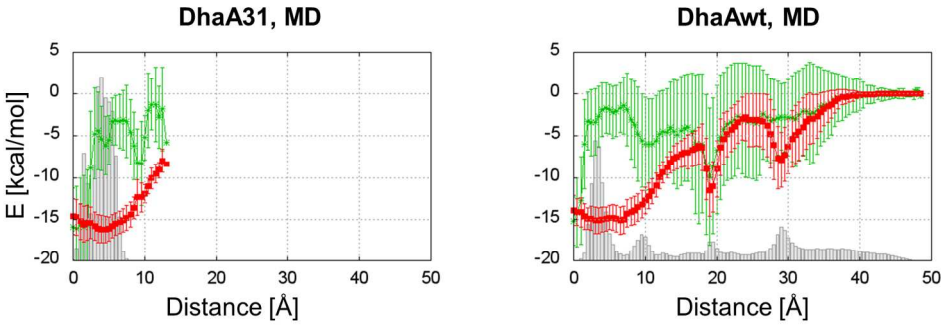


Figure 7

656x227mm (72 x 72 DPI)

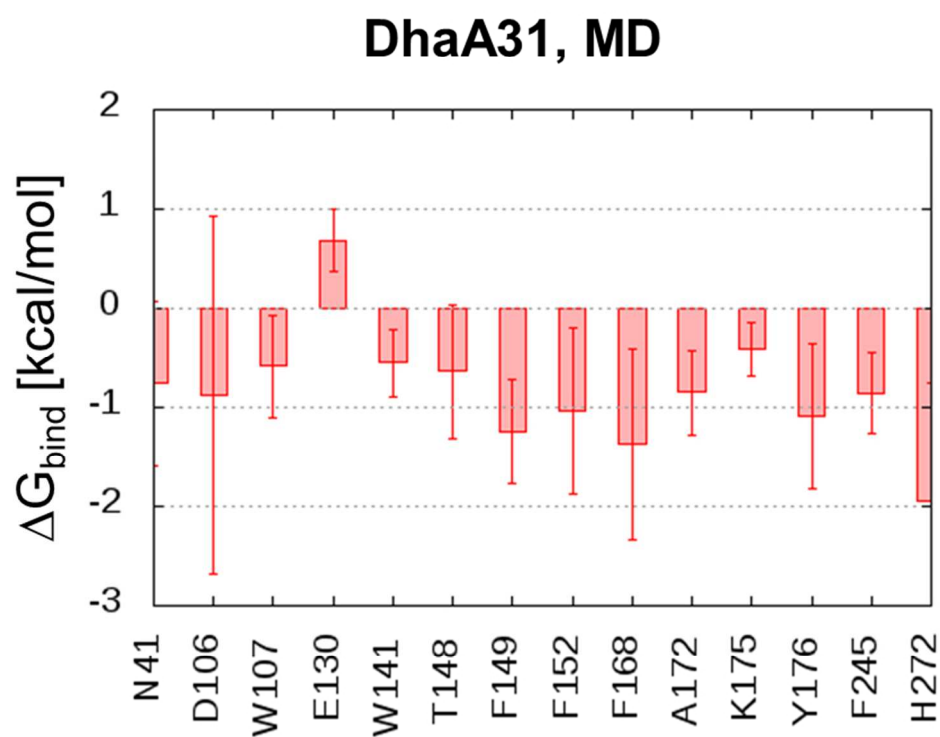


Figure 8

314x250mm (72 x 72 DPI)

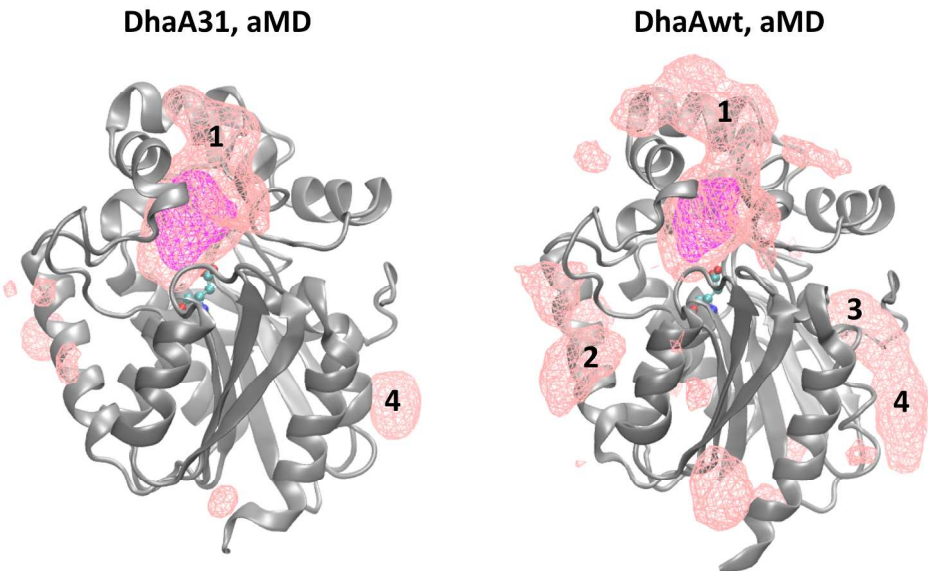


Figure 9

237x138mm (300 x 300 DPI)

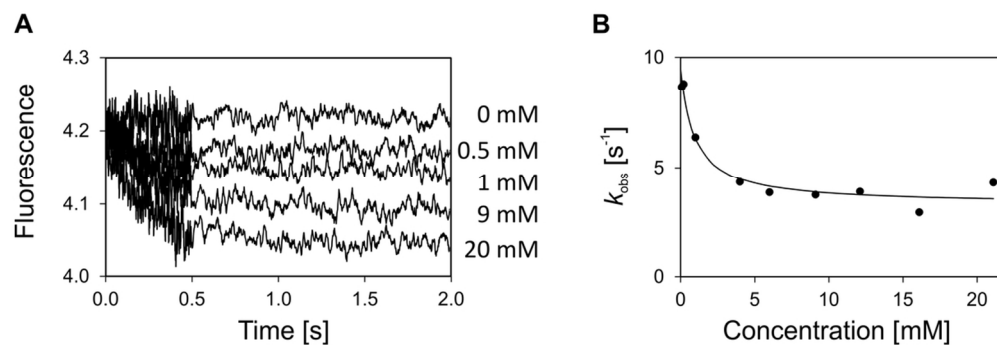
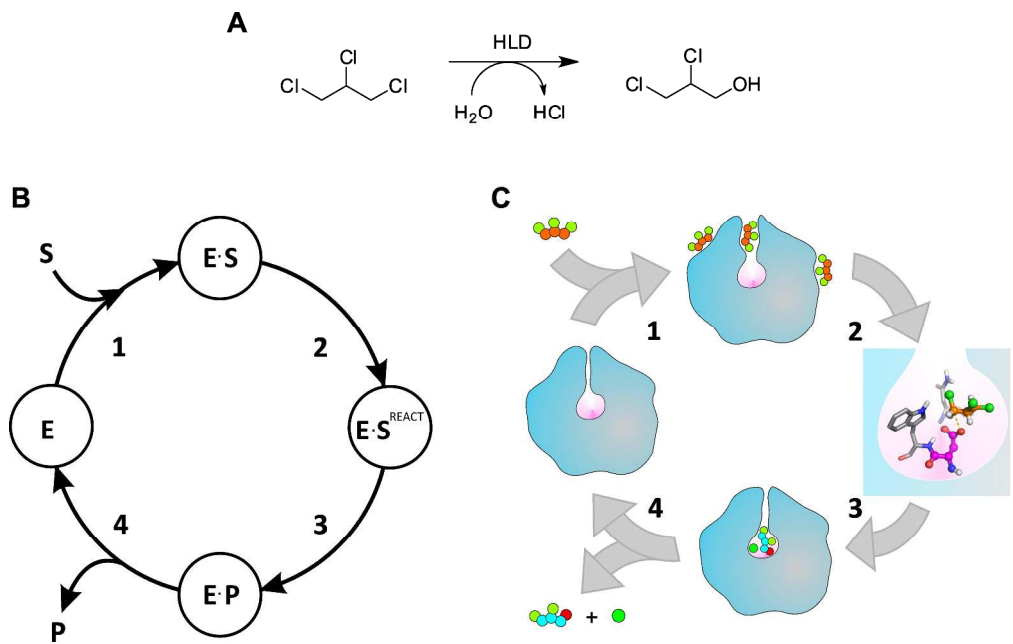


Figure 10

103x35mm (300 x 300 DPI)



Scheme 1

375x235mm (300 x 300 DPI)