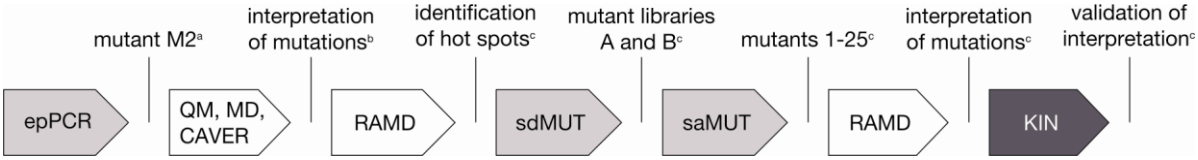
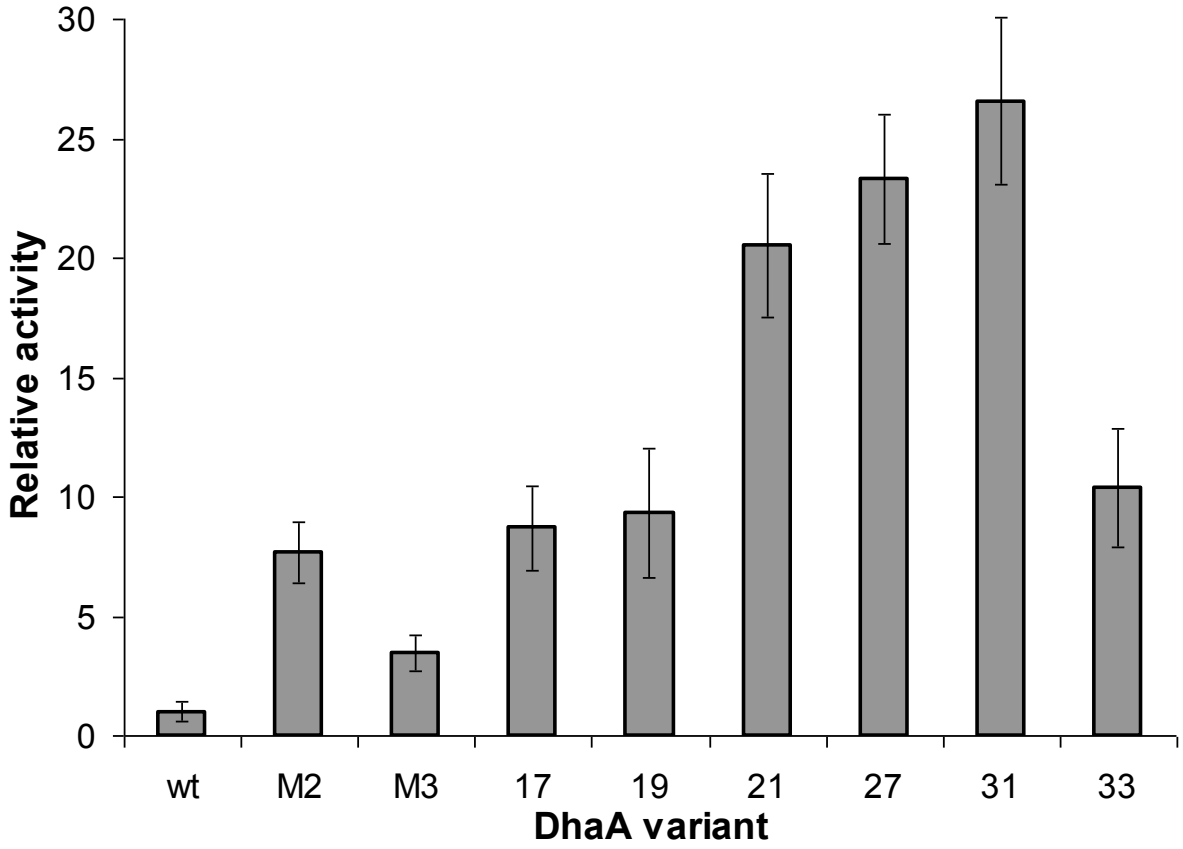
Supplementary Information

Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate

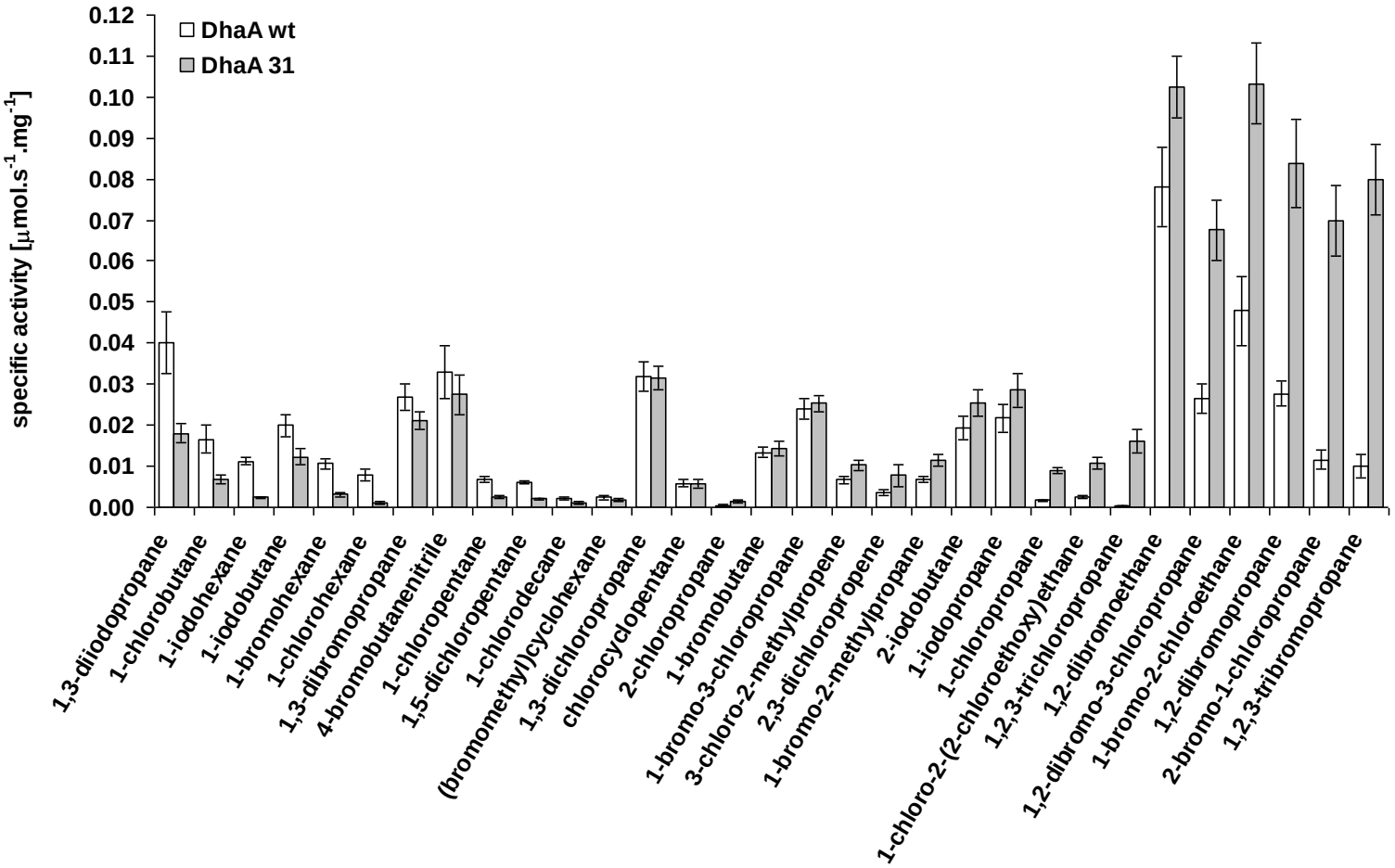
Martina Pavlova, Martin Klvana, Zbynek Prokop, Radka Chaloupkova, Pavel Banas, Michal Otyepka, Rebecca C. Wade, Masataka Tsuda, Yuji Nagata & Jiri Damborsky



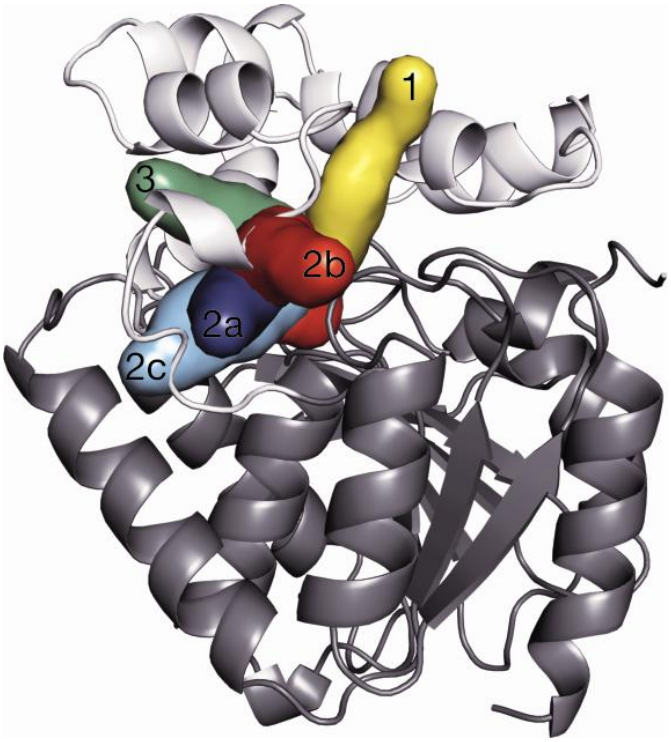
Supplementary Figure 1 Flowchart of rational design and directed evolution of DhaA to enhance its activity towards TCP. epPCR – error prone PCR; QM – Quantum Mechanics; MD – classical Molecular Dynamics; RAMD – Random Acceleration Molecular Dynamics; sdMUT – side-directed mutagenesis; saMUT – saturation mutagenesis; KIN – pre-steady state kinetics; experimental studies are indicated by gray and black arrows; theoretical studies are indicated by white arrows; [a] studies by Bosma *et al.* [1] and Gray *et al.* [2]; [b] study by Banas *et al.* [3]; [c] this study.



Supplementary Figure 2 Relative activities of wild-type DhaA, M2, M3 and the six variants with the highest activity towards TCP expressed as fold changes of specific activity relative to wild-type DhaA (0.0020 ± 0.0008 μmol s^{-1} mg^{-1}).



Supplementary Figure 3 Specific activities of the wild-type DhaA and the mutant 31 with various halogenated substrates (1,3-diiodopropane, **15**; 1-chlorobutane, **16**; 1-iodohexane, **17**; 1-iodobutane, **18**; 1-bromohexane, **19**; 4-bromobutanenitrile, **20**; 1-chloropentane, **21**; 1,5-dichloropentane, **22**; 1-chlorodecane, **23**; (bromomethyl)cyclohexane, **24**; 1,3-dichloropropane, **25**; chlorocyclopentane, **26**; 2-chloropropane, **27**; 1-bromobutane, **28**; 1-bromo-3-chloropropane, **29**; 3-chloro-2-methylprop-1-ene, **30**; 2,3-dichloroprop-1-ene, **31**; 1-chloro-2-methylpropane, **32**; 2-iodobutane, **33**; 1-iodopropane, **34**; 1-chloropropane, **35**; 1-chloro-2-(2-chloroethoxy)ethane, **36**; 1,2-dibromo-3-chloropropane, **37**; 1-bromo-2-chloroethane, **38**; 1,2-dibromopropane, **39**; 2-bromo-1-chloropropane, **40**; 1,2,3-tribromopropane, **41**; **1**, **6**, **7**, **9**).



Supplementary Figure 4 Product release pathways observed in computer simulations with wild-type DhaA and its variants. Cartoon model of wild-type DhaA: gray - main domain, white - cap domain; surface representation - RAMD pathways. See Supplementary Table 1 online for assignment of access pathways to individual variants.

Supplementary Table 1 Product release pathways observed in computer simulations with wild-type DhaA and its variants

DhaA	No. of RAMD simulations	Pathway frequency					
		1	2a	2b	2c	3	No. exits
wt	24	71% (17)	8 % (2)	0 %	0 %	4% (1)	17% (4)
M2	8	50% (4)	13% (1)	0 %	0 %	0 %	37% (3)
21	8	50% (4)	0 %	0 %	13 % (1)	0 %	37% (3)
27	7	29% (2)	0 %	14 % (1)	0 %	14 % (1)	43% (3)
31	8	13% (1)	0 %	0 %	0 %	0 %	87% (7)

The frequency at which product release along each pathway was observed in RAMD simulations is given as a percentage and the corresponding number is shown in parenthesis. 1 – main tunnel, 2a – slot tunnel.

Supplementary Table 2 Active clones of DhaA variants obtained from directed evolution

Protein	Residue*						Frequency§	Relative activity‖
DhaA	135	141	176	245	246	273		
wt†	I	W	C	V	L	Y	-	1
M2†	I	W	**Y**	V	L	**F**	-	4
M3†	I	**F**	**Y**	V	L	**F**	-	5
17†	**F**	**F**	**Y**	**M**	**I**	**F**	9¶	24
18	**Y**	**F**	**Y**	**F**	**I**	**F**	11¶	15
19†	**Y**	**F**	**Y**	**M**	**I**	**F**	3¶	26
20	**F**	**F**	**Y**	**F**	**I**	**F**	3¶	21
21†	**L**	**F**	**Y**	**F**	**I**	**F**	1	25
22	I	**F**	**Y**	**F**	**I**	**F**	1	22
23	**F**	**F**	**Y**	**M**	L	**F**	4¶	18
24	I	**F**	**Y**	**M**	L	**F**	1	12
25	**F**	W	**Y**	**M**	L	**F**	4¶	22
26	**Y**	W	**Y**	**F**	**I**	**F**	11¶	23
27†‡	**V**	W	**Y**	**F**	**I**	**F**	1	39
28	**Y**	W	**Y**	**L**	**T**	**F**	1	22
29	**F**	W	**Y**	**I**	L	**F**	5	11
30	**Y**	W	**Y**	**M**	**I**	**F**	3¶	13
31†	**F**	W	**Y**	**F**	**I**	**F**	3¶	24
32	**F**	W	**Y**	**M**	**I**	**F**	9¶	16
33†	**C**	W	**Y**	**Y**	L	**F**	1	28
34	**F**	W	**Y**	**L**	**I**	**F**	2	17
35	**F**	W	**Y**	**F**	**V**	**F**	2	23
36	**F**	W	**Y**	**C**	L	**F**	1	20
37	**Y**	W	**Y**	**I**	L	**F**	1	13
38	**F**	W	**Y**	**Y**	**I**	**F**	1	5
39	**Y**	W	**Y**	**M**	**M**	**F**	1	12
40	**F**	W	**Y**	**I**	**M**	**F**	1	5
41	**Y**	W	**Y**	**M**	L	**F**	1	16

Library A, 25-42; Library B, 17-24

* Diversified residues are indicated in bold.

† Clones selected for detailed characterisation.

‡ Contains the spontaneous mutation R254G.

§ Frequency of combination of mutations at positions 135, 245, and 246.

¶ Combinations were found in both libraries (Library A and Library B).

‖ Specific activities of crude extracts with TCP (wt, 0.014 nmol s^{-1} mg^{-1} of total protein).

Supplementary Table 3 Oligonucleotide primers[*] used for mutagenesis

Name	Sequence (5`→ 3`)	Application
dhaASD*Bam*HI	CGGGGATCCTAAGGAGGAAATCGAAATGTCAGAAATCGGTACAGG	Amplification of *dhaA*
dhaAHis*Hin*dIII	GCCAAGCTTCTA*GTGATGGTGATGGTGATG*GAGTGCGGGGAGCCAGCG	Amplification of *dhaA*
dhaAC176Yfw	T**A**CGTCGTCCGT CCGCTTAC	Site-directed mutagenesis of *dhaA*
dhaAC176Yrev	TTTCGGGAGCGCACCCTC	Site-directed mutagenesis of *dhaA*
dhaAY273Ffw	T**T**CCTCCAGGAAGACAACCC	Site-directed mutagenesis of *dhaA*
dhaAY273Frev	GTGCAATCCCGGGGC	Site-directed mutagenesis of *dhaA*
dhaAW141Ffw	GGACGAAT**TT**CCGGAATTCG	Site-directed mutagenesis of *dhaA*
dhaAW141Frev	CACGTCGGGATAGGCC	Site-directed mutagenesis of *dhaA*
dhaA135fw[†]	GGCCT**NNK**CCGACGTG	Saturation mutagenesis of *dhaA*
dhaA135rev	GGATGAATTCCATACATGCAATAC	Saturation mutagenesis of *dhaA*
dhaA245-6fw[†]	**NNK**ATCCCCCCGGCCGAAG	Saturation mutagenesis of *dhaA*
dhaA245-6rev[†]	**M**N**N**GCCGGGTGTGCCCCAG	Saturation mutagenesis of *dhaA*

[*]Recognition sites for restriction enzymes used for cloning are underlined, the Shine-Dalgarno sequence is indicated by the gray box, the sequence for the additional six-histidyl tail is shown in italics and the introduced mutations in bold.
[†]N = A, G, C or T; K = G or T.

SUPPLEMENTARY METHODS

Molecular modelling and simulation. Random acceleration molecular dynamics (RAMD) simulation enables the discovery of product release pathways in protein structures and identification of residues lining protein tunnels [4,5]. RAMD is an enhanced sampling technique that makes the egress of a product from a buried enzyme active site observable in computationally accessible simulation times [6]. RAMD is like a standard molecular dynamics simulation, except that an additional force is applied to the centre of mass of the ligand in a randomly chosen direction. After a user-defined number of timesteps, the distance travelled by the ligand is compared to a threshold parameter. If the ligand does not reach the threshold distance, a new, randomly chosen, direction is given to the force acting on its centre of mass, otherwise the force direction is maintained. The process is iterated until the ligand has been released into the bulk solvent or a specified maximum simulation time (1 ns in this work) is reached. The access of TCP to the active site was not simulated since we are not aware of any reliable modelling technique that could be used for this purpose. We assume that the accessibility of pathways allowing the exchange of ligands between the buried active site and bulk solvent can be addressed by simulating product release. The initial orientation of the product, DCL, in the active site of wild-type DhaA was modelled using AUTODOCK 3.05 [7]. The starting structures for RAMD simulations were snapshots obtained after 2.0 to 2.8 ns equilibration molecular dynamics (EMD) following heating from 0 to 300 K during the first 200 ps. The simulations were performed with the SANDER module of AMBER 8 [8]. The different RAMD simulations were run by (i) starting from different snapshots from EMD, (ii) varying the RAMD parameters (0.04 or 0.05 kcal mol^{-1} $Å^{-2}$ for the force constant and 0.002 or 0.004 Å for the threshold distance criterion evaluated after 10 time steps of 2 fs), or (iii) a combination of the two. A modified version of AMBER 8 with the RAMD method implemented was used for the simulations of product release at 300 K. Models of *R*- and *S*-DCL (**42**, **43**, respectively) were built using PyMol 0.99 (DeLano Scientific, USA) and geometrically optimised by the AM1 method implemented in MOPAC 2000 using the following keywords: SCFCRT=1D-12 EF GNORM=0.0001 STEP=15 POINTS=12 LET PRECISE [9] and using GAUSSIAN 94 with the restricted Hartree-Fock method and a 6-31G* basis set [10]. Partial atomic charges were fitted to reproduce the electrostatic potential calculated with GAUSSIAN using the RESP module of AMBER 8. Both EMD and RAMD were carried out using the AMBER94 force field [11], with the protein and crystallographic

water molecules immersed in a rectangular box of TIP3P model water molecules under constant pressure (1 atm), with the use of periodic boundary conditions, PME and SHAKE. The structures of the mutant proteins were modelled using the mutagenesis module of the program PyMol 0.99 (DeLano Scientific, USA). Molecular dynamics simulations were also conducted with the substrate TCP to study the effect of water molecules in the active site on reactivity. The parameter settings of these calculations were the same as those used for DCL. The simulations consisted of 2 ns of EMD followed by 2 ns of production MD. The presence of NACs along the calculated production MD trajectories was evaluated every 0.5 ps using the two-parameter geometric condition proposed by Hur and Bruice for this reaction class: (i) the distance between the nucleophilic oxygen and the attacked carbon atom ≤ 3.41 Å, and (ii) the angle formed by the nucleophilic oxygen, the attacked carbon and the leaving chlorine > 157 degrees [12].

Materials and DNA manipulations. The enzymes used for DNA manipulations were obtained from Takara Bio (Kyoto, Japan), Toyobo (Osaka, Japan) and New England Biolabs (Beverly, USA). The cloning and mutagenesis primers were obtained from Hokkaido System Science (Hokkaido, Japan). The strains used in this study were *Escherichia coli* strains DH5α, BL21 and XJb (Zymo Research, Orange, USA). The plasmid pUC18 (Takara Bio, Kyoto, Japan) was used for basic cloning manipulations. The ampicillin-resistant pAQN vector [13] was used for overexpression of enzyme variants. Oligonucleotide primers were designed according to the nucleotide sequence of *dhaA* (GenBank accession no. AF060871) using the primers listed in **Supplementary Table 3** online. Established methods were employed for the following procedures: preparation of plasmid DNA, digestion of plasmids and PCR-amplified DNA fragments with restriction endonucleases, ligation, agarose gel electrophoresis, and transformation of *E. coli* cells [14]. The nucleotide sequences were determined by the dideoxy chain termination method with an automated ABI Prism 310 DNA sequencer (Applied Biosystems, Foster City, USA). The recognition sites for suitable restriction enzymes were added to the forward and reverse primers for cloning. A Shine-Dalgarno sequence (TAAGGAGG) was added to the forward primer for overexpressing the protein product in *E. coli*, and the primer dhaAHisHindIII was used for adding a six-histidyl tail to the C-terminal of the protein product of *dhaA*.

Saturation mutagenesis. Saturation mutagenesis at amino acid positions 135 and 245-6 was carried out using a set of degenerate synthetic oligonucleotides. Two oligonucleotide primer

pairs were designed to create two separate sub-libraries DhaA135 (diversity 32) and DhaA245-6 (diversity 1056). The basic procedure utilised a double-stranded DNA vector with a gene of interest (pAQN-M2, pAQN-M3) and synthetic primers containing the desired mutations. The coding strand primers contained NNK at the position to be mutagenized, where N was an equal mixture of all four nucleotides and K was an equal mixture of G and T. The plasmids pAQN-*dhaAHisM2* and pAQN-*dhaAHisM3* were used as templates in inverted PCR, in which initial denaturation of the reaction components at 94°C for 2 min was followed by 30 amplification cycles, consisting of 94°C for 15 s, 56°C/65°C for 30 s (DhaA135/DhaA245-6), and 68°C for 3 min with a final 10 min incubation at 68°C. The entire plasmid was amplified. The product was treated with *Dpn*I to digest the parental DNA template, leaving the mutation-containing *in vitro* synthesized DNA. *E. coli* DH5α cells were transformed with the nicked vector DNA incorporating the desired mutation. All transformants were used for the stock bulk sublibrary and five-fold inoculation of 3 ml Luria Broth (LB) media with appropriate antibiotic led to isolation of plasmids for each sublibrary. The plasmids were purified from colonies using a BioRad Miniprep kit. The isolated plasmids were mixed in equal amounts. Each sublibrary was digested with *Hin*dIII and *Nru*I restriction enzymes simultaneously at 37°C for 2 h. A small fragment of DNA (330 bp) from the sub-library DhaA245-6 was cloned into the digested sub-library DhaA135 (4600 bp), resulting in a final mutated DhaA library (DhaA135+245-6), which was electroporated into *E. coli* XJb cells (1 mm cuvette, 1.8 kV, 200 Ω, 25 μF). The recovered transformed cells were plated onto LB agar plates supplemented with ampicillin, and the cells with the best clone from previous rounds of screening, and positive and negative controls were transformed separately. Agar plates were incubated overnight at 37°C.

Circular Dichroism (CD) spectroscopy. CD spectra were recorded at room temperature (22°C) using a Jasco J-810 spectropolarimeter (Jasco, Tokyo, Japan). Data were collected from 185 to 260 nm, at 100 nm/min, 1 s response time and 2 nm bandwidth using a 0.1 cm quartz cuvette. The concentration of tested proteins was in the range of 0.26-0.30 mg/ml. Each spectrum shown is the average of ten individual scans and is corrected for absorbance caused by the buffer. Collected CD data are expressed in terms of the mean residue ellipticity (Θ_{MRE}).

Steady-state kinetics and pre-steady state burst analysis. Michaelis-Menten kinetic constants of DhaA wild-type and mutants with TCP were determined by previously described initial velocity measurements [15]. The substrate concentration was assayed by a Trace GC 2000

gas chromatography system equipped with a flame ionization detector (Thermo Finnigan, San Jose, USA), and a DB-FFAP capillary column (30 m x 0.25 mm x 0.25 mm; J&W Scientific, Folsom, USA). The method previously described by Iwasaki *et al.* [16] was used for determining the product concentration. Rapid quench-flow experiments were performed at 37°C in a glycine buffer (pH 8.6) using a QFM 400 rapid quench-flow instrument (BioLogic, Claix, France). The reaction was started by rapidly mixing 70 µl enzyme and 70 µl substrate solutions, then quenched with 100 µl 0.8 M H_2SO_4 after time intervals ranging from 2 ms to 8 s. The quenched mixture was directly injected into 0.5 ml of ice-cold diethyl ether with 1,2-dichloroethane as an internal standard. After extraction, the diethyl ether layer containing non-covalently bound substrate and alcohol product was collected, dried on a short column containing anhydrous Na_2SO_4 and analysed using a Trace 2000 gas chromatograph equipped with an MS detector (Thermo Finnigan, San Jose, USA) and a DB-FFAP capillary column (J&W Scientific, Folsom, USA). The amount of chloride in the water phase was measured using an 861 Advanced Compact ion chromatograph equipped with a METROSEP A Supp 5 column (Metrohm, Herisau, Switzerland). The steady-state kinetic constants K_m and k_{cat} were calculated using the computer program Origin 6.1 (OriginLab, Massachusetts, USA). The kinetic data were fitted using the same computer program Origin 6.1 to linear or burst models described by equation $c = A * (1-\exp(-k_B*t)) + k_L*t$, where A is the amplitude of burst, k_B is the burst (exponential) phase rate and k_L is the steady-state phase rate.

Supplementary References

1. Bosma, T., Damborsky, 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.* **68**, 3582-3587 (2002).
2. Gray, K.A. *et al.* Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv. Synth. Catal.* **343**, 607-616 (2001).
3. Banas, P., Otyepka, M., Jerabek, P., Petrek, M. & Damborsky, J. Mechanism of enhanced conversion of 1,2,3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling. *J. Comput.-Aid. Mol. Design* **20**, 375-383 (2006).
4. Winn, P.J., Ludemann, S.K., Gauges, R., Lounnas, V. & Wade, R.C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad. Sci. USA* **99**, 5361-5366 (2002).
5. Luedemann, S.K., Lounnas, V. & Wade, R.C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.* **303**, 797-811 (2000).
6. Schleinkofer, K., Sudarko, Winn, P.J., Ludemann, S.K. & Wade, R.C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep.* **6**, 584-589 (2005).
7. Morris, G.M. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639-1662 (1998).
8. Case, D.A. *et al.* (University of California, San Francisco, 1997).
9. Stewart, J.J.P. MOPAC - A semiempirical molecular-orbital program. *J. Comput.-Aid. Mol. Design* **4**, 1-45 (1990).
10. Frisch, M.J. *et al.* (Gaussian, Inc., Pittsburgh PA, 1995).
11. Cornell, W.D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197 (1995).
12. Hur, S., Kahn, K. & Bruice, 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. USA* **100**, 2215-2219 (2003).
13. Nagata, Y., Hynkova, K., Damborsky, J. & Takagi, M. Construction and characterization of histidine-tagged haloalkane dehalogenase (LinB) of a new substrate class from a γ-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26. *Prot. Express. Purif.* **17**, 299-304 (1999).
14. Sambrook, J. & Russell, D.W. *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press, New York, 2001).
15. Chaloupkova, R. *et al.* Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel. *J. Biol. Chem.* **278**, 52622-52628 (2003).
16. Iwasaki, I., Utsumi, S. & Ozawa, T. New colorimetric determination of chloride using mercuric thiocyanate and ferric ion. *Bull. Chem. Soc. Japan* **25**, 226 (1952).