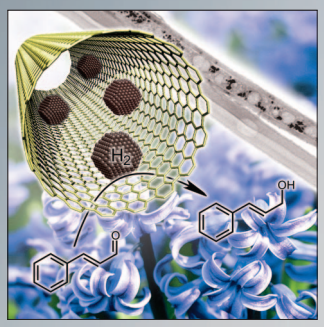
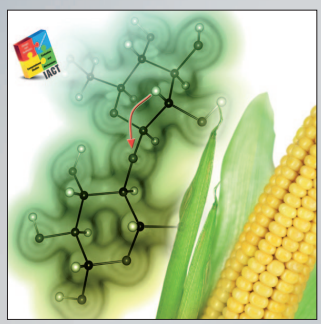
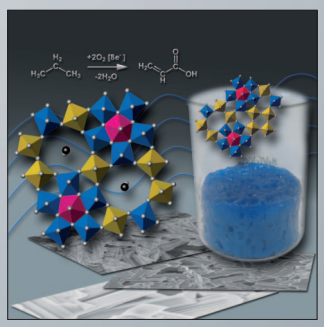
Heterogeneous & Homogeneous & Bio-

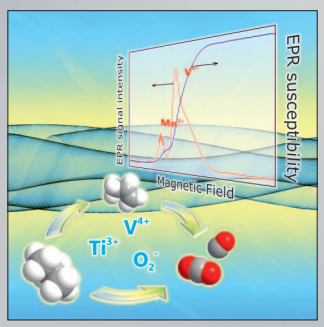
CHEMCATCHEM

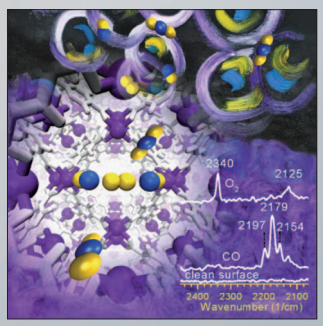
CATALYSIS

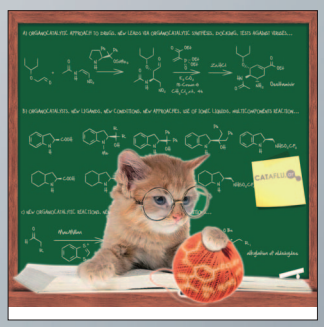


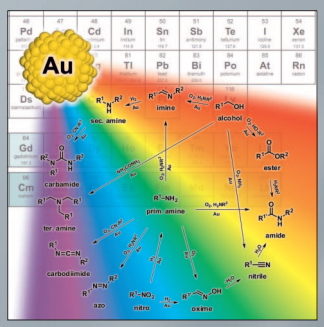


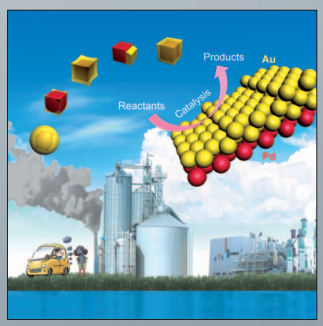


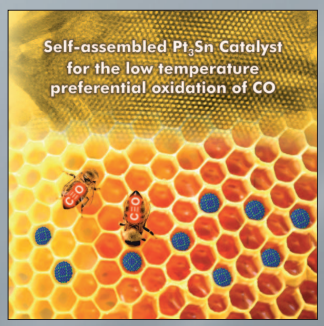


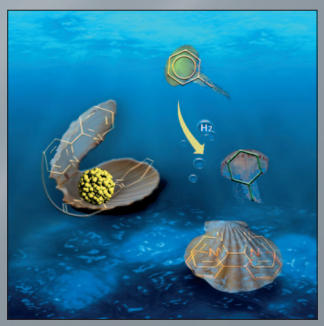












Reprint

ChemPubSoc Europe

WILEY-VCH

Supported by:
GECATS
GERMAN CATALYSIS SOCIETY

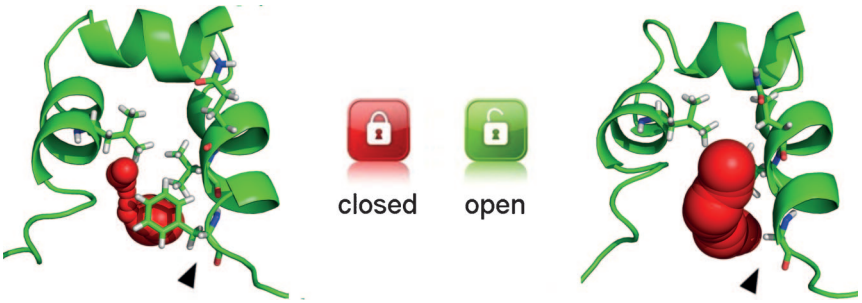
Heterogeneous & Homogeneous & Bio-
CHEMCATCHEM
CATALYSIS

Table of Contents

V. Liskova, D. Bednar, T. Prudnikova, P. Rezacova, T. Koudelakova, E. Sebestova, I. K. Smatanova, J. Brezovsky, R. Chaloupkova, J. Damborsky**

648 – 659

Balancing the Stability–Activity Trade-Off by Fine-Tuning Dehalogenase Access Tunnels



Tuning enzyme access tunnels: The catalytic activity and stability of enzymes can be inversely related. It is shown that haloalkane dehalogenases with high stability and activity can be created by fine-tuning the diameter and dynamics of the access tunnels that connect buried active sites to the surrounding bulk solvent.

WILEY-VCH

ChemPubSoc Europe

DOI: 10.1002/cctc.201402792

CHEMCATCHEM Full Papers

Balancing the Stability–Activity Trade-Off by Fine-Tuning Dehalogenase Access Tunnels

Veronika Liskova,[a, b] David Bednar,[a, b] Tatyana Prudnikova,[c, d] Pavlina Rezacova,[e, f] Tana Koudelakova,[a] Eva Sebestova,[a] Ivana Kuta Smatanova,[c, d] Jan Brezovsky,[a] Radka Chaloupkova,*[a] and Jiri Damborsky*[a, b]

A variant of the haloalkane dehalogenase DhaA with greatly enhanced stability and tolerance of organic solvents but reduced activity was created by mutating four residues in the access tunnel. To create a stabilised enzyme with superior catalytic activity, two of the four originally modified residues were randomised. The resulting mutant F176G exhibited 32- and 10-times enhanced activity towards 1,2-dibromoethane in buffer and 40% DMSO, respectively, upon retaining high stability. Structural and molecular dynamics analyses demonstrated that the new variant exhibited superior activity because the F176G mutation increased the radius of the tunnel's mouth and the mobility of α-helices lining the tunnel. The new variant's tunnel was open in 48% of trajectories, compared to 58% for the wild-type, but only 0.02% for the original four-point variant. Delicate balance between activity and stability of enzymes can be manipulated by fine-tuning the diameter and dynamics of their access tunnels.

Introduction

Enzymes are natural catalysts with great potential for industrial applications. Most natural enzymes are not tolerant of harsh conditions such as those associated with extremes of pH, elevated temperatures, high salinity, or the presence of organic solvents. The development of thermodynamically and kinetically stable enzymes that retain high activity under harsh operating conditions has thus been a major but challenging goal in protein engineering over the last few decades. The structural stability of enzymes is usually maintained through non-covalent interactions including hydrogen bonds, salt bridges, hydrophobic interactions and van der Waals forces, all of which help to enhance the robustness of these biocatalysts. As such, the most common strategy for increasing enzyme stability is to introduce new interactions that increase the activation energy of enzyme denaturation.[1] However, many protein mutagenesis studies have shown that stability and function are often tightly related; mutations that increase stability often reduce function and vice-versa.[2–6] This negative correlation between enzyme stability and functionality is known as the stability–activity trade-off.

Several different protein engineering techniques can be used to manipulate the properties of enzymes, depending on the property of interest and the available information on the structure–function relationship of the enzyme. In general, there are three main strategies for developing modified enzymes: rational design, directed evolution and semi-rational design.[7] Rational design uses various computational tools to identify key residues and predicts the effects of mutations based on knowledge of tertiary structure and structure–function relationships. This approach usually generates limited numbers of mutants, necessitating only a modest amount of laboratory work on screening and selection.[8–10] In contrast, directed evolution uses random mutagenesis to generate mutant libraries with mutations across the entire gene sequence. These libraries are then screened or selected to identify improved variants. This strategy does not require structural information about the target protein, but an efficient screening

[a] V. Liskova, D. Bednar, Dr. T. Koudelakova, Dr. E. Sebestova, Dr. J. Brezovsky, Dr. R. Chaloupkova, Prof. J. Damborsky
Loschmidt 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)
E-mail: radka@chemi.muni.cz
jiri@chemi.muni.cz

[b] V. Liskova, D. Bednar, Prof. J. Damborsky
International Clinical Research Center
St. Anne's University Hospital
Pekarska 53, 656 91 Brno (Czech Republic)

[c] Dr. T. Prudnikova, Assoc. Prof. I. K. Smatanova
Faculty of Science
University of South Bohemia in Ceske Budejovice
Branisovska 31, 370 05 Ceske Budejovice (Czech Republic)

[d] Dr. T. Prudnikova, Assoc. Prof. I. K. Smatanova
Institute of Nanobiology and Structural Biology
Academy of Sciences of the Czech Republic
Zamek 136, 373 33 Nove Hrady (Czech Republic)

[e] Dr. P. Rezacova
Institute of Molecular Genetics
Academy of Sciences of the Czech Republic
Videnska 1083, 142 20 Prague 4 (Czech Republic)

[f] Dr. P. Rezacova
Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic
Flemingovo nam. 2, 166 10 Prague 6 (Czech Republic)

Supporting Information for this article is available on the WWW under http://dx.doi.org/10.1002/cctc.201402792.

or selection assay is essential for the identification of interesting hits.[11,12] Semi-rational design combines the benefits of the directed evolution and rational design approaches. Specific residues or segments of the enzyme's structure are identified by the rational approach and then subjected to mutagenesis to create focused libraries that do not require laborious screening. At the same time, these small focused libraries have much higher hit rates than those generated in directed evolution, while including mutations that would be unlikely to be identified or explored on the basis of computational results.[13–16]

Haloalkane dehalogenases (HLDs; EC 3.8.1.5) are enzymes that catalyse the hydrolytic cleavage of carbon–halogen bonds in halogenated hydrocarbons to yield the corresponding alcohol, a proton and a halide.[17] HLDs have potential applications in bioremediation,[18,19] decontamination,[20] industrial biocatalysis,[21,22] biosensing[23–25] and cell imaging.[26–30] However, their utility in these applications is limited by their low stability and activity under the harsh conditions that are often required.[31–33]

A stabilised variant of the HLD DhaA from *Rhodococcus rhodochrous* NCIMB 13064[34] has been created by Gene Site Saturation Mutagenesis (GSSM).[35] Compared to wild-type DhaA, the resulting highly stable ten-point DhaA mutant (named DhaA63 in Ref. [33]) was 30000 times more capable of refolding after denaturation at 55 °C. However, its catalytic efficiency in aqueous buffer was six times lower than that of the wild-type enzyme. Several other DhaA variants[33] with substantially improved stability and an ability to tolerate the presence of the organic co-solvent dimethylsulfoxide (DMSO) were obtained by a combination of random mutagenesis and focused directed evolution.[33] Detailed biochemical and structural analysis of these variants revealed that their stabilisation was largely owing to mutations in the residues that form the access tunnel, which connects the buried active site of enzymes to the surrounding solvent. Although the tunnel mutants exhibited increased stability and retained catalytic activity in 40 vol.% DMSO, their activity in buffer solutions was low. The DhaA80 variant, which carries four of the ten substitutions found in DhaA63 (T148L, G171Q, A172V and C176F), exhibited 4000-fold higher kinetic stability than the wild-type enzyme in 40 vol.% DMSO and remained stable at temperatures up to 16.4 °C higher than those inactivating the wild-type. However, the catalytic activity of engineered DhaA80 towards 1,2-dibromoethane in a buffer solution was reduced by two orders of magnitude. The stabilisation of DhaA80 was a result of the introduction of four residues at the opening of the access tunnel, three of which were bulky and hydrophobic. These mutations led to enhanced intramolecular hydrophobic packing at the tunnel opening and possibly prevented the destabilisation of the protein structure by admission of DMSO molecules to the active site.[33]

This work aimed to address these issues by enhancing the catalytic activity of the highly stable DhaA80 variant in buffer solutions. Mutagenesis targeting two of the four tunnel-mouth residues that were replaced to create DhaA80 (V172 and F176) led to the identification of the DhaA106 variant, which differs from DhaA80 in only a single residue (F176G) but exhibits 32 times higher catalytic activity in a buffer, sacrificing only 4 °C of thermal stability. Moreover, DhaA106 exhibited enhanced activity towards 26 out of 30 tested halogenated compounds and thus replicates the substrate specificity of the wild-type enzyme. Crystallographic analysis followed by molecular modelling revealed that enhanced catalytic activity of DhaA106 was caused by the increase in the diameter of the access tunnel and the mobility of the adjacent secondary structure elements.

Results

Rational design of two focused libraries

Compared to the wild-type enzyme, the DhaA80 variant carries four substitutions (T148L, G171Q, A172V and C176F) in the tunnel-lining residues. These were introduced during a combined mutation and screening campaign that tested the activity of the mutated enzymes towards 1,2-dibromoethane in 40 vol.% DMSO. One of these tunnel residues (Q171) exhibited the particularly high variability in the study of Gray et al.[35] and was randomised in our previous study with the goal of further enhancing the enzyme's stability.[33] Several of the resulting variants exhibited improved thermostability at the expense of catalytic activity. Previous saturation mutagenesis of T148 with the degenerate codon NNK provided only the substitution by leucine in two independent libraries. We therefore performed mutagenesis experiments targeting the other two tunnel residues (V172 and F176), using DhaA80 as a template. To minimise losses of enzyme robustness, the stabilisation potentials of all possible combinations of substitutions, including self-mutations in the two target positions, were investigated by using FoldX.[36] Neutral or stabilising effects were predicted for 15 of the 800 possible variants (Table S1 in the Supporting Information). Val, Ile and Leu were stabilising or neutral residues in position 172 whereas Leu, Met, Trp and Phe were identified as the stabilising or neutral residues in position 176. Degenerate codons encoding all of the strongly stabilising residues identified in the two positions were designed using the CASTER v2.0.[37] The resulting degenerate codons VTY and WKS were used to construct a smart library in which positions 172 and 176 were simultaneously saturated (library I). In addition, a second library (library II) was constructed in which the codon for position 176, that had not been previously mutated by GSSM,[35] was replaced with the degenerate codon NNK, which encodes all of the standard amino acid residues.

Screening of libraries and biochemical characterisation of variant DhaA106

Both libraries were constructed by using QuickChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, USA). Colonies of the libraries were screened by using a pH indicator-based colorimetric assay optimised for the presence of the organic co-solvent DMSO.[38] In total, the activities of 142 and 94 colonies from libraries I and II, respectively, were tested against 1,2-dibromoethane in 52 vol.% DMSO. The number of

colonies tested in each case represented at least 95% of the total variation in each library. Two positive hits were identified in each library. The hit with the greatest improvement in activity relative to that of the template in the presence of DMSO and in buffer solution was the single-point mutant F176G, which was designated DhaA106 (Table S2). This variant was expressed in a larger volume of *Escherichia coli* BL21 cells, purified to homogeneity and subjected to detailed biochemical and structural analysis. Its properties were compared to those of the template (DhaA80), the wild-type enzyme (DhaA) and the previously described highly thermostable and solvent-tolerant DhaA63 variant.

Circular dichroism spectroscopy in the far-UV spectral region was used to assess the impact of the single-point F176G mutation on the folding and secondary structure of DhaA106. The circular dichroism spectrum of DhaA106 was identical to those of the other DhaA variants (Figure S1 in the Supporting Information): all of the enzymes' spectra exhibited a single positive peak at 195 nm and two negative maxima at 208 and 222 nm, features that are characteristic of α-helical content.[39] Thermal denaturation experiments were performed to test the effect of the F176G mutation on the thermal stability of DhaA106. The melting temperature of DhaA106 was 62.7 °C, which is approximately 4 °C and 6 °C lower than those of DhaA80 and DhaA63, respectively. However, it is 12 °C higher than the value for DhaA (Table 1). This result was not unexpected because the F176G mutation replaces a hydrophobic and sterically demanding phenylalanine residue in the enzyme's tunnel mouth with small glycine residue, eliminating stabilising hydrophobic and van der Waals interactions. The phenylalanine residue was previously estimated to increase the melting temperature by 5 °C relative to that of the wild-type enzyme.[33]

Table 1. Melting temperatures of DhaA variants.

Variants	T_m [°C]
DhaA	50.4 ± 0.3[a]
DhaA63	68.3 ± 0.3[a]
DhaA80	66.8 ± 0.2[a]
DhaA106	62.7 ± 0.1

[a] Data from Ref. [33].

The specific activities of the purified DhaA variants were tested against 1,2-dibromoethane in buffer solution and buffer solutions containing either 40 vol.% or 52 vol.% DMSO (Figure 1, Table S3). In the absence of DMSO, the specific activities of DhaA106 were 32- and 9-fold higher than those of DhaA80 and DhaA63, respectively. Whereas the activity of DhaA106 in a pure buffer was significantly greater than that of the template DhaA80, it was 70% lower than that of wild-type DhaA (Figure 1a). Interestingly, activity tests in the presence of DMSO revealed that DhaA106 was the most active tested variant (Figure 1b,c). Its activity in 40 vol.% DMSO was 10 times greater than that of the template and twice as high as that of the most temperature- and solvent-resistant variant DhaA63.

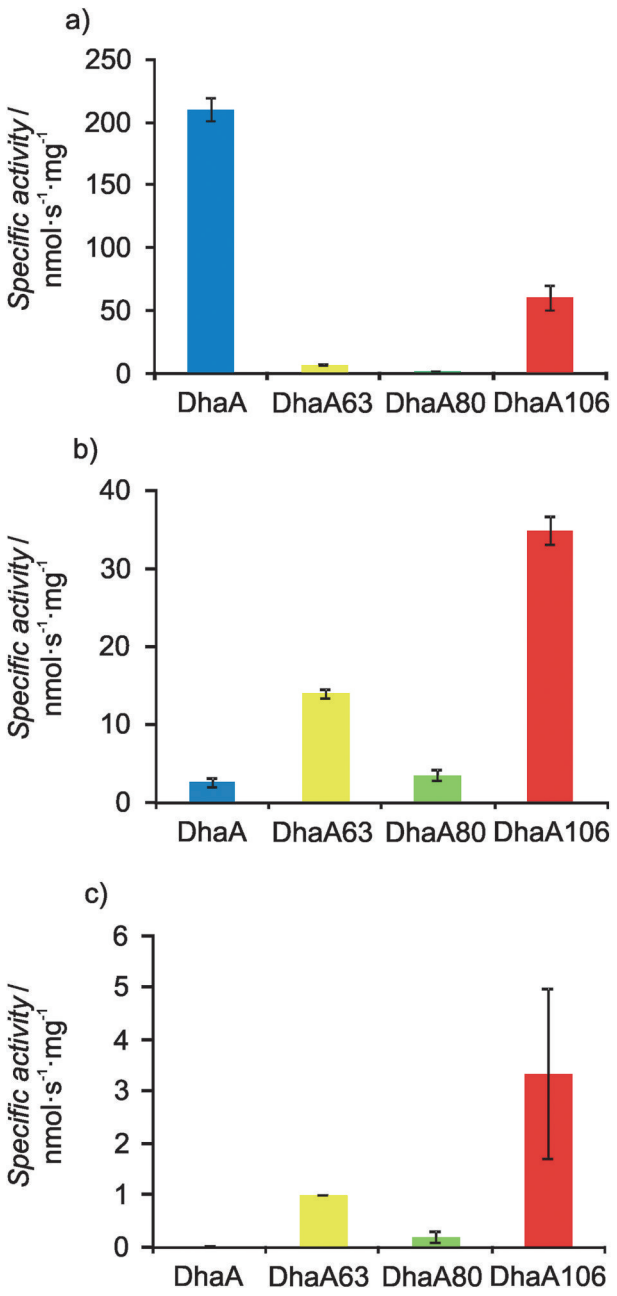


Figure 1. Specific activities of DhaA variants towards 1,2-dibromoethane at 37 °C a) in aqueous buffer, b) in 40 vol.% DMSO and c) in 52 vol.% DMSO.

Raising the DMSO concentration in the reaction buffer to 52 vol.% significantly reduced the activity of all DhaA variants but did not change their order of activity in terms of co-solvent tolerance (DhaA106 > DhaA63 > DhaA80 > DhaA). These results imply that the F176G mutation in the mouth of the enzyme's access tunnel significantly enhanced its activity in aqueous environments and in the presence of organic solvents.

To better understand the origin of this change in activity, steady-state kinetic constants were determined for conversion of 1,2-dibromoethane by DhaA106 and compared to those for DhaA80, DhaA63 and DhaA (Figure 2, Tables 2 and 3). In an aqueous environment, the F176G mutation significantly in-

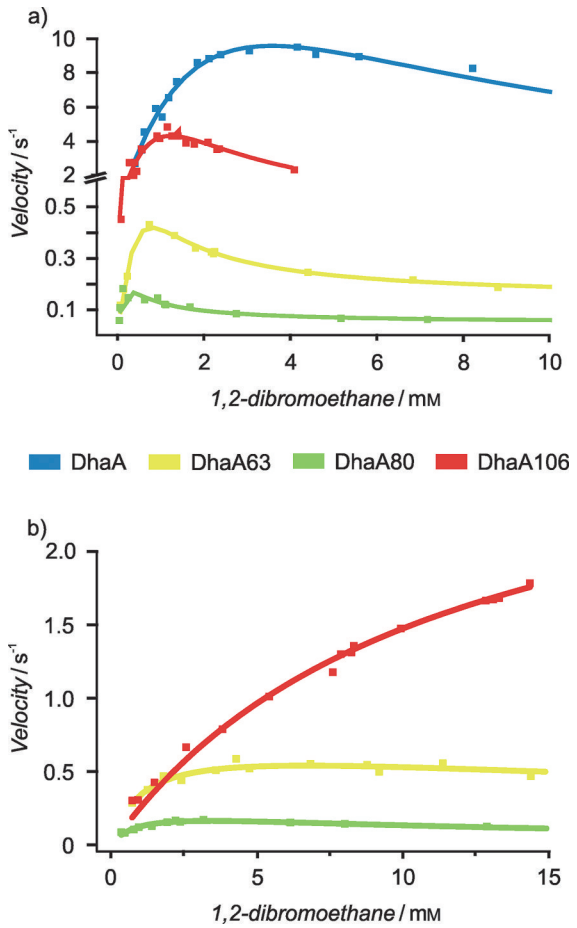


Figure 2. Steady-state kinetic profiles of DhaA variants at 37 °C a) in a buffer and b) in 40 vol.% DMSO [Note the different scales in (a) and (b)]. The kinetic profile of DhaA in 40 vol.% DMSO was measured under different conditions from those for the other variants (the duration of the experiment was limited to 3 min) because of its very low stability in the experimental environment.

Table 2. Steady-state kinetic parameters of DhaA variants with 1,2-dibromoethane in buffer solution.

Variant	K_m [mM]	K_{si} [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mM^{-1}]
DhaA	3.56 ± 0.67[a]	3.56 ± 0.70[a]	28.67 ± 3.95[a]	8.05 ± 2.63[a]
DhaA63	1.70 ± 1.28[a]	0.24 ± 0.20[a]	2.23 ± 1.46[a]	1.31 ± 1.85[a]
DhaA80	0.13 ± 0.09[a]	0.41 ± 0.33[a]	0.34 ± 0.14[a]	2.62 ± 2.89[a]
DhaA106	0.89 ± 0.08[b]	1.28 ± 0.11	11.01 ± 0.23	12.37 ± 1.37

[a] Data from Ref. [33]. [b] Cooperativity with Hill coefficient $n = 1.36 \pm 0.13$.

Table 3. Steady-state kinetic parameters of DhaA variants with 1,2-dibromoethane in 40 vol.% DMSO.

Variant	K_m [mM]	K_{si} [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mM^{-1}]
DhaA	ND[a]	ND[a]	ND[a]	ND[a]
DhaA63	1.08 ± 0.26[b]	41.44 ± 14.14[b]	0.72 ± 0.06[b]	0.66 ± 0.22[b]
DhaA80	0.88 ± 0.16[b]	13.14 ± 2.73[b]	0.25 ± 0.02[b]	0.28 ± 0.07[b]
DhaA106	11.17 ± 1.38	NA[c]	3.14 ± 0.21	0.28 ± 0.05

[a] ND = data could not be collected under the comparable conditions because of the protein instability. [b] Data from Ref. [33]. [c] NA = not applicable.

creased the enzyme's catalytic rate (relative to DhaA80), suppressed substrate inhibition and reduced the free enzyme's affinity for 1,2-dibromoethane. The turnover number of DhaA106 in buffer was 32- and 5-fold higher than those for DhaA80 and DhaA63, respectively, and 3-fold lower than that of DhaA (Figure 2a, Table 2). Adding 40 vol.% DMSO to the reaction mixture reduced the k_{cat} values for all of the studied DhaA variants while increasing their K_m and K_{si} constants, implying that the organic co-solvent acts as a mixed inhibitor (Figure 2b, Table 3). Notably, DMSO very strongly reduced the affinity of free DhaA106 and the enzyme–substrate complex for 1,2-dibromoethane. DhaA106 consequently exhibited no substrate inhibition and had the highest catalytic rate of the tested variants in the presence of DMSO (Table 3).

The specific activity of DhaA106 was tested further using a set of 30 halogenated substrates (Figure 3, Table S4) to determine whether the F176G mutation affected its activity in aqueous buffer towards substrates other than 1,2-dibromoethane. The activity of DhaA106 was greater (between 2 and 49 times higher) than that of DhaA80 for all tested compounds. The enzyme was most active towards multi-substituted C2–C3 bromoalkanes including 1,2-dibromoethane, 1,2-di-

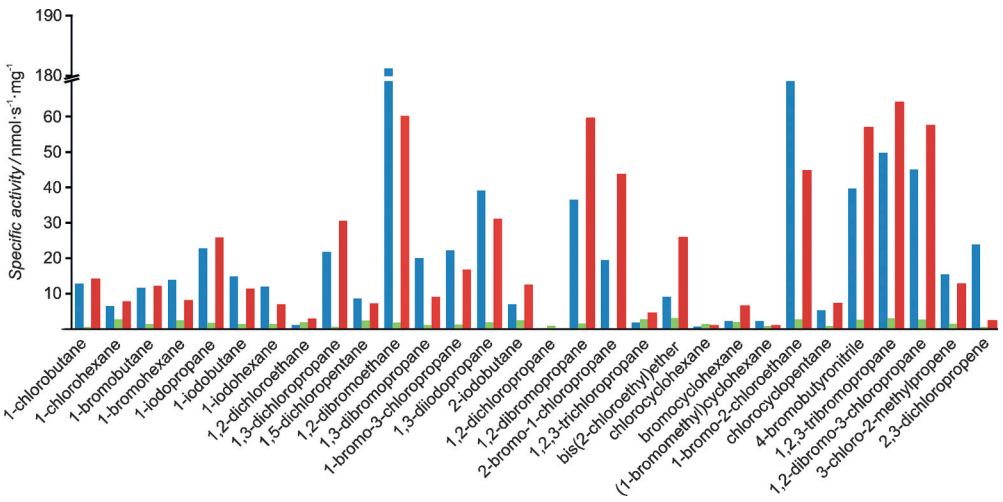


Figure 3. Substrate–specificity profiles of DhaA (blue), DhaA80 (green) and DhaA106 (red) based on their activity in aqueous buffer towards 30 halogenated hydrocarbons. The graph presents the specific activity of each enzyme towards the individual substrates at 37 °C.

bromopropane, 1,2,3-tribromopropane and 1,2-dibromo-3-chloropropane.

In addition, the activity of DhaA106 was comparable to or greater than that of DhaA for more than half of the tested substrates. Principal component analysis (PCA) using transformed activity data set was used to explore the relationships between the individual DhaA variants (Figure S2). A similar analysis demonstrated that wild-type HLD enzymes cluster into four distinct substrate specificity groups (SSGs).[40] Like DhaA and DhaA80, DhaA106 was found to belong to SSG-I (Figure S2a). Enzymes in SSG-I are robust catalysts with high activity towards brominated ethanes and propanes, and detectable activity towards poorly degradable compounds such as 1,2-dichloroethane, 1,2-dichloropropane and 1,2,3-trichloropropane.[40]

Although all of the tested DhaA variants belong to the same SSG, their substrate preferences differed to some extent as demonstrated by their different positions on the PCA score plot (Figure S2a). The relative activity of DhaA106 was more than one order of magnitude greater than that of DhaA80 for several substrates including 1-chlorobutane, 1,3-dichloropropane, 1,2-dibromoethane, 1,2-dibromopropane, 4-bromobutyronitrile, 1,2,3-tribromopropane and 1,2-dibromo-3-chloropropane. Unlike DhaA, DhaA106 exhibited decreased preference for substrates with longer alkyl chains such as 1-bromohexane, 1-iodohexane and (1-bromomethyl)-cyclohexane, as well as disubstituted C2–C3 haloalkanes such as 1,2-dibromoethane, 1,3-dibromopropane and 2,3-dichloropropene.

Crystallographic analysis of DhaA106

The structure of DhaA106 was solved at the resolution of 1.69 Å (Table S5) by molecular replacement using the structure of DhaA14 (PDB ID 3G9X)[41] as a search model. The resulting diffraction data enabled the localisation of residues 4–295, showing that the enzyme exists as a monomer in the crystal with a solvent content of approximately 41.96%. As expected, the overall structure of DhaA106 resembles that of DhaA,[41,42] consisting of an α/β-hydrolase core domain and a helical cap domain (Figure S3). The core domain is formed by a central twisted eight-stranded β-sheet (mostly parallel, with a single antiparallel β2-strand) surrounded by six α-helices. The cap domain consists of five α-helices linked by six loop insertions. The active site is located in a predominantly hydrophobic cavity, at the interface between the core and the cap domains, connected to the protein surface by two access tunnels. Visual inspection of the crystal structure revealed that the F176G mutation changed the diameter of the main access tunnel and the intramolecular contacts between its hydrophobic residues.

Molecular dynamics and analysis of access tunnels

Molecular dynamics (MD) simulations were performed to further explore the structural basis of the enhanced catalytic activity and reduced thermostability of DhaA106. Two independent 200 ns long simulations were run for each of DhaA, DhaA63, DhaA80 and DhaA106. CAVER 3.01[43] was then used to analyse 100 000 snapshots from each simulation to identify the access tunnels and to provide information on the opening and closing of the access tunnel as well as time-resolved changes in bottleneck radii. Aside from the main access tunnels, slot tunnels were identified in the structures of all studied enzymes. The slot tunnels showed less favourable geometric parameters than the main access tunnels, as deduced from the tunnel width, length and curvature (Table S6), indicating that the main tunnel acts as the preferred pathway for transport of studied substrates and products.

The main access tunnel of DhaA was detected in 94% of the snapshots taken during the simulation and was open in 58% of the snapshots. The average tunnel bottleneck radius was 1.5 Å, with a maximum value of 3.1 Å (Figure 4, Table S7). The secondary structures of the DhaA cap domain exhibited substantial flexibility: the distance between the two helices situated on the opposite sides of the tunnel (quantified in terms of the distance between the Cα atoms of residues F144 and C176) ranged from 7.0 to 14.0 Å (Figure S4).

The access tunnels of DhaA63 and DhaA80 exhibited very similar properties to each other. The main access tunnel was detectable in 6% and 1% of the snapshots for DhaA63 and DhaA80, respectively. Similarly, the tunnel was only open in 0.05% of the DhaA63 snapshots and 0.02% of those for DhaA80. These mutants had identical average bottleneck radii of 1.1 Å, with a maximal radius of 1.7 Å (Figure 4, Table S7). The separation of the helices in the cap domain was somewhat more constrained than in the wild-type protein, ranging from 8.0 to 13.0 Å in both cases (Figure S4). Both of these variants have four bulky residues in the main access tunnel that are not present in the wild-type enzyme and which serve to restrict the tunnel's opening upon making the cap domain more rigid. The C176F mutation in the tunnel entrance was particularly important in narrowing the main pathway to the active site in these two enzymes.

The modified access tunnel of DhaA106 is strikingly similar to that of wild-type DhaA. The F176G mutation in DhaA106 created a void in the tunnel entrance that is not present in DhaA63 and DhaA80, and enhanced the mobility of the cap domain's secondary elements. The tunnel was detected in 86% of the snapshots and was open in 48% of them. The average tunnel bottleneck radius was 1.4 Å, with a maximum value of 2.8 Å (Figure 4, Table S7). The distance between helices ranged from 7.5 to 14.0 Å (Figure S4). The F176G mutation removed some of the contacts between residues within the access tunnel of DhaA106, explaining its lower thermodynamic stability than that of the template DhaA80. However, because DhaA106 retains three stabilising mutations (T148L, G171Q and A172V), it is much more stable than the wild-type enzyme.

Discussion

The development of new approaches for the rational engineering of stable catalysts that retain catalytic activity is a key challenge in protein engineering.[44] This work aimed to improve the catalytic activity of a recently constructed highly stable and solvent-resistant DhaA80[33] while minimising losses of

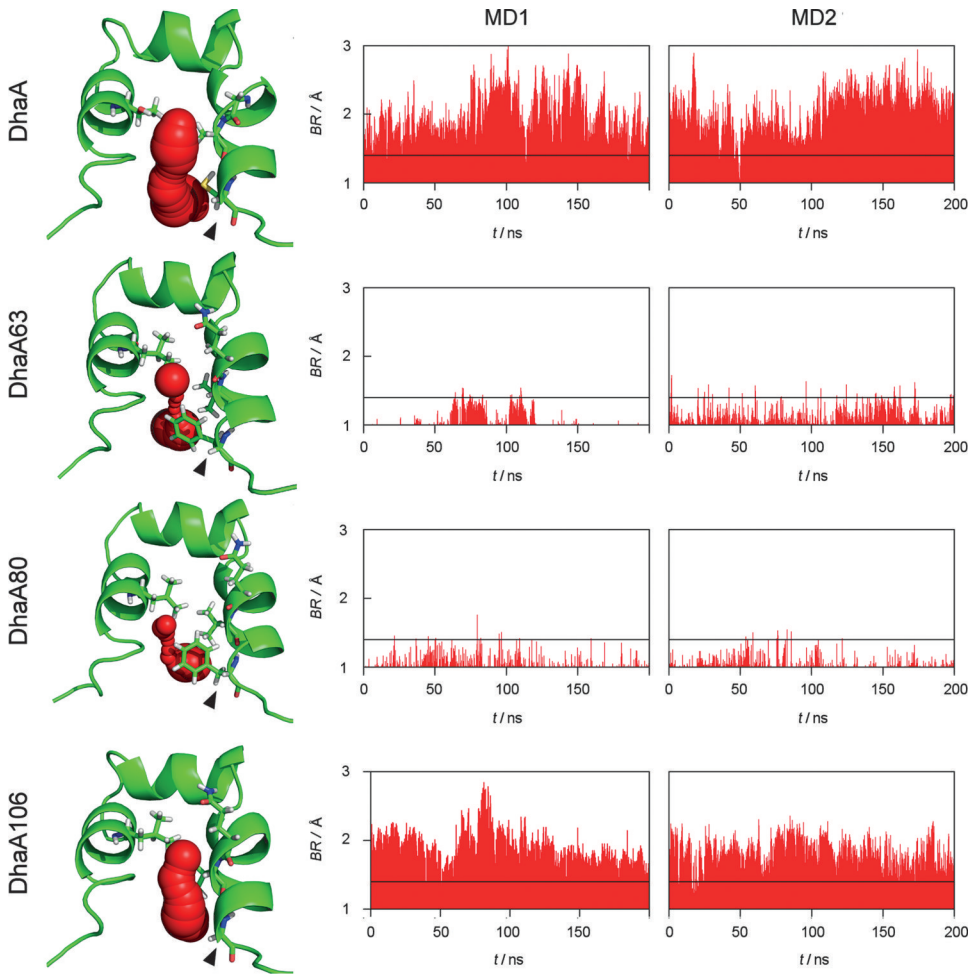


Figure 4. Visualisation of the representative structures of the main tunnels in the cap domains of the studied enzymes and changes in the tunnel bottlenecks over time. Left: PyMOL 1.5 visualisations of the cap domain residues (green cartoon) and the tunnel (indicated by the red spheres). The side chains of the residues located at the 148, 171, 172 and 176 positions (i.e., the positions mutated in DhaA63) in each enzyme are represented by sticks and the location of residue 176 in DhaA106 is indicated by a black arrowhead. Right: The evolution of the bottleneck radius (BR) in two independent 200 ns long MD simulations. The black horizontal lines indicate the threshold radius (1.4 Å) above which the tunnel was considered to be open. The tunnels were analysed by using CAVER 3.01.[43]

thermodynamic stability. To this end, the effects of all possible substitutions in the targeted tunnel positions (F176 and V172) were evaluated by using the computational tool FoldX.[36] A smart saturation mutagenesis library was then constructed featuring enzyme variants incorporating every possible combination of the predicted stabilising or neutral access tunnel substitutions (library I). In addition, a second library was constructed in which one of the target positions in the tunnel mouth (F176) was randomized by using site-saturation mutagenesis (library II). The best variant was DhaA106, which was obtained by site-saturation mutagenesis and exhibited significantly enhanced activity in the presence and absence of DMSO. The catalytic activity of DhaA106 towards 1,2-dibromoethane in buffer solution and 40 vol.% DMSO was 32- and 10 times higher than that of DhaA80, and its melting temperature (which reflects its thermodynamic stability) was only 4 °C lower. DhaA106 also exhibited significantly enhanced activity (relative to DhaA80) towards 26 of 29 additional halogenated compounds, showing similar levels of activity to wild-type DhaA.

Sequencing of DhaA106 revealed that it contained a substitution that would be difficult to design rationally. The variant carries a small glycine residue in the tunnel mouth in place of a bulky phenylalanine. Glycine contains a hydrogen atom as its side chain, giving it much more conformational freedom than other amino acids. Its low steric demand means that adjacent residues have much more flexibility than they would otherwise.[45,46] As high flexibility is often associated with low stability in proteins, one common strategy for enhancing their stability is to rigidify their most flexible regions.[1,47–50] However, it is necessary to maintain a balance between stability and flexibility to retain the protein's biological functionality. Stability ensures an appropriate geometry for ligand binding and prevents denaturation under physiological conditions, and flexibility is necessary to allow catalysis at a metabolically appropriate rate.[51–53]

In this work, replacing a bulky phenylalanine in the tunnel mouth of DhaA80 with the smallest amino acid glycine led to a variant (DhaA106) in which the intramolecular hydrophobic packing of the tunnel residues was partially disrupted, reducing the protein's stability ($\Delta T_m = -4$ °C). However, this also increased the flexibility and mobility of the two α-helices lining the main tunnel, increasing the chance of the tunnel being in an open state. Consequently, the variant was more catalytically active than the template. The main access tunnel of DhaA106 was identified and found to be open in 86 and 48%, respectively, of the molecular dynamics snapshots that were analysed; the corresponding values for the template enzyme DhaA80 were only 1 and 0.02%. The frequency of tunnel opening in DhaA106 was comparable to that of DhaA, the main tunnel or which was identifiable in 94% of its snapshots and open in 58%. We hypothesise that reopening of the access tunnel facilitates the admission of the substrate to the active site or the release of the product, whereas the remaining three bulky and hydrophobic muta-

tions in the access tunnel[33] significantly increase the enzyme's thermodynamic stability relative to that of the wild-type ($\Delta T_m = 12\,^{\circ}C$).

It has previously been shown that modifying the size, physico-chemical properties and dynamics of access tunnels by protein engineering can change the catalytic activity, substrate specificity, enantioselectivity and stability of HLDs[8,33,54–56] and many other enzymes with buried active sites[57] such as cytochrome P450s,[58–63] β-glucosidases,[64] lipases,[15,65–69] esterases[70] and epoxide hydrolases.[47,71,72]

Tunnel-mouth engineering was shown to have profound effects on the activity and specificity of the enzyme LinB from *Sphingobium japonicum* UT26.[8] The residue L177, located in the tunnel opening at the position corresponding to F176 in DhaA80, was selected for saturation mutagenesis on the basis of structural and phylogenetic analyses. The effects of the resulting mutations on the variants' catalytic activities greatly differed for individual substrates.[8] Similar findings have been reported for epoxide hydrolases from *Agrobacterium radiobacter* AD1[72] and *Aspergillus niger* M200,[71] in which the engineering of a single amino acid in the tunnel mouth led to improved enzyme activity and enantioselectivity. As with DhaA106, the catalytic activity of LinB variants was generally increased by introducing a small non-polar amino acid at position 177, whereas the introduction of bulky aromatic or charged residues dramatically reduced activity towards most substrates, including 1,2-dibromoethane. The small side chain of the introduced glycine residue in the tunnel mouth of the LinB L177G mutant was proposed to increase the radius of the tunnel mouth, thereby facilitating substrate entry and product release. Conversely, the bulkier side chain of the introduced tryptophan residue in the LinB L177W variant presumably blocked the mouth of the enzyme's main access tunnel, reducing its catalytic activity.[8] A detailed analysis of 1,2-dibromoethane passage through the access tunnel of the LinB L177W mutant confirmed that this mutation significantly reduced the rate of product release.[73] Moreover, mutation in position 177 of LinB significantly affected its thermal stability.[56] Similar observations were also reported for β-glucosidase from *Trichoderma reesei* the activity and stability of which were significantly affected by mutations in the substrate entrance region.[64]

The influence of the tunnel-lining residues on the catalytic activity of DhaA towards 1,2,3-trichloropropane (TCP) has been studied extensively.[35,54,55,74,75] Independently performed error-prone polymerase chain reaction (PCR) experiments generated two double point mutants, G3D+C176F[35] and C176Y+Y273F[74] the activities of which towards TCP are 4- and 3.5 times greater than that of the wild-type, respectively. Interestingly, both mutants carried bulky residues (tyrosine or phenylalanine) in the 176 position whereas the wild-type enzyme has a comparatively small cysteine residue in this position. The mutants thus have much narrower access tunnel entrances.[54] MD simulations and structure-based enzyme design identified the 176 position and another four access tunnel residues as crucial for the activity of DhaA. Mutagenesis at these positions yielded a variant the catalytic activity and efficiency of which towards TCP were 32- and 26 times higher, respectively, than those of the wild-type. This variant had bulky aromatic residues at four of the five targeted positions, which restricted the access of water molecules to the active site cavity. The rate-limiting step of TCP conversion in the resulting variant was shifted from carbon–halogen bond cleavage to the release of the reaction products.[55]

Sealing the access tunnel of DhaA with bulky residues has previously been identified as a viable strategy for enhancing its thermodynamic stability and resistance to the organic co-solvent DMSO.[33] The introduction of four bulky hydrophobic residues into the access tunnel yielded a DhaA variant (DhaA80) with a closed tunnel exhibiting enhanced intramolecular packing. This modification prevents destabilisation of the protein's structure owing to the admission of DMSO into the active site. Rigidifying and narrowing the tunnel in this way shields the interior of the protein from the organic solvent but presumably also makes the exchange of substrate and product molecules between the active site and bulk solvent more difficult. Stabilising the protein in this way therefore reduces its activity, demonstrating the need to strike a careful balance between protecting the buried active site from solvent molecules (which may cause denaturation or compete with the desired substrate) and retaining sufficient flexibility for catalytic activity.

In this study, we showed that it was possible to create a DhaA variant with a superior trade-off between activity and stability relative to that seen in DhaA80. This was achieved by replacing a sterically demanding phenylalanine residue at the mouth of the main access tunnel with a small glycine residue. This substitution enhanced the flexibility of the two α-helices that form the tunnel but did not strongly reduce protein resistance to organic co-solvents and tolerance towards elevated temperatures because other bulky hydrophobic residues inside the tunnel were retained. A potentially viable alternative strategy for balancing activity and stability in this case would be to introduce a molecular gate in the access tunnel. Molecular gates are dynamic protein structures that regulate substrate access to the active site and product release while preventing the access of undesirable solvent molecules and synchronising processes occurring in distant parts of a protein.[76] Introducing a gate should protect the enzyme against irreversible inactivation under harsh conditions but maintaining good catalytic performance. Additionally, the auxiliary slot tunnels could be subjected to optimisation for further improving the stability of DhaA106 without significantly compromising its activity.

Conclusions

We have demonstrated that the catalytic performance of the thermodynamically robust, but less active, haloalkane dehalogenase variant DhaA80 can be greatly enhanced by fine-tuning the geometry and dynamics of its access tunnel. A single-point mutation (F176G) in the tunnel mouth yielded the new variant DhaA106, which exhibits greater flexibility in the secondary structure elements that form the access tunnel. The activities of DhaA106 towards 1,2-dibromoethane in buffer solution and in 40 vol.% DMSO were 32- and 10 times higher,

respectively, than those of the template enzyme DhaA80. On the other hand, its melting temperature (which reflects its thermodynamic stability) was reduced by only 4 °C. The high stability of the template enzyme was preserved because DhaA106 retains three previously introduced bulky residues in the tunnel interior that provide good hydrophobic packing and prevent solvent molecules from accessing the active site. In addition to its enhanced activity towards 1,2-dibromoethane, DhaA106 also exhibited enhanced activity towards 26 out of 29 other halogenated compounds. These results suggest that a fine balance between tunnel flexibility and tight hydrophobic packing, as well as a precisely engineered tunnel diameter are important for HLD activity and stability. Tunnel residues are thus good targets for modification if seeking to balance the activity and stability of catalysts with buried active sites.

Experimental Section

Library design and predicting the effects of mutations on enzyme stability

The structure of DhaA80 (PDB ID 4F60) was downloaded from the RCSB PDB database.[77] The structure was prepared for analysis by removing ligands and water molecules. Missing atoms in side chains were added by using the <RepairPDB> module of FoldX.[36] The stability effects of all possible double point mutations in positions F176 and V172 of DhaA80 were estimated by using the FoldX <BuildModel> module.[36] Two variants differing in the specified order of mutations were considered for each double point mutant (e.g., V172A, F176A and F176A, V172A for the F176A+V172A mutant). Calculations were performed 5 times for each variant following the recommended protocol (pH 7, temperature 298 K, ion strength 0.050 M, VdWDesign 2). All stabilised ($\Delta\Delta G \leq -1$ kcal mol^{-1}) and neutral (-1 kcal mol$^{-1} < \Delta\Delta G \leq 1$ kcal mol^{-1}) mutants were selected and the frequencies of individual residues at target positions were counted. Suitable degenerate codons for saturation mutagenesis were chosen using the CASTER v2.0 program.[37] The degenerate codons were selected to encode all frequent residues from the double point mutants without producing an excessively large library.

Library construction

Saturation mutagenesis was performed using the QuickChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, USA). Positions 172 and 176 of DhaA80 were saturated simultaneously by using the following oligonucleotides (Sigma Aldrich, St. Louis, USA):

5′-GCTTTCATCGAGCAAVTYCTCCCGAAAWKSGTCGTCCGTCCGCTTACG-3′ (forward),
5′-CGTAAGCGGACGGACGACSMWTTTCGGGAGRABTTGCTCGATGAAAGC-3′ (reverse).

Position 176 was independently saturated by using a pair of oligonucleotides (Sigma Aldrich, St. Louis, USA):

5′-CGAGCAAGTGCTCCCGAAANNKGTCGTCCGTCCGCTTAC-3′ (forward),
5′-GTAAGCGGACGGACGACMNNTTTCGGGAGCACTTGCTCG-3′ (reverse).

The entire plasmid pAQN:: dhaA80His6 served as a template for PCR and was amplified according to the manufacturer's protocol. PCR was performed by using 50 μL reaction mixtures containing template DNA (10 ng), the oligonucleotides (5 pmol each), and 0.2 mM dNTPs in Phusion HF buffer with 1.5 mM $MgCl_2$ and 1 U of Phusion DNA Polymerase. PCR proceeded under the following conditions: 30 s at 95 °C, and then 18 cycles of 30 s at 95 °C, 60 s at 55 °C and 300 s at 68 °C; followed by 10 min at 72 °C. PCR products were then treated with the methylation-dependent endonuclease DpnI for 60 min at 37 °C. The resulting plasmids were transformed into *E. coli* XJb(DE3) cells (ZymoResearch, Orange, USA) by using the standard electroporation protocol.[78] Ten candidates from each library were randomly selected for sequencing.

Cultivation in microtiter plates (MTP) and preparation of lysates

MTP wells filled with Luria-Bertani (LB) medium (150 μL) with ampicillin added to a final concentration of 100 μg mL^{-1} were inoculated with the single colonies by using sterile tooth-picks. Four wells were inoculated with *E. coli* XJB pAQN:: dhaA80His6 cells to serve as positive controls for basal activity measurement and another four wells were inoculated with *E. coli* XJB carrying an empty vector (pAQN) to serve as negative controls in the library screening. Cultures were grown overnight at 37 °C at 200 rpm. After 14 h of cultivation ($OD_{600} = 0.4$), 50 μL culture samples from each cultivation plates was added to 50 μL of 30 vol% glycerol in new 96-well plates to create a replica plate for storage. Fresh LB medium with ampicillin (100 μL), L-arabinose at a final concentration of 3 mM and IPTG at a final concentration of 0.5 mM were added to each well of the cultivation plate and incubated at 30 °C at 200 rpm for 4 h. Cells were harvested and frozen at −80 °C.

Library screening

Library screening was performed by using the modified pH colorimetric assay described by Holloway et al.[38] The assay is based on the detection of the protons produced during the dehalogenation reaction. After 10 min at RT, a 50 μL volume of the lysis buffer (1 mM HEPES, 20 mM Na_2SO_4 and 1 mM EDTA, pH 8.2) was added to each well of the defrosted plates. Cell debris was removed from the lysate by centrifugation at 1600 g for 20 min after 1 h incubation at 100 rpm at RT. A 20 μL volume of the lysate was transferred into each well of a new MTP and a 180 μL volume of assay buffer (52 vol.% DMSO, 1 mM HEPES, 20 mM Na_2SO_4 and 1 mM EDTA, pH 8.2) containing 1,2-dibromoethane (DBE, 9.3 mM) was added. The substrate was incubated in the reaction buffer at 37 °C for 30 min before starting the reaction. The MTP plate was sealed carefully with a lid and parafilm. The reaction mixture was then diluted by using a buffer solution containing the pH indicator phenol red (1 mM HEPES, 20 mM Na_2SO_4 and 1 mM EDTA, 50 μg mL^{-1} phenol red, pH 8.2) for detection after 14 h of dehalogenation. The change in the color of the pH indicator was estimated by spectrophotometry at 540 nm as described by Holloway et al.[38]

Expression and purification of proteins

Recombinant plasmids with the DhaA variants were transformed into *E. coli* BL21(DE3). For overexpression, cells were grown at 37 °C to an optical density (OD_{600}) of ≈0.6 in LB medium (1 L) containing ampicillin (100 μg mL^{-1}). Protein expression was induced by adding IPTG to a final concentration of 0.5 mM in LB medium and

the temperature was decreased to 20 °C. Cells were harvested by centrifugation for 10 min at 3700 g after overnight cultivation. During harvesting, cells were washed once with 50 mM phosphate buffer with 10% glycerol (pH 7.5) and then resuspended in equilibrating purification buffer (16.4 mM K_2HPO_4, 3.6 mM KH_2PO_4, 500 mM NaCl, 10 mM imidazole, pH 7.5). Harvested cells were kept at −80 °C. Defrosted cells were disrupted by sonication with a Hielscher UP200S ultrasonic processor (Hielscher Ultrasonics, Teltow, Germany) and C-terminus His-tagged enzymes were purified to homogeneity by using Ni-NTA Superflow Cartridges (Qiagen, Hilden, Germany) as described previously.[33] The eluted proteins were dialysed against 50 mM phosphate buffer (pH 7.5). Protein concentrations were determined by using the Bradford reagent (Sigma–Aldrich, St. Louis, USA) with bovine serum albumin as a standard. The purity of the resulting proteins was checked by SDS polyacrylamide gel electrophoresis in 15% polyacrylamide gels. The gels were stained with Coomassie brilliant blue R-250 dye (Fluka, Buchs, Switzerland) and the molecular mass of the proteins was determined by using the Protein Molecular Weight Marker (Fermentas, Burlington, Canada).

Circular dichroism (CD) spectroscopy

CD spectra were recorded at 20 °C by using a Chirascan spectropolarimeter (Applied Photophysics, Leatherhead, United Kingdom) equipped with a Peltier thermostat. Data were collected from 185 to 260 nm, at 100 nm min^{-1} with a 1 s response time and 2 nm bandwidth using a 0.1 cm quartz cuvette. Each spectrum shown is the average of five individual scans and was corrected for the buffer's absorbance. Collected CD data were expressed in terms of the mean residue ellipticity. The thermal unfolding of the enzymes was followed by monitoring their ellipticity at 222 nm during heating from 20 to 80 °C at a rate of 1 °C min^{-1}, with a resolution of 0.1 °C. The resulting thermal denaturation curves were roughly normalised to represent signal changes between approximately 1 and 0 and fitted to sigmoidal curves using Origin 6.1 (OriginLab, Northampton, USA). Melting temperatures (T_m) were calculated as the midpoints of the enzymes' normalised thermal transitions.

Activity assay

Enzymatic activity was assayed by using the colorimetric method developed by Iwasaki et al.[79] The release of halide ions was analysed spectrophotometrically at 460 nm by using a SUNRISE microplate reader (Tecan, Grödig/Salzburg, Austria) after reaction with mercuric thiocyanate and ferric ammonium sulfate. The reactions were performed at 37 °C in 25 mL Reacti-flasks closed by Mininert valves. The reaction mixtures contained 1,2-dibromoethane dissolved in 100 mM glycine buffer (pH 8.6, 10 mL), 60 mM glycine buffer with 40 vol.% DMSO (10 mL) or 48 mM glycine buffer with 52 vol.% DMSO (10 mL). The reactions were initiated by addition of enzyme and monitored by periodically withdrawing of 1 mL samples from the reaction mixture, immediately mixing them with 35% nitric acid (0.1 mL) to terminate the reaction, and analysing the quenched samples spectrophotometrically. Dehalogenation activities were quantified as rates of product formation over time. Each activity was measured in 3–5 independent replicates and expressed as mean values with a standard error.

Principal component analysis (PCA)

A matrix containing the activity data for nine wild-type HLDs and two mutant HLDs with 30 substrates was analysed by PCA.[80] The aim of the analysis was to uncover relationships between individual HLDs based on their activities towards the standardised set of substrates.[40] PCA was performed by using STATISTICA 10.0 (StatSoft, Tulsa, USA). The raw data were log-transformed and weighted relative to the individual enzyme's activity towards other substrates prior to performing PCA to better discern the enzyme specificity profiles.[40] These transformed data were used to identify substrate specificity groups, i.e., groups of enzymes that exhibited similar specificity profiles regardless of their overall specific activities.

Steady-state kinetics

Steady-state kinetic constants were determined at 37 °C in 25 mL Reacti-flasks closed by Mininert valves by using the method previously described by Iwasaki et al.[79] The reaction mixtures contained 1,2-dibromoethane dissolved in 100 mM glycine buffer (pH 8.6, 10 mL) or 60 mM glycine buffer with 40 vol.% DMSO (10 mL). The activity measurements were performed by using at least twelve different substrate concentrations (0.2–20 mM). The initial concentration of 1,2-dibromoethane was determined by gas chromatography using a Trace GC 2000 (Finnigen, San Jose, USA) equipped with a flame ionisation detector and a DB-FFAP 30 m × 0.25 mm × 0.25 μm capillary column (J&W Scientific, Folsom, USA). The reaction was started by adding the enzyme. Samples were periodically withdrawn over a 60 min measurement period, immediately quenched by mixing with 35% nitric acid (0.1 mL) and then analysed. All data points corresponded to the mean of three independent replicates. Kinetic parameters were determined by nonlinear curve fitting of the resulting data points using Origin 6.1 (OriginLab, Northampton, USA) by the following equation for Michaelis–Menten kinetics [Eq. (1)] and Hill equation including substrate inhibition [Eq. (2)],[81,82] in which K_m is the Michaelis constant, $K_{0.5}$ is the substrate concentration at which half-maximal velocity is achieved according to the cooperativity model, n is the Hill coefficient, K_{si} is the inhibition constant and k_{cat} is the catalytic constant:

$$\frac{v}{V_{lim}} = \frac{[S]}{K_m + [S]} \quad (1)$$

$$\frac{v}{V_{lim}} = \frac{[S]^n}{K_{0.5}^n + [S]^n\left(1 + \frac{[S]}{K_{si}}\right)} \quad (2)$$

Crystallographic analysis

Crystals of DhaA106 were obtained by the vapour diffusion method in a sitting drop at RT. Crystals were grown from the drop prepared by mixing 2 μL volume of the protein (10.6 mg mL^{-1} in 50 mM Tris–HCl pH 7.5) with a 2 μL volume of precipitant solution (0.1 M sodium acetate trihydrate pH 4.8, 0.2 M ammonium acetate and 35% wt./vol. PEG 4000) and equilibrated against reservoir solution (300 μL). Diffraction data were collected at 100 K by using a home-source X-ray diffraction station (rotating anode Nonius FR591, Bruker-Nonius) equipped with a MAR345 detector (1.542 Å monochromatic fixed wavelength), at a resolution of 1.69 Å. The diffraction data were processed using the XDS program.[83] The structure of DhaA106 was solved by the molecular-replacement method using the program MOLREP[84] and the structure of DhaA14[85] (PDB ID 3G9X) as the search model. Model refinement

was performed by using the program REFMAC 5[86] from the CCP4 package (Collaborative Computational Project, Number 4, 1994), interspersed with manual adjustments using Coot.[87] The quality of the model with respect to the experimental data was assessed by using the program SFCHECK.[88] All-atom contacts in the refined structure of DhaA106 were validated by using the internal tools of Coot[87] and the MOLPROBITY service.[89]

Preparation of protein structures for simulations

The structures of DhaA and DhaA80 were downloaded from the RCSB PDB database (PDB ID 4E46 and 4F60), whereas the structure of DhaA106 was obtained within this study (PDB ID 4WCV). All structures were prepared for analysis by removing ligands and water molecules. Missing atoms in side chains were added by using the <RepairPDB> module of FoldX.[36] Repaired structures were minimised by Rosetta's minimize_with_cst application. Both backbone and side chains optimisation was enabled, the distance for full atom pair potential was set to 9 Å, and standard weights for energy function with a constraint weight of 1 were used. The output of the minimisation process was processed by using the script convert_to_cst_file.sh to create a constraint file.[90] Protocol 16 incorporating the backbone flexibility within the ddg_monomer module of Rosetta was applied to create a model of DhaA63 with default settings.[90] All four structures were protonated using the H++ server at pH 7.5.[91] Water molecules from the respective crystal structures were added to the systems. In the case of DhaA63, non-overlapping water molecules from the crystal structure of DhaA80 were used. Cl^- and Na^+ ions were added to a final concentration of 0.1 M using the Tleap module of AMBER 12.[92] Using the same module, an octahedron set of TIP3P water molecules[93] was added such that all atoms within the system were at least 10 Å from the octahedron's surface.

MD simulations

Energy minimisation and MD simulations were performed by using the PMEMD module of AMBER12[92] with the ff10 force field.[94] Initially, the investigated systems were minimised by 500 steps of steepest descent followed by 500 steps of conjugate gradient over five rounds with decreasing harmonic restraints. The restraints were applied as follows: 500 kcal mol^{-1} $Å^{-2}$ on all heavy atoms of the protein, and then 500, 125, 25 and 0 kcal mol^{-1} $Å^{-2}$ on backbone atoms only. The subsequent MD simulations employed periodic boundary conditions, using the particle mesh Ewald method for treatment of interactions beyond 10 Å cut-off,[95,96] and a 2 fs time step with the SHAKE algorithm to fix all bonds containing hydrogens.[97] Equilibration simulations consisted of two steps: (i) 20 ps of gradual heating from 0 to 300 K at constant volume, using a Langevin thermostat with a collision frequency of 1.0 ps^{-1}, and with harmonic restraints of 5.0 kcal mol^{-1} $Å^{-2}$ on the positions of all protein atoms, and (ii) 2000 ps of unrestrained MD at 300 K using the Langevin thermostat at a constant pressure of 1.0 bar using a pressure coupling constant of 1.0 ps. Finally, two separate 200 ns long production MD simulations were run for each system using the same settings as the second step of MD equilibration. Coordinates were saved at intervals of 2 ps and the resulting trajectories were analysed by using the Cpptraj module of AMBER12, and visualised by using Pymol 1.5 (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) and VMD 1.9.1.[98]

Tunnels analysis

Tunnels were analysed by using CAVER 3.01.[43] 100 000 snapshots sampled every 2 ps from 200 ns MD simulations were used as input structures. Each atom in the structure was approximated by 12+1 spheres. The tunnel search was performed by using a probe radius of 1.0 Å and its opening (i.e. ability to accommodate water molecules) was assessed by using a 1.4 Å probe; these values correspond to the program's default settings. 100 000 randomly selected tunnels were clustered into 25 clusters by using hierarchical average link clustering with a clustering threshold of 5. The remaining tunnels were assigned to these clusters by using supervised machine learning. The starting point initially specified by ND2 atom of Asn 41, OD2 atom of Asp 106, NE1 atom of Trp 107 and NE2 atom of His 272 was automatically optimised to prevent its collision with protein atoms.

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic (P207/12/0775), the Czech Ministry of Education of the Czech Republic (LO1214 and LH14027) and the European Regional Development Fund (ICRC CZ.1.05/1.1.00/02.0123). J.B. was supported by the "Employment of Best Young Scientists for International Cooperation Empowerment" (CZ.1.07/2.3.00/30.0037) project co-financed by the European Social Fund and the state budget of the Czech Republic. MetaCentrum and CERIT-SC are acknowledged for providing access to computing facilities (LM2010005 and CZ.1.05/3.2.00/08.0144). The authors would like to express thanks to Tatsiana Holubeva for help with enzyme crystallisation.

Keywords: alkanes · enzyme catalysis · halogenation · molecular dynamics · protein engineering

[1] Y. Xie, J. An, G. Yang, G. Wu, Y. Zhang, L. Cui, Y. Feng, *J. Biol. Chem.* **2014**, *289*, 7994 – 8006.
[2] B. M. Beadle, B. K. Shoichet, *J. Mol. Biol.* **2002**, *321*, 285 – 296.
[3] E. M. Meiering, L. Serrano, A. R. Fersht, *J. Mol. Biol.* **1992**, *225*, 585 – 589.
[4] C. Schreiber, A. M. Buckle, A. R. Fersht, *Structure* **1994**, *2*, 945 – 951.
[5] V. L. Thomas, A. C. McReynolds, B. K. Shoichet, *J. Mol. Biol.* **2010**, *396*, 47 – 59.
[6] X. Wang, G. Minasov, B. K. Shoichet, *J. Mol. Biol.* **2002**, *320*, 85 – 95.
[7] T. Davids, M. Schmidt, D. Böttcher, U. T. Bornscheuer, *Curr. Opin. Chem. Biol.* **2013**, *17*, 215 – 220.
[8] R. Chaloupková, J. Sýkorová, Z. Prokop, A. Jesenská, M. Monincová, M. Pavlová, M. Tsuda, Y. Nagata, J. Damborský, *J. Biol. Chem.* **2003**, *278*, 52622 – 52628.
[9] V. G. H. Eijsink, A. Bjørk, S. Gåseidnes, R. Sirevåg, B. Synstad, B. van den Burg, G. Vriend, *J. Biotechnol.* **2004**, *113*, 105 – 120.
[10] J. T. Holland, J. C. Harper, P. L. Dolan, M. M. Manginell, D. C. Arango, J. A. Rawlings, C. A. Apblett, S. M. Brozik, *PLoS One* **2012**, *7*, e37924.
[11] H. Zhao, K. Chockalingam, Z. Chen, *Curr. Opin. Biotechnol.* **2002**, *13*, 104 – 110.
[12] P. Dalby, *Curr. Opin. Struct. Biol.* **2011**, *21*, 473 – 480.
[13] M. Lehmann, M. Wyss, *Curr. Opin. Biotechnol.* **2001**, *12*, 371 – 375.
[14] R. A. Chica, N. Doucet, J. N. Pelletier, *Curr. Opin. Biotechnol.* **2005**, *16*, 378 – 384.
[15] H. B. Brundiek, A. S. Evitt, R. Kourist, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2012**, *51*, 412 – 414; *Angew. Chem.* **2012**, *124*, 425 – 428.
[16] H. J. Wijma, R. J. Floor, D. B. Janssen, *Curr. Opin. Struct. Biol.* **2013**, *23*, 588 – 594.
[17] J. Damborský, E. Rorije, A. Jesenská, Y. Nagata, G. Klopman, W. J. Peijnenburg, *Environ. Toxicol. Chem.* **2001**, *20*, 2681 – 2689.

[18] M. Stucki, G. Thuer, *Environ. Sci. Technol.* **1995**, *29*, 2339 – 2345.
[19] K. Skopelitou, N. Georgakis, R. Efrose, E. Flemetakis, N. E. Labrou, *J. Mol. Catal. B* **2013**, *97*, 5 – 11.
[20] Z. Prokop, F. Oplustil, J. DeFrank, J. Damborský, *Biotechnol. J.* **2006**, *1*, 1370 – 1380.
[21] A. Westerbeek, W. Szymański, H. J. Wijma, S. J. Marrink, B. L. Feringa, D. B. Janssen, *Adv. Synth. Catal.* **2011**, *353*, 931 – 944.
[22] Z. Prokop, Y. Sato, J. Brezovsky, T. Mozga, R. Chaloupkova, T. Koudelakova, P. Jerabek, V. Stepankova, R. Natsume, J. G. E. van Leeuwen, D. B. Janssen, J. Florian, Y. Nagata, T. Senda, J. Damborsky, *Angew. Chem. Int. Ed.* **2010**, *49*, 6111 – 6115; *Angew. Chem.* **2010**, *122*, 6247 – 6251.
[23] S. Bidmanova, R. Chaloupkova, J. Damborsky, Z. Prokop, *Anal. Bioanal. Chem.* **2010**, *398*, 1891 – 1898.
[24] S. Bidmanova, A. Hlavacek, J. Damborsky, Z. Prokop, *Sens. Actuators B* **2012**, *161*, 93 – 99.
[25] S. Bidmanova, J. Damborsky, Z. Prokop, *J. Biocatal. Biotransformation* **2013**, *2*, 1 – 7.
[26] S. Mazzucchelli, M. Colombo, P. Verderio, E. Rozek, F. Andreata, E. Galbiati, P. Tortora, F. Corsi, D. Prosperi, *Angew. Chem. Int. Ed.* **2013**, *52*, 3121 – 3125; *Angew. Chem.* **2013**, *125*, 3203 – 3207.
[27] G. Los, R. Learish, N. Karassina, C. Zimprich, M. G. Mcdougall, L. P. Encell, R. Freidman-Ohana, M. Wood, G. Vidugiris, K. Zimmerman, P. Otto, S. Berhstock, D. H. Klaubert, K. Wood, *Cell Notes* **2006**, 10 – 14.
[28] G. V. Los, K. Wood in *Methods in Molecular Biology: High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery, Vol. 356* (Eds.: D. L. Taylor, J. R. Haskins, K. Giuliano) Humana Press, Inc, Totowa, **2006**, pp. 195 – 208.
[29] G.. V. Los, L. P. Encell, M. G. Mcdougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, K. V. Wood, *ACS Chem. Biol.* **2008**, *3*, 373 – 382.
[30] G. V. Los, A. Drazins, N. Karassina, C. Zimprich, R. Learish, M. G. Mcdougall, L. P. Encell, R. Friedman-Ohana, M. Wood, G. Vidugiris, K. Zimmerman, P. Otto, D. H. Klaubert, K. V. Wood, *Cell Notes* **2005**, 2 – 6.
[31] R. Chaloupkova, Z. Prokop, Y. Sato, Y. Nagata, J. Damborsky, *FEBS J.* **2011**, *278*, 2728 – 2738.
[32] V. Stepankova, J. Damborsky, R. Chaloupkova, *Biotechnol. J.* **2013**, *8*, 719 – 729.
[33] T. Koudelakova, R. Chaloupkova, J. Brezovsky, Z. Prokop, E. Sebestova, M. Hesseler, M. Khabiri, M. Plevaka, D. Kulik, I. K. Smatanova, P. Rezacova, R. Ettrich, U. T. Bornscheuer, J. Damborsky, *Angew. Chem. Int. Ed.* **2013**, *52*, 1959 – 1963; *Angew. Chem.* **2013**, *125*, 2013 – 2017.
[34] A. N. Kulakova, M. J. Larkin, L. A. Kulakov, *Microbiology* **1997**, *143*, 109 – 115.
[35] K. A. Gray, T. H. Richardson, K. Kretz, J. M. Short, F. Bartnek, R. Knowles, L. Kan, P. E. Swanson, D. E. Robertson, *Adv. Synth. Catal.* **2001**, *343*, 607 – 617.
[36] R. Guerois, J. E. Nielsen, L. Serrano, *J. Mol. Biol.* **2002**, *320*, 369 – 387.
[37] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891 – 903.
[38] P. Holloway, J. T. Trevors, H. Lee, *J. Microbiol. Methods* **1998**, *32*, 31 – 36.
[39] S. Y. Venyaminov, J. T. Yang in *Circular Dichroism and the Conformational Analysis of Biomolecules*, (Ed.: G. D. Fasman), Plenum Press, New York and London, **1996**, pp. 69-108.
[40] T. Koudelakova, E. Chovancova, J. Brezovsky, M. Monincova, A. Fortova, J. Jarkovsky, J. Damborsky, *Biochem. J.* **2011**, *435*, 345 – 354.
[41] A. Stsiapanava, J. Dohnalek, J. Gavira, M. Kuty, T. Koudelakova, J. Damborsky, I. K. Smatanova, *Acta Crystallogr. Sect. D* **2010**, *66*, 962 – 969.
[42] J. Newman, T. S. Peat, R. Richard, L. Kan, P. E. Swanson, J. A. Affholter, I. H. Holmes, J. F. Schindler, C. J. Unkefer, T. C. Terwilliger, *Biochemistry* **1999**, *38*, 16105 – 16114.
[43] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, J. Damborsky, *PLoS Comput. Biol.* **2012**, *8*, e1002708.
[44] J. Damborsky, J. Brezovsky, *Curr. Opin. Chem. Biol.* **2014**, *19*, 8 – 16.
[45] B. X. Yan, Y. Q. Sun, *J. Biol. Chem.* **1997**, *272*, 3190 – 3194.
[46] H. Neurath, *J. Am. Chem. Soc.* **1943**, *65*, 2039 – 2041.
[47] M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem. Int. Ed.* **2006**, *45*, 7745 – 7751; *Angew. Chem.* **2006**, *118*, 7909 – 7915.
[48] G. Feller, *J. Phys. Condens. Matter* **2010**, *22*, 323101.
[49] H. Jochens, D. Aerts, U. T. Bornscheuer, *Protein Eng. Des. Sel.* **2010**, *23*, 903 – 909.
[50] M. T. Reetz, P. Soni, L. Fernández, Y. Gumulya, J. D. Carballeira, *Chem. Commun.* **2010**, *46*, 8657 – 8658.
[51] P. Fields, *Comp. Biochem. Physiol. Part A* **2001**, *129*, 417 – 431.
[52] S. Cesarini, C. Bofill, F. I. J. Pastor, M. T. Reetz, P. Diaz, *Process Biochem.* **2012**, *47*, 2064 – 2071.
[53] C. L. Tsou, *Science* **1993**, *262*, 380 – 381.
[54] P. Banás, M. Otyepka, P. Jerábek, M. Petrek, J. Damborský, *J. Comput.-Aided. Mol. Des.* **2006**, *20*, 375 – 383.
[55] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, *Nat. Chem. Biol.* **2009**, *5*, 727 – 733.
[56] J. Damborsky, Z. Prokop, T. Koudelakova, V. Stepankova, R. Chaloupkova, A. Gora, E. Chovancova, J. Brezovsky, *Method of Thermostabilisation of a Protein and/or Stabilisation Towards Organic Solvents*, **2013**, US 8,580,932.
[57] Z. Prokop, A. Gora, J. Brezovsky, R. Chaloupkova, V. Stepankova, J. Damborsky in *Protein Engineering Handbook* (Eds.: S. Lutz, U. T. Bornscheuer), Wiley-VCH, Weinheim, **2012**, pp. 421 – 464.
[58] G. Smith, S. Modi, I. Pillai, L. Lian, M. J. Sutcliffe, M. P. Pritchard, T. Friedberg, G. C. K. Roberts, C. R. Wolf, *Biochem. J.* **1998**, *331*, 783 – 792.
[59] T. Nguyen, M. Tychopoulos, F. Bichat, C. Zimmermann, J. Flinois, M. Diry, E. Ahlberg, M. Delaforge, L. Corcos, P. Beaune, P. Dansette, F. Andre, I. de Waziers, *Mol. Pharmacol.* **2008**, *73*, 1122 – 1133.
[60] D. Fishelovitch, S. Shaik, H. J. Wolfson, R. Nussinov, *J. Phys. Chem. B* **2009**, *113*, 13018 – 13025.
[61] K. K. Khan, Y. Q. He, T. L. Domanski, J. R. Halpert, *Mol. Pharmacol.* **2002**, *61*, 495 – 506.
[62] Z. Wen, J. Baudry, M. R. Berenbaum, M. A. Schuler, *Protein Eng. Des. Sel.* **2005**, *18*, 191 – 199.
[63] A. B. Carmichael, L. L. Wong, *Eur. J. Biochem.* **2001**, *268*, 3117 – 3125.
[64] H.-L. Lee, C.-K. Chang, W.-Y. Jeng, A. H.-J. Wang, P.-H. Liang, *Protein Eng. Des. Sel.* **2012**, *25*, 733 – 740.
[65] J. Schmitt, S. Brocca, R. D. Schmid, J. Pleiss, *Protein Eng.* **2002**, *15*, 595 – 601.
[66] Z. Qian, J. R. Horton, X. Cheng, S. Lutz, *J. Mol. Biol.* **2009**, *393*, 191 – 201.
[67] Z. Marton, V. Léonard-Nevers, P.-O. Syrén, C. Bauer, S. Lamare, K. Hult, V. Tranc, M. Graber, *J. Mol. Catal. B* **2010**, *65*, 11 – 17.
[68] M. Z. Kamal, T. A. S. Mohammad, G. Krishnamoorthy, N. M. Rao, *PLoS One* **2012**, *7*, e35188.
[69] H. Brundiek, S. K. Padhi, R. Kourist, A. Evitt, U. T. Bornscheuer, *Eur. J. Lipid Sci. Technol.* **2012**, *114*, 1148 – 1153.
[70] A. Schließmann, A. Hidalgo, J. Berenguer, U. T. Bornscheuer, *ChemBioChem* **2009**, *10*, 2920 – 2923.
[71] M. Kotik, V. Stepánek, P. Kyslík, H. Maresová, *J. Biotechnol.* **2007**, *132*, 8 – 15.
[72] B. Van Loo, J. H. L. Spelberg, J. Kingma, T. Sonke, M. G. Wubbolts, D. B. Janssen, *Chem. Biol.* **2004**, *11*, 981 – 990.
[73] L. Biedermannová, Z. Prokop, A. Gora, E. Chovancová, M. Kovács, J. Damborsky, R. C. Wade, *J. Biol. Chem.* **2012**, *287*, 29062 – 29074.
[74] T. Bosma, G. Stucki, D. B. Janssen, *Appl. Environ. Microbiol.* **2002**, *68*, 3582 – 3587.
[75] M. Klvana, M. Pavlova, T. Koudelakova, R. Chaloupkova, P. Dvorak, Z. Prokop, A. Stsiapanava, M. Kuty, I. Kuta-Smatanova, J. Dohnalek, P. Kulhanek, R. C. Wade, J. Damborsky, *J. Mol. Biol.* **2009**, *392*, 1339 – 1356.
[76] A. Gora, J. Brezovsky, J. Damborsky, *Chem. Rev.* **2013**, *113*, 5871 – 5923.
[77] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235 – 242.
[78] J. Sambrook, D. W. Russel, *Molecular Cloning, A Laboratory Manual, 3rd ed.*, Cold Spring Harbor Laboratory Press, New York, **2001**.
[79] I. Iwasaki, S. Utsumi, T. Ozawa, *Bull. Chem. Soc. Jpn.* **1952**, *25*, 226 – 226.
[80] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37 – 52.
[81] I. H. Segel, *Enzyme Kinetics*, Wiley, New York, **1993**.
[82] V. Stepankova, M. Khabiri, J. Brezovsky, A. Pavelka, J. Sykora, M. Amaro, B. Minofar, Z. Prokop, M. Hof, R. Ettrich, R. Chaloupkova, J. Damborsky, *ChemBioChem* **2013**, *14*, 890 – 897.
[83] W. Kabsch, *J. Appl. Crystallogr.* **1993**, *26*, 795 – 800.
[84] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022 – 1025.
[85] A. Stsiapanava, R. Chaloupkova, A. Fortova, J. Brynda, M. S. Weiss, J. Damborsky, I. K. Smatanova, *Acta Crystallogr. Sect. F* **2011**, *67*, 253 – 257.

[86] G. N. Murshudov, A. A. Vagin, *Acta Crystallogr. Sect. D* **1997**, *53*, 240 – 255.
[87] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D* **2010**, *66*, 486 – 501.
[88] A. A. Vaguine, J. Richelle, S. J. Wodak, *Acta Crystallogr. Sect. D* **1999**, *55*, 191 – 205.
[89] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. Sect. D* **2010**, *66*, 12 – 21.
[90] E. H. Kellogg, A. Leaver-Fay, D. Baker, *Proteins* **2011**, *79*, 830 – 838.
[91] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, A. Onufriev, *Nucleic Acids Res.* **2005**, *33*, W368 – W371.
[92] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 12, **2012**, University of California, San Francisco.
[93] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926 – 935.
[94] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712 – 725.
[95] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089 – 10092.
[96] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577 – 8593.
[97] J. P. Ryckaert, G. Ciccotti, H. J. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327 – 341.
[98] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33 – 38.

Received: October 2, 2014
Published online on January 22, 2015