



HAL
open science

Substrate specificity of haloalkane dehalogenases

Tana Koudelakova, Eva Chovancova, Jan Brezovsky, Marta Monincova,
Andrea Fortova, Jiri Jarkovsky, Jiri Damborsky

► **To cite this version:**

Tana Koudelakova, Eva Chovancova, Jan Brezovsky, Marta Monincova, Andrea Fortova, et al..
Substrate specificity of haloalkane dehalogenases. *Biochemical Journal*, 2011, 435 (2), pp.345-354.
10.1042/BJ20101405 . hal-00581534

HAL Id: hal-00581534

<https://hal.science/hal-00581534>

Submitted on 31 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Substrate specificity of haloalkane dehalogenases

Tana KOUDELAKOVA^{*1}, Eva CHOVANCOVA^{*1}, Jan BREZOVSKY^{*}, Marta MONINCOVA^{*}, Andrea FORTOVA^{*}, Jiri JARKOVSKY[†] and Jiri DAMBORSKY^{*2}

*Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment, Faculty of Science, Masaryk University, Kamenice 5/A13, 62500 Brno, Czech Republic, †Institute of Biostatistics and Analyses, Faculty of Medicine and Faculty of Science, Masaryk University, Kamenice 126/3, 62500 Brno, Czech Republic.

SYNOPSIS

An enzyme's substrate specificity is one of its most important characteristics. The quantitative comparison of broad-specificity enzymes requires the selection of a homogenous set of substrates for experimental testing, determination of substrate specificity data, and analysis using multivariate statistics. We describe a systematic analysis of the substrate specificities of nine wild-type and four engineered haloalkane dehalogenases. The enzymes were characterized experimentally using a set of 30 substrates selected using statistical experimental design from a set of nearly 200 halogenated compounds. Analysis of the activity data showed that the most universally-useful substrates in the assessment of haloalkane dehalogenase activity are 1-bromobutane, 1-iodopropane, 1-iodobutane, 1,2-dibromoethane and 4-bromobutanenitrile. Functional relationships among the enzymes were explored using Principal Component Analysis. Analysis of the untransformed specific activity data revealed that the overall activity of wild-type haloalkane dehalogenases decreases in the following order: LinB~DbjA > Dh1A~DhaA~DbeA~DmbA > DatA~DmbC~DrbA. After transforming the data, we were able to classify haloalkane dehalogenases into four substrate specificity groups. These functional groups are clearly distinct from the evolutionary subfamilies, suggesting that phylogenetic analysis cannot be used to predict the substrate specificity of individual haloalkane dehalogenases. Structural and functional comparisons of wild-type and mutant enzymes revealed that the architecture of the active site and the main access tunnel significantly influences the substrate specificity of these enzymes, but is not its only determinant. The identification of other structural determinants of the substrate specificity remains a challenge for further research on haloalkane dehalogenases.

Short (page heading) title: Substrate specificity of haloalkane dehalogenases

Key words: haloalkane dehalogenase, substrate specificity, Principal Component Analysis, phylogenetic analysis, protein engineering.

Abbreviations used: HLD, haloalkane dehalogenase; PC, Principal Component; PCA, Principal Component Analysis; SSG, substrate specificity group.

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed (e-mail jiri@chemi.muni.cz).

INTRODUCTION

Enzymes are biological catalysts that are essential components of every biological system and are valuable in biotechnology. The key functional characteristics of an enzyme are its catalytic activity towards different substrates and its substrate specificity, i.e. the range of substrates it can convert. As such, the identification of enzymes that efficiently catalyse new chemical reactions or display novel substrate specificities is of great scientific and practical interest. The traditional way of isolating novel biocatalysts is a time-consuming multistep process involving enrichment of organisms from a natural resource, construction of a genomic library, cloning of the library into a host organism, screening for appropriate activity, protein purification and biochemical characterization. This process has been greatly accelerated by the development of new techniques in molecular biology and bioinformatics, including high-throughput techniques for screening mutant and metagenomic libraries, methods for the *in silico* identification of potential targets using sequence database searches and bioinformatics tools, and various novel approaches to protein engineering [1, 2].

Haloalkane dehalogenases (HLDs, EC 3.8.1.5) are enzymes that catalyze hydrolytic cleavage of the carbon-halogen bond in a wide range of halogenated compounds. They have a number of potential practical applications, including roles in industrial biocatalysis [3, 4], bioremediation [5], detoxification [6], biosensing [7] and molecular imaging [8]. The properties of several HLDs have been improved by directed evolution [9-12]. A substantial body of knowledge concerning the structure and function of HLDs also allows construction of modified enzymes by rational design [4, 13, 14].

Structurally, HLDs belong to the α/β -hydrolase superfamily. Their active site is buried in the predominantly hydrophobic cavity at the interface of the α/β -hydrolase core domain and the helical cap domain, and is connected to the bulk solvent by access tunnels [4, 15-18]. The active-site residues that are essential for catalysis are referred to as the catalytic pentad, and comprise a nucleophilic aspartate, a basic histidine, an aspartic or glutamic acid moiety that serves as a general acid, and either two tryptophans or a tryptophan-asparagine pair that serve to stabilise the leaving halide ion [19]. The HLD family currently includes 14 distinct enzymes with experimentally confirmed dehalogenation activity [20-27]. A recent analysis of the sequences and structures of these HLDs and their homologs divided the family into three phylogenetic subfamilies, HLD-I, HLD-II and HLD-III, which differ mainly in the composition of the catalytic pentad and cap domain [19].

To date, HLDs have been isolated from bacterial strains originating from the soil [23, 24, 27, 28], sea water [22, 27], obligatory animal pathogens [20, 21], plant symbionts [26] and plant parasites [29]. While the biological function of many HLDs remains unknown, those that were isolated from bacteria inhabiting contaminated soil are known to be involved in metabolic pathways that enable the host organisms to utilize halogenated compounds as carbon sources [23, 28, 30]. HLDs catalyze the hydrolysis of chlorinated, brominated and iodinated alkanes, alkenes, cycloalkanes, alcohols, carboxylic acids, esters, ethers, epoxides, amides and nitriles [4, 31, 32] and are thus broad-specificity enzymes, exhibiting miscellaneous activity across a wide range of substrate classes. The substrate specificity of HLDs can be described in terms of a quantitative profile of their specific activities with respect to a set of specific substrates. Quantitative comparisons of such specificity profiles can be used to identify appropriate catalysts for practical applications and to further our understanding of the relationships between the enzymes in terms of their function, structure and evolution.

This study focused on the comparison and classification of the substrate specificities of nine members of the HLD family. A functional classification of the HLDs was carried out using Principal Component Analysis (PCA) and the classification thus derived was compared with one derived on the basis of the enzymes' evolutionary relationships. The purpose of this

comparison was to see whether the substrate specificity of individual HLDs reflects the evolution of the family and thus could be predicted from established phylogenetic classifications. Factors influencing the substrate specificity of HLDs were assessed by structural and functional comparison of wild-type and mutant enzymes. This study also identifies 'universal' substrates converted by all of the enzymes examined as well as 'preferred' and 'characteristic' substrates for individual substrate specificity groups. Such knowledge will be useful for the selection of appropriate biocatalysts for specific biotechnological applications and the development of platforms for screening HLD activity in different hosts, environments, or *in vitro* samples.

EXPERIMENTAL

Materials

All halogenated compounds used were of at least 95% purity, and were purchased from Sigma-Aldrich (St. Louis, MO, U.S.A.).

Preparation of enzymes and activity assay

The wild-type HLDs examined were: DatA from *Agrobacterium tumefaciens* C58 [29], DbeA from *Bradyrhizobium elkani* USDA94 (T. Prudnikova, P. Rezacova, Z. Prokop, T. Mozga, Y. Sato, M. Kutý, Y. Nagata, J. Damborsky, I. Kuta-Smatanova, R. Chaloupkova, unpublished work), DbjA from *Bradyrhizobium japonicum* USDA110 [26], DhaA from *Rhodococcus rhodochrous* NCIMB 13064 [24], DhIA from *Xanthobacter autotrophicus* GJ10 [23], DmbA [21] and DmbC [22] from *Mycobacterium bovis* 5033/66, DrbA from *Rhodopirellula baltica* SH1 [22] and LinB from *Sphingobium japonicum* UT26 [25] (Supplementary Table S1). Mutant enzymes were constructed by rational design or focused directed evolution and include DbeA1 and DbeA2, which carry the insertions $_{142}\text{Val-Ala-Glu-Glu-Gln-Asp-His-Ala-Glu}_{143}$ and $_{141}\text{Glu-Val-Ala-Glu-Glu-Gln-Asp-His-Ala}_{142}$, respectively (R. Chaloupkova, T. Mozga, Y. Sato, T. Prudnikova, T. Koudelakova, E. Chovancova, P. Rezacova, Y. Nagata, I. Kuta-Smatanova and J. Damborsky, unpublished work); DbjA Δ , from which the $_{139}\text{His-Thr-Glu-Val-Ala-Glu-Glu}_{147}$ residues were deleted [4]; and DhaA31, which incorporates the substitutions Ile135Phe, Cys176Tyr, Val245Phe, Leu246Ile, Tyr273Phe [11] (Supplementary Table S2). His-tagged enzymes were heterogeneously expressed in *Escherichia coli* or *Mycobacterium smegmatis* strains using appropriate vectors and purified to homogeneity using immobilized metal affinity chromatography as described elsewhere [11, 13, 21, 22, 26]. Specific activities of HLDs towards the set of 30 halogenated substrates were taken from previous studies [11, 22, 29] (T. Prudnikova, P. Rezacova, Z. Prokop, T. Mozga, Y. Sato, M. Kutý, Y. Nagata, J. Damborsky, I. Kuta-Smatanova, R. Chaloupkova, unpublished work) or determined under the conditions used in those studies for DbjA, DhaA, DhIA, DmbA and LinB (Supplementary Table S3). Enzyme concentration was estimated using Bradford reagent (Sigma-Aldrich, St. Louis, MO, U.S.A.) with bovine serum albumin as a standard. Specific activity was measured using reagents containing mercuric thiocyanate and ferric ammonium sulphate; the halide ions released during the dehalogenase reaction were quantified by an endpoint spectrophotometric measurement [33]. Reactions were carried out in 100 mM glycine buffer (pH 8.6) in 25 cm³ Microflasks closed by Mininert valves (Alltech, Deerfield, IL, U.S.A.) at 37°C. The initial experimental concentration of the halogenated substrates in the reaction mixture was established on gas chromatograph GC Trace 2000 (Thermo Fisher Scientific, Waltham, MA, U.S.A.) equipped with flame ionization detector and capillary column DB-FFAP 30 m \times 0.25 mm \times 0.25 μm (J&W Scientific, Folsom, CA, U.S.A.) (Supplementary Table S3). Samples were periodically withdrawn with a 1 cm³ syringe

(Hamilton, Reno, NV, U.S.A.) during 40 min measurement after the initiation of the reaction by the addition of an enzyme. All withdrawn samples were immediately mixed with 35% nitric acid to stop the reaction. The reagents with mercuric thiocyanate and ferric ammonium sulphate were subsequently added to the collected samples and absorbance of the final mixture was measured in a microtiter plate at 460 nm by Sunrise spectrophotometer (Tecan, Männedorf, Switzerland). Spontaneous hydrolysis of substrates in buffer was tested in the abiotic control. The specific activities were quantified by an initial linear slope of the increasing halide concentration in time after the subtraction of spontaneous hydrolysis. Kinetic constants of all nine wild-type HLDs towards 1-chlorobutane (**4**) or 1-iodobutane (**29**) were collected from the literature or measured as described in the Supplementary methods (Supplementary Table S4).

Statistical analyses

A matrix containing the activity data for the nine wild-type HLDs towards 30 substrates (Supplementary Tables S3 and S5) was analyzed by PCA to uncover relationships between individual HLDs (cases) and their substrates (variables). PCA of the data matrix X allows it to be expressed as the product of two new matrices plus a noise matrix of residuals: $X = TP' + E$ [34, 35]. The score matrix T (9 HLDs \times 30 substrates) summarizes the X -variables, the loading matrix P' (number of Principal Components \times 30 substrates) shows the influence of individual variables on the projection model, and the residual matrix E (9 HLDs \times 30 substrates) quantifies the differences between the original values and the projections. The underlying principles of PCA can be visualised by considering its geometrical interpretation [35, 36]. It is impossible to imagine nine points, representing the activity of individual HLDs, distributed in thirty-dimensional space. PCA projects these points onto a lower-dimensional subspace, and establishes a reduced set of new orthogonal coordinates called Principal Components (PCs). PCs are fitted to points in multidimensional space by the least squares method, such that the first PC is aligned in the direction of maximum variance in the data set, the second is aligned in the direction of the maximum remaining variance, and so on. The coordinate values of individual cases in the new coordinate system are called scores (t), and the projection of the data points onto the two-dimensional plain defined by any two PCs is called a score plot. The cosines of the angles between a given PC and the axes defined by the original variables are called loadings (p), and represent the contributions of the original variables to a particular PC. PCA was conducted using the Statistica 8.0 software package (StatSoft, Tulsa, OK, U.S.A.). Two PCAs were performed. In the first, the raw data concerning individual enzymes' specific activities towards particular substrates were used as the primary input data. In the second, the raw data were log-transformed and weighted relative to the individual enzyme's activity towards other substrates prior to analysis, in order to better discern individual enzymes' specificity profiles. Thus: (i) each specific activity value was incremented by 1 unit to avoid logarithmic transformation of zero values; (ii) the log of this new value was taken; and (iii) this log value was then divided by the sum of all the log values for that particular enzyme to give a log-transformed, weighted measure of that enzyme's activity towards that specific substrate relative to its activity towards all of the other substrates considered. These transformed data were used to identify enzymes with interesting or unusual specificity profiles, without regard to their overall specific activity. The score plots obtained from the analysis of these log-transformed data were used to classify the HLDs into substrate specificity groups; substrates that were important in defining individual groups were identified from the loading plots. The coordinates of individual enzymes in the space defined by the biologically significant PCs arising from this analysis were used to calculate a matrix of Euclidean distances. This matrix was in turn used to construct a dendrogram to characterize the similarities of individual HLDs

in terms of their substrate specificity profiles. The dendrogram was generated using the neighbour-joining (NJ) method [37], as implemented in the the DISTTREE program in the VANILLA v1.2 software package [38].

Phylogenetic analysis

Phylogenetic analysis of HLDs was carried out as previously described [19]. Briefly, all of the available sequences of HLDs and their closest homologs were gathered from the NCBI non-redundant protein database [39] using PSI-BLAST database searches [40]. HLDs were separated from other related protein families by clustering using CLANS [41]. A multiple sequence alignment of HLDs was constructed using MUSCLE v3.5 [42], and was then manually refined using the BioEdit v7.0.1 sequence editor [43]. Selected regions of the alignment were used to estimate a suitable evolutionary model and parameters by PROTTEST [44] and then for phylogenetic reconstruction by the NJ method. A distance matrix for NJ inference was generated using the MLDIST program of the VANILLA v1.2 package according to the WAG model of amino acid substitution [45]. The resulting phylogenetic tree was rooted by outgroup analysis. A Mantel test, performed using version 2.11.1 of the 'R' environment for statistical computing and graphics [46], was used to investigate the correlation between the matrices of the HLDs' phylogenetic distances and the matrix of Euclidean distances obtained from the PCA comparing substrate specificity profiles of wild-type HLDs.

RESULTS

Characterization of wild-type and engineered HLDs with a homogenous set of substrates

The substrate specificities of nine wild-type and four mutant HLDs with respect to a homogenous set of 30 substrates (Table 1) were studied and quantitatively compared. This substrate set was selected using statistical experimental design from 194 potential HLD substrates to sample entire space of 28 different physico-chemical properties (Supplementary Methods, Supplementary Table S6). An identical set of substrates and assay conditions was used for the characterization of all of the enzymes; otherwise the subsequent statistical analysis of the data obtained would have been less reliable. All of the HLDs examined exhibited good relative activities towards 1-bromobutane (**18**), 1-iodopropane (**28**), 1-iodobutane (**29**), 1,2-dibromoethane (**47**) and 4-bromobutanenitrile (**141**). 1,2-dichloroethane (**37**), 1,2-dichloropropane (**67**), 1,2,3-trichloropropane (**80**), chlorocyclohexane (**115**) and (bromomethyl)cyclohexane (**119**) were found to be generally poor substrates for the HLDs examined (Figure 1).

Table 1, Figure 1 here

Functional classification of wild-type HLDs

The matrix of the HLDs' untransformed specific activities towards the various substrates was subjected to analysis using PCA. Three biologically significant PCs were identified, which together accounted for 79% of the variance in the primary dataset (Supplementary Table S7). PC1 ranked the enzymes according to the magnitude of their overall activity towards the tested substrates: LinB~DbjA > Dh1A~DhaA~DbeA~DmbA > DatA~DmbC~DrbA (Supplementary Figure S1). LinB and DbjA were generally the most active of the HLDs analyzed; their specific activities were two to three orders of magnitude greater than those of DmbC and DrbA. PC2 and PC3 further separated the HLDs; specifically, these components identified three enzymes with unique specific activities towards several substrates (Supplementary Figure S2). DmbA exhibits high activity towards 2-iodobutane (**64**), 1-

chloro-2-(2-chloroethoxy)ethane (**111**) and chlorocyclopentane (**138**). DbjA exhibits exceptionally high activity towards 2-bromo-1-chloropropane (**76**) and chlorocyclopentane (**138**), and also catalyses the dehalogenation of the highly resistant substrates 1,2-dichloropropane (**67**) and 1,2,3-trichloropropane (**80**). DhIA possesses exceptional activity towards the chlorinated substrates 1,2-dichloroethane (**37**) and 1,3-dichloropropane (**38**).

Analysis of the untransformed data revealed that enzymes with similar overall activities can nevertheless have divergent activity profiles; similarities in the magnitude of two enzymes' overall activities can obscure interesting and potentially useful differences in their reactivity towards specific substrates. In such cases, data pretreatment methods can be used to facilitate interpretation of datasets by emphasizing biologically relevant information [47]. To this end, the activity data were log-transformed and weighted as described in the Experimental section, in order to minimize complications arising from differences in the enzymes' absolute catalytic proficiency and to emphasize the differences in their specificity profiles (Figure 1, Supplementary Table S5). The top three biologically significant PCs from this second model accounted for 62% of the variance in the transformed dataset (Supplementary Table S7). PCs quantify how individual HLDs act with all tested 30 substrates, resulting in the clusters of HLDs with similar specificity profiles. On the basis of the model, the HLDs were divided into four substrate specificity groups (SSGs) (Figure 2A, left): (i) SSG-I comprising DbjA, DhaA, DhIA and LinB, (ii) SSG-II containing DmbA, (iii) SSG-III containing DrbA and (iv) SSG-IV comprising DataA, DbeA and DmbC. This classification of the HLDs was primarily due to differences in their position along PC1 and PC2 (Figure 2A, left); the classification of DrbA and DmbA into separate groups was justified by the difference in their position along PC3 (Supplementary Figure S4A). HLDs in the same SSG exhibited common substrate preferences that differentiated them from HLDs in other groups (Table 2, Figure 2, left). HLDs in SSG-I are characterized primarily by their catalytic robustness. Their activity can be detected towards most of the tested substrates. All members are active towards at least one of poorly degradable compounds: 1,2-dichloroethane (**37**), 1,2-dichloropropane (**67**), 1,2,3-trichloropropane (**80**) and chlorocyclohexane (**115**). Enzymes in SSG-II, SSG-III and SSG-IV are more selective for specific halogenated compounds, differentiating them from other SSGs. The substrate specificity profile of SSG-II is similar to SSG-I, as obvious for example from good conversion of 1,2-dibromoethane (**47**) and 1-bromo-2-chloroethane (**137**). On the other hand, the substrate specificity of SSG-II is unique due to good activity towards otherwise not preferred substrates and inactivity towards 1,3-diiodopropane (**54**). DrbA from SSG-III possesses extremely low or zero activity towards all tested compounds. Unique preference for 1-chlorobutane (**4**) and inactivity with otherwise good substrates represent SSG-III characteristics. SSG-IV is mainly characterized by preference for terminally substituted brominated and iodinated propanes and butanes.

Figure 2, Table 2 here

Functional classification of mutant HLDs

In addition to the wild-type enzymes, four mutants (DbeA1, DbeA2, DbjA Δ and DhaA31; see Supplementary Table S5) were examined. The incorporation of these enzymes' specificity data generated a new PCA model, whose top three biologically-significant PCs accounted for 58% of the total variance in the dataset (Supplementary Table S7). The incorporation of the data on the mutant enzymes did not affect the proposed functional classification of the HLDs (Figure 2A, right), demonstrating the robustness of the model constructed for the wild-type enzymes. The engineered HLDs were found to cluster in the same SSG as their 'parent' enzymes.

The most pronounced difference between a mutant and its 'parent' in terms of substrate specificity was observed with DhaA31. Relative to DhaA, DhaA31 exhibits decreased relative activity towards longer substrates such as 1-bromohexane (20), 1-iodohexane (31) and increased relative activity towards 2-iodobutane (64), 1,2-dibromopropane (72) and 1,2,3-trichloropropane (80). Compared to DbjA, DbjA Δ exhibited a loss of activity towards 2-iodobutane (64) and decreased relative activity towards 2-bromo-1-chloropropane (76). However, it also exhibited a gain of activity towards 1,2-dibromo-3-chloropropane (155) and enhanced relative activity towards 1,3-diiodopropane (54), 1,2-dibromopropane (72) and 1-bromo-2-chloroethane (137). Relative to the 'parent' enzyme, the DbeA1 and DbeA2 mutants exhibited improved relative activity towards 1-bromobutane (18), 1-bromohexane (20) and 1,3-dibromopropane (48), and reduced relative activity towards 3-chloro-2-methylprop-1-ene (209), 1,5-dichloropentane (40), 2-iodobutane (64), and 1,3-diiodopropane (54), which was the best substrate for the 'parent' DbeA.

Comparison of functional and evolutionary classifications of HLDs

The examined dataset included representatives of all three HLD phylogenetic subfamilies. The HLD-I subfamily was represented by a single enzyme, Dh1A. The HLD-II subfamily was represented by six enzymes (DatA, DbeA, DbjA, DhaA, DmbA and LinB), while the HLD-III subfamily was represented by DmbC and DrbA. DbeA and DbjA have a protein sequence identity of 71% and are thus the most closely related pair of enzymes in the dataset, followed by DmbA and LinB, which show 68% protein sequence identity (Supplementary Table S8). On the other hand, the DbjA-DrbA and Dh1A-DmbA pairs both have mutual protein sequence identities of 19%, and are thus the most dissimilar pairs of enzymes in the dataset.

A comparison of the phylogenetic tree with the substrate specificity dendrogram revealed that members of the same phylogenetic subfamily are spread across different SSGs (Figure 3). Dh1A from the HLD-I subfamily is in SSG-I, along with the HLD-II subfamily members DbjA, DhaA and LinB. DmbA did not cluster together with its close relative LinB; instead, it forms a separate cluster, SSG-II. The other two HLD-II members – DbeA and DatA – are in SSG-IV together with DmbC from the HLD-III subfamily; the second representative of HLD-III, DrbA, is in its own specificity group, SSG-III. The Mantel test further confirmed the absence of a statistically significant correlation between the enzymes' evolutionary relationships and their substrate specificity profiles ($r_s = -0.286$; $p = 0.915$).

Figure 3 here

DISCUSSION

The biochemical characterization of nine HLDs with a set of 30 halogenated substrates, followed by multivariate statistical analysis, phylogenetic inference and structural comparisons allowed us to investigate the relationships between the structure, function and evolution of this broad-specificity family of enzymes. The analysis of the substrate preferences of individual wild-type HLDs revealed that 1-bromobutane (18), 1-iodopropane (28), 1-iodobutane (29), 1,2-dibromoethane (47) and 4-bromobutanenitrile (141) are good substrates for all nine enzymes. These 'universal' substrates are suitable for screening or biochemical characterization of putative HLDs.

Substrate specificities of individual HLDs

PCA carried out with untransformed data ranked the enzymes according to their absolute activities along PC1. LinB and DbjA possess the highest activities of all the analyzed HLDs, and are thus the most suitable family members for mechanistic studies [4, 13] and biotechnological applications [4-7]. At the other end of the spectrum, DatA, and especially

DmbC and DrbA, have very low specific activities towards most of the tested substrates. In the case of DrbA and DmbC, this may be related to the unique composition of their catalytic pentad, Asp-His-Asp+Asn-Trp, [19] or their highly oligomeric structures [22]. Their low activity values may also reflect incompatibility with the selected class of substrates. Nonetheless, DrbA, which originates from a marine organism, exhibited good catalytic efficiency and high relative activity towards 1-iodobutane (**29**) [22] (Supplementary Tables 4 and 5); this compound is produced by marine algae, along with other iodinated compounds [48]. The low activity of DatA may be due to its unusual active site, in which a tyrosine takes the place of the tryptophan located next to the nucleophile. This tryptophan residue is typically involved in stabilising the leaving halide in HLDs and is highly conserved in other members of the HLD family [19]. To date, DatA is the only characterized HLD to feature this exchange.

The distribution of HLDs along PC2 and PC3 highlighted certain unique functional properties of individual enzymes, such as the ability to convert resistant organic compounds or high activity towards specific substrates. Knowledge of these important catalytic properties is useful when selecting HLDs for use as biocatalysts or biosensing components. Relative to other HLDs, DhIA possesses a uniquely high activity towards 1,2-dichloroethane (**37**). DhIA is naturally produced by 1,2-dichloroethane-degrading microorganisms [23], which have already been used successfully in a full-scale groundwater treatment plant [5]. DbjA is the only characterized HLD to exhibit significant activity towards the persistent compound 1,2-dichloropropane (**67**) (Supplementary Table S3). DhaA and DbjA can also convert the highly toxic environmental pollutant 1,2,3-trichloropropane (**80**), albeit at a slow rate (Supplementary Table S3). Enzymes having at least some activity towards a target compound can be further optimized by protein engineering [9]; the catalytic efficiency of DhaA towards 1,2,3-trichloropropane (**80**) has recently been improved by a factor of 26 by means of directed evolution [11], resulting in an efficient catalyst for biotechnological applications.

Functional and evolutionary classifications of HLDs

To examine the similarities and differences in the substrate specificities of the wild-type and mutant enzymes, the raw data were subjected to a transformation to suppress the obfuscating effects of the different absolute activities of individual enzymes. The wild-type HLDs were divided into four SSGs; SSG-I consisted of DbjA, DhaA, DhIA and LinB. It has been suggested that these enzymes belong in different specificity classes in previous studies [26, 49, 50]; our analysis suggests that the previously observed differences between the members of this group are relatively insignificant if one considers a broader range of enzymes and substrates. The common feature of the SSG-I enzymes is their catalytic robustness. In particular, the SSG-I members exhibited measurable activity towards most of the chlorinated compounds, suggesting that these HLDs can effectively stabilize a chloride leaving group. Kinetic analysis with 1-chlorobutane (**4**) revealed that SSG-I enzymes also exhibit higher turnover numbers than other tested HLDs (Supplementary Table S4). The highest specific activities observed with this group of enzymes were obtained with brominated ethanes and propanes. This is consistent with earlier studies, which showed that dibrominated compounds having low lowest unoccupied molecular orbital energies are efficiently and rapidly dehalogenated by HLDs [32, 51, 52].

Three SSG-I members, DbjA, DhaA and LinB, belong to the HLD-II subfamily. Notably, the substrate specificity profiles of these three enzymes are more similar to that of DhIA from the HLD-I subfamily than to those of the other three HLD-II members in the dataset, namely DmbA (which we classified into SSG-II), DatA, and DbeA (both classified into SSG-IV). The fact that DbeA and DmbA enzymes were not classified into SSG-I alongside their close evolutionary relatives DbjA and LinB demonstrates that a close evolutionary relationship

between two HLDs does not necessarily imply that they will have similar activity and specificity profiles. At the time of writing, only 14 members of the HLD family have been experimentally characterized and shown to be dehalogenation-competent. However, a recent sequence database search and bioinformatics analysis identified more than 200 putative members of this family (E. C., unpublished work).

These results indicate that it is not possible to predict the substrate specificity of putative HLDs solely on the basis of sequence similarities with experimentally characterized family members. This observation is in correspondence with previous observations that a subtle change in the key active-site residues can lead to modulation or even a switch of enzyme substrate specificity [53, 54]. Several mutants of LinB carrying a single point substitution at the opening of the access tunnel have been reported to have modified activities towards various halogenated substrates [13]. Similarly, a few mutations in the specificity-determining regions of HLDs have led to changes in substrate specificity during laboratory [55] and natural [56] evolution. We speculate that the incongruence between the phylogenetic and functional classifications of HLDs reflects a certain 'plasticity' of these enzymes. This would enable the host organisms to quickly evolve the capacity to convert novel substrates, which is essential for the adaptation of bacteria to various living environments.

Statistical analysis of the merged data set of the wild-type and mutant dehalogenases demonstrated that developed PCA model can be used for classification of characterized members of the HLD family (Supplementary Figure S4). The prediction of the specificity group can be made for any newly isolated HLD with determined specificity profile using the protocols described in the Methods section.

Structural determinants of substrate specificity in HLDs

We have previously proposed that the substrate specificity of individual HLDs is influenced by the architecture of their active-site cavities and the anatomy of their access tunnels [11, 13, 17, 51, 57, 58]. The active-site cavities of DbjA [4] and LinB [17] are the largest of all HLDs whose structure is known, and both enzymes do indeed perform well with bigger substrates such as monohalogenated butanes, pentanes, hexanes, cyclopentanes and cyclohexanes. The large active sites are also consistent with these enzymes' very broad substrate specificity. The cavities of DhaA [16] and DmbA [18] are smaller and therefore cannot so readily accommodate these larger substrates. The smallest and most occluded active-site cavity is that of DhIA [15, 59]; it is optimized for its 'natural' substrate, 1,2-dichloroethane (**37**). Notably, this enzyme shows enhanced activity towards other small substrates [59]. The key role of the access tunnels in controlling the specificity of HLDs was strikingly demonstrated in a recent directed evolution experiment which sought to improve the activity of DhaA towards 1,2,3-trichloropropane (**80**) [11]. A DhaA31 mutant carrying five substitutions in its access tunnels was prepared; this increased the occlusion of its active site, restricting the access of water molecules to the active site. In turn, the exclusion of water enhances the stability of the activated complex, enhancing the activity of the mutant towards halogenated ethanes and propanes. However, the mutant also exhibits decreased activity towards longer haloalkanes such as hexanes, presumably due to the steric hindrance between the alkyl chains of substrates and the large hydrophobic residues introduced in the access tunnels.

Comparative analysis of closely related HLDs provided further insight into the structural determinants of their substrate specificity. DmbA exhibits 68% sequence identity with LinB, but despite this these two enzymes were classified into different SSGs. While their catalytic residues are positioned identically [17, 18], there are significant differences in the anatomies of their active-site cavities and main access tunnels, which might be responsible for the observed differences in substrate specificity (Figure 4A). DbeA and DbjA exhibit 71% sequence identity and provide a second example of a pair of closely-related enzymes with

different kinetic properties and substrate specificities (Supplementary Tables S4 and S3). Compared to DbeA, DbjA carries an insertion of nine amino acids between the main and cap domains [4]. A Structural comparison of the two revealed that their active-site cavities and main access tunnels are structurally similar; the main structural difference lies in the conformational behaviour of His139, which is located in close proximity to the insertion (Figure 4B). His139 adopts two alternative conformations in DbjA [4], but only one conformation in DbeA (T. Prudnikova, P. Rezacova, Z. Prokop, T. Mozga, Y. Satō, M. Kutý, Y. Nagata, J. Damborsky, I. Kuta-Smatanova, R. Chaloupkova, unpublished work). The role of the conformational behaviour of His139 in controlling the enzyme's specificity was probed using the deletion mutant DbjA Δ . His139 adopts only one conformation in DbjA Δ , resembling that observed in DbeA (Figure 4C). However, the mutant DbjA Δ retains the substrate specificity of DbjA, demonstrating that His139 does not play an essential role in controlling substrate specificity. This conclusion was further supported by an experiment using two mutants of DbeA: DbeA1 and DbeA2. These mutants were constructed to mimic the active site and the main access tunnel of DbjA (R. Chaloupkova, T. Mozga, Y. Satō, T. Prudnikova, T. Koudelakova, E. Chovancova, P. Rezacova, Y. Nagata, I. Kuta-Smatanova and J. Damborsky, unpublished work). As was the case with DbjA Δ , their substrate specificity profiles were more similar to their 'parent' enzyme, DbeA, than to the target protein DbjA.

Figure 4 here

By comparing wild-type and mutant HLDs, we were able to address an intriguing question – namely, whether it is possible to interconvert the substrate specificity of two HLDs by modifying their active-site cavities and main access tunnels. Even when the mutants had identical active-site and main tunnel residues to those observed in the target enzyme, switches in the mutants' substrate specificity were not detected. The mutants all exhibited similar substrate specificity to their respective 'parent' enzymes, and were classified into the same SSGs as their 'parents' by PCA, indicating that our mutations did not target one or more of the key determinants of HLD substrate specificity. Thus, the interconversion of substrate specificity remains one of the challenges for the rational design of HLDs. In addition to re-engineering of the active site and the main access tunnel, it may also be necessary to modify auxiliary access tunnels or tunnel openings [60], the distribution of charges on the protein's surface [61, 62], protein solvation [14], or protein dynamics [63].

ACKNOWLEDGEMENTS

We gratefully acknowledge Tomas Mozga, Radka Chaloupkova, Tatyana Prudnikova, Pavlina Rezacova and Ivana Kuta-Smatanova for providing their unpublished data for analysis and thank Hana Moskalikova for her help with the measurement of catalytic constants. Financial support from the Ministry of Education, Youth and Sports of the Czech Republic (grants LC06010 to T.K., MSM0021622412 to J.J., MSM0021622413 to J.B. and CZ.1.05/2.1.00/01.0001 to A.F.) and the Grant Agency of the Czech Academy of Sciences (grant IAA401630901 to J.D.) is also gratefully acknowledged.

REFERENCES

- 1 Bornscheuer, U. T. (2005) Trends and challenges in enzyme technology. *Adv. Biochem. Eng. Biotechnol.* **100**, 181-203
- 2 Whisstock, J. C. and Lesk, A. M. (2003) Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.* **36**, 307-340

- 3 Swanson, P. (1999) Dehalogenases applied to industrial-scale biocatalysis. *Curr. Opin. Biotechnol.* **10**, 365-369
- 4 Prokop, Z., Sato, Y., Brezovsky, J., Mozga, T., Chaloupkova, R., Koudelakova, T., Jerabek, P., Stepankova, V., Natsume, R., van Leeuwen, J. G., Janssen, D. B., Florian, J., Nagata, Y., Senda, T. and Damborsky, J. (2010) Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering. *Angew. Chem. Int. Ed.* **49**, 6111-6115
- 5 Stucki, G. and Thuer, M. (1995) Experiences of a large-scale application of 1,2-dichloroethane degrading microorganisms for groundwater treatment. *Environ. Sci. Technol.* **29**, 2339-2345
- 6 Prokop, Z., Oplustil, F., DeFrank, J. and Damborsky, J. (2006) Enzymes fight chemical weapons. *Biotechnol. J.* **1**, 1370-1380
- 7 Bidmanova, S., Chaloupkova, R., Damborsky, J. and Prokop, Z. (2010) Development of an enzymatic fiber-optic biosensor for detection of halogenated hydrocarbons. *Anal. Bioanal. Chem.* **398**, 1891-1898
- 8 Los, G., Encell, L., McDougall, M., Hartzell, D., Karassina, N., Zimprich, C., Wood, M., Learish, R., Ohane, R., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D., Bulleit, R. and Wood, K. (2008) HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373-382
- 9 Bosma, T., Damborsky, J., Stucki, G. and Janssen, D. (2002) Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. *Appl. Environ. Microbiol.* **68**, 3582-3587
- 10 Gray, K., Richardson, T., Kretz, K., Short, J., Bartnek, F., Knowles, R., Kan, L., Swanson, P. and Robertson, D. (2001) Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv. Synth. Catal.* **343**, 607-617
- 11 Pavlova, M., Klvana, M., Prokop, Z., Chaloupkova, R., Banas, P., Otyepka, M., Wade, R. C., Tsuda, M., Nagata, Y. and Damborsky, J. (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.* **5**, 727-733
- 12 Pikkemaat, M. G. and Janssen, D. B. (2002) Generating segmental mutations in haloalkane dehalogenase: a novel part in the directed evolution toolbox. *Nucleic Acids Res.* **30**, E35
- 13 Chaloupkova, R., Sykorova, J., Prokop, Z., Jesenska, A., Monincova, M., Pavlova, M., Tsuda, M., Nagata, Y. and Damborsky, J. (2003) Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel. *J. Biol. Chem.* **278**, 52622-52628

- 14 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. and Damborsky, J. (2009) Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations. *J. Mol. Biol.* **392**, 1339-1356
- 15 Verschueren, K. H., Seljee, F., Rozeboom, H. J., Kalk, K. H. and Dijkstra, B. W. (1993) Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* **363**, 693-698
- 16 Newman, J., Peat, T. S., Richard, R., Kan, L., Swanson, P. E., Affholter, J. A., Holmes, I. H., Schindler, J. F., Unkefer, C. J. and Terwilliger, T. C. (1999) Haloalkane dehalogenases: structure of a *Rhodococcus* enzyme. *Biochemistry* **38**, 16105-16114
- 17 Marek, J., Vevodova, J., Smatanova, I. K., Nagata, Y., Svensson, L. A., Newman, J., Takagi, M. and Damborsky, J. (2000) Crystal structure of the haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26. *Biochemistry* **39**, 14082-14086
- 18 Mazumdar, P. A., Hulecki, J. C., Cherney, M. M., Garen, C. R. and James, M. N. G. (2008) X-ray crystal structure of *Mycobacterium tuberculosis* haloalkane dehalogenase Rv2579. *Biochim. Biophys. Acta* **1784**, 351-362
- 19 Chovancova, E., Kosinski, J., Bujnicki, J. M. and Damborsky, J. (2007) Phylogenetic analysis of haloalkane dehalogenases. *Proteins* **67**, 305-316
- 20 Jesenska, A., Bartos, M., Czernekova, V., Rychlik, I., Pavlik, I. and Damborsky, J. (2002) Cloning and expression of the haloalkane dehalogenase gene *dhmA* from *Mycobacterium avium* N85 and preliminary characterization of DhmA. *Appl. Environ. Microbiol.* **68**, 3724-3730
- 21 Jesenska, A., Pavlova, M., Strouhal, M., Chaloupkova, R., Tesinska, I., Monincova, M., Prokop, Z., Bartos, M., Pavlik, I., Rychlik, I., Möbius, P., Nagata, Y. and Damborsky, J. (2005) Cloning, biochemical properties, and distribution of mycobacterial haloalkane dehalogenases. *Appl. Environ. Microbiol.* **71**, 6736-6745
- 22 Jesenska, A., Monincova, M., Koudelakova, T., Hasan, K., Chaloupkova, R., Prokop, Z., Geerlof, A. and Damborsky, J. (2009) Biochemical characterization of haloalkane dehalogenases DrbA and DmbC, representatives of a novel subfamily. *Appl. Environ. Microbiol.* **75**, 5157-5160
- 23 Keuning, S., Janssen, D. B. and Witholt, B. (1985) Purification and characterization of hydrolytic haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10. *J. Bacteriol.* **163**, 635-639
- 24 Kulakova, A. N., Larkin, M. J. and Kulakov, L. A. (1997) The plasmid-located haloalkane dehalogenase gene from *Rhodococcus rhodochrous* NCIMB 13064. *Microbiology* **143**, 109-115
- 25 Nagata, Y., Hynkova, K., Damborsky, J. and Takagi, M. (1999) Construction and characterization of histidine-tagged haloalkane dehalogenase (LinB) of a new substrate

- class from a gamma-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26. *Protein. Expr. Purif.* **17**, 299-304
- 26 Sato, Y., Monincova, M., Chaloupkova, R., Prokop, Z., Ohtsubo, Y., Minamisawa, K., Tsuda, M., Damborsky, J. and Nagata, Y. (2005) Two rhizobial strains, *Mesorhizobium loti* MAFF303099 and *Bradyrhizobium japonicum* USDA110, encode haloalkane dehalogenases with novel structures and substrate specificities. *Appl. Environ. Microbiol.* **71**, 4372-4379
- 27 Chan, W. Y., Wong, M., Guthrie, J., Savchenko, A. V., Yakunin, A. F., Pai, E. F. and Edwards, E. A. (2010) Sequence- and activity-based screening of microbial genomes for novel dehalogenases. *Microb. Biotechnol.* **3**, 107-120
- 28 Nagata, Y., Nariya, T., Ohtomo, R., Fukuda, M., Yano, K. and Takagi, M. (1993) Cloning and sequencing of a dehalogenase gene encoding an enzyme with hydrolase activity involved in the degradation of gamma-hexachlorocyclohexane in *Pseudomonas paucimobilis*. *J. Bacteriol.* **175**, 6403-6410
- 29 Hasan, K., Fortova, A., Koudelakova, T., Chaloupkova, R., Ishitsuka, M., Nagata, Y., Damborsky, J. and Prokop, Z. Biochemical characteristics of the novel haloalkane dehalogenase DatA isolated from the plant pathogen *Agrobacterium tumefaciens* C58. *Appl. Environ. Microbiol.* doi: 10.1128/AEM.02109-10
- 30 Curragh, H., Flynn, O., Larkin, M. J., Stafford, T. M., Hamilton, J. T. and Harper, D. B. (1994) Haloalkane degradation and assimilation by *Rhodococcus rhodochrous* NCIMB 13064. *Microbiology* **140**, 1433-1442
- 31 Damborsky, J., Rorije, E., Jesenska, A., Nagata, Y., Klopman, G. and Peijnenburg, W. J. (2001) Structure-specificity relationships for haloalkane dehalogenases. *Environ. Toxicol. Chem.* **20**, 2681-2689
- 32 Schanstra, J. P., Kingma, J. and Janssen, D. B. (1996) Specificity and kinetics of haloalkane dehalogenase. *J. Biol. Chem.* **271**, 14747-14753
- 33 Iwasaki, I., Utsumi, S. and Ozawa, T. (1952) New colorimetric determination of chloride using mercuric thiocyanate and ferric ion. *Bull. Chem. Soc. Jpn.* **25**, 226
- 34 Eriksson, L., Antti, H., Gottfries, J., Holmes, E., Johansson, E., Lindgren, F., Long, I., Lundstedt, T., Trygg, J. and Wold, S. (2004) Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics. *Anal. Bioanal. Chem.* **380**, 419-429
- 35 Wold, S., Esbensen, K. and Geladi, P. (1987) Principal component analysis. *Chemometr. Intell. Lab.* **2**, 37-52
- 36 Eriksson, L., Johansson, E., Kettaneh-Wold, N. and Wold, S. (1999) Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS), pp. 43-68, Umetrics, Umea


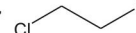
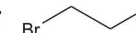
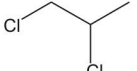
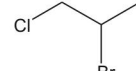
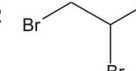


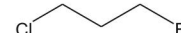


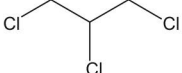
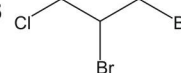
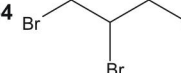
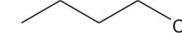

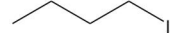



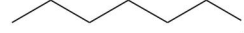




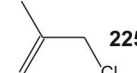
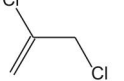
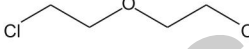
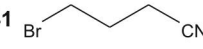
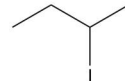
- 37 Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425
- 38 Drummond, A. and Strimmer, K. (2001) PAL: an object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics* **17**, 662-663
- 39 Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Yaschenko, E. and Ye, J. (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **37**, D5-D15
- 40 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402
- 41 Frickey, T. and Lupas, A. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702-3704
- 42 Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797
- 43 Hall, T. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95-98
- 44 Abascal, F., Zardoya, R. and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105
- 45 Whelan, S. and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691-699
- 46 Hornik, K. and Leisch, F. (2005) R version 2.1.0. *Comp. Stat.* **20**, 197-202
- 47 van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. and van der Werf, M. J. (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142
- 48 Gribble, G. W. (2003) The diversity of naturally produced organohalogens. *Chemosphere* **52**, 289-297
- 49 Damborsky, J., Nyandoroh, M. G., Nemeč, M., Holoubek, I., Bull, A. T. and Hardman, D. J. (1997) Some biochemical properties and the classification of a range of bacterial haloalkane dehalogenases. *Biotechnol. Appl. Biochem.* **26**, 19-25
- 50 Nagata, Y., Miyauchi, K., Damborsky, J., Manova, K., Ansorgova, A. and Takagi, M. (1997) Purification and characterization of a haloalkane dehalogenase of a new substrate

- class from a gamma-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26. Appl. Environ. Microbiol. **63**, 3707-3710
- 51 Kmunicek, J., Hynkova, K., Jedlicka, T., Nagata, Y., Negri, A., Gago, F., Wade, R. C. and Damborsky, J. (2005) Quantitative analysis of substrate specificity of haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26. Biochemistry **44**, 3390-3401
- 52 Bosma, T., Pikkemaat, M. G., Kingma, J., Dijk, J. and Janssen, D. B. (2003) Steady-state and pre-steady-state kinetic analysis of halopropane conversion by a *Rhodococcus* haloalkane dehalogenase. Biochemistry **42**, 8047-8053
- 53 Penning, T. M. and Jez, J. M. (2001) Enzyme redesign. Chem. Rev. **101**, 3027-3046
- 54 Morley, K. L. and Kazlauskas, R. J. (2005) Improving enzyme properties: when are closer mutations better? Trends Biotechnol. **23**, 231-237
- 55 Pries, F., van den Wijngaard, A. J., Bos, R., Pentenga, M. and Janssen, D. B. (1994) The role of spontaneous cap domain mutations in haloalkane dehalogenase specificity and evolution. J. Biol. Chem. **269**, 17490-17494
- 56 Ito, M., Prokop, Z., Klvana, M., Otsubo, Y., Tsuda, M., Damborsky, J. and Nagata, Y. (2007) Degradation of beta-hexachlorocyclohexane by haloalkane dehalogenase LinB from gamma-hexachlorocyclohexane-utilizing bacterium *Sphingobium* sp. MI1205. Arch. Microbiol. **188**, 313-325
- 57 Silberstein, M., Damborsky, J. and Vajda, S. (2007) Exploring the binding sites of the haloalkane dehalogenase Dh1A from *Xanthobacter autotrophicus* GJ10. Biochemistry **46**, 9239-9249
- 58 Damborsky, J. and Brezovsky, J. (2009) Computational tools for designing and engineering biocatalysts. Curr. Opin. Chem. Biol. **13**, 26-34
- 59 Kmunicek, J., Luengo, S., Gago, F., Ortiz, A. R., Wade, R. C. and Damborsky, J. (2001) Comparative binding energy analysis of the substrate specificity of haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10. Biochemistry **40**, 8905-8917
- 60 Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J. and Damborsky, J. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinformatics **7**, 316
- 61 Russell, A. J. and Fersht, A. R. (1987) Rational modification of enzyme catalysis by engineering surface charge. Nature **328**, 496-500
- 62 de Kreijl, A., van den Burg, B., Venema, G., Vriend, G., Eijssink, V. G. H. and Nielsen, J. E. (2002) The effects of modifying the surface charge on the catalytic activity of a thermolysin-like protease. J. Biol. Chem. **277**, 15432-15438
- 63 Ota, N. and Agard, D. A. (2001) Enzyme specificity under dynamic control II: Principal component analysis of α -lytic protease using global and local solvent boundary conditions. Prot. Sci. **10**, 1403-1414

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20101405

Accepted Manuscript

Table 1 The set of thirty substrates used to test the substrate specificity of the haloalkane dehalogenases.

	Chlorinated	Chlorinated & brominated	Brominated	Iodinated
Ethanes	37 	137 	47 	
Propanes	67 	76 	72 	28 
	38 	52 	48 	54 
	80 	155 	154 	
Butanes	4 		18 	29 
Pentanes	40 			
Hexanes	6 		20 	31 
Cyclic alkanes	138  115 		117  119 	
Propenes	209  225 			
Miscellaneous alkanes	111 		141 	64 

Accepted Manuscript

Table 2 Characteristics of the haloalkane dehalogenase substrate specificity groups (SSGs).

'Substrates characterizing SSG' are compounds important for defining the specified SSG. 'Preferred substrates' are compounds converted at high relative rates by the members of the specified SSG under the described conditions. Key: n.a. – not applicable, ↑ - high relative activity, ↓ - low relative activity, × - no activity under the used conditions.

SSG: Enzyme	Converted substrates	Substrates characterizing SSG	Preferred substrates
SSG-I: DbjA DhaA DhlA LinB	26 - 29	n.a.	1,2-dibromoethane (47) 1,3-dibromopropane (48) 1-bromo-3-chloropropane (52) 1-bromo-2-chloroethane (137) 4-bromobutanenitrile (141)
SSG-II: DmbA	23	↑ 2-iodobutane (64) ↑ 1-chloro-2-(2-chloroethoxy)ethane (111) ↑ chlorocyclopentane (138) ↓ 1-bromohexane (20) ↓ 1-iodohexane (31) ↓ 1,2-dibromopropane (72) × 1,3-diiodopropane (54)	2-iodobutane (64) 1-chloro-2-(2-chloroethoxy)ethane (111)
SSG-III: DrbA	17	↑ 1-chlorobutane (4) × 1,3-dibromopropane (48) × 1-bromo-3-chloropropane (52) × 1,2,3-tribromopropane (154)	1-chlorobutane (4) 1-iodobutane (29) 1,3-diiodopropane (54) 2,3-dichloroprop-1-ene (225)
SSG-IV: DatA DbeA DmbC	21 - 25	↑ 1-bromobutane (18) ↑ 1,3-dibromopropane (48) ↓ 1-bromo-2-chloroethane (137) ↓ 2,3-dichloroprop-1-en (225)	1-bromobutane (18) 1-iodopropane (28) 1,3-dibromopropane (48) 1-bromo-3-chloropropane (52)

Accepted Manuscript

FIGURE CAPTIONS

Figure 1 Substrate specificity profiles of haloalkane dehalogenases.

Transformation of the primary data suppressed the differences in the enzymes' absolute activities and allowed comparison of the HLDs' substrate specificity profiles. Colour coding corresponds to individual substrate specificity groups: red corresponds to SSG-I, yellow to SSG-II, green to SSG-III, and blue to SSG-IV. The values higher than 0.2 are depicted by truncated cones for the clarity.

Figure 2 Principal Component Analysis of the transformed specific activity data.

A colour version of this figure is available as Supplementary Figure S3. (A) Clustering of HLDs based on their activity with tested substrates in t_1/t_2 score plots. t_1 and t_2 are the PC scores of individual HLDs in the plane defined by PC1 and PC2. Depicted score plots are the two-dimensional windows into thirty-dimensional space explaining 47% and 46% data variance in the transformed datasets, respectively. Wild-type HLDs (*left*) were divided into four substrate specificity groups: SSG-I, SSG-II, SSG-III and SSG-IV. Clusters are depicted by grey ovals. Analysis of merged data for both wild-type and mutant enzymes (*right*) reveals that the specific activities of mutant HLDs were altered, but the mutants were clustered within the same specificity groups as their 'parent' enzymes. Arrows indicate changes in substrate specificity caused by mutations. (B) Distribution of variables in p_1/p_2 loading plots corresponding to t_1/t_2 in (A). p_1 and p_2 are the PC loadings of corresponding variables (relative activities with particular substrates). The values quantify contributions of individual variables to given PCs. The variables contributing similar information are grouped together. The variables localized further from the origin possess the stronger effect on PC than the variables localized closer to the origin of the plot. Comparison of a score and a loading plot enables identification of variables (relative activities with particular substrates) responsible for clustering of enzymes.

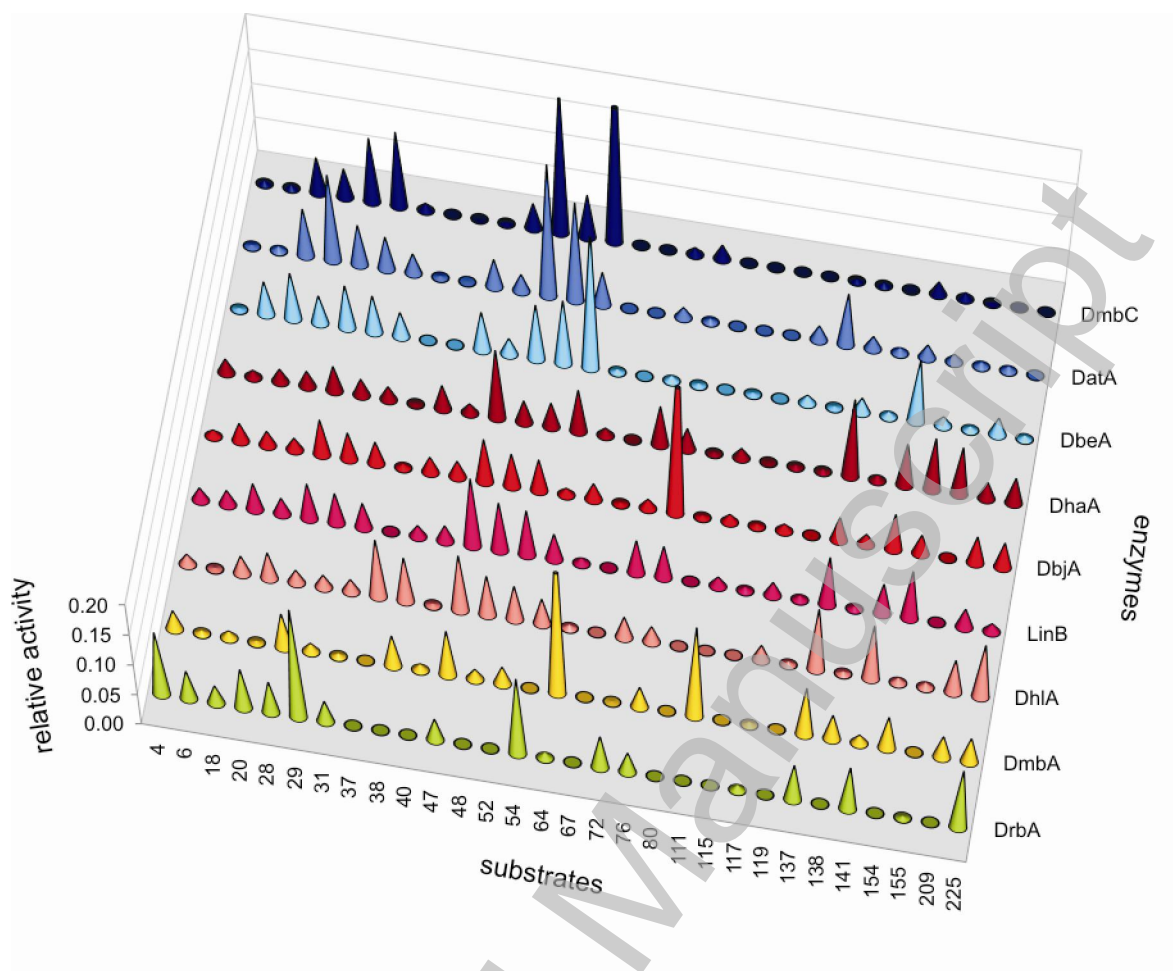
Figure 3 Comparison of evolutionary and substrate specificity relationships of nine wild-type haloalkane dehalogenases.

A colour version of this figure is available as Supplementary figure S5. (A) The phylogenetic tree indicates that the haloalkane dehalogenases are subdivided into three phylogenetic subfamilies: HLD-I, HLD-II and HLD-III. Individual subfamilies are marked by a black line. (B) The dendrogram indicates that the HLDs can be classified into four substrate specificity groups: SSG-I, SSG-II, SSG-III and SSG-IV.

Figure 4 Stereoview of the structural comparison of the active-site cavities and main access tunnels of closely related haloalkane dehalogenases.

The residues making-up the active-site and the tunnel are shown. A nucleophile is labelled by a star. Aligned residues with distinct structure or orientation are depicted in sticks. Tunnels are represented by grey surface. (A) Comparison of DmbA (blue) and LinB (green) – the orientation of catalytic residues is identical, while three tunnel residues possess significant differences in the structure. (B) Comparison of DbeA (blue) and DbjA (green) – the active-site cavities and the main access tunnel are structurally similar, but conformational behaviour of His139 differs for individual cases; flexible His139 of DbjA adopts two different orientations). (C) Comparison of DbeA (blue) and DbjA Δ (green) – the active-site cavity and the main access tunnel of DbjA Δ resembles DbeA; His139 adopts only one conformation in both structures.

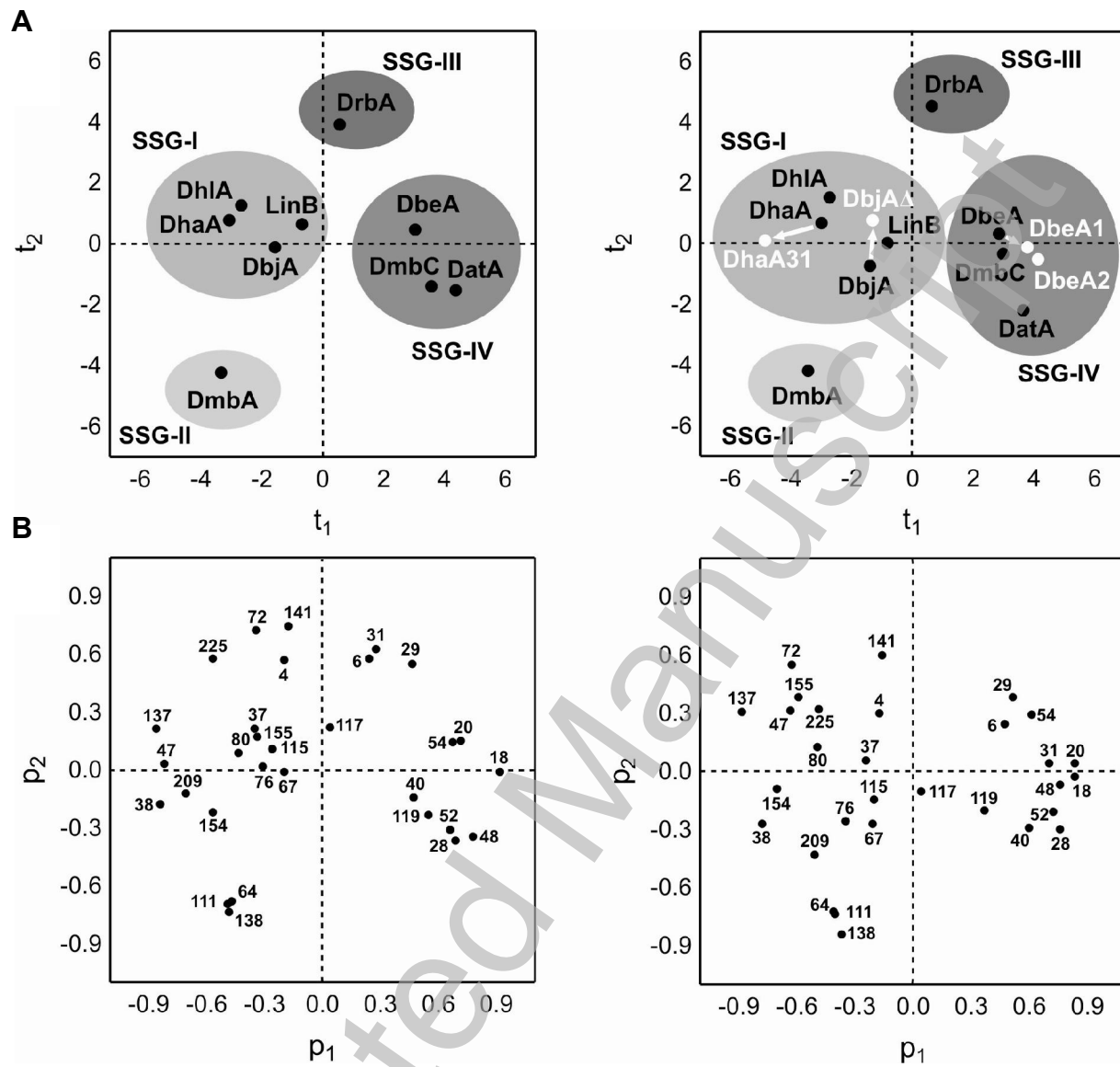
Figure 1



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20101405

Accepted Manuscript

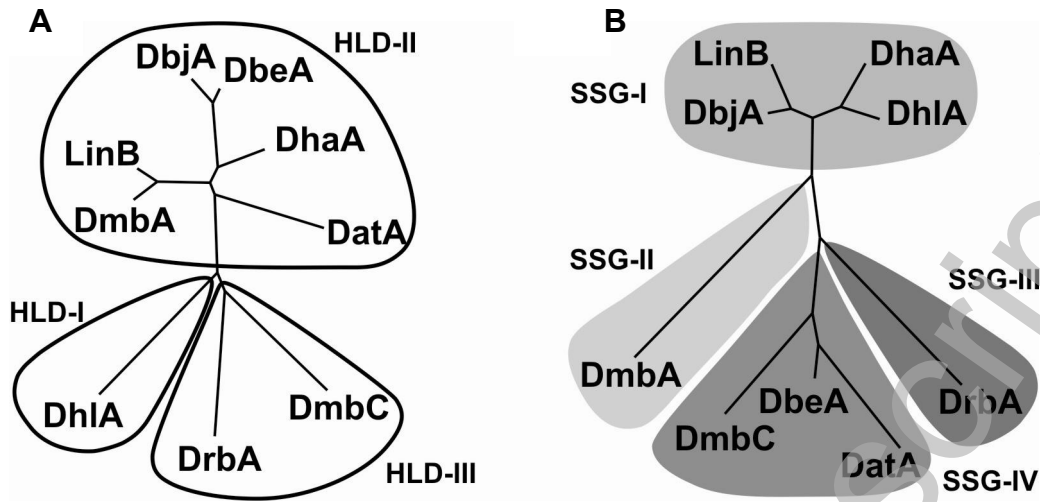
Figure 2



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20101405

Accepted Manuscript

Figure 3



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20101405

Accepted Manuscript

Figure 4

