PROTEINS: Structure, Function, and Bioinformatics 67:305–316 (2007)

Phylogenetic Analysis of Haloalkane Dehalogenases

Eva Chovancová,[1] Jan Kosinski,[2] Janusz M. Bujnicki,[2] and Jiří Damborský[1]*
[1]*Loschmidt Laboratories, Faculty of Science, Masaryk University, Brno, Czech Republic*
[2]*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, Poland*

ABSTRACT **Haloalkane dehalogenases (HLDs) are enzymes that catalyze the cleavage of carbon–halogen bonds by a hydrolytic mechanism. Although comparative biochemical analyses have been published, no classification system has been proposed for HLDs, to date, that reconciles their phylogenetic and functional relationships. In the study presented here, we have analyzed all sequences and structures of genuine HLDs and their homologs detectable by database searches. Phylogenetic analyses revealed that the HLD family can be divided into three subfamilies denoted HLD-I, HLD-II, and HLD-III, of which HLD-I and HLD-III are predicted to be sister-groups. A mismatch between the HLD protein tree and the tree of species, as well as the presence of more than one HLD gene in a few genomes, suggest that horizontal gene transfers, and perhaps also multiple gene duplications and losses have been involved in the evolution of this family. Most of the biochemically characterized HLDs are found in the HLD-II subfamily. The dehalogenating activity of two members of the newly identified HLD-III subfamily has only recently been confirmed, in a study motivated by this phylogenetic analysis. A novel type of the catalytic pentad (Asp-His-Asp+Asn-Trp) was predicted for members of the HLD-III subfamily. Calculation of the evolutionary rates and lineage-specific innovations revealed a common conserved core as well as a set of residues that characterizes each HLD subfamily. The N-terminal part of the cap domain is one of the most variable regions within the whole family as well as within individual subfamilies, and serves as a preferential site for the location of relatively long insertions. The highest variability of discrete sites was observed among residues that are structural components of the access channels. Mutations at these sites modify the anatomy of the channels, which are important for the exchange of ligands between the buried active site and the bulk solvent, thus creating a structural basis for the molecular evolution of new substrate specificities. Our analysis sheds light on the evolutionary history of HLDs and provides a structural framework for designing enzymes with new specificities. Proteins 2007;67:305–316.** © **2007 Wiley-Liss, Inc.**

Key words: conservation; evolution; haloalkane dehalogenase; phylogenetic analysis; sequence; structure

INTRODUCTION

Haloalkane dehalogenases (HLDs) are enzymes that catalyze the hydrolytic cleavage of carbon–halogen bonds, yielding a primary alcohol, a proton, and a halide (EC 3.8.1.5). To date, HLD activity has been experimentally confirmed in only about a dozen different proteins. HLDs have broad substrate specificities, which nevertheless differ between individual members of the family. These enzymes are able to convert a wide spectrum of substrates including halogenated alkanes, cycloalkanes, alkenes, ethers, alcohols, ketones, and cyclic dienes. Several HLDs have been shown to be involved in biodegradation pathways of important environmental pollutants.[1–10] Furthermore, these enzymes have also been found to be present in pathogenic bacteria[11,12] and rhizobial bacteria,[13] where their function remains unknown.

Structurally HLDs belong to the α/β-hydrolase superfamily.[14–17] The three-dimensional structure of three HLDs has been solved, revealing two common domains: the α/β-hydrolase core domain (which is conserved in members of the α/β-hydrolase superfamily) and a helical cap domain. The α/β-hydrolase fold is composed of an eight-stranded mostly parallel β-sheet flanked by α-helices, and serves as a scaffold for the main catalytic residues. The cap domain composed of a few helices inserted into the catalytic domain, usually C-terminally to β-strand 6, has been found in the structure of many α/β-hydrolases (not only HLDs) and is known to influence the substrate specificity of these enzymes. The active site cavity is located between the main domain and the cap domain. A catalytic pentad of residues that is essential for hydrolysis has been identified and it includes Asp (nucleophile), His (base), Asp or Glu (catalytic acid), and two halide-stabilizing residues, Trp and Trp or Asn.[18,19]

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: Centre for Biocatalysis and Biotransformation; Grant number: LC06010; Grant sponsor: INCHEMBIOL (Czech Ministry of Education); Grant number: MSM0021622412; Grant sponsor: KONTAKT; Grant number: CZ-PL 25; Grant sponsor: EU FP5; Grant number: QLK6-CT-2002-90363; Grant sponsors: EMBO&HHMI (Young Investigator Program).

*Correspondence to: Jiří Damborský, Loschmidt Laboratories, Faculty of Science, Masaryk University, Kamenice 5/A4, 625 00 Brno, Czech Republic. E-mail: jiri@chemi.muni.cz

Received 10 August 2006; Revised 23 September 2006; Accepted 23 October 2006

Published online 12 February 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21313

In addition to the experimentally characterized HLDs, many other proteins have been suggested to belong to this protein family, based mainly on sequence similarities.[20] However, it is unclear whether they represent true members of the family in either an evolutionary or functional sense. The aim of this study was to analyze the phylogeny of the HLD family to establish relationships among its individual members, delineate major lineages, and infer the evolutionary history of HLDs. Knowledge of the evolutionary history of HLDs and other α/β-hydrolase families should help attempts to elucidate their structure–function relationships, in particular the development of currently observed enzymatic activities based on a common structural scaffold. The evolutionary classification scheme should also serve as a useful platform to identify and classify new family members and to guide predictions concerning their catalytic, biochemical, and structural properties. It should be emphasized that cloning, expression, and the biochemical characterization of new enzymes identified through the systematic analysis of the HLD family may lead to the discovery of biocatalysts with novel characteristics that are suitable for practical applications. Our analysis could also be used to guide protein design and to engineer proteins with new dehalogenating activities for industrial purposes.

MATERIALS AND METHODS

Comparative Sequence Analysis

Sequences of individual datasets were clustered using CLANS (CLuster ANalysis of Sequences), a Java utility that applies a version of the Fruchterman–Reingold graph layout algorithm.[21] CLANS uses the *P*-values of high-scoring segment pairs obtained from an $N \times N$ BLAST search, to compute attractive and repulsive forces between each sequence pair in a user-defined dataset. A three-dimensional representation is obtained by randomly seeding sequences in space. The sequences are then moved within this environment according to the force vectors resulting from all pairwise interactions and the process is repeated to convergence. Default parameters and varying *P*-value thresholds were used in our analysis. The initial multiple sequence alignments of selected sequences were performed by MUSCLE v3.5.[22] Output alignments were refined manually using the BioEdit v7.0.1 sequence editor[23] to minimize gaps, particularly in the regions of regular secondary structure known from the experimentally solved three-dimensional structures and predicted by bioinformatic methods (see below). Partial or engineered sequences, sequences with incomplete catalytic triad, as well as poorly aligned regions of proteins lacking sufficient conservation, were excluded from further analyses.

Phylogenetic Analysis

Multiple sequence alignments were used for the selection of suitable evolutionary models and parameters by PROTTEST[24] and then for phylogenetic reconstructions by the maximum likelihood (ML) and neighbor-joining (NJ) methods. The ML analysis was performed by PHYML[25] using the WAG model of amino acid substitution[26] and was based on the preliminary NJ tree generated by BIONJ.[27] Distance matrices for the NJ inferences[28] were generated by the MLDIST program of the VANILLA v1.2 package[29] according to the WAG model. Confidence levels of output trees were estimated by bootstrapping the data 1000 times. The resulting phylogenetic trees were rooted either by introducing outgroup sequences or by midpoint rooting. In the case of the whole HLD family, three different outgroups identified by cluster analysis were used, and thus phylogenetic trees were calculated for three different datasets. Each dataset included HLD sequences and sequences from one of the outgroups. Four-cluster likelihood mapping analysis[30] implemented in the TREE-PUZZLE v5.2 package[31] and four-cluster analysis[32] by the PHYLTEST program[33] were used to test the tree topologies obtained from the phylogenetic analyses.

Analysis of Residue Conservation

Multiple sequence alignment was used to estimate the level of conservation of individual sites in HLDs. Normalized evolutionary rates for each amino acid site of the alignment were calculated by the CONSURF 3.0 server[34] according to the WAG model of evolution.[26] The evolutionary rates were further used to identify regions with significantly higher or lower level of conservation. For this purpose, a statistical test was performed by a simulation in which the mean conservation of each defined region of the sequence was compared with the mean conservation of 10,000,000 different randomly generated sets of residues of the same size. Regions ranging in size from 2 to 9 residues were analyzed. Two null hypotheses were tested for each such region: the mean evolutionary rate of the analyzed region consisting of n residues is either not higher (first hypothesis) or not lower (second hypothesis) than the mean evolutionary rate of the random set of n residues drawn without replacement from the set of all residues. Subsequently, for each region, a *P*-value expressing the support for the null hypothesis was calculated from the equation: $P = t_c/t$, where t_c indicates the number of times that the mean evolutionary rate of the analyzed region was not higher (or not lower) than the mean of the random set, and t indicates the total number of iterations (here 10,000,000). An arbitrary selected cut-off of 5% was used. If the *P*-value was lower, the null hypothesis was rejected and the given region was regarded as significantly less (or more) conserved than could be expected by chance. Regions of different lengths were analyzed and a majority consensus was mapped onto the protein surface. These analyses were performed both for the entire HLD family and, separately, for each subfamily.

Homology Modeling

Secondary structure prediction and tertiary fold-recognition was carried out via the GeneSilico meta-server gateway.[35] Secondary structure was predicted using PSIPRED,[36] PROFsec,[37] PROF,[38] SABLE,[39] JNET,[40] JUFO,[41] and SAM-T02.[42] Solvent accessibility for individ-

ual residues was predicted with SABLE[39] and JPRED.[43] Fold-recognition analysis (alignment of the query sequence to known protein structures) was carried out using FFAS03,[44] SAM-T02,[42] 3DPSSM,[45] BIOINBGU,[46] FUGUE,[47] mGenTHREADER,[48] and SPARKS.[49] The fold-recognition alignments reported by these methods were compared, evaluated, and ranked by the Pcons server.[50] Accordingly, fold-recognition alignments to the structures of highly scored templates were used as starting points for homology modeling using the FRankenstein's monster approach,[51] as described previously.[52] The first set of models were built with MODELLER,[53] based on unrefined FR alignments. The quality of local structure in these preliminary models was assessed by VERIFY3D[54] via the COLORADO3D server.[55] All these models were superimposed and a hybrid model was constructed from fragments conserved in more than 50% of models, while the nonconsensus regions were built from fragments with the highest local VERIFY3D scores. The hybrid model was not refined directly, but superimposed onto the template structures to recreate the sequence alignment, which was then used to build a new model. This new model was re-evaluated using VERIFY3D to identify segments of secondary structure with poor scores (segments that exhibited consensus in the first step or good scores in the second step were not modified in subsequent steps). For each of the nonconsensus and poorly scored regions, a number of alternative models were built by locally shifting target-template alignments. The models were evaluated again and the best scoring segments were recombined and then the whole procedure of structural recombination, regeneration, and modification of alignments, model building, and evaluation was iterated until the score could not be significantly improved.

RESULTS

Sequence Database Searches and Clustering of α/β-Hydrolases and Haloalkane Dehalogenases

Sequences of α/β-hydrolases for which experimentally solved three-dimensional structures are available were obtained from SCOP.[56] In addition, 14 putative α/β-hydrolases were identified by DALI[57] and Fatcat[58] searches of the protein data bank (PDB)[59] and added to the dataset. Redundant sequences were discarded and the final dataset included 115 sequences of α/β-hydrolases with known three-dimensional structure (Supplementary Table SI). Sequences of this dataset were clustered using CLANS to identify α/β-hydrolase families that are closely related to HLDs and to separate them confidently from other α/β-hydrolases (data not shown). The sequences of HLDs and their closest homologs with a known structure were used as queries (Supplementary Table SI) for the PSI-BLAST searches[60] of the nr database run until convergence with an *e*-value threshold of 10^{-10}, to identify all members with no known structure. Sequences with more than 90% identity were removed using the ExPASy tool for decreasing redundancy (http://www.expasy.org/tools/redundancy/), yielding a final set of 3442 proteins.

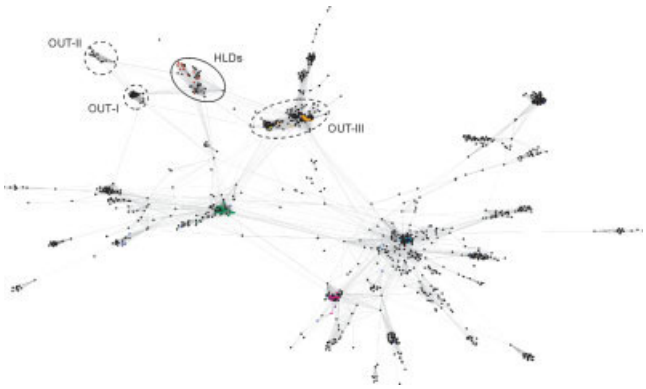


Fig. 1. Cluster analysis of a subset of sequences obtained from PSI-BLAST searches, performed at a cutoff *P*-value of 10^{-25}. Biochemically characterized proteins include haloalkane dehalogenases (HLDs), cytosolic epoxide hydrolases (EPHX), fluoroacetate dehalogenases (FDH), perhydrolases (PH), carbon–carbon bond hydrolases (C−C), and various carboxylic ester hydrolases (EST). Sequences of outgroup-I (OUT-I), outgroup-II (OUT-II), and outgroup-III (OUT-III) were used for rooting haloalkane dehalogenase phylogenetic trees. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Cluster analysis of the 3442 sequences of α/β-hydrolases with sequence similarity to HLDs was carried out with CLANS using the stringent *P*-value threshold of 10^{-25}. This revealed the subdivision of this set into several clusters. Haloalkane dehalogenases were localized toward the edge of the largest cluster composed of over 1300 sequences. Biochemically characterized proteins of this cluster include HLDs, various carboxylic ester hydrolases [carboxylesterases, arylesterases, methylesterases, lipases, enol-lactone hydrolases, dihydrocoumarin hydrolases, poly(3-hydroxyalkanoate) depolymerases], fluoroacetate dehalogenases, cytosolic epoxide hydrolases, luciferases, perhydrolases, and carbon–carbon bond hydrolases (see Fig. 1). Inter-relationships among all sequences within this cluster were investigated by varying the *P*-value threshold for "attraction" between individual sequences. The final set, considered hereafter as the HLD family, comprised 44 sequences. Three clusters were finally identified as being the most closely related to HLDs: two families currently lacking experimentally characterized proteins (outgroup-I and outgroup-II) and the family of cytosolic epoxide hydrolases and fluoroacetate dehalogenases (outgroup-III). These three families were later used as alternative outgroups for rooting the phylogenetic tree of the HLD family.

Phylogenetic Analysis of Haloalkane Dehalogenases

The phylogenetic trees of HLDs were inferred from the maximum likelihood and neighbor-joining analyses (see Materials and Methods section for details). The topology of the trees agreed with the results of clustering, implying that the HLD family should be subdivided into three main subfamilies, termed as HLD-I, HLD-II, and HLD-III. In analyses employing three different outgroups, the

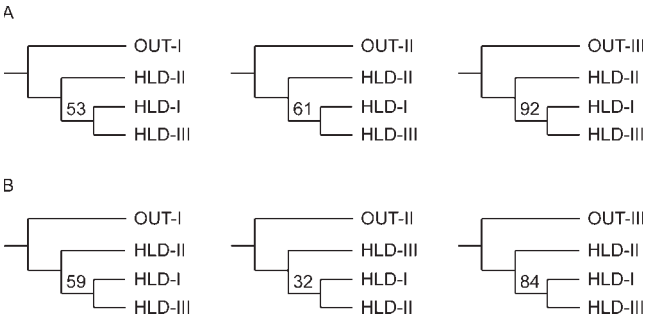


Fig. 2. Results of the outgroup analysis represented by schema of HLD phylogenetic trees indicating inter-relationships among individual HLD subfamilies (HLD-I, HLD-II, and HLD-III). Outgroup analysis was performed using three alternative outgroups (OUT-I, OUT-II, and OUT-III) and phylogenetic trees were calculated by neighbor-joining (**A**) and maximum-likelihood (**B**) methods. Numbers above branches indicate bootstrap support values for given sister-group relationships.

root of the tree was placed within the branch connecting the HLD-II subfamily with the rest of the family (see Fig. 2). Similar results were obtained by midpoint-rooting of the tree. Four-cluster likelihood mapping employing HLD-I, HLD-II, and HLD-III and either of the outgroup families provided further support for this scenario and revealed preferential grouping of the HLD-I and HLD-III subfamilies. The branching pattern of HLD-II and HLD-III grouped together against HLD-I obtained only low support. However, it was not possible to completely rule out a third hypothesis suggesting a sister-group relationship between the HLD-I and HLD-II subfamilies. The results of four-cluster analysis agreed with the four-cluster likelihood mapping in that the topologies corresponding to sister-group relationships of HLD-I with HLD-III, or HLD-I with HLD-II, were not significantly different. Thus, both alternative positions of the root are indicated in the phylogenetic tree of HLDs (see Fig. 3).

All three subfamilies include experimentally characterized HLDs. The HLD-I subfamily segregates into two subgroups. This subdivision is also apparent from the sequence alignment (Supplementary Figure S1). Subgroup IA includes the experimentally confirmed haloalkane dehalogenase, DhlA, from *Xanthobacter autotrophicus*,[1] while subgroup IB is represented by the mycobacterial dehalogenases DmbB from *Mycobacterium bovis* and *M. tuberculosis*[12] and DhmA from *M. avium*.[20] The HLD-II subfamily includes the following experimentally characterized haloalkane dehalogenases: LinB from *Sphingobium japonicum*,[61] DmbA from *M. bovis* and *M. tuberculosis*,[12] DmsA from *M. smegmatis* (unpublished data), DhaA from *Rhodococcus sp.*,[7] DatA from *Agrobacterium tumefaciens* (Nagata, personal communication), DbjA from *Bradyrhizobium japonicum*,[13] DmlA from *Mesorhizobium loti*,[13] and surprisingly, an enzyme with luciferase activity. Luciferase from the sea pansy, *Renilla reniformis*,[62] clearly falls into a well-defined cluster together with LinB, DmbA, DmsA, a protein from an environmental sample, and proteins from the purple sea urchin, *Strongylocentrotus purpuratus*. The HLD-III subfamily currently includes two proteins that have been empirically shown to possess low dehalogenating activity, i.e. DmbC from *M. bovis* and *M. tuberculosis,* and DrbA from *Rhodopirellula baltica* (unpublished data). We note that the HLD-III subfamily is not as well-defined as HLD-I and HLD-II. This is primarily due to the uncertain position of DmbC and three putative proteins from *Jannaschia sp.*, *Nocardia farcinica*, and *Burkholderia cenocepacia* within the tree of HLDs. In most of the trees, these proteins group together with the HLD-III subfamily. However, this grouping had relatively low statistical support. Also, the sequence alignment in certain regions indicates significant differences between these proteins and other members of the HLD-III subfamily. The characteristics of all three subfamilies are summarized in Table I.

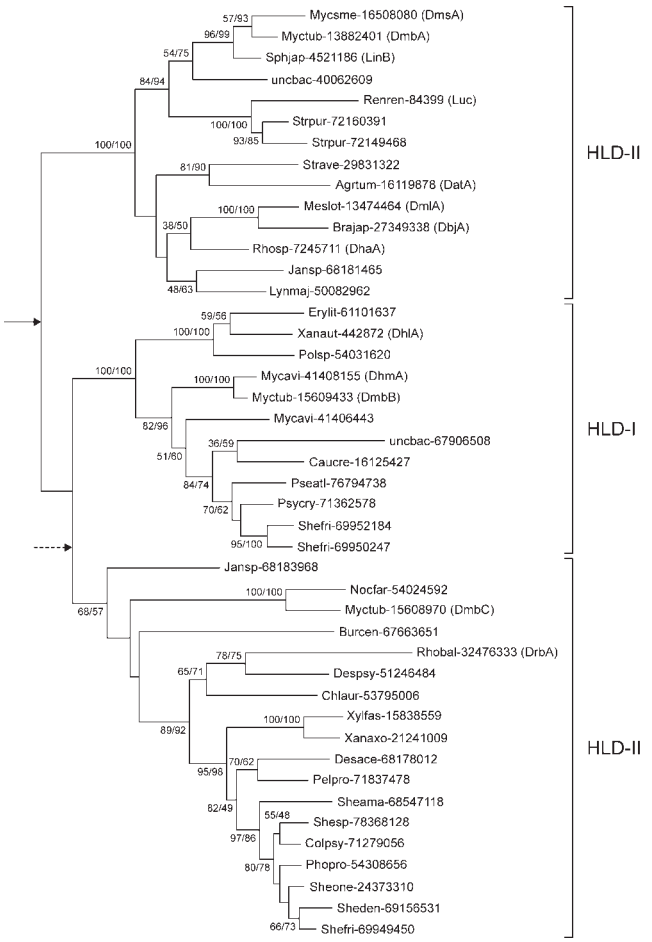


Fig. 3. Phylogenetic tree of haloalkane dehalogenases calculated by the maximum likelihood method. Bootstrap support values obtained from both neighbor-joining and maximum likelihood reconstructions are depicted above branches. Values that were not higher than 50% for any of the methods used are not shown. The tree is rooted based on the results of outgroup analysis and its probable root is indicated by the solid arrow. An alternative root position is indicated by the dotted arrow. The subdivision of the haloalkane dehalogenase family into three subfamilies (HLD-I, HLD-II, and HLD-III) is indicated.

Sequence and Structure Comparisons

Three available experimental structures of HLDs were compared; those of DhlA (PDB ID 1EDE[63]), DhaA (PDB

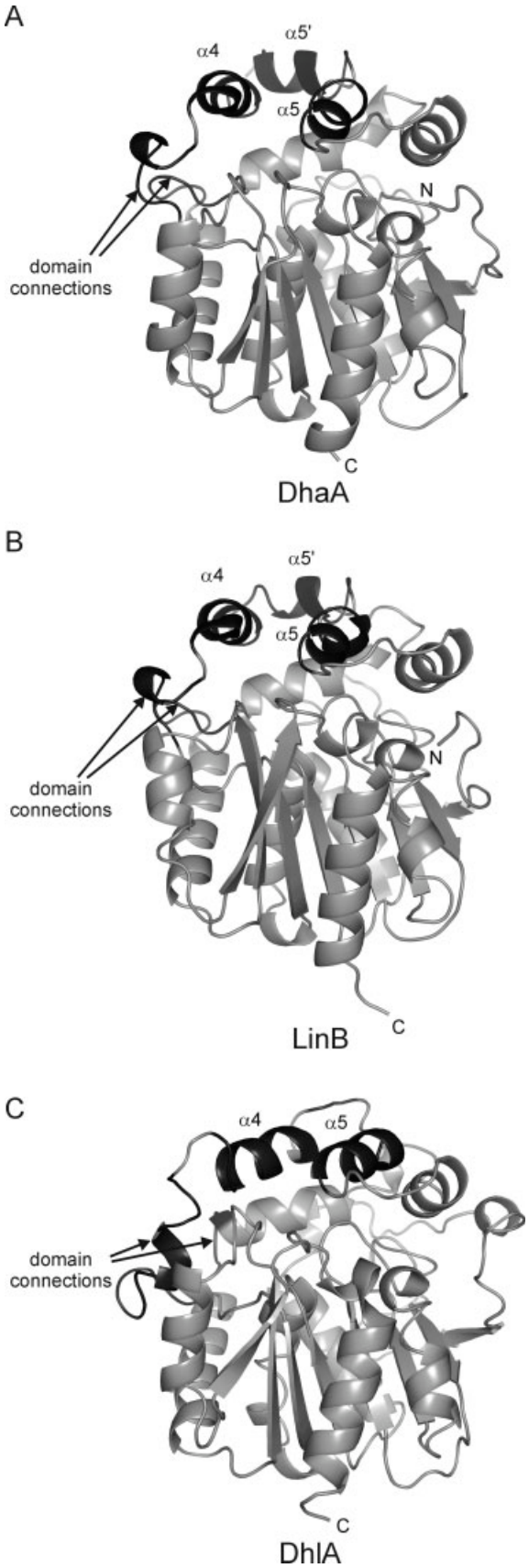
TABLE I. List of Experimentally Characterized Haloalkane Dehalogenases

Subfamily	Protein	Organism	GI number	Substrates	Catal. pentad	Reference
HLD-I	DhlA	*Xanthobacter autotrophicus*	442872	Small, terminally halogenated	Asp-His-Asp+Trp-Trp	1
	DhmA	*Mycobacterium avium*	41408155			12
	DmbB	*Mycobacterium tuberculosis*	15609433			12
HLD-II	LinB	*Sphingobium japonicum*	4521186	Larger, β-substituted	Asp-His-Glu+Asn-Trp	61
	DmbA	*Mycobacterium tuberculosis*	13882401			12
	DmsA	*Mycobacterium smegmatis*	16508080			Unpublished data
	DhaA	*Rhodococcus* sp.	7245711			7
	DbjA	*Bradyrhizobium japonicum*	27349338			13
	DmlA	*Mesorhizobium loti*	13474464			13
	DatA	*Agrobacterium tumefaciens*	16119878			Nagata, personal communication
HLD-III	DrbA	*Rhodopirellula baltica*	32476333	Unknown	Asp-His-Asp+Asn-Trp	Unpublished data
	DmbC	*Mycobacterium tuberculosis*	15608970			Unpublished data

ID 1BN6[64]), and LinB (PDB ID 1IZ7[65]). HLD-I and HLD-II subfamilies differ mainly in the cap domain. The second helix in the cap domain (α5′), which is common to HLD-II subfamily members [Fig. 4(a,b)] and can also be predicted for HLD-III members, is not present in the structure of the HLD-I protein DhlA [Fig. 4(c)]. Moreover, the spatial arrangement of helices α4 and α5 is different in the cap domains of DhlA and HLD-II enzymes. In addition, enzymes from different HLD subfamilies exhibit structural divergence in the loop regions that connect the cap to the main domain, and in the C-terminal part of the main domain. Analysis of the multiple sequence alignment also reveals differences within subfamilies. For example, DbjA exhibits an 11 residue-long insertion in the N-terminus of the cap domain that is not present in other proteins from the HLD-II subfamily. In the sequence of DhlA, a long insertion of 10 amino acids is situated in an analogous region. Furthermore, two very long insertions, of 34 and 24 residues, are present in the sequence of a HLD-I subfamily member, uncbac-67906508, obtained from an environmental sample. The former is localized in the N-terminus of the cap domain, while the latter follows helix α5.

Different compositions of the catalytic pentad were identified for each of the HLD subfamilies (see Fig. 5). However, three residues of the pentad are identical in all subfamilies and thus would be expected to fulfill identical functions, allowing us to extrapolate our knowledge of the catalytic mechanism to the HLD-III subfamily; i.e. nucleophile—Asp (D108 of LinB), catalytic base—His (H272 of LinB), and one of the halide stabilizing residues—Trp (W109 of LinB). HLDs possess two types of catalytic acid, Asp and Glu.[18] In the HLD-I subfamily, the catalytic acid, Asp, is located in the loop following β-strand 7 (D260 of DhlA), whereas HLD-II subfamily members contain a Glu in the loop following β-strand 6 (E186 of LinB). Based on the sequence alignment, the Asp corresponding to the catalytic acid of the HLD-I subfamily was identified and predicted to fulfill this function in the HLD-III subfamily. Similarly, HLDs differ both in the type and location of one halide-stabilizing residue.[18] HLD-I members employ Trp (W175 of DhlA) located in helix α4, whereas HLD-II members use Asn (N38 of LinB) located in the loop following β-strand 3. In the sequences of all HLD-III subfamily members, we also found Asn in the position corresponding to the HLD-II-like halide-stabilizing Asn. Moreover, in some HLD-III members, we found a Trp residue corresponding to the halide-stabilizing amino acid in HLD-I members.

The conservation profile in Figure 6(a) shows that the α/β core is the most conserved region of the HLDs, and surface residues are the least conserved. Regions having the highest sequence variability were found in the N-terminal part of the cap domain and in all three helices within the C-terminal region of the main domain [Fig. 6(b)]. Nine residues are fully conserved in all analyzed sequences. The catalytic base (H272 of LinB) and the catalytic nucleophile (D108 of LinB), a Gly that participates in stabilizing the catalytic water (G37 of LinB) and a His



located N-terminally to this residue (H36 of LinB). Four conserved residues are located within a highly conserved loop following β-strand 4: (G65, G67, D62, and S69 of LinB). The last fully conserved residue is a nucleophile +2 Gly (G110 of LinB). Some of the most variable positions include three sites within the N-terminal main domain (A5, G54, and G99 of LinB), five cap domain sites (F143, Q146, E161, Q172, and E184 of LinB), two sites located in helix α8 (D226 and S232 of LinB), two in helix α9 (R254 and D255 of LinB), and finally eight positions within the C-terminal domain (A285, A288, and R291-A296 of LinB) [Fig. 6(c)]. Identical analyses of conservation were performed for each subfamily [Fig. 6(d)].

DISCUSSION

There are currently a dozen known members of the haloalkane dehalogenase family with experimentally confirmed dehalogenase activity. Traditionally, HLDs were classified according to their substrate specificity.[66] At least four different classes of HLDs have been proposed, namely DhlA, LinB, and DhaA enzymes[18,67] and, more recently, the DbjA enzyme.[13] However, such classification is problematic due to insufficient biochemical characterization of the majority of HLDs. Therefore, we decided to adopt a phylogenetic approach to assess relationships within the HLD family and to establish an objective classification of these enzymes.

To find the root of a HLD phylogenetic tree and establish the direction of evolutionary changes, it was first necessary to find sequences closely related to HLDs that could be used as an outgroup. We clustered HLDs and their homologs using the criterion of pairwise sequence similarity to delineate the "core" HLD family as well as the most closely related, but clearly distinct, protein families. Of the three such families we identified, only one includes experimentally characterized proteins, the epoxide hydrolases and fluoroacetate dehalogenases, while the other two comprise only uncharacterized proteins. This distribution of proteins in the sequence space provides support for the hypothesis that HLDs, together with epoxide hydrolase and fluoroacetate dehalogenase families, have a common ancestor that diverged from other α/β-hydrolases. Not only do these enzymes share significant structural similarities, but of all 363 experimentally characterized α/β-hydrolases included in this study, only these possess Asp as the catalytic nucleophile, suggesting that this residue is the synapomorphy of the haloalkane dehalogenase/epoxide hydrolase/fluoroacetate dehalogenase clade.

Fig. 4. Structures of three haloalkane dehalogenases determined by protein crystallography. DhaA (**A**) and LinB (**B**) proteins are representatives of subfamily HLD-II, whereas DhlA (**C**) belongs to the subfamily HLD-I. The main structural differences discussed in this paper are highlighted, namely, helix α5′ lost in HLD-I, the differently arranged helices α4 and α5 and the connections of the main and cap domains.

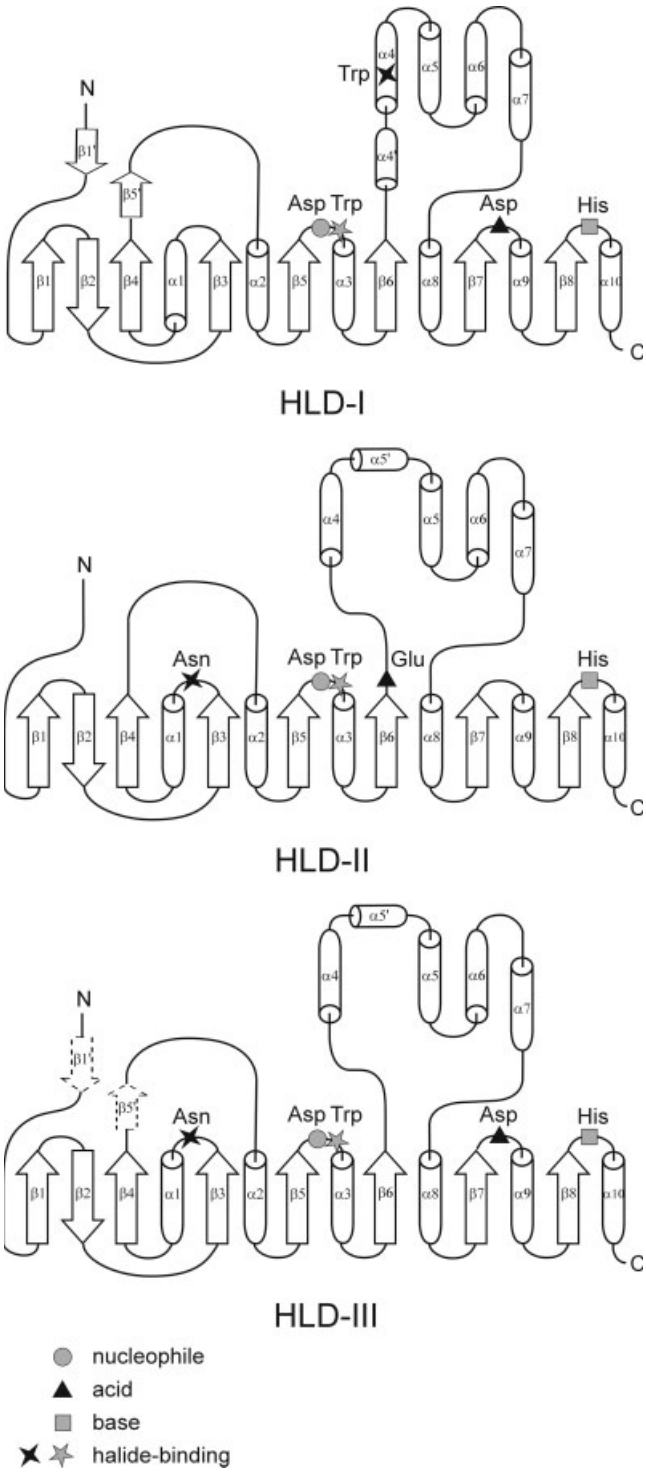


Fig. 5. The topological arrangement of secondary structure elements in individual haloalkane dehalogenase subfamilies (HLD-I, HLD-II, and HLD-III). Positions of catalytic pentad residues are indicated by symbols. Nucleophile, catalytic base, and one halide-stabilizing residue are conserved among all subfamilies (gray), whereas the catalytic acid and second halide-stabilizing residue differ among subfamilies (black).

Both outgroup-I and outgroup-II only include putative proteins identified in genome sequencing projects. While for outgroup-I the catalytic triad is preserved and consequently these proteins could potentially serve as hydrolases, the catalytic nucleophile is replaced by Gly in all members of outgroup-II. A homology model was constructed for a representative member of this group to identify whether another residue could potentially compensate for this change. However, the candidate residues Asp/Ser/Cys were not found in a suitable position in the presumptive active site. Sequence conservation within outgroup-II, as well as between this group and other sequences in our dataset, suggests that these proteins are not degenerate and do exhibit a conserved ligand-binding and perhaps catalytic activity. However, they either catalyze a reaction other than hydrolysis or they employ a completely different reaction mechanism. It is also possible that they may serve as receptors or transporters.

Based on our phylogenetic analyses, we propose that the HLD family should be divided into three subfamilies. Previous classifications based on substrate specificities suggested at least four classes. However, three of these former classes (DhaA, LinB, and DbjA) belong to the same subfamily delineated in this work (HLD-II). The other previous specificity class (DhlA) falls within part of the HLD-I subfamily. DhlA has very different substrate specificity from HLD-II proteins, being active with smaller substrates (Table I). The proposed HLD-III subfamily was not included in previous classifications as no experimentally characterized protein was available for this subfamily. However, experiments ongoing in our laboratory have confirmed that the DrbA and DmbC enzymes have weak dehalogenating activity (unpublished data). Differences in substrate specificity of HLDs are attributed to differences in composition, geometry and size of the active site, the halide-stabilizing residues, and in the entrance channels connecting the active site with the protein surface.[65]

Crystal structures are available for DhlA[63] from the HLD-I subfamily, and for DhaA[64] and LinB[65] from the HLD-II subfamily. There is currently no structure available for the HLD-III subfamily, however, some information about the composition of these proteins can be deduced from the sequence alignment and homology models. Similarly to HLD-II members, the presence of the $\alpha5'$ helix in the cap domain was predicted for the HLD-III subfamily. Both alternative positions of the root in the phylogenetic tree of HLDs support the presence of $\alpha5'$ helix in the ancestor of HLDs [Fig. 7(a)]. We therefore propose that the HLD-I subfamily has lost the $\alpha5'$ helix, rather than that both the HLD-II and HLD-III subfamilies independently acquired it. The different spatial arrangement and high sequence divergence of helices $\alpha4$ and $\alpha5$ in HLD subfamilies may be a consequence of the $\alpha5'$ helix loss. The helix loss hypothesis is further supported by the composition of the cap domain of epoxide hydrolases, the closest relatives of HLDs with a known structure in which the helix is also present. Based on the alignment, the presence of the $\alpha5'$ helix is also suggested

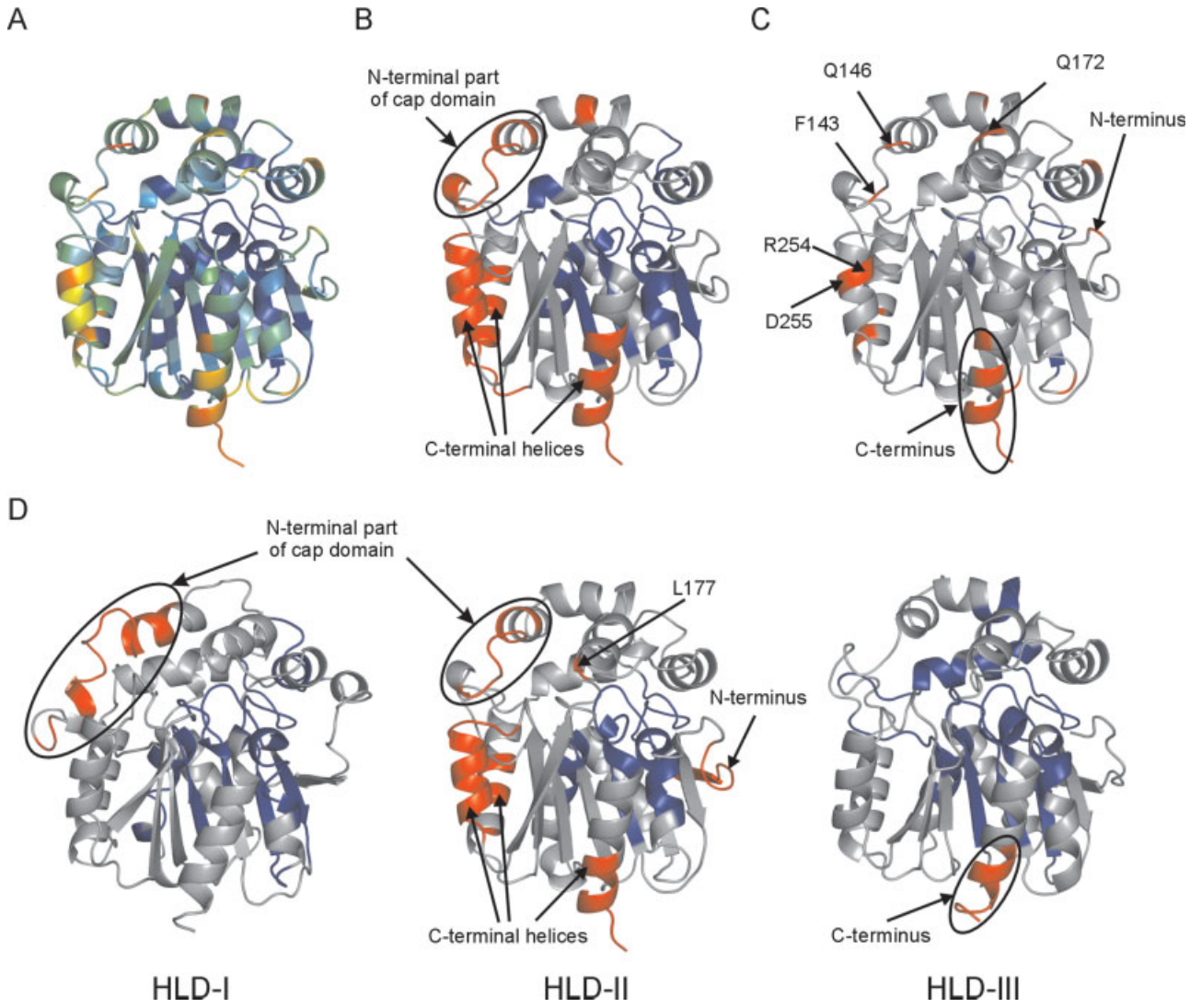


Fig. 6. Conservation profiles of the haloalkane dehalogenase family. Highly conserved regions are colored in blue, highly variable in red. The LinB structure was used as a representative of the entire family (**A**–**C**). (A) Conservation was assigned by the CONSURF server. (B) Regions of the highest conservation and the highest variability in haloalkane dehalogenase sequences as indicated by the statistical test. (C) The twenty most conserved and the most variable sites within haloalkane dehalogenase family. (**D**) Conservation profiles of individual haloalkane dehalogenase subfamilies (HLD-I, HLD-II, and HLD-III). Experimentally determined structures of DhlA, LinB, and a homology model of DrbA enzymes were used as representatives of subfamilies, HLD-I, HLD-II, and HLD-III, respectively.

for two experimentally uncharacterized outgroups, providing further support for the presence of this helix in the ancestor of HLDs. The cap domain and a uteroglobin-like structure composed of four helices have been previously proposed to have a common origin.[68,69]

To date, two different catalytic pentads have been proposed for haloalkane dehalogenases[18]: Asp-His-Asp+Trp-Trp for the HLD-I subfamily and Asp-His-Glu+Asn-Trp for the HLD-II subfamily. In this study, we have identified a new subfamily of HLDs with a novel catalytic pentad composed of Asp-His-Asp+Asn-Trp [Fig. 7(b)]. Other authors have hypothesized that repositioning of a catalytic acid from β-strand 6 (Glu of HLD-II) to the β-strand 7 (Asp of HLD-I) occurred during the molecular adaptation of HLDs to the substrate 1,2-dichloroethane.[70] For HLD-III, we have found Asp in a position corresponding to that in HLD-I, but Glu, characteristic of HLD-II, was not present. It appears that the repositioning of a catalytic acid occurred just after the evolutionary separation of the HLD-II subfamily from the ancestor of HLD-I and HLD-III. It therefore seems unlikely that repositioning of the catalytic acid has been an adaptation to dehalogenation of 1,2-dichloroethane, which was unknown in nature until the industrial revolution and the repositioning must have occurred much earlier. Moreover, members of outgroup-I, outgroup-II, and some proteins of outgroup-III also contain a catalytic acid identical to that of HLD-I and HLD-III. This suggests that the ancestor of HLDs probably had a catalytic acid in the position following β-strand 7. All members of the HLD-III subfamily contain

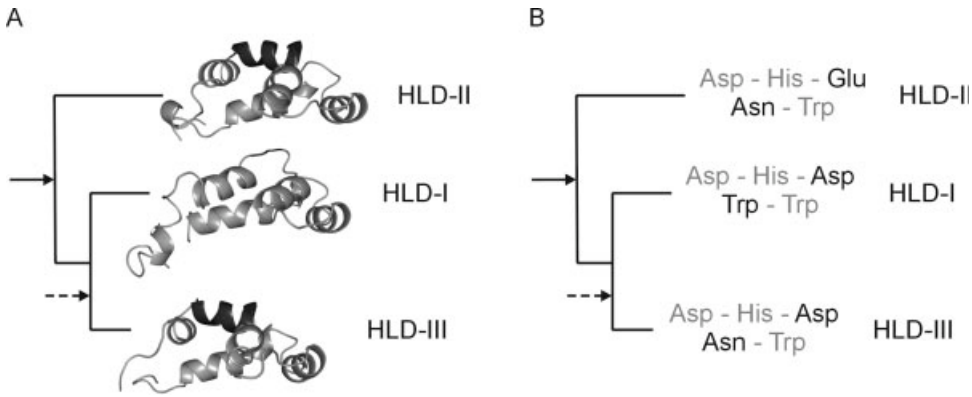


Fig. 7. Evolution of the cap domain (**A**) and the catalytic pentad (**B**) within the haloalkane dehalogenase family. The tree is rooted based on the results of outgroup analysis and its probable root is indicated by the solid arrow. An alternative root position is indicated by the dotted arrow. The region of helix $\alpha5'$ lost in the HLD-I subfamily is highlighted. The catalytic pentad is composed of the catalytic triad of a nucleophile-base-catalytic acid and a pair of halide-stabilizing residues. The nucleophile, catalytic base and one halide-stabilizing residue are conserved among all subfamilies (gray), whereas the catalytic acid and second halide-stabilizing residue differ among subfamilies (black).

Asn in their sequence at the site corresponding to the halide-stabilizing Asn of HLD-II. Moreover, some proteins of this subfamily also have a Trp corresponding to the halide-stabilizing Trp present in the HLD-I subfamily. However, the spatial location of this residue is different from that seen in HLD-I members due to the presence of the $\alpha5'$ helix in the HLD-III cap domain. Therefore, it is probably not correctly oriented to fulfill a halide-stabilization function in this subfamily. Also, the functionality of the Trp could be dependent on the loss of the $\alpha5'$ helix within the cap domain of the HLD-I subfamily, leading to repositioning of its side-chain in space.

To assess the importance of individual amino acid residues within the HLD family as a whole, we analyzed sequence conservation by calculating their respective evolutionary rates. High conservation of a particular site usually indicates its importance for maintaining structural or functional properties of the protein. As expected, both β-strands in the protein core and the loops carrying catalytic residues are highly conserved among HLDs. In addition to the catalytic loops, a loop following β-strand 4 was also found to be strongly conserved, but the reason for this remains unclear. The precise determination of highly variable sites was complicated by ambiguities in the alignment of variable regions. Nevertheless, it is important to delineate variable residues in the region involved in substrate binding, since high mutability may also be a result of selection pressure on variants with potentially new substrate specificities. Such variable positions are particularly suitable targets for protein engineering by site-directed mutagenesis. In HLDs, three highly variable sites, corresponding to F143, Q146, and Q172 of LinB, form the walls of the access channels. It has been shown that replacement of channel residues can effectively change the size of the channel and substrate specificity of the haloalkane dehalogenase from *S. japonicum* UT26.[71] Sites corresponding to R254 and D255 of LinB are located on the protein surface close to the entrance of the channel. We speculate that the high variability of these five positions reflects the adaptation of HLDs to different substrates by modifying the channel size, shape, and location.

In addition to the variability of individual residues, we also considered the variability of entire protein segments. The N-terminal part of the cap domain and all three helices in the C-terminus of the main domain were found to be among the most variable parts of HLDs. The high variability of the last C-terminal helix is not surprising as it is located at the very end of the sequence, which is not involved in functionally important interactions, and its conservation is, therefore, not essential. The other two C-terminal helices participate in the formation of the surface close to the entrance to the channel leading to the active site. However, it is unclear whether they participate in substrate binding or if their variability is in some way connected with participation in channelling of the substrate into the active site. The N-terminal part of the cap domain is also one of the most variable regions, not only between, but also within, the HLD subfamilies. The participation of the variable sites within this region in the modulation of channel anatomy and substrate specificity has been discussed above. Previously, this region has been proposed to influence the substrate specificity of HLDs[72] and was consequently identified as a suitable target for mutagenesis likely to lead to modifications of the substrate specificity of DhlA. The importance of the N-terminal part of the cap domain for the substrate specificity of DhlA was further supported by the results of a COMBINE analysis that identified seven highly significant enzyme–substrate interactions in this region.[73] This region has been also shown to be the most flexible part of HLD structures by displaying concerted functionally relevant motions.[74] Three proteins have also been found to carry a unique long insertion within this region. A 10 amino acid insertion was identified in the DhlA sequence and was attributed to adaptation for the conversion of

1,2-dichloroethane by this enzyme.[75] An insertion of 11 residues in the sequence of the DbjA enzyme is of different evolutionary origin to that of DhlA, but it is located in the same region. Two very long insertions of 34 and 24 residues in the sequence of uncbac-67906508 are localized in the region analogous to that of epoxide hydrolases. Long insertions in the cap domain of epoxide hydrolases increase the size of the active site cavity as well as the size of the entrance channel, and similar insertions in HLDs would be expected to have similar effects. The N-terminal part of the cap domain was found to be highly variable within the HLD-I and HLD-II subfamilies when they were separately analyzed. In the case of the HLD-III subfamily, only the very C-terminal part of the sequence was found to be significantly more variable than the rest of the protein. This is not surprising, since the overall similarity of most sequences within the HLD-III subfamily is higher than that seen in the other two subfamilies and consequently no significantly preferred region for mutation was found within the sequences of HLD-III members. For the HLD-II subfamily, the site corresponding to L177 of LinB was identified as being highly variable. L177 is located at the opening of the channel and its mutagenesis has previously been shown to yield protein variants with modified substrate specificity.[71] Our study indicates that variability of this site represents a subfamily-specific adaptation mechanism.

All experimentally confirmed HLDs are of bacterial origin. Putative HLDs identified by our analysis are also predominantly from bacterial species. The only exceptions are putative proteins from *S. purpuratus*. However, these proteins are closely related to *Renilla* luciferases and may possess luciferase activity instead of dehalogenase activity. Genes encoding HLDs have been identified in the genomes of various Proteobacteria and Actinobacteria species and also in at least one member of each of the Planctomycetes, Cyanobacteria, and Chloroflexi (Supplementary Table SII). It seems very likely that as the number of completely sequenced genomes increases, HLDs will be identified in a wider spectrum of organisms. Our phylogenetic HLD tree does not agree with the established taxonomy of the host organisms. For example, highly similar *dhaA* genes were found in distantly related species of *Rhodococcus* and *Pseudomonas*, suggesting that horizontal gene transfer has been the driving force in the spread of these enzymes among bacteria.[76] In some species, more than one putative dehalogenase gene was identified. For example, *Shewanella frigidimarina* carries two putative HLD genes belonging to the HLD-I subfamily that exhibit 77% sequence identity, indicating that some divergence has already occurred since duplication. This suggests that paralogization, potentially followed by subfunctionalization and adaptation to different substrates, have also played roles in the evolution of HLDs. Interestingly, *S. frigidimarina* also possesses an additional putative HLD of subfamily HLD-III. *Jannaschia* sp. possesses two putative HLD genes, each belonging to a different subfamily, HLD-II and HLD-III. *M. tuberculosis* has even three different HLDs with confirmed dehalogenating activity, namely, DmbA, DmbB, and DmbC from subfamilies HLD-II, HLD-I, and HLD-III, respectively. Thus, the presence of multiple dehalogenases in several bacteria has probably arisen from a combination of duplications and independent horizontal gene transfers. It is important to note that the physiological function of HLDs in most of the bacteria analyzed in this work is not currently understood. This phylogenetic study should provide a useful framework for, and stimulate, both comparative biochemical analyses and further efforts to elucidate the function of HLDs in the natural environment of their hosts.

REFERENCES

1. Keuning S, Janssen DB, Witholt B. Purification and characterization of hydrolytic haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10. J Bacteriol 1985;163:635–639.
2. Yokota T, Omori T, Kodama T. Purification and properties of haloalkane dehalogenase from *Corynebacterium* sp. strain m15-3. J Bacteriol 1987;169:4049–4054.
3. Scholtz R, Leisinger T, Suter F, Cook AM. Characterization of 1-chlorohexane halidohydrolase, a dehalogenase of wide substrate range from an *Arthrobacter* sp. J Bacteriol 1987;169:5016–5021.
4. Nagata Y, Miyauchi K, Damborsky J, Manova K, Ansorgova A, Takagi M. Purification and characterization of haloalkane dehalogenase of a new substrate class from a γ-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26. Appl Environ Microbiol 1997;63:3707–3710.
5. Janssen DB, Gerritse J, Brackman J, Kalk C, Jager D, Witholt B. Purification and characterization of a bacterial dehalogenase with activity toward halogenated alkanes, alcohols and ethers. Eur J Biochem 1988;171:67–72.
6. Sallis PJ, Armfield SJ, Bull AT, Hardman DJ. Isolation and characterization of a haloalkane halidohydrolase from *Rhodococcus erythropolis* Y2. J Gen Microbiol 1990;136:115–120.
7. Kulakova AN, Larkin MJ, Kulakov LA. The plasmid-located haloalkane dehalogenase gene from *Rhodococcus rhodochrous* NCIMB 13064. Microbiology 1997;143:109–115.
8. Poelarends GJ, Wilkens M, Larkin MJ, van Elsas JD, Janssen DB. Degradation of 1,3-dichloropropene by *Pseudomonas cichorii* 170. Appl Environ Microbiol 1998;64:2931–2936.
9. Poelarends GJ, van Hylckama Vlieg JET, Marchesi JR, Freitas dos Santos LM, Janssen DB. Degradation of 1,2-dibromoethane by *Mycobacterium* sp. strain GP1. J Bacteriol 1999;181:2050–2058.
10. Kumari R, Subudhi S, Suar M, Dhingra G, Raina V, Dogra C, Lal S, van der Meer JR, Holliger C, Lal R. Cloning and characterization of *lin* genes responsible for the degradation of hexachlorocyclohexane isomers by *Sphingomonas paucimobilis* strain B90. Appl Environ Microbiol 2002;68:6021–6028.
11. Jesenska A, Sedlacek I, Damborsky J. Dehalogenation of haloalkanes by *Mycobacterium tuberculosis* H37Rv and other mycobacteria. Appl Environ Microbiol 2000;66:219–222.
12. Jesenska A, Pavlova M, Strouhal M, Chaloupkova R, Tesinska I, Monincova M, Prokop Z, Bartos M, Pavlik I, Rychlik I, Mobius P, Nagata Y, Damborsky J. Cloning, biochemical properties, and distribution of mycobacterial haloalkane dehalogenases. Appl Environ Microbiol 2005;71:6736–6745.
13. Sato Y, Monincova M, Chaloupkova R, Prokop Z, Ohtsubo Y, Minamisawa K, Tsuda M, Damborsky J, Nagata Y. Two rhizobial strains, *Mesorhizobium loti* MAFF303099 and *Bradyrhizobium japonicum*, USDA110, encode haloalkane dehalogenases with novel structures and substrate specificities. Appl Environ Microbiol 2005;71:4372–4379.
14. Bugg TDH. Diverse catalytic activities in the α β-hydrolase family of enzymes: activation of H_2O, HCN, H_2O_2, and O_2. Bioorg Chem 2004;32:367–375.
15. Holmquist M. α/β-hydrolase fold enzymes: structures, functions and mechanisms. Curr Protein Pept Sci 2000;1:209–235.
16. Nardini M, Dijkstra BW. α/β hydrolase fold enzymes: the family keeps growing. Curr Opin Struct Biol 1999;9:732–737.

17. Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, Sussman JL, Verschueren KHG, Goldman A. The α/β hydrolase fold. Protein Eng Des Sel 1992;5:197–211.
18. Damborsky J, Koca J. Analysis of the reaction mechanism and substrate specificity of haloalkane dehalogenases by sequential and structural comparisons. Protein Eng Des Sel 1999;12:989–998.
19. Janssen DB. Evolving haloalkane dehalogenases. Curr Opin Chem Biol 2004;8:150–159.
20. Jesenska A, Bartos M, Czernekova V, Rychlik I, Pavlik I, Damborsky J. Cloning and expression of the haloalkane dehalogenase gene *dhmA* from *Mycobacterium avium* N85 and preliminary characterization of DhmA. Appl Environ Microbiol 2002;68: 3724–3730.
21. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 2004;20:3702–3704.
22. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–1797.
23. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1999;41:95–98.
24. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 2005;21:2104–2105.
25. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52:696–704.
26. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 2001;18:691–699.
27. Gascuel O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol 1997;14: 685–695.
28. Saitou N, Nei M. The neighbor-joining method—a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–425.
29. Drummond A, Strimmer K. PAL: an object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics 2001;17:662–663.
30. Strimmer K, vonHaeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 1997;94:6815–6819.
31. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 2002;18:502–504.
32. Rzhetsky A, Kumar S, Nei M. 4-cluster analysis—a simple method to test phylogenetic hypotheses. Mol Biol Evol 1995;12:163–167.
33. Kumar S. PHYLTEST: a program for testing phylogenetic hypotheses, Version 2.0. University Park, PA: Pennsylvania State University; 1996.
34. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 2005;33:W299–W302.
35. Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. Nucleic Acids Res 2003;31:3305–3307.
36. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16:404–405.
37. Rost B, Yachdav G, Liu JF. The PredictProtein server. Nucleic Acids Res 2004;32:W321–W326.
38. Ouali M, King RD. Cascaded multiple classifiers for secondary structure prediction. Protein Sci 2000;9:1162–1176.
39. Adamczak R, Porollo A, Meller J. Accurate prediction of solvent accessibility using neural networks-based regression. Proteins 2004;56:753–767.
40. Cuff JA, Barton GJ. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins 2000;40:502–511.
41. Meiler J, Baker D. Coupled prediction of protein secondary and tertiary structure. Proc Natl Acad Sci USA 2003;100:12105–12110.
42. Karplus K, Karchin R, Draper J, Casper J, Mandel-Gutfreund Y, Diekhans M, Hughey R. Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. Proteins 2003;53:491–496.
43. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. Bioinformatics 1998;14:892–893.
44. Jaroszewski L, Rychlewski L, Godzik A. Improving the quality of twilight-zone alignments. Protein Sci 2000;9:1487–1496.
45. Kelley LA, MacCallum RM, Sternberg MJE. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 2000;299:499–520.
46. Fischer D. Hybrid fold recognition: combining sequence derived properties with evolutionary information. Pac Symp Biocomput 2000;119–130.
47. Shi JY, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 2001;310: 243–257.
48. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287: 797–815.
49. Zhou HY, Zhou YQ. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. Proteins 2004;55: 1005–1013.
50. Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A. Pcons: a neural-network-based consensus predictor that improves fold recognition. Protein Sci 2001;10:2354–2362.
51. Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM. A "Frankenstein's monster" approach to comparative modeling: merging the finest fragments of fold-recognition models and iterative model refinement aided by 3D structure evaluation. Proteins 2003;53:369–379.
52. Feder M, Bujnicki JM. Identification of a new family of putative PD-(D/E)XK nucleases with unusual phylogenomic distribution and a new type of the active site. BMC Genomics 2005;6:21.
53. Fiser AS, Sali A. MODELLER: generation and refinement of homology-based protein structure models. Methods Enzymol 2003;374:461–491.
54. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with 3-dimensional profiles. Nature 1992;356:83–85.
55. Sasin JM, Bujnicki JM. COLORADO3D, a web server for the visual analysis of protein structures. Nucleic Acids Res 2004;32: W586–W589.
56. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP—a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
57. Holm L, Sander C. Protein-structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
58. Ye YZ, Godzik A. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res 2004;32:W582–W585.
59. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
60. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
61. Nagata Y, Nariya T, Ohtomo R, Fukuda M, Yano K, Takagi M. Cloning and sequencing of a dehalogenase gene encoding an enzyme with hydrolase activity involved in the degradation of hexachlorocyclohexane in *Pseudomonas paucimobilis*. J Bacteriol 1993;175:6403–6410.
62. Lorenz WW, McCann RO, Longiaru M, Cormier MJ. Isolation and expression of a cDNA encoding *Renilla reniformis* luciferase. Proc Natl Acad Sci USA 1991;88:4438–4442.
63. Verschueren KHG, Kingma J, Rozeboom HJ, Kalk KH, Janssen DB, Dijkstra BW. Crystallographic and fluorescence studies of the interaction of haloalkane dehalogenase with halide ions. Studies with halide compounds reveal a halide binding site in the active site. Biochemistry 1993;32:9031–9037.
64. Newman J, Peat TS, Richard R, Kan L, Swanson PE, Affholter JA, Holmes IH, Schindler JF, Unkefer CJ, Terwilliger TC. Haloalkane dehalogenase: structure of a *Rhodococcus* enzyme. Biochemistry 1999;38:16105–16114.
65. Marek J, Vevodova J, Kuta-Smatanova I, Nagata Y, Svensson LA, Newman J, Takagi M, Damborsky J. Crystal structure of

the haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26. Biochemistry 2000;39:14082–14086.
66. Damborsky J, Nyandoroh MG, Nemec M, Holoubek I, Bull AT, Hardman DJ. Some biochemical properties and classification of a range of bacterial haloalkane dehalogenases. Biotechnol Appl Biochem 1997;26:19–25.
67. Damborsky J, Rorije E, Jesenska A, Nagata Y, Klopman G, Peijnenburg WJGM. Structure-specificity relationships for haloalkane dehalogenases. Environ Toxicol Chem 2001;20:2681–2689.
68. Callebaut I, Poupon A, Bally R, Demaret JP, Housset D, Delettre J, Hossenlopp P, Mornon JP. The uteroglobin fold. Uteroglobin/Clara cell protein family. Ann N Y Acad Sci 2000;923:90–112.
69. Russell RB, Sternberg MJE. Two new examples of protein structural similarities within the structure-function twilight zone. Protein Eng Des Sel 1997;10:333–338.
70. Krooshof GH, Kwant EM, Damborsky J, Koca J, Janssen DB. Repositioning the catalytic triad aspartic acid of haloalkane dehalogenase: effects on stability, kinetics, and structure. Biochemistry 1997;36:9571–9580.
71. Chaloupkova R, Sykorova J, Prokop Z, Jesenska A, Monincova M, Pavlova M, Tsuda M, Nagata Y, Damborsky J. Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel. J Biol Chem 2003;278:52622–52628.
72. Pries F, VandenWijngaard AJ, Bos R, Pentenga M, Janssen DB. The role of spontaneous cap domain mutations in haloalkane dehalogenase specificity and evolution. J Biol Chem 1994;269: 17490–17494.
73. Kmunicek J, Luengo S, Gago F, Ortiz AR, Wade RC, Damborsky J. Comparative binding energy analysis of the substrate specificity of haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10. Biochemistry 2001;40:8905–8917.
74. Otyepka M, Damborsky J. Functionally relevant motions of haloalkane dehalogenases occur in the specificity-modulating cap domains. Protein Sci 2002;11:1206–1217.
75. Pikkemaat MG, Janssen DB. Generating segmental mutations in haloalkane dehalogenase: a novel part in the directed evolution toolbox. Nucleic Acids Res 2002;30:e35.
76. Poelarends GJ, Kulakov LA, Larkin MJ, van Hylckama Vlieg JET, Janssen DB. Roles of horizontal transfer and gene integration in evolution of 1,3-dichloropropene- and 1,2-dibromoethane-degradative pathways. J Bacteriol 2000;182:2191–2199.