university of groningen

umcg:

University of Groningen

Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities

Janssen, D.B.; Dinkla, I.J.T.; Poelarends, G.J.; Terpstra, P

Published in:
Environmental Microbiology

DOI:
10.1111/j.1462-2920.2005.00966.x

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2005

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Janssen, D. B., Dinkla, I. J. T., Poelarends, G. J., & Terpstra, P. (2005). Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities: evolution and distribution of novel enzyme activities. *Environmental Microbiology*, *7*(12), 1868 - 1882. https://doi.org/10.1111/j.1462-2920.2005.00966.x

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 06-11-2026

Environmental Microbiology (2005) **7**(12), 1868–1882 doi:10.1111/j.1462-2920.2005.00966.x

Minireview

Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities

Dick B. Janssen,[1]* Inez J. T. Dinkla,[1] Gerrit J. Poelarends[1] and Peter Terpstra[2]
[1]*Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, the Netherlands.*
[2]*Section Medical Biology, University Medical Centre Groningen, University of Groningen, A. Deusinglaan 1, 9713 AV Groningen, the Netherlands.*

Summary

Bacterial dehalogenases catalyse the cleavage of carbon-halogen bonds, which is a key step in aerobic mineralization pathways of many halogenated compounds that occur as environmental pollutants. There is a broad range of dehalogenases, which can be classified in different protein superfamilies and have fundamentally different catalytic mechanisms. Identical dehalogenases have repeatedly been detected in organisms that were isolated at different geographical locations, indicating that only a restricted number of sequences are used for a certain dehalogenation reaction in organohalogen-utilizing organisms. At the same time, massive random sequencing of environmental DNA, and microbial genome sequencing projects have shown that there is a large diversity of dehalogenase sequences that is not employed by known catabolic pathways. The corresponding proteins may have novel functions and selectivities that could be valuable for biotransformations in the future. Apparently, traditional enrichment and metagenome approaches explore different segments of sequence space. This is also observed with alkane hydroxylases, a category of proteins that can be detected on basis of conserved sequence motifs and for which a large number of sequences has been found in isolated bacterial cultures and genomic databases. It is likely that ongoing genetic adaptation, with the recruitment of silent sequences into functional catabolic routes and evolution of substrate range by mutations in structural genes, will further enhance the catabolic potential of bacteria toward synthetic organohalogens and ultimately contribute to cleansing the environment of these toxic and recalcitrant chemicals.

Received 25 May, 2005; accepted 19 October, 2005. *For correspondence. E-mail d.b.janssen@rug.nl; Tel. (+31) 503 634 209; Fax (+31) 503 634 165.

Introduction: degradable and recalcitrant compounds

The recalcitrance of many synthetic chemicals to biodegradation is mainly due to a lack of enzymes that can carry out critical steps in a catabolic pathway. This especially holds for low-molecular weight halogenated compounds. These xenobiotic chemicals are relatively water soluble and bioavailable, and theoretically could be converted by short metabolic routes to intermediates that support cellular growth under aerobic conditions. Yet, no organisms have been found that oxidatively degrade and use as a carbon source important environmental chemicals such as chloroform, trichloroethylene, 1,1,1-trichloroethane, 1,2-dichloropropane and 1,2,3-trichloropropane (Fig. 1). Attempts to obtain enrichments or pure cultures that aerobically grow on these chemicals have met no success. However, some other halogenated chemicals are easily biodegradable, and cultures that utilize chloroacetate, 2-chloropropionate and 1-chlorobutane can be readily enriched from almost any soil sample (Leisinger, 1996; van Agteren *et al.*, 1998). For still other compounds, degradative organisms can be isolated, but only after prolonged adaptation or if a suitable inoculum is used in which the catabolic activity that is searched for apparently is already enriched due to pre-exposure to halogenated chemicals in the environment. Compounds of this class of intermediate degradability include dichloromethane, 1,2-dichloroethane and the nematocides 1,2-dibromoethane and 1,3-dichloropropene (Fig. 1).

With halogenated compounds, an obvious critical step in a potential biodegradation pathway is dehalogenation. Biochemical research with organisms that grow on halogenated compounds has shown that a broad range of dehalogenases exists, both for aliphatic and aromatic compounds. The enzymes represent different reaction

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd

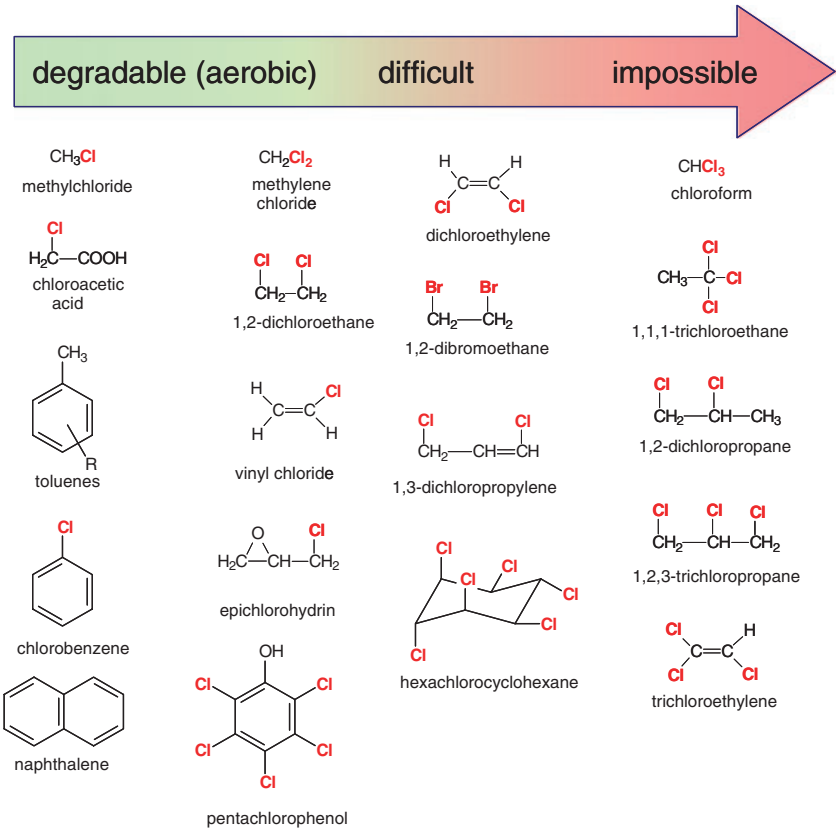


Fig. 1. Degradability of some (halogenated) aliphatic and aromatic compounds. The substrates on the left have often found to be readily degradable, and organisms that utilize them for growth are easily isolated from soil or water. The compounds on the right have repeatedly been found to be recalcitrant and enrichment cultures were negative. The other compounds (middle) may be biodegradable, but organisms are not ubiquitous and success of enrichment very much depends on conditions and choice of inoculum.

types, and they may or may not use cofactors for activity. Several X-ray structures have been solved, and these suggest that most of these enzymes have evolved specifically to catalyse a dehalogenation reaction, and are not just enzymes that carry out a certain reaction with a natural compound and fortuitously also cleave carbon-halogen bonds in a xenobiotic organohalogen compound due to catalytic promiscuity. The substrate range of a dehalogenase often determines whether or not a compound is degradable, although the accumulation of toxic intermediates after an initial dehalogenation step may also be the cause of poor metabolism, especially with substrates carrying more than one halogen (van Hylckama Vlieg *et al.*, 2000). In view of the key role of dehalogenases, several questions arise. How do new dehalogenases arise in nature? What are their catalytic mechanisms and how do they influence the range of compounds that can be converted? How are these enzymes distributed in environmental organisms?

This review addresses some of these issues, by describing examples of catalytic mechanisms of dehalogenases that were obtained from X-ray crystallography. From these mechanisms, conserved functional sequence motifs can be derived, and by using these in searches of databases, an estimate is obtained about the abundance of some key dehalogenases in bacteria. The results are compared with the situation for alkane hydroxylases, which are the key enzymes in the initial steps of alkane metabolism. Sequence information and comparison of catabolic gene clusters also shed light on the mechanisms by which new catabolic activities may evolve and become distributed.

Dehalogenase mechanisms

Five dehalogenases that are members of different protein superfamilies have been studied by X-ray crystallography. This has provided detailed insight in their catalytic mechanisms and evolutionary relationships. For several other dehalogenases, structural and mechanistic insight has been obtained by comparing their sequences with other members of the same phylogenetic family, often in combination with mutagenesis experiments. From this work, it has become clear that most dehalogenases belong to protein superfamilies that harbour both dehalogenases and proteins that carry out completely different reactions (de Jong and Dijkstra, 2003; Fig. 2).

The first example is provided by the haloalkane dehalogenases, with three solved X-ray structures (DhlA, DhaA and LinB in Table 1). These proteins possess a α/β-hydrolase fold main domain, with a cap domain on top of it and the active site located in between the two domains

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

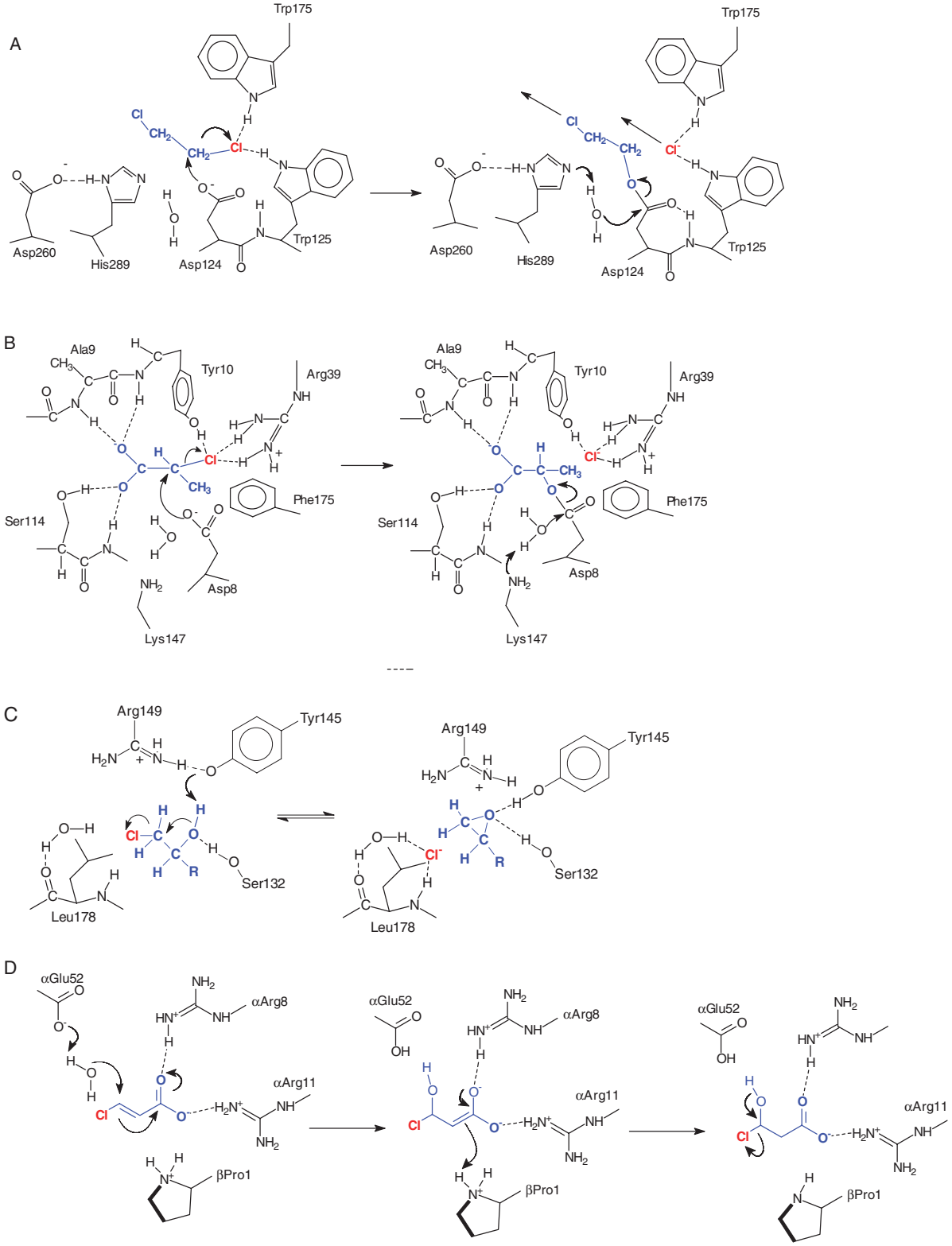


Fig. 2. Common dehalogenase mechanisms for aliphatic organochlorine compounds.
A. Hydrolytic dehalogenation catalysed by an α/β-hydrolase fold type haloalkane dehalogenase (DhlA).
B. Hydrolytic dehalogenation catalysed by HAD-type haloacid dehalogenase (DhlB).
C. Dehalogenation by an SDR-type halohydrin dehalogenase (HheC).
D. Chloroacrylic acid dehalogenation by a 4-OT type chloroacrylic acid dehalogenase (CaaD).

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

(Verschueren *et al*., 1993). Their properties have recently been reviewed (Janssen, 2004). The α/β-hydrolase structural fold is also found in lipases, acetylcholinesterases, esterases, lactonases, epoxide hydrolases and others, showing that the haloalkane dehalogenases belong to a protein superfamily of which the members carry out diverse reactions, mostly with non-halogenated compounds. The key residues for identification of new haloalkane dehalogenase sequences are the catalytic pentad amino acids: a nucleophilic aspartate, a conserved histidine base close to the C-terminus, an acidic residue located in the sequence between the Asp and His, and two residues involved in binding of the halide. Of these five residues, the nucleophilic Asp, the His and one halide-binding Trp positioned next to the nucleophilic Asp are fully conserved (Fig. 2A).

A rather common group of aliphatic dehalogenases are the haloacid dehalogenases. These have been divided in group I and group II enzymes, of which the latter are well characterized (Hisano *et al*., 1996; Ridder *et al*., 1999). They define the so-called HAD superfamily of hydrolases. The enzymes (DhlB, HAD-Ps in Table 1) have a nucleophilic aspartate close to the N-terminus and share structural similarity with the phosphatase domain found in many proteins, including eukaryotic transporter proteins. The conserved nucleophilic aspartate is involved in forming a covalent intermediate by a nucleophilic substitution, similar to what was found in haloalkane dehalogenases. For database analysis, other important catalytic residues are an arginine at a position (Nu + *c*. 30), which is involved in halide binding, and a lysine further downstream that acts as the catalytic base needed for cleavage of the covalent intermediate (Fig. 2B).

A third family of aliphatic dehalogenases is formed by the halohydrin dehalogenases, of which the structure has been investigated recently (de Jong *et al*., 2003).The catalytic mechanism of these enzymes is somewhat similar to that of members of the short-chain dehydrogenase-reductase (SDR) superfamily of proteins. Both the dehalogenases and the SDR enzymes possess a conserved catalytic triad for proton abstraction from the hydroxyl group of the substrate (van Hylckama Vlieg *et al*., 2001). In the case of SDR proteins, the negative charge developing on the hydroxyl oxygen is transferred via a hydride to the $NAD(P)^+$ cofactor; instead, in the dehalogenases it is passed on to the halogen at a neighbouring carbon

Table 1. Number of positive hits obtained with protein sequences of dehalogenases and alkane hydroxylases as queries in Blast-P searches against protein sequence databases.

Query protein	GenBank GI number	Super family	Filtered by conserved pattern or residues	No of identical genes found in isolates	No. of homologues in microbial genomes	No. of homologues in Archae genomes	No. of homologues in Sargasso Sea
DcmA	482502	GST	126H-W*	11 (> 96%)	2 (33–22%)	0	4 (30–25%)
DhlA	729681	α/β-HF	124D-W-G, 289H*	12 (100%)	15 (50–26%)	1 (30%)	35 (41–21%)
DhaA	3114657	α/β-HF	106D-W-G, 272H*	7 (> 98%)	28 (53–23%)	1 (26%)	44 (49–22%)
LinB	9789853	α/β-HF	108D-W-G, 273F*	3 (> 99%)	10 (69–25%)	0	43 (65–21%)
DhlB	3122178	HAD	8D, 39R, 147K#	7 (49–39%)	39 (57–25%)	3 (26–24%)	119 (52–23%)
HAD-Ps	3122176	HAD	10D, 41R, 151K#	12 (77–39%)	35 (55–23%)	4 (27–23%)	115 (50–24%)
HheA	15213645	SDR	135S-X(7,17)-Y-X(3)-R*	2 (99%), 3 (49–48%)	6 (30–21%)	0	4 (31–21%)
HheB	15209119	SDR	126S-X(7,17)-Y-X(3)-R*	1 (98%)	16 (45–19%)	1 (27%)	10 (53–21%)
HheC	15213643	SDR	132S-X(7,17)-Y-X(3)-R*	2 (100–92%), 2 (49%)	6 (30–20%)	0	5 (31–21%)
CaaD	10637969	4-OT	9R-X-X-R*	0	0	0	6 (35–29%)
LinA	51859616	DH	25D, 73H#	3 (> 99%)	4 (30–27%)	0	3 (31–25%)
CbzA	2392484	ECH	90H, 145D#	6 (86–51%)	21 (41–26%)	1 (28%)	48 (37–27%)
AtzA	32455822	DA	60H-X-H, 243H, 327D*	4 (> 98%)	134 (36–20%)	12 (31–27%)	20 (32–19%)
TriA	42558845	DA	76H-X-H, 251H, 287H*	2 (> 99%)	156 (36–21%)	21 (26–20%)	32 (34–21%)
PcpC	22417110	GST	12S-X-C*	1 (96%)	2 (37–23%)	0	49 (41–19%)
AlkB	113639	AH	269 N-Y-X-E-H-Y-G*	9 (100–77%)	21 (50–32%)	0	2 (82%), 49 (50–25%)
AlkM	2623971	AH	281 N-Y-X-E-H-Y-G*	12 (90–62%)	1 (100%), 20 (50–32%)	0	51 (56–28%)

Protein sequences (column 2) were compared by BlastP (Altschul *et al*., 1990; 1997) to the NCBI whole microbial database (295 proteomes, April 2005) and to the NCBI environmental Sargasso sea proteins (Venter *et al*., 2004). Hits with an E-score < 0.01 were considered homologues. Output lists were filtered by looking for conserved residues in a multiple alignment (#, column 4) or by using PHI-BlastP with a conserved pattern (* column 4). If uncertain, sequences in the low-similarity region were used as Blast queries against the Uniprot database to confirm their identity. Results in columns 5–8 give the number of homologues followed by the percentage range of amino acid identities. In some cases very high scores are mentioned separately. AlkB, AlkM: only hits containing motif C (see text) were counted. AlkM scores 100% in one case in column 6 because the whole genome of the original host has been sequenced.

GST, glutathione transferase; α/β-HF, α/β-hydrolase fold family; HAD, haloacid dehalogenase; SDR, short chain dehydrogenase reductase; 4-OT, 4-oxalocrotonate tautomerase; DH, dehydratase; ECH, enoyl CoA hydratase; DA, deaminase; AH, alkane hydroxylase. The abbreviations for the proteins and their functions are given in the text.

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

atom, which is displaced by an intramolecular substitution mechanism, resulting in the formation of an epoxide (Fig. 2C, HheA, HheB, HheC). Thus, no covalent intermediate is formed during dechlorination catalysis by this enzyme.

Another group of recently characterized dehalogenating proteins is formed by the chloroacrylic acid dehalogenases (CaaD in Table 1), which are present in bacteria that degrade the nematocide 1,3-dichloropropene. These proteins can dehalogenate a substrate in which the halogen is bound to an sp^2-hybridized carbon atom, and there are *cis*- and *trans*-specific enzymes (Poelarends and Whitman, 2004). Structures have recently been solved, and they confirmed what was expected on basis of the sequence, i.e. that the enzymes belong to the tautomerase superfamily of proteins. The members of this superfamily are characterized by a ring-like hexameric structure, which is formed by a trimer of dimers. In some members of the tautomerase superfamily, such as 4-oxalocrotonate tautomerase (4-OT), the hexamers are composed of identical peptides, while in other members, such as *trans*-3-CaaD, they are composed of different but similar peptide chains. Still other members of the superfamily, like *cis*-3-CaaD, have a trimeric quaternary structure. In these dehalogenases, the catalytic mechanism does not involve the formation of a covalent enzyme-substrate intermediate (de Jong *et al*., 2004). Instead, the mechanism of *trans*-3-CaaD is similar to that of a hydratase, with a key role for the N-terminal proline of three of the subunits, which acts as the general acid catalyst that protonates the C-2 carbon atom in *trans*-3-chloroacrylic acid (Fig. 2D) (Azurmendi *et al*., 2004). In CaaD, no clear halide-binding site has yet been detected and it is possible that these enzymes mechanistically carry out only a hydration reaction to form an unstable halohydrin intermediate that rapidly dechlorinates.

For several other aliphatic dehalogenases, no structures have been solved, but mechanistic information can be deduced from similarity to enzymes that are well characterized (Leisinger *et al*., 1994). For example, dichloromethane dehalogenase (DcmA) catalyses the conversion of dichloromethane to formaldehyde in a glutathione-dependent reaction. Catalysis was proposed to proceed via the formation of a reactive S-(chloromethyl)glutathione intermediate that stays bound to the enzyme and decomposes by solvolysis, which may be the rate-limiting step (Stourman *et al*., 2003). Site-directed mutagenesis of DcmA demonstrated the essentiality of a serine residue in the N-terminal part of the protein, a residue also shown to be critical for catalysis in θ-class glutathione transferases (Vuilleumier and Leisinger, 1996). This conserved catalytic serine may play a role in enhancing the nucleophilicity of the glutathione thiol.

An interesting type of aliphatic dehalogenase is the enzyme (LinA) that is responsible for the first step in the bacterial degradation of lindane (α-hexachlorocyclohexane). In a peculiar reaction, HCl is eliminated, converting the substrate to pentachlorocyclohexene. A structure has not been solved, but a mechanism was predicted on basis of the stereochemistry of the reaction and low but significant sequence similarity to scytalone dehydratase. The reaction is proposed to involve abstraction of an axial proton, with concomitant anti-elimination of a *trans*-axial chloride from the adjacent carbon atom (Nagata *et al*., 2001; Trantirek *et al*., 2001). A conserved His-Asp pair was proposed to be involved in proton abstraction, but details about halogen binding and stabilization are still lacking.

If we consider aromatic compounds, it is again possible to distinguish several phylogenetically and mechanistically different dehalogenases. The carbon-halogen bond between a halogen and an arenic or vinylic carbon atom is more difficult to cleave than the one between a halogen and an sp^3-hybridized carbon atom. Therefore, most haloaromatics are dehalogenated after ring cleavage. However, a nucleophilic displacement of aromatic halogens occurs in 4-chlorobenzoate degradation. This is the fifth type of dehalogenase for which a structure has been solved (Benning *et al*., 1998). The compound is first activated by conjugation to coenzyme A, after which a hydrolytic dehalogenase (CbzA) that belongs to the enoyl hydratase superfamily, displaces the halogen by a nucleophilic addition-elimination mechanism. No distinct halide-binding site is present in the X-ray structure, but a nucleophilic aspartate and a histidine involved in catalysis were identified (Zhang *et al*., 2001). Another aromatic dehalogenase is tetrachlorohydroquinone dehalogenase. This enzyme catalyses the replacement of chlorine by hydrogen in tetra- and trichlorohydroquinone during the degradation of pentachlorophenol by *Sphingbium chlorophenolicum* (Kiefer and Copley, 2002). The protein (PcpC in Table 1) bears sequence and mechanistic similarity to glutathione transferases.

A specific hydrolytic dehalogenase (AtzA) is also involved in the bacterial degradation of atrazine. This enzyme is related to melamine deaminase (TriA) and the AtzA and TriA proteins differ only by nine amino acid substitutions (Seffernick *et al*., 2001).

Origin and distribution of dehalogenase gene sequences

Analysis of the genetic organization of biodegradation pathways provides insight into the genetic processes that led to their evolution. It appears that catabolic genes for xenobiotic compounds are often associated with transposable elements and insertion sequences. They are also fre-

© 2005 The Authors

quently located on transmissible plasmids. One striking example of a mobile element that has assisted catabolic genes in their dissemination is IS*1071*. This insertion element flanks the haloacetate dehalogenase gene dehH2 on plasmid pUO1 in *Moraxella* sp. strain B (Kawasaki *et al.*, 1992), the haloalkane dehalogenase gene *dhaA* on the chromosome in *P. pavonaceae* 170 (Poelarends *et al.*, 2000a), the atrazine degradative genes *atzA*, *atzB* and *atzC* on plasmid pADP-1 in *Pseudomonas* sp. ADP (Wackett, 2004), the aniline degradative genes on plasmid pTDN1 in *Pseudomonas putida* UCC22 (Fukumori and Saint, 1997), and presumably also the p-sulfobenzoate degradative genes on plasmids pTSA and pPSB in *Comamonas testosteroni* strains T-2 and PSB-4 respectively (Junker and Cook, 1997). These observations clearly indicate that gene mobilization between and within replicons is an important process during genetic adaptation. It also suggests that genes that are involved in biodegradation of xenobiotics were recruited from a 'pre-industrial' gene pool by integration, transposition, homologous recombination and mobilization. Association of dehalogenase sequences with mobile genetic elements has also been observed in other cases, for example with haloacetate dehalogenases (Slater *et al.*, 1985; Thomas *et al.*, 1992; van der Ploeg *et al.*, 1995), with dichloromethane dehalogenase (Schmid-Appert *et al.*, 1997), and with γ-hexachlorocyclohexane dehalogenase (Dogra *et al.*, 2004).

Although insight in the sequence of events that led to the current genetic make-up of biodegradation pathways is lacking, the general nature of some of the processes involved is understood. Less is known about the origin of the structural genes that encode critical enzymes, such as dehalogenases, and the degree of divergence that occurred during evolution of the current sequences. The possibility to rapidly evolve new enzyme selectivities by laboratory evolution is known since the 1970s, most notably through the work of P.H. Clarke and coworkers who showed that the substrate specificity of *Pseudomonas aeruginosa* amidase can be changed by mutagenesis and selection on plates (Betz *et al.*, 1974; Paterson and Clarke, 1979). This approach was called experimental enzyme evolution and is conceptually similar to directed evolution.

Theoretically, it is possible that a current gene for a specific critical (dehalogenase) reaction was already present in the pre-industrial gene pool. Alternatively, there could be a short evolutionary pathway that led from an unknown pre-existing gene to the gene as we currently find it in a biodegradation pathway. It has even been suggested that a new sequence for an enzyme acting on a synthetic compound could evolve through the activation of an unused alternative open reading frame of a pre-existing internal repetitious coding sequence (Ohno, 1984). Here it should be noted that the similarity of a dehalogenase to members of an enzyme superfamily that catalyse other reactions generally does not provide information about the process of adaptation to xenobiotic compounds. The level of sequence similarity that exists between a dehalogenase and other proteins in a phylogenetic family is usually less than 50% (Table 1). Therefore, the time of divergence should be much earlier than a century ago, and the process of divergence thus cannot be related to the introduction of industrial chemicals into the environment. If the dehalogenases and other critical enzymes that occur in catabolic pathways have undergone recent mutations, there should be closely related sequences in nature that differ from the current enzymes by only a few mutations. No such primitive dehalogenase has yet been detected, with the notable exception of TriA, the enzyme that dehalogenates the herbicide atrazine (vide infra).

Another issue is the function of the pre-industrial dehalogenase or dehalogenase-like sequences from which the current catabolic systems with their activated and mobilized genes originate. The original genes may have been involved in the dehalogenation of naturally occurring halogenated compounds, of which there are many (Gribble, 1998). Such proteins may fortuitously also have been active with a xenobiotic halogenated substrate, just because of their lack of substrate specificity. Alternatively, the evolutionary precursor of a dehalogenase may have catalysed a different reaction that has some mechanistic similarity to dehalogenation, in which case the original enzyme may or may not have shown some dehalogenation activity due to catalytic promiscuity. Finally, the precursor genes for dehalogenases may have been silent or cryptic genes, with no clear function for the pre-industrial host (Hall *et al.*, 1983). In all cases, the gene could have become functional in a dehalogenation pathway as a result of the fortuitous ability to catalyse dehalogenation of a xenobiotic compound, possibly after acquisition of some mutations.

One way to obtain information about the evolutionary origin of dehalogenase genes is to compare the dehalogenase sequences that have been detected in different bacterial cultures. If closely related sequences are present, this would make it possible to identify sequence differences and to determine the effect of the mutations on substrate selectivity. Another approach is to search for sequences that are closely related to dehalogenases in databases of sequenced genomes. These organisms have, with some recent exceptions, no known history of degradation of halogenated compounds. Currently the whole sequence of more than 300 bacterial genomes is available. If we find closely related sequences in these databases, they could define evolutionary ancestors of the current dehalogenases. A further potential source of new dehalogenase sequences could come from metagenome

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

libraries. Several of these libraries have been prepared, and some have been used for massive sequencing (Venter *et al*., 2004), whereas others have been explored for the presence of new biotransformation enzymes.

Analysis of sequence databases such as Pfam and of the literature revealed that in several cases the same dehalogenase sequence has repeatedly been detected in organisms that have been isolated from different geographical locations. Thus, identical or almost identical haloalkane dehalogenases, dichloromethane dehalogenases and atrazine chlorohydrolases have been detected in isolates from different areas (Table 1). This suggests that either these sequences have been recruited from the pre-industrial gene pool only once, after which they became distributed by horizontal transmission, or they have been repeatedly recruited from an identical pre-industrial sequence. By itself, the worldwide distribution of identical sequences is common in the bacterial world, as 100% sequence identity is normal for specific genes within different strains of the same species. However, it is currently unknown by what mechanisms a possible process of global distribution of recently evolved catabolic genes may occur.

The structural and mechanistic analysis of dehalogenases makes it possible to define sequence fingerprints that allow the identification of genes that are phylogenetically related in genomic databases. The current version of the NCBI databases lists more than 250 proteome sequences of bacteria, and 31 archaeal proteomes. When these were searched for 15 different dehalogenase sequences that have been detected in organisms isolated on halogenated compounds as a carbon source, it appeared that not a single closely related counterpart of dehalogenase genes was present in these sequenced bacterial genomes (Table 1). Even close homologues are absent, including for those sequences that have repeatedly been detected in identical form in independently isolated bacteria. For example, both dichloromethane dehalogenase and 1,2-dichloroethane dehalogenase have been detected in essentially identical form in more than 10 independently isolated bacterial cultures that originate from different geographical locations, but no close homologue is present in the sequenced microbial genomes. Similarly, Blast searches with dehalogenase sequences against the sequences of environmental DNA obtained in the Sargasso Sea sequencing project (Venter *et al*., 2004) again indicated that no close homologues are present in the 1.2 million new genes that were discovered in this massive DNA sequencing project (Table 1). Thus, sequences that have repeatedly been acquired by batch enrichment were not detected in the environmental DNA. Marchesi and Weightman (2003), working with haloacid dehalogenases (HAD superfamily), also described that a bias is introduced by enrichment and isolation. Both for group I dehalogenases and for group II (HAD) enzymes, the diversity and phylogenetic grouping of dehalogenase sequences obtained via pure cultures significantly deviated from what was observed by analysis of DNA directly isolated from the samples that were used for enrichment, such as activated sludge.

At the same time, it appears that more distantly related dehalogenase sequences are quite common. For example, large numbers of putative haloalkane dehalogenase and haloacid dehalogenase sequences are present in the whole-genome database and in the Sargasso Sea database (Table 1). The identification of these sequences is based on detailed insight in structure–function relationships and the fact that several homologous sequences in these protein families have been detected that indeed encode functional dehalogenases. The exact activity of the proteins encoded by putative dehalogenase sequences is difficult to predict. Jesenska and colleagues (2002) have shown that at least one active dehalogenase is encoded by the *M. tuberculosis* genome, but its physiological function is unclear. In the HAD superfamily of enzymes, at least eight different sequences have been found to encode functional haloacid dehalogenases, and the environmental diversity may be large (Marchesi and Weightman, 2003). In other cases, e.g. with CaaD and hexachlorocyclohexane dechlorinase (LinA), it is more difficult to predict which of the homologues do encode dehalogenating enzymes because the sequence similarity is low and the residues involved in dehalogenation have been less well identified.

Haloalkane dehalogenase from *Xanthobacter autotrophicus* (DhlA)

Haloalkane dehalogenase (DhlA) was originally discovered in *X. autotrophicus* GJ10, a nitrogen-fixing hydrogen bacterium that was enriched with 1,2-dichloroethane as the sole carbon source. Subsequently, identical DhlA-encoding genes have been discovered in several other strains of *X. autotrophicus*, isolated in the Netherlands and in Germany, and in isolates of *Ancylobacter aquaticus* obtained with 1,2-dichloroethane or chloroethylvinyl ether as the growth substrate (van den Wijngaard *et al*., 1992). Recently, a strain of *Xanthobacter flavus* was isolated in South Korea, and this organism also possessed an identical dehalogenase (Song *et al*., 2004). In fact, DhlA is still the only known hydrolytic haloalkane dehalogenase that operates in 1,2-dichloroethane degrading bacteria; no variants have been described, and at least 12 identical copies of this gene have been obtained from different environmental isolates. The question arises: where does this enzyme that has such a remarkable activity with a very stable xenobiotic compound come from? Did

© 2005 The Authors

it pre-exist at the time industrial production of 1,2-dichloroethane started, or did it evolve from an ancestor that converted some halogenated or non-halogenated compound?

Insight into this issue has been obtained by inspection of the wild-type sequence and by experimental evolution methods. The N-terminal part of the cap domain of the dehalogenase harbours two short tandem sequence repeats (at the DNA level: one perfect 15 bp repeat and one 9 bp repeat carrying one substitution). These repeats occurred in a segment of the protein that was observed to be the target of substitutions, deletions and generation of new repeats when mutants were selected that possessed an enhanced activity toward 1-chlorohexane (Pries *et al*., 1994). Thus, experimental evolution leads to duplications and other mutations in a part of the sequence where the wild-type already harbours sequence repeats. This observation led to the hypothesis that the sequence repeats in the wild-type *dhlA* gene were the result of recent evolutionary events that caused adaptation to 1,2-dichloroethane. An insertion is detected in the same region of the cap domain when a sequence alignment is performed between DhlA and its most similar homologue present in the bacterial genomes database, which independently confirms that the sequence encoding the N-terminal part of the cap domain is a target for adaptive mutations (Fig. 3).

Assuming that the short repeats present in the cap domain sequence of the wild-type dehalogenase were indeed generated recently in a pre-industrial dehalogenase, one can postulate a DhlA sequence for this hypothetical primitive dehalogenase (Pikkemaat and Janssen, 2002) (Fig. 3B). The primitive *dhlA* sequence predicted by this so-called retro genetics approach has been constructed *in vitro*, an it appeared to encode a protein with significant activity toward bromoalkanes, but was completely inactive toward 1,2-dichloroethane. Thus, the ancestral dehalogenase may have been a debrominating enzyme. Subsequently, we created with the incremental truncation method (ITCHY) a library of derivatives of the primitive DhlA that carried random direct repeats in the same region where the wild-type sequence harboured repeats. Some of the repeat-carrying variants indeed had evolved activity with 1,2-dichloroethane. All these mutations influenced the region of the cap domain of haloalkane dehalogenase proximal to the tryptophan that donates a hydrogen bond to the leaving halide (Trp175) (Pikkemaat and Janssen, 2002).

These results are all in agreement with the hypothesis that the current 1,2-dichloroethane dehalogenase has evolved by a short evolutionary pathway from a pre-existing (pre-industrial) haloalkane dehalogenase that was active with brominated but not with chlorinated compounds. Association with transmissible plasmids and other mobile genetic elements has facilitated the worldwide spread of the evolved dehalogenase.

Rhodococcus haloalkane dehalogenase (DhaA)

Another abundant type of haloalkane dehalogenase is the one from *Rhodococcus erythropolis* (DhaA). Mutually identical copies of the *dhaA* genes have been detected in different organisms, of which the taxonomy has been rather confusing due to errors in classification. The first *dhaA* sequence was determined by Kulakova and colleagues (1997), using a strain of *Rhodococcus* (*Rhodococcus rhodochrous* NCIMB 13064) that was isolated on 1-chlorobutane. Later sequencing of several other haloalkane dehalogenase genes from 1-chlorobutane-, 1-chlorohexane- and 1,6-dichlorohexane-degrading gram-positive organisms, some of which had been isolated before strain NCIMB 13064, showed that these possessed identical *dhaA* sequences. Poelarends and colleagues (2000b) have reclassified several organisms and using 16S rRNA gene sequencing they showed that *R. erythropolis* Y2 (England), *R. rhodochrous* NCIMB13064 (N. Ireland), *Corynebacterium* sp. strain m15 (Japan), *Arthrobacter* strain HA1 (Switzerland), strain GJ70 (the Netherlands, originally called *Acinetobacter*) and strain TB2 (USA) should all be classified as *R. erythropolis*. All these strains possessed the same haloalkane dehalogenase, and in all cases the dehalogenase gene was preceded by the same invertase gene sequence and a regulatory gene, and on the downstream side an alcohol dehydrogenase and an aldehyde dehydrogenase encoding gene. Thus, this catabolic cluster is highly conserved and distributed worldwide in closely related gram-positive bacteria. The structure of this DhaA type haloalkane dehalogenase was solved by X-ray crystallography, using a protein with two substitutions (A172V and A292G) compared with the original DhaA sequence of strain NCIMB 13064 (Newman *et al*., 1999). In a recent directed evolution study, the substrate specificity of DhaA has been modified to enhance conversion of the highly recalcitrant chemical 1,2,3-trichloropropopane (Bosma *et al*., 2002).

This reservoir of haloalkane dehalogenase-producing rhodococci may have been the source for catabolic pathways that are active with more exotic haloalkanes (Fig. 4). In the first place, enrichment of 1,2-dibromoethane degrading bacteria has yielded, after much patience, a culture that can slowly grow with 1,2-dibromoethane as sole carbon source. This *Mycobacterium* strain produces a haloalkane dehalogenase that is almost identical to DhaA, but there are three substitutions (C176F, P248S, Y272F), and, most remarkably, on the C-terminal side the enzyme is 14 amino acids longer due to an in-frame fusion of the 3′ end of the dehalogenase gene with 42 bases that

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

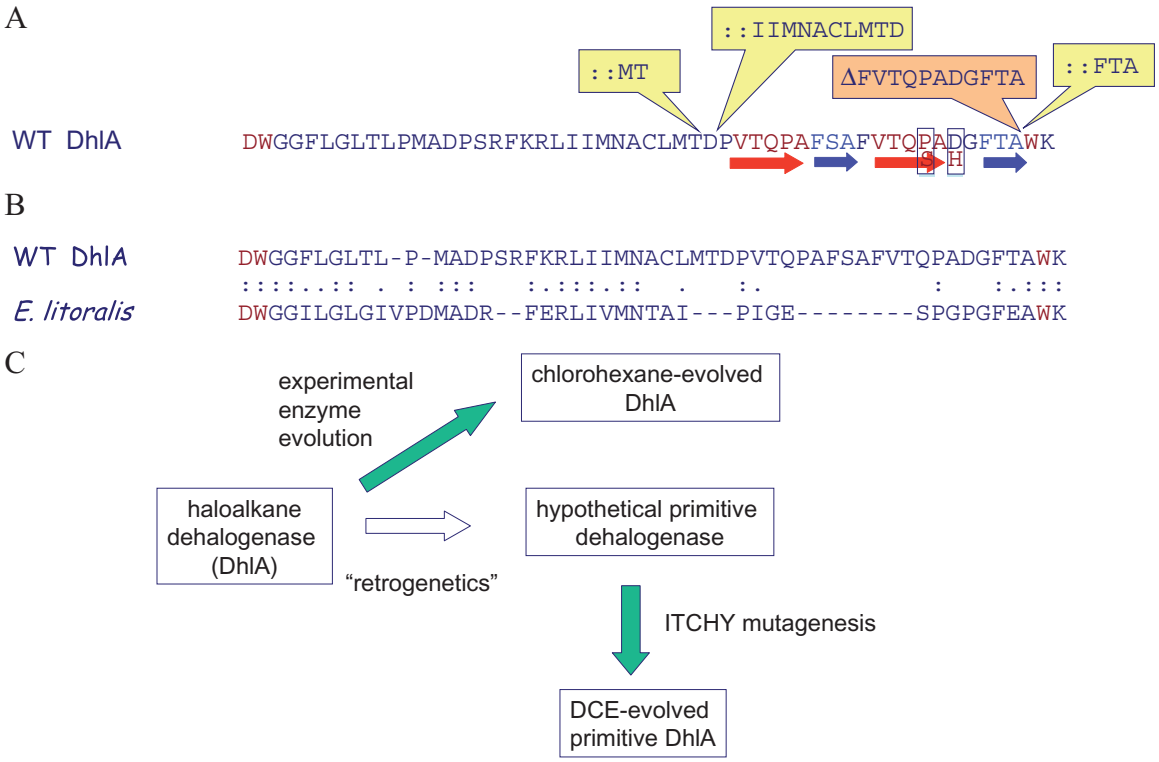


Fig. 3. Evolution of 1,2-dichloroethane dehalogenase activity. The N-terminal part of the cap domain of wild-type DhlA harbours two short tandem sequence repeats (shown as arrows), which might be signs of recent genetic adaptation to 1,2-dichloroethane. This idea is supported by the following observations. First, as shown in panel A, changes are observed in this part of the cap domain (shown in boxes, three tandem duplications, a large deletion and two substitutions) when DhlA is forced to evolve dehalogenase activity toward 1-chlorohexane, a substrate not used by the wild-type enzyme (experimental enzyme evolution; Pries *et al.*, 1994). Second, in a pairwise sequence alignment the closest homologue of DhlA in the database (a putative protein from *Erythrobacter litoralis* HTCC2594) shows a gap precisely in this region of the cap domain (panel B). Third, from the current 1,2-dichloroethane dehalogenase sequence (DhlA) a pre-industrial sequence (primitive) was proposed and constructed. The encoded protein is inactive with 1,2-dichloroethane (DCE), but does convert 1,2-dibromoethane (DBE) (panel C). When the primitive dehalogenase is subjected to ITCHY mutagenesis, a technique that allows the introduction of random repeats and deletions in the N-terminal part of the cap domain, some mutants carrying repeats (D2, IC12 and 3B2) had evolved 1,2-dichloroethane dehalogenase activity (Pikkemaat and Janssen, 2002) (panel D). On the basis of these findings, we propose that a primitive dehalogenase with a shorter stretch of sequence in the N-terminal region of the cap (as observed in the *E. litoralis* putative dehalogenase) was recruited from the pre-industrial environmental gene pool, and evolved into the current DhlA by a short evolutionary pathway that included generation of short duplications and substitutions.

encode the last 14 amino acids of a halohydrin dehalogenase (*hheB* gene) (Poelarends *et al.*, 1999). Thus, this chimeric dehalogenase is 307 amino acids long instead of the 293 amino acids of the standard *Rhodocccus* enzyme. The fusion does not seem to be necessary for the catalytic activity of the protein: 1,2-dibromoethane is an excellent substrate for both DhaA variants. Instead, it happens to be there as a fortuitous result of a gene fusion event that occurred during the evolution of the haloalkane dehalogenase gene region.

Other striking differences between the organization of the *Mycobacterium* GP1 *dhaA* region and that of the original *Rhodococcus* exist in the regulatory gene that represses DhaA formation in *R. erythropolis* when no

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

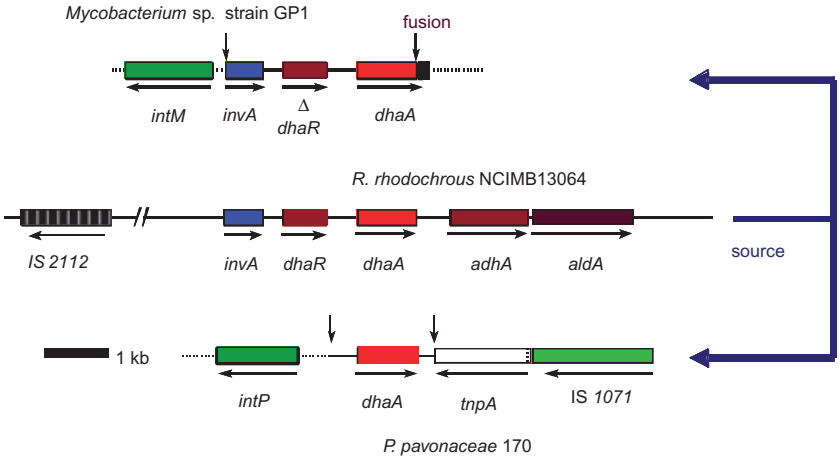


Fig. 4. Evolution and distribution of DhaA-type haloalkane dehalogenase sequences. The proposed starting point is the widely distributed gene cluster detected in rhodococci (middle). Adaptation to 1,2-dibromoethane (*Mycobacterium* GP1) involves inactivation of the regulatory gene (*dhaR*), association with an integrase sequence (*intM*), and loss of the alcohol and aldehyde dehydrogenase genes (*adhA* and *aldA* respectively). Recombination sites are shown by vertical arrows. Somewhere during these processes, a fortuitous fusion with a segment of a *hheB* gene has occurred. Adaptation to 1,3-dichloropropene (*P. pavonaceae*) is proposed to involve loss of the regulatory gene, loss of the dehydrogenase genes, association with an integrase type sequence (*intP*) and with an insertion element carrying a transposase gene (IS*1071*), and mobilization to the genome of a gram-negative *Pseudomonas* (Poelarends *et al.*, 2000a).

inducer is present (Poelarends *et al.*, 2000a). This regulatory gene is inactivated in *Mycobacterium* strain GP1 by a small deletion, which is required to allow gene expression because 1,2-dibromoethane does not act as an inducer. Here, we see another sign of recent pathway evolution: constitutive expression of a structural gene due to the absence of a functional regulatory gene. Dehalogenation of a xenobiotic organohalogen compound requires a catalytic protein that can convert the compound: the dehalogenase. Regulated expression requires a second protein that recognizes and responds to the xenobiotic chemical: the regulatory protein. Constitutive expression of dehalogenase genes has been demonstrated for atrazine chlorohydrolase (AtzA), 1,2-dichloroethane dehalogenase (DhlA), L-2-haloacid dehalogenase (DhlB), *trans*-3-CaaD, γ-hexachlorocyclohexane dehydrochlorinase (LinA) and 1,3,4,6-tetrachloro-1,4,-cyclohexadiene chlorohydrolase (LinB), suggesting that evolution is recent and has not yet led to systems for regulating gene expression.

Also in the 1,3-dichloropropene degrader *Pseudomonas pavonaceae* a *dhaA*-type haloalkane dehalogenase gene is present, indicating that it has recently been transferred from the Gram-positive *Rhodococcus* to a Gram-negative bacterium. In the *Pseudomonas* the whole regulatory gene has been lost (Fig. 4; Poelarends *et al.*, 2000a). Furthermore, the alcohol dehydrogenase and aldehyde dehydrogenase genes have disappeared. However, the haloalkane dehalogenase gene is associated with a putative integrase sequence, which is also the case in *Mycobacterium* sp. GP1. It is well possible that gene acquisition by integrases plays an important role during acquisition of new catabolic gene clusters for xenobiotic compounds. It is tempting to speculate that the fortuitous fusion that we see between the *dhaA* gene and a segment of the *hheB* gene in strain GP1 has something to do with the activity of an integron-like gene acquisition system.

Atrazine chlorohydrolase (AtzA)

The bacterial degradation of the herbicide atrazine by *Pseudomonas* ADP starts with a hydrolytic dechlorination, catalysed by an enzyme called atrazine chlorohydrolase (AtzA) (Fig. 5). Identical atrazine chlorohydrolase genes have been obtained from different sources, including an *Arthrobacter* from China (Cai *et al.*, 2003), strains from France (Rousseaux *et al.*, 2001), an undescribed β-proteobacterium strain CDB21 from Japan of which the sequence was deposited by Iwasaki and colleagues (EMBL Accession number AB194097), and four other isolates from different locations including an *Alcaligenes*, a *Ralstonia* and an *Agrobacterium* sp. (de Souza *et al.*, 1998). The protein belongs to the amidohydrolase superfamily, which also houses the next two enzymes of the atrazine catabolic pathway: hydroxyatrazine ethylaminohydrolase (AtzB) and N-isopropylammelide N-isopropylaminohydrolase (AtzC) (de Souza *et al.*, 1996). Other members of the amidohydrolase superfamily are triazine deaminase, hydantoinase, melamine deaminase, cytosine deaminase and phosphotriesterase. All these $(\alpha\beta)_8$ barrel proteins show rather low sequence identities in pairwise comparisons, with the exception of AtzA and melamine

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

(2,4,6-triamino-1,3,5-triazine) deaminase (TriA), which are 98% identical (Seffernick *et al*., 2001). Thus, AtzA and TriA are closely related, indicating very recent evolutionary divergence. Both atrazine and melamine are xenobiotic compounds, but melamine has been in use for a longer period of time than atrazine. Accordingly, it was proposed that TriA may have functioned as a precursor for AtzA. The proteins differ only by nine out of 475 amino acids, and using gene shuffling a number of hybrids has been obtained from which determinants of the substrate specificity could be identified (Raillard *et al*., 2001). It appeared that a wide range of activities could be obtained in such a library of hybrids, but leaving group specificity was mainly determined by residue 328. An Asn at this position favoured dechlorination, whereas an Asp gave an enzyme that could hydrolytically displace a broader range of substituents from the 2-position of 2-substituted 4,6-dialkylamino-triazines, albeit with a relatively low catalytic rate. Dechlorination activity may thus have evolved from deamination activity.

An overlap in these two activities was indeed observed for another member of the amidohydrolase superfamily, AtzB, which is the second enzyme in the normal atrazine degradation pathway (Fig. 5). AtzB is catalytically promiscuous and can not only catalyse a deamination reaction, but also is able to catalyse the dechlorination of the substrate analogue 2-chloro-4-amino-6-hydroxy-s-triazine, which is an intermediate in an alternative atrazine degradation pathway (Fig. 5) (Boundy-Mills *et al*., 1997).

These observations suggest that atrazine hydrolase and melamine deaminase very recently diverged from a pre-existing hydrolase that mainly acted as a deaminase or may have had both dehalogenase and deaminase activity. It was proposed that the possibility to achieve variations in substrate range in a key enzyme by an organism is a measure of the potential to recruit new catabolic activities (Wackett, 2004. In general, catalytic promiscuity could play an important role in the evolution of new enzyme activities (Aharoni *et al*., 2005).

Alkane hydroxylase (AlkB, AlkM)

Aliphatic hydrocarbons are introduced into the environment in large quantities both by human activities and by natural processes. Plants, for instance, can produce (odd-length) n-alkanes as part of mixtures of waxes. Thus, they cannot be regarded as xenobiotic compounds, even though most cases of contamination of surface soils with high levels of alkanes are caused by industrial processing of petroleum. In many different environments bacteria have been exposed to these compounds, and one would expect that evolution of enzymes that can hydroxylate alkanes has occurred over a longer period of time than with dehalogenases that act on exotic compounds.

The degradation of alkanes usually starts with the oxidation of the terminal carbon atom yielding an alcohol. Conversion of medium- and long-chain alkanes is often performed by integral membrane hydroxylases of which the best described member is AlkB of *P. putida* GPo1 (also referred to as *Pseudomonas oleovorans* GPo1) that converts C_5 to C_{12} alkanes. Both the sequence and genetic organization of the gene encoding AlkB, as well as the biochemical characteristics of the protein have been studied extensively (van Beilen *et al*., 1994). AlkB contains six membrane-spanning helices and forms an ω-hydroxylase system that also consists of a rubredoxin (AlkG) and a rubredoxin reductase (AlkT) involved in electron transport. AlkB belongs to a large superfamily of proteins that also includes non-haem integral membrane desaturases, epoxidases, acetylenases, conjugases, ketolases, decarbonylases and methyl oxidases. These proteins all contain eight conserved histidines that are located on the cytoplasmic site of the protein and are essential for positioning

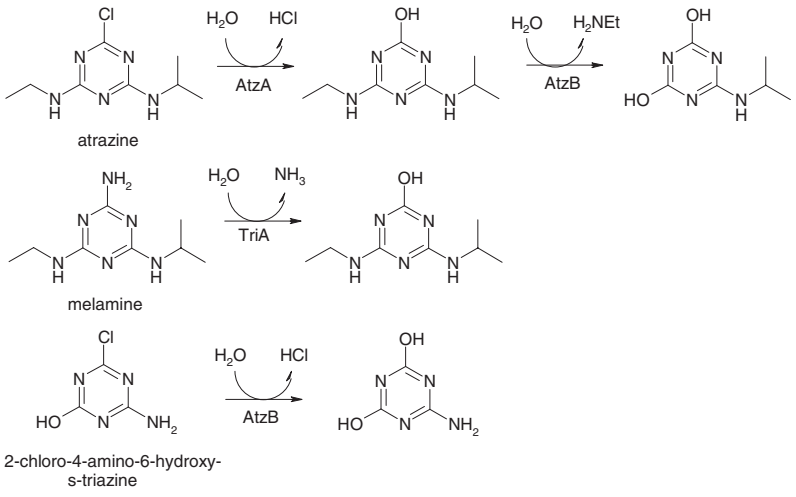


Fig. 5. Degradation of atrazine by AtzA and AtzB. These hydrolases both are members of the amidohydrolase superfamily. AtzA is highly similar to TriA, the enzyme that deaminates the amino analogue of atrazine called melamine. There are only nine substitutions and a short evolutionary pathway may be involved in adaptation of TriA to atrazine (Seffernick *et al*., 2001). AtzB catalyses the second step in atrazine degradation, which is a deamination reaction, but it can also catalyse a dechlorination reaction with the atrazine soil metabolite 2-chloro-4,6-diamino-s-triazine to yield ammelide (Boundy-Mills *et al*., 1997). These observations suggest that dechlorination and deamination activities are evolutionarily closely related.

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

of the Fe ions in the di-iron active site. The conserved histidines are situated in three motifs ($HX_{3-4}H$, $HX_{2-3}HH$ and $H/QX_{2-3}HH$) and are here referred to as motif A, B and D (Shanklin *et al*., 1994). In alkane hydroxylases an additional histidine-containing motif (NYXEHYG), referred to as motif C, was identified (Smits *et al*., 1999). The genes encoding alkane oxidation in *P. putida* GPo1 are located on the OCT-plasmid in two operons. The *alkBF-GHJKL* operon encodes the alkane hydroxylase (AlkB), two rubredoxins, an aldehyde dehydrogenase, an alcohol dehydrogenase, an acyl-CoA synthetase and an outer membrane protein of unknown function. The *alkST* operon encodes a regulator of expression and a rubredoxin reductase.

A similar well-characterized alkane hydroxylase system is present in *Acinetobacter* sp. strain ADP1 (Ratajczak *et al*., 1998). It converts long-chain alkanes ranging from C_{12} to C_{16} and the sequence has the same conserved motifs for iron binding. As just mentioned, the *alk* genes of strain GPo1 are plasmid-localized, but in strain ADP1 the alkane hydroxylase genes are present at different locations on the chromosome. This likely correlates with the presence or absence of selection pressure in the environment from which the organisms were isolated: strain GPo1 was isolated from oil, while strain ADP1 has no record of exposure to high levels of alkanes before or during its isolation. On the other hand, both AlkB and AlkM expression are regulated, which is evolutionary more advanced that what was observed with several dehalogenases, which are often constitutively expressed.

Using the *alkB* gene sequence it was possible to develop probes that could be applied for detecting genes related to *alkB* in bacteria isolated from oil-contaminated environments or from uncontaminated soils. Many positives have been found, in some cases with a high sequence identity (> 70%) to the *alkB* gene (Sotsky *et al*., 1994; Smits *et al*., 1999; Vomberg and Klinner, 2000; Whyte *et al*., 2002a). One of the alkane hydroxylase sequences obtained in this way belongs to *Pseudomonas aerofaciens* and shows 95% identity to AlkB. Other homologues of *alkB* were detected in known alkane-degrading strains such as *P. putida* P1 and *Alcanivorax borkumensis* AP1. Thus, the *alkB* gene or close homologues appear to occur quite often, and nine different alkane-degrading cultures have been described to possess highly similar alkane hydroxylases (Table 1). For the *P. aeruginosa* PAO1 *alkB* variant more than 10 identical sequences were found in *P. aeruginosa* strains from clinical and soil samples.

Alkane hydroxylase systems that are less related to AlkB or AlkM also occur, for example in gram-positive bacteria. van Beilen and colleagues (2002) detected quite diverse alkane hydroxylases in organisms from a trickle bed reactor that degraded alkanes. Diversity was also observed by Whyte and colleagues (2002b), who studied the hydroxylase system of an alkane-degrading *R. rhodochrous* and a *R. erythropolis* strain. The organisms contain three to five distinguished alkane hydroxylase systems that all are only distantly related to AlkB, but almost identical between the two different *Rhodococcus* strains. There are also variations in genetic organization, and one of the *Rhodococcus* systems has all three alkane hydroxylase components clustered in a single putative operon.

Whole genome sequencing of bacteria revealed a significant number of putative alkane hydroxylases. Both for AlkB of *P. putida* GPo1 and for AlkM, hits were obtained with sequence identity percentages up to 50%. Using growth complementation assays in different hosts, the functionality of more than 18 different putative alkane hydroxylases was proven, including those from *Burkholderia cepacia* RR10, *P. aeruginosa* PAO1 and *Mycobacterium tuberculosis* H37Rv (van Beilen *et al*., 1994; 2003; Smits *et al*., 2002). When searching for putative alkane hydroxylase genes in environmental DNA (Venter *et al*., 2004), a large number of *alkB* and *alkM* homologues were again detected, including two sequences that were 82% similar to *alkB*. These are the closest homologues of any of the sequences tested in Table 1. Apparently, alkane hydroxylases are highly abundant in the microbial population in the Sargasso Sea. Of these alkane hydroxylases, apart from the two hits just mentioned, only a few were closely related to AlkB or AlkM or other *alk* homologues that have been detected in isolates. This suggests that a large number of alkane hydroxylase sequences are present in the environment. The conservation of the catalytically important motifs described above suggests that most of these sequences encode functional enzymes.

The above observations indicate that the diversity of functional alkane hydroxylases is much broader than the diversity of dehalogenases that act on xenobiotic substrates. Thus, whereas very few solutions have evolved for the incorporation of dehalogenase genes in catabolic pathways for nematocides such as 1,2-dibromoethane and 1,2-dichloroethane, several different alkane hydroxylase genes have been recruited in operons for alkane degradation. This correlates with the more widespread occurrence of alkane degraders and the easier degradability of n-alkanes as compared with many halogenated compounds.

Conclusions: accession and exploration of sequence space

One of the striking observations concerning the distribution and evolution of key catabolic genes is that identical sequences have repeatedly been detected in organisms that are enriched on xenobiotic halogenated substrates as

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

a carbon source. This holds for dichloromethane dehalogenase (DcmA), haloalkane dehalogenases (DhlA, DhaA, LinB) and atrazine chlorohydrolase (AtzA). Probably, the number of solutions that nature has found to degrade these compounds is very small, and horizontal distribution occurs faster than generation of new pathways. Indeed, the dehalogenase genes are often associated with integrase genes, invertase genes, or insertion elements, and they are usually localized on mobile plasmids. Recent mobilization of a dehalogenase gene across the gram-border has been suggested in at least one case.

It remains unknown to what degree the current active dehalogenases differ from their proposed recent evolutionary ancestors. Only for 1,2-dichloroethane dehalogenase and atrazine chlorohydrolase short evolutionary pathways have been suggested that could describe the generation of a functional dehalogenase from a closely related precursor sequence that in pre-industrial times may have acted on a different substrate. The diversity of such specialized dehalogenases that act on xenobiotic compounds seems to be more restricted than that of enzymes converting alkanes, which are easier to degrade and for which organisms are more common.

Even though the same dehalogenase sequences are detected in organisms that are isolated in different geographical areas, and in some cases even on different substrates, they are not the most abundant dehalogenase sequences identified in whole genome sequencing projects and massive random sequencing. Thus, it appears that enrichment techniques explore a different segment of sequence space than massive sequencing of environmental DNA. The presence of large numbers of unexplored functional sequences in genomic databases suggests that the biotransformation scope of microbial systems has an enormous potential for further growth.

Acknowledgements

We thank the present and former colleagues of our labs for contributing with discussions and experimental work to our research on dehalogenase mechanisms and enzyme evolution.

References

van Agteren, M.H., Keuning, S., and Janssen, D.B. (1998) *Handbook on Biodegradation and Biological Treatment of Hazardous Organic Compounds.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Aharoni, A., Gaidukov, L., Khersonsky, O.McQ., Gould, S., Roodveldt, C., and Tawfik, D.S. (2005) The 'evolvability' of promiscuous protein functions. *Nat Genet* **37:** 73–76.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Altschul, S.F., Madden, T.L., Schaffer, 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.

Azurmendi, H.F., Wang, S.C., Massiah, M.A., Poelarends, G.J., Whitman, C.P., and Mildvan, A.S. (2004) The roles of active-site residues in the catalytic mechanism of *trans*-3-chloroacrylic acid dehalogenase: a kinetic, NMR, and mutational analysis. *Biochemistry* **43:** 4082–4091.

van Beilen, J.B., Wubbolts, M.G., and Witholt, B. (1994) Genetics of alkane oxidation by *Pseudomonas oleovorans*. *Biodegradation* **5:** 161–174.

van Beilen, J.B., Smits, T.H., Whyte, L.G., Schorcht, S., Rothlisberger, M., Plaggemeier, T., *et al.* (2002) Alkane hydroxylase homologues in Gram-positive strains. *Environ Microbiol* **4:** 676–682.

van Beilen, J.B., Li, Z., Duetz, W.A., Smits, T.H.M., and Witholt, B. (2003) Diversity of alkane hydroxylase systems in the environment. *Oil Gas Sci Technol* **58:** 427–440.

Benning, M.M., Wesenberg, G., Liu, R., Taylor, K.L., Dunaway-Mariano, D., and Holden, H.M. (1998) The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. strain CBS-3. *J Biol Chem* **273:** 33572–33579.

Betz, J.L., Brown, P.R., Smyth, M.J., and Clarke, P.H. (1974) Evolution in action. *Nature* **247:** 261–264.

Bosma, T., Damborsky, J., Stucki, G., and Janssen, D.B. (2002) Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. *Appl Environ Microbiol* **68:** 3582–3587.

Boundy-Mills, K.L., de Souza, M.L., Mandelbaum, R.T., Wackett, L.P., and Sadowsky, M.J. (1997) The *atzB* gene of *Pseudomonas* sp. strain ADP encodes the second enzyme of a novel atrazine degradation pathway. *Appl Environ Microbiol* **63:** 916–923.

Cai, B., Han, Y., Liu, B., Ren, Y., and Jiang, S. (2003) Isolation and characterization of an atrazine-degrading bacterium from industrial wastewater in China. *Lett Appl Microbiol* **36:** 272–276.

Dogra, C., Raina, V., Pal, R., Suar, M., Lal, S., Gartemann, K.H., *et al.* (2004) Organization of *lin* genes and IS6100 among different strains of hexachlorocyclohexane-degrading *Sphingomonas paucimobilis*: evidence for horizontal gene transfer. *J Bacteriol* **186:** 2225–2235.

Fukumori, F., and Saint, C.P. (1997) Nucleotide sequences and regulational analysis of genes involved in conversion of aniline to catechol in *Pseudomonas putida* UCC22 (pTDN1). *J Bacteriol* **179:** 399–408.

Gribble, G.W. (1998) Naturally occurring organohalogen compounds. *Acc Chem Res* **31:** 141–152.

Hall, B.G., Yokoyama, S., and Calhoun, D.H. (1983) Role of cryptic genes in microbial evolution. *Mol Biol Evol* **1:** 109–124.

Hisano, T., Hata, Y., Fujii, T., Liu, J.Q., Kurihara, T., Esaki, N., and Soda, K. (1996) Crystal structure of L-2-haloacid dehalogenase from *Pseudomonas* sp. YL. An alpha/beta hydrolase structure that is different from the alpha/beta hydrolase fold. *J Biol Chem* **271:** 20322–20330.

van Hylckama Vlieg, J.E.T., Poelarends, G.J., Mars, A.E.,

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

and Janssen, D.B. (2000) Detoxification of reactive intermediates during microbial metabolism of halogenated compounds. *Curr Opin Microbiol* **3:** 257–262.

van Hylckama Vlieg, J.E.T., Tang, L., Lutje Spelberg, J.H., Smilda, T., Poelarends, G.J., Bosma, T., *et al.* (2001) Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases. *J Bacteriol* **183:** 5058–5066.

Janssen, D.B. (2004) Evolving haloalkane dehalogenases. *Curr Opin Chem Biol* **8:** 150–159.

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.

de Jong, R.M., and Dijkstra, B.W. (2003) Structure and mechanism of bacterial dehalogenases: different ways to cleave a carbon-halogen bond. *Curr Opin Struct Biol* **13:** 722–730.

de Jong, R.M., Tiesinga, J.J., Rozeboom, H.J., Kalk, K.H., Tang, L., Janssen, D.B., and Dijkstra, B.W. (2003) Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD (P) H binding site. *EMBO J* **22:** 4933–4944.

de Jong, R.M., Brugman, W., Poelarends, G.J., Whitman, C.P., and Dijkstra, B.W. (2004) The X-ray structure of trans-3-chloroacrylic acid dehalogenase reveals a novel hydration mechanism in the tautomerase superfamily. *J Biol Chem* **279:** 11546–11552.

Junker, F., and Cook, A.M. (1997) Conjugative plasmids and the degradation of arylsulfonates in *Comamonas testosteroni*. *Appl Environ Microbiol* **63:** 2403–2410.

Kawasaki, H., Tsuda, K., Matsushita, I., and Tonomura, K. (1992) Lack of homology between two haloacetate dehalogenase genes encoded on a plasmid from *Moraxella* sp. strain B. *J Gen Microbiol* **138:** 1317–1323.

Kiefer, P.M., and Copley, S.D. (2002) Characterization of the initial steps in the reductive dehalogenation catalyzed by tetrachlorohydroquinone dehalogenase. *Biochemistry* **41:** 1315–1322.

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.

Leisinger, T. (1996) Biodegradation of chlorinated aliphatic compounds. *Curr Opin Biotechnol* **7:** 295–300.

Leisinger, T., Bader, R., Hermann, R., Schmid-Appert, M., and Vuilleumier, S. (1994) Microbes, enzymes and genes involved in dichloromethane utilization. *Biodegradation* **5:** 237–248.

Marchesi, J.R., and Weightman, A.J. (2003) Comparing the dehalogenase gene pool in cultivated alpha-halocarboxylic acid-degrading bacteria with the environmental metagene pool. *Appl Environ Microbiol* **69:** 4375–4382.

Nagata, Y., Mori, K., Takagi, M., Murzin, A.G., and Damborsky, J. (2001) Identification of protein fold and catalytic residues of gamma-hexachlorocyclohexane dehydrochlorinase LinA. *Proteins* **45:** 471–477.

Newman, J., Peat, T.S., Richard, R., Kan, L., Swanson, P.E., Affholter, J.A., *et al.* (1999) Haloalkane dehalogenases: structure of a *Rhodococcus* enzyme. *Biochemistry* **38:** 16105–16114.

Ohno, S. (1984) Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence. *Proc Natl Acad Sci USA* **81:** 2421–2425.

Paterson, A., and Clarke, P.H. (1979) Molecular basis of altered enzyme specificities in a family of mutant amidases from *Pseudomonas aeruginosa*. *J Gen Microbiol* **114:** 75–85.

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–E35.

van der Ploeg, J., Willemsen, M., van Hall, G., and Janssen, D.B. (1995) Adaptation of *Xanthobacter autotrophicus* GJ10 to bromoacetate due to activation and mobilization of the haloacetate dehalogenase gene by insertion element IS*1247*. *J Bacteriol* **177:** 1348–1356.

Poelarends, G.J., and Whitman, C.P. (2004) Evolution of enzymatic activity in the tautomerase superfamily: mechanistic and structural studies of the 1,3-dichloropropene catabolic enzymes. *Bioorg Chem* **32:** 376–392.

Poelarends, G.J., van Hylckama Vlieg, J.E.T., Marchesi, J.R., Freitas dos Santos, L.M., and Janssen, D.B. (1999) Degradation of 1,2-dibromoethane by *Mycobacterium* sp. strain GP1. *J Bacteriol* **181:** 2050–2058.

Poelarends, G.J., Kulakov, L.A., Larkin, M.J., van Hylckama Vlieg, J.E.T., and Janssen, D.B. (2000a) Roles of horizontal gene transfer and gene integration in evolution of 1,3 dichloropropene- and 1,2 dibromoethane-degradative pathways. *J Bacteriol* **182:** 2191–2199.

Poelarends, G.J., Zandstra, M., Bosma, T., Kulakov, A., Larkin, M.J., Marchesi, J.R., *et al.* (2000b) Haloalkane-utilizing *Rhodococcus* strains isolated from geographically distinct locations possess a highly conserved gene cluster encoding haloalkane catabolism. *J Bacteriol* **182:** 2725–2731.

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.

Raillard, S., Krebber, A., Chen, Y., Ness, J.E., Bermudez, E., Trinidad, R., *et al.* (2001) Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. *Chem Biol* **8:** 891–898.

Ratajczak, A., Geissdorfer, W., and Hillen, W. (1998) Alkane hydroxylase from *Acinetobacter* sp. strain ADP1 is encoded by *alkM* and belongs to a new family of bacterial integral-membrane hydrocarbon hydroxylases. *Appl Environ Microbiol* **64:** 1175–1179.

Ridder, I.S., Rozeboom, H.J., Kalk, K.H., and Dijkstra, B.W. (1999) Crystal structures of intermediates in the dehalogenation of haloalkanoates by L-2-haloacid dehalogenase. *J Biol Chem* **274:** 30672–30678.

Rousseaux, S., Hartmann, A., and Soulas, G. (2001) Isolation and characterisation of new Gram-negative and Gram-positive atrazine degrading bacteria from different French soils. *FEMS Microbiol Ecol* **36:** 211–222.

Schmid-Appert, M., Zoller, K., Traber, H., Vuilleumier, S., and Leisinger, T. (1997) Association of newly discovered IS

© 2005 The Authors
Journal compilation © 2005 Society for Applied Microbiology and Blackwell Publishing Ltd,

elements with the dichloromethane utilization genes of methylotrophic bacteria. *Microbiology* **143:** 2557–2567.

Seffernick, J.L., de Souza, M.L., Sadowsky, M.J., and Wackett, L.P. (2001) Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different. *J Bacteriol* **183:** 2405–2410.

Shanklin, J., Whittle, E., and Fox, B.G. (1994) Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry* **33:** 12787–12794.

Slater, J.H., Weightman, A.J., and Hall, B.G. (1985) Dehalogenase genes of *Pseudomonas putida* PP3 on chromosomally located transposable elements. *Mol Biol Evol* **2:** 557–567.

Smits, T.H., Rothlisberger, M., Witholt, B., and van Beilen, J.B. (1999) Molecular screening for alkane hydroxylase genes in Gram-negative and Gram-positive strains. *Environ Microbiol* **1:** 307–317.

Smits, T.H., Balada, S.B., Witholt, B., and van Beilen, J.B. (2002) Functional analysis of alkane hydroxylases from gram-negative and gram-positive bacteria. *J Bacteriol* **184:** 1733–1742.

Song, J.S., Lee, D.H., Lee, K., and Kim, C.K. (2004) Genetic organization of the *dhlA* gene encoding 1,2-dichloroethane dechlorinase from *Xanthobacter flavus* UE15. *J Microbiol* **42:** 188–193.

Sotsky, J.B., Greer, C.W., and Atlas, R.M. (1994) Frequency of genes in aromatic and aliphatic hydrocarbon biodegradation pathways within bacterial populations from Alaskan sediments. *Can J Microbiol* **40:** 981–985.

de Souza, M.L., Sadowsky, M.J., and Wackett, L.P. (1996) Atrazine chlorohydrolase from *Pseudomonas* sp. strain ADP: gene sequence, enzyme purification, and protein characterization. *J Bacteriol* **178:** 4894–4900.

de Souza, M.L., Seffernick, J., Martinez, B., Sadowsky, M.J., and Wackett, L.P. (1998) The atrazine catabolism genes *atzABC* are widespread and highly conserved. *J Bacteriol* **180:** 1951–1954.

Stourman, N.V., Rose, J.H., Vuilleumier, S., and Armstrong, R.N. (2003) Catalytic mechanism of dichloromethane dehalogenase from *Methylophilus* sp. strain DM11. *Biochemistry* **42:** 11048–11056.

Thomas, A.W., Slater, J.H., and Weightman, A.J. (1992) The dehalogenase gene *dehI* from *Pseudomonas putida* PP3 is carried on an unusual mobile genetic element designated DEH. *J Bacteriol* **174:** 1932–1940.

Trantirek, L., Hynkova, K., Nagata, Y., Murzin, A., Ansorgova, A., Sklenar, V., and Damborsky, J. (2001) Reaction mechanism and stereochemistry of gamma-hexachlorocyclohexane dehydrochlorinase LinA. *J Biol Chem* **276:** 7734–7740.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

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.

Vomberg, A., and Klinner, U. (2000) Distribution of *alkB* genes within n-alkane degrading bacteria. *J Appl Microbiol* **89:** 339–348.

Vuilleumier, S., and Leisinger, T. (1996) Protein engineering studies of dichloromethane dehalogenase/ glutathione S-transferase from *Methylophilus* sp. strain DM11. Ser12 but not Tyr6 is required for enzyme activity. *Eur J Biochem* **239:** 410–417.

Wackett, L.P. (2004) Evolution of enzymes for the metabolism of new chemical inputs into the environment. *J Biol Chem* **279:** 41259–41262.

Whyte, L.G., Schultz, A., van Beilen, J.B., Luzc, A.P., Pellizaric, V., Labbé, D., and Greer, C.W. (2002a) Prevalence of alkane monooxygenase genes in Arctic and Antarctic hydrocarbon-contaminated and pristine soils. *FEMS Microbiol Ecol* **41:** 141–150.

Whyte, L.G., Smits, T.H., Labbe, D., Witholt, B., Greer, C.W., and van Beilen, J.B. (2002b) Gene cloning and characterization of multiple alkane hydroxylase systems in *Rhodococcus* strains Q15 and NRRL B-16531. *Appl Environ Microbiol* **68:** 5933–5942.

van den Wijngaard, A.J., van der Kamp, K.W., van der Ploeg, J., Pries, F., Kazemier, B., and Janssen, D.B. (1992) Degradation of 1,2-dichloroethane by *Ancylobacter aquaticus* and other facultative methylotrophs. *Appl Environ Microbiol* **58:** 976–983.

Zhang, W., Wei, Y., Luo, L., Taylor, K.L., Yang, G., Dunaway-Mariano, D., *et al.* (2001) Histidine 90 function in 4-chlorobenzoyl-coenzyme a dehalogenase catalysis. *Biochemistry* **40:** 13474–13482.

© 2005 The Authors