Diversity and Biocatalytic Potential of Epoxide Hydrolases Identified by Genome Analysis†

Bert van Loo,1‡ Jaap Kingma,1 Michael Arand,2 Marcel G. Wubbolts,3 and Dick B. Janssen1* Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, The Netherlands; Institute of Pharmacology and Toxicology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; and DSM Pharma Chemicals, Advanced Synthesis, Catalysis & Development, DSM Research, P.O. Box 18, 6160 MD Geleen, The Netherlands

Received 5 September 2005/Accepted 19 January 2006

Epoxide hydrolases play an important role in the biodegradation of organic compounds and are potentially useful in enantioselective biocatalysis. An analysis of various genomic databases revealed that about 20% of sequenced organisms contain one or more putative epoxide hydrolase genes. They were found in all domains of life, and many fungi and actinobacteria contain several putative epoxide hydrolase-encoding genes. Multiple sequence alignments of epoxide hydrolases with other known and putative α/β-hydrolase fold enzymes that possess a nucleophilic aspartate revealed that these enzymes can be classified into eight phylogenetic groups that all contain putative epoxide hydrolase domains. To determine their catalytic activities, 10 putative bacterial epoxide hydrolase genes and 2 known bacterial epoxide hydrolase genes were cloned and overexpressed in Escherichia coli. The production of active enzyme was strongly improved by fusion to the maltose binding protein (MalE), which prevented inclusion body formation and facilitated protein purification. Eight of the 12 fusion proteins were active toward one or more of the 21 epoxides that were tested, and they converted both terminal and nonterminal epoxides. Four of the new epoxide hydrolases showed an uncommon enantiopreference for meso-epoxides and/or terminal aromatic epoxides, which made them suitable for the production of enantiopure (S,S)-diols and (R)-epoxides. The results show that the expression of epoxide hydrolase genes that are detected by analyses of genomic databases is a useful strategy for obtaining new biocatalysts.

Enantiopure epoxides and vicinal diols are valuable intermediates in the synthesis of a number of pharmaceutical compounds. Epoxide hydrolases (EC 3.3.2.3) catalyze the conversion of epoxides to the corresponding diols. If they are enantioselective, they can be used to produce enantiopure epoxides by means of kinetic resolution (5). In the past, when only epoxide hydrolases from mammalian sources were known (12), the use of epoxide hydrolases in biocatalysis was hampered by their poor availability and insufficient catalytic performance, such as a low turnover rate or poor enantioselectivity. The potential for biocatalytic application of epoxide hydrolases was significantly increased with the discovery of microbial epoxide hydrolases (41), which are easier to produce in large quantities. The cloning and overexpression of several enantioselective epoxide hydrolases, e.g., from Agrobacterium radiobacter (35), Aspergillus niger (3), and potato plants (40), not only facilitated large-scale production of these enzymes but also made it possible to improve their biocatalytic properties by site-directed or random mutagenesis (34, 36, 43).

Since many microbial genome sequences are available in the public domain, it is useful to screen these databases for genes that might encode new enzymes with interesting properties.

† Supplemental material for this article may be found at http://aem.asm.org/.
‡ Present address: Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, CB2 1GA Cambridge, United Kingdom.

* Corresponding author. Mailing address: Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands; Phone: 31 (0)50 364086; Fax: 31 (0)50 3634165; E-mail: D.B Janssen@rug.nl.

Novel epoxide hydrolases can be identified by performing a BLAST search of the genomic databases, using amino acid sequences of known epoxide hydrolases as queries. This approach will result in putative epoxide hydrolases but also in amino acid sequences from structurally and mechanistically related enzymes, such as esterases and dehalogenases (33), which can be filtered out using conserved epoxide hydrolase sequence motifs that define the active site (Fig. 1). The putative epoxide hydrolase-encoding genes can subsequently be cloned and overexpressed in a host with no endogenous epoxide hydrolase activity, such as Escherichia coli, and the activity of the encoded proteins can be tested.

Most epoxide hydrolases for which sequence information is presently available are members of the α/β-hydrolase fold family, to which lipases, esterases, and haloalkane dehalogenases also belong. These enzymes consist of a main domain that is composed of a central β-sheet surrounded by α-helices and a variable cap domain positioned on top of the substrate binding site (Fig. 1) (33). They use a catalytic mechanism that involves an Asp/Ser/Cys-His-Asp/Glu (nucleophile-histidine-acid) catalytic triad located on top loops of the main domain. The positions of these residues are structurally conserved in the order nucleophile-histidine-acid (Fig. 1A and B) (33), but in the primary amino acid sequence the order is nucleophile-acid-histidine, with the acid located either in front of the cap domain (Fig. 1C, position 3a) or after the cap domain (Fig. 1C, position 3b).

In the case of α/β-hydrolase fold epoxide hydrolases, the catalytic triad nucleophile is an invariable aspartate that opens the epoxide ring by nucleophilic attack (Fig. 1A) (4, 35). The ring-opening reaction is assisted by two conserved tyrosines.
that are located in the cap domain (Fig. 1C) (6, 36). The resulting alkyl-enzyme intermediate is subsequently hydrolyzed by a water molecule that is activated by a histidine that functions as a proton acceptor and, in turn, is assisted by the acidic residue (Fig. 1B). The negative charge that develops on the carbonyl oxygen of the nucleophilic aspartate during hydrolysis of the alkyl-enzyme intermediate is stabilized by two backbone amides that are contributed by the residue following the catalytic nucleophile and residue X in a conserved H-G-X-P motif (Fig. 1C). This motif is located between strand 3a and helix α1 (Fig. 1C, position 5) and is conserved in haloalkane dehalogenases and epoxide hydrolases. Residue X is usually an aromatic residue in epoxide hydrolases, whereas it is an asparagine or a glutamate in haloalkane dehalogenases. The side chain of the amino acid at this position lines the active site (31, 51). Between the H-G-X-P motif and the catalytic nucleophile, there is a conserved G-X-Sm-X-S/T motif of unknown function (Fig. 1C, position 6).

This paper describes the screening of various genomic databases for epoxide hydrolases of the α/β-hydrolase fold family. Based on phylogenetic analysis of the resulting sequences and comparison to other α/β-hydrolase fold enzymes that have
a nucleophilic aspartate, the epoxide hydrolases were divided into different phylogenetic groups. Ten of these putative epoxide hydrolases, together with two known bacterial epoxide hydrolases, were cloned and overexpressed in *E. coli* and subsequently tested for their biocatalytic potential.

**MATERIALS AND METHODS**

**Materials.** The epoxides that were used in this study are indicated in Fig. 2. The same numbering is used for the epoxides and the corresponding diols, with the extension “a” for the diol. Epoxides 1, 2, 7, 9, 10, 20, and 25 were synthesized from *A. nidulans*, and enantiopure diols 20a and 23a, and L-epinephrine were obtained from Acros (Geel, Belgium). Epoxide 5 was obtained from Enzis (Groningen, The Netherlands), and epoxides 6, 11, 12, 15 to 17, 19, and 23 were obtained from Aldrich. Furthermore, epoxides 8 and 14 were obtained from Lancaster (Frankfurt am Main, Germany), epoxides 13, 21, and 24 were obtained from Fluka, epoxide 12 was obtained from Tokyo Kasei, and epoxide 22 was obtained from Sigma. Epoxides 1, 3, and 4 were synthesized from *para*-methoxybenzaldehyde, *para*-chlorobenzaldehyde, and *meta*-chlorobenzaldehyde, respectively. Chiral high-performance liquid chromatography (HPLC) columns were obtained from D incarceration. A G-TA capillary column was obtained from Altech, and a CP Chirasil Dex CB column was obtained from Chrompack. All DNA-modifying enzymes and ampicillin were obtained from Roche Molecular Biochemicals, except for *E. coli* DNA polymerase (Stratagene).

**Cloning of epoxide hydrolase genes.** Genes encoding putative epoxide hydrolases were amplified from respective sources of genomic DNA (see below), using the primer pairs described in the supplemental material, and subsequently cloned into the pMAL-c2x plasmid vector. Upon bacterial expression of these constructs, epoxide hydrolases were obtained as fusion proteins with a maltose binding protein (MalE) domain linked to the N terminus of the epoxide hydrolase domain via a polyasparagine linker and a small decapeptide. The following DNA sources served as templates: whole-cell material from an overnight culture (Busch, Bfech1, Npueh1, Npueh2, and Ppueh), genomic DNA added up to 0.05 mg ml⁻¹ (Draeh, Rpaeh2, Scoeh6, and Tfeuh), and plasmid DNA (AraEchA [43], Corech [29], and MtupEF). Primers were used at 0.4 nM in a reaction mixture with 0.2 mM concentration of each deoxynucleoside triphosphate and 0.025 U *Pfu* polymerase. The reactions with high G + C content were performed in the presence of 5% (vol/vol) dimethyl sulfoxide (DMSO). For genes with no high G + C content, the temperature program used was 15 min at 94°C without polymerase, followed by 30 cycles of 60 s at 94°C, 45 s at 58°C, and 70 s at 72°C and a final step of 4 min at 72°C. For genes with high G + C content, the temperature program used was 15 min at 94°C without polymerase, followed by 30 cycles of 60 s at 95°C, 45 s at 68°C, and 70 s at 72°C and a final step of 4 min at 72°C. For PCR amplification of MtupEF the following program was used: 15 min at 94°C without polymerase, followed by 30 cycles of 60 s at 94°C, 45 s at 50°C, and 70 s at 68°C and a final step of 4 min at 68°C. The PCR products were digested with restriction enzymes (see the supplemental material for details) and subsequently ligated into *BamHI-HindIII*, *EcoRI-HindIII*, *EcoRI-PstI*, or *XbaI-HindIII*-digested pMAL-c2x plasmid DNA, using T4 DNA ligase. The ligation mixtures were transformed into *E. coli* TOP10 cells by electroporation. The transformants were plated on LB medium containing ampicillin. Colonies were checked for inserts by using PCR with *Taq* polymerase and with colony material as the template. Positive colonies were used to inoculate 5 ml of liquid LB medium and grown overnight at 37°C. Plasmid sequences that contained the catalytic triad, at least one ring-opening tyrosine, and motifs 5 and 6 were annotated as putative epoxide hydrolases. All putative epoxide hydrolase sequences were aligned, together with the known epoxide hydrolase sequences and a number of known and putative haloalkane and haloacetate dehalogenases that also belong to the *α/β*-hydrolase fold family and possess an invariable asparagine as the catalytic nucleophile. The resulting phylogenetic tree was displayed using TreeView and edited. The multiple sequence alignments were displayed as bit scores for each position, using the WebLogo sequence generator (http://weblogo.berkeley.edu).

**FIG. 2.** Epoxide substrates used in this study. 1, *para*-methylstyrene oxide; 2, styrene oxide; 3, *para*-chlorostyrene oxide; 4, *meta*-chlorostyrene oxide; 5, *para*-nitrostyrene oxide; 6, *cis*-stilbene oxide; 7, *trans*-stilbene oxide; 8, chalcone-*α,β*-epoxide; 9, phenyl glycidyl ether; 10, *para*-nitrophenyl glycidyl ether; 11, epichlorohydrin; 12, epibromohydrin; 13, vinyl oxirane; 14, tert-butylvinylcarbinol; 15, 1,2-epoxybutane; 16, 1,2-epoxypentane; 17, 1,2-epoxyhexane; 18, 1,2-epoxyheptane; 19, 1,2-epoxyoctane; 20, *cis*-2,3-epoxybutane; 21, *trans*-2,3-epoxybutane; 22, fosfomycin; 23, cyclohexene oxide; 24, limonene-1,2-epoxide; 25, 3,4-epoxytetrahydrofuran.
DNA was extracted using a plasmid purification kit from Roche, and the inserts were sequenced.

Protein expression and purification. Induction conditions for the expression of fusion proteins were optimized by adding up to 0.4 mM IPTG to a 5-mL culture with an optical density at 600 nm (OD600) of ~0.2, followed by overnight incubation at various temperatures ranging from 8 to 37°C. The cells were harvested and resuspended in 0.4 mL TEDANG buffer (20 mM Tris-HCl, pH 7.4, 1 mM EDTA, 1 mM dithiothreitol, 0.02% [wt/vol] Na2S, 200 mM NaCl, 10% [vol/vol] glycerol) at 4°C. After brief sonication, a cell extract (CFE) was obtained by centrifugation. Cell extracts were analyzed for expression by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The levels of expression were determined from Coomassie-stained gels with the GelPro Analyzer program. Total protein contents were determined using the Bradford reagent. For Bsueth, expression was optimized further by transforming the pMAL-c2x-Bsueth construct into E. coli BL21(DE3)/pLyS3 Rosetta, which has elevated levels of rare tRNAs that could enhance expression.

Preparative-scale production of proteins was achieved by induction of 1- to 3-liter cultures at an OD600 of ~0.2 with 0.4 mM IPTG, followed by overnight incubation at temperatures ranging from 8 to 30°C. Cells were harvested by centrifugation, washed, and resuspended in TEDANG buffer at 4°C. The cells were lysed by sonication, and cell extracts were obtained by centrifugation at 200,000 × g for 90 min.

Purification of proteins from CFE was achieved using amylase resin to selectively bind the fusion proteins. Resin with bound fusion protein was subsequently poured into a column, and after the column was washed with TEDANG buffer, the fusion protein was eluted with the same buffer containing 10 mM maltose. Fractions containing active fusion protein were pooled and concentrated. The protein content of the purified enzyme fraction was determined by the Bio-Rad assay. The extinction coefficient was calculated from the amino acid sequence of the Maie-EH fusion proteins, using Lasergene-Protein.

Spectrophotometric epoxide hydrolase assays. All continuous spectrophotometric measurements were performed on a Kontron Uvikon 930 UV/VIS spectrophotometer. Epoxide hydrolase activity toward para-nitro styrene oxide (pNSO; epoxide 5 in Fig. 2) and para-nitrophenyl glycidyl ether (pNPGE; epoxide 6 in Fig. 2) was determined in 100 mM Tris-SO4, pH 7.5, as described previously (43). Errors in values of initial activities calculated from the decrease (pNSO, 310 nm) or increase (pNPGE, 350 nm) in absorbance were <10%, as indicated by duplicate measurements. The detection limit for these conversions was 0.002 μmol min−1 mg−1.

The steady-state kinetic parameters of the purified novel epoxide hydrolases were determined by measuring progress curves for the conversion of enantiopure pNSO (epoxide 5) and pNPGE (epoxide 6). The substrate was added at a concentration of up to 0.5 mM, with a final concentration of DMSO of <1% (vol/vol). A suitable amount of purified protein was added to start the reaction. The spectrophotometric conversion traces for the separate enantiomers of epoxides 5 and 10 were directly fitted to Michaelis-Menten kinetics, as described before, to obtain kcal/Km and kcal/Km values (36, 43). In case the Km was high and the substrate concentration used, the spectrophotometric traces were fitted according to first-order kinetics, and the first-order rate constant equals kcal/Km. In this case, only lower limits of kcal and Km were obtained. More details on the fitting procedures used can be found in the supplemental material. Errors in values for kcal, Km, and kcal/Km were <10%, as indicated by duplicate measurements.

Substrate profiling of the various epoxide hydrolases was done using the adrenaline test (46). Initial testing of various vicinal diols in a reaction with IO4− and adrenaline revealed that adenochrome was rapidly produced with all diols that are formed from the epoxides shown in Fig. 2, except for diols derived from glycidyl ethers (epoxides 9 and 10), which do not react fast enough with IO4−. The adenochrome formed by the reaction with remaining IO4− was determined from the resulting data, as described before (26, 43).

**RESULTS AND DISCUSSION**

Phylogenetic analysis. To identify the suitability of genomic sequences as a source of new biocatalytically useful epoxide hydrolases, we first made an inventory of potential targets for overexpression in E. coli by screening various eukaryotic and prokaryotic genome databases for the presence of putative epoxide hydrolase-encoding sequences. The amino acid sequences of epoxide hydrolases for which the activity is known were used as query sequences. In total, 438 sequenced organisms were screened for the presence of putative epoxide hydrolases (Table 1). The obtained hits were manually screened for the sequence motifs that are characteristic of epoxide hydrolases (Fig. 1C) in order to filter out false-positive results. All sequences that contained the catalytic triad, at least one ring-opening tyrosine, and the H-G-X-P and G-X-Sm-X-S/T motifs were annotated as putative epoxide hydrolases. In this way, 292 amino acid sequences of putative epoxide hydrolases were identified in 93 different organisms for which no epoxide hydrolase activity had previously been reported. These sequences were subsequently aligned with the amino acid sequences of other α/β-hydrolase fold enzymes that have the conserved nucleophilic aspartate, including 42 epoxide hydrolases with known activity, halohaline dehalogenases, and fluorocatechate dehalogenases. Analysis of the multiple sequence alignment revealed that 18 hydrolases that were annotated in the databases as epoxide hydrolases most likely are fluorocatechate dehalogenases (see below). The remaining 274 putative epoxide hydrolases originated from 91 taxonomically different organisms (Table 1). From this set, 35 sequences were removed before further phylogenetic analysis since they were
100% identical to sequences from closely related organisms. A multiple sequence alignment was done with the remaining 239 putative epoxide hydrolases found in sequencing projects, 6 putative epoxide hydrolases from other sources, 42 known epoxide hydrolases, 19 (putative) fluoroacetate dehalogenases, and 8 (putative) haloalkane dehalogenases.

The database search results showed that 20% of the organisms that have been sequenced thus far contain one or more putative epoxide hydrolases. However, in some phylogenetic (sub)classes, such as actinobacteria and fungi, >50% of the sequenced organisms contain at least one putative epoxide hydrolase, whereas in other (sub)classes, such as firmicutes and chlamydiae, they are hardly present (Table 1). Most microbial epoxide hydrolases described in the literature originate from actinobacteria (41) and fungi (39), and their epoxide hydrolase activity was determined with whole cells (39, 41). Since many of the sequenced actinobacteria and fungi appear to contain multiple sequence alignment was done with the remaining 239 putative epoxide hydrolases from other sources, 42 known epoxide hydrolases described in the literature originate from actinobacteria (8, 9). This is most strikingly demonstrated by the fact that the amino acid sequences of the epoxide hydrolases from A. radiobacter AD1 (35) and A. niger (3) are <10% identical, but the α/β-hydrolase fold part of their X-ray structures and their active sites are virtually superimposable (31, 51). Most of the phylogenetic groups of epoxide hydrolases consist of homologs from evolutionarily very distant organisms. Since the overall identity within each group is rather low, and since the high identities that do occur are between enzymes that originated from closely related organisms, the majority of these genes are probably orthologs that were separated during speciation. Since bacteria are biased toward deleting genes that are no longer needed (28), several specialized microorganisms such as lactic acid bacteria probably lost their epoxide hydrolase genes during evolution. Indeed, most epoxide hydrolases occur in organisms with large genomes (>8 Mb), even though the size of their collective genomes is similar to that of the small-genome organisms (<2 Mb) (Fig. 4). Lateral transfer may play a role in some cases, as suggested by the situation encountered in Mycobacterium tuberculosis, which contains 11 α/β-fold hydrolases (3 dehalogenases and 8 epoxide hydrolases) in a 4-Mb genome that harbors less than 4,000 genes in total.

In order to investigate whether the phylogenetic study performed here was biased by the available sequenced genomes, a BLAST search with the same 42 query sequences of epoxide hydrolases with known activity was done on the open reading frames that were identified in the Sargasso Sea environmental
genome sequencing project (44). In total, 61 complete putative epoxide hydrolase sequences were retrieved, which were all different from the 42 known epoxide hydrolases. A multiple sequence alignment including the 61 environmental sequences yielded the same eight phylogenetic groups. Most of the environmental sequences clustered either with other environmental sequences or with putative epoxide hydrolases from Burkholderia spp., which are very abundant in the Sargasso Sea (44). Since many of the sequences retrieved from the Sargasso Sea probably are missing part of their N and/or C terminus, they were not included in further phylogenetic studies.

Characteristics of phylogenetic groups. The members of phylogenetic group 1 epoxide hydrolases originated predominantly from proteobacteria (>70% of the group members). This group includes an epoxide hydrolase that was recently identified in an environmental gene library (49). The members of subgroup 1A are, on average, 40 amino acids longer than the other members of this group. Using SignalP software (http://www.cbs.dtu.dk/services/SignalP), these extra amino acids were identified as an N-terminal signal peptide for protein excretion. A more thorough analysis of the various characteristic signal peptide regions suggested that these proteins are secreted via a twin arginine translocation pathway (1). This pathway usually translocates completely folded proteins, including cofactor-dependent enzymes, across cellular membranes (14).

Group 2 contains epoxide hydrolases from bacterial, archaeal, and eukaryotic origins. Some of the eukaryotic epoxide hydrolases originated from multicellular organisms. Thus far, only two classes of eukaryotic epoxide hydrolases from multicellular organisms are known, which are the microsomal/juvenile hormone epoxide hydrolases (group 5) and the cytosolic/plant epoxide hydrolases (group 8). The putative eukaryotic epoxide hydrolases that cluster in group 2 have an N-terminal extension of unknown function and are therefore somewhat longer than the enzymes in this group that are of prokaryotic origin (Table 2). Group 2 contains a relatively large number of putative epoxide hydrolases that originated from cyanobacteria (20%).

Group 3 consists predominantly of putative epoxide hydrolases from actinobacteria, β-proteobacteria, and fungi. About
one-third of the group members consist of an epoxide hydrolase domain (N-terminal) and a C-terminal domain that is homologous to SDR proteins (Table 2). The epoxide hydrolase domain of the fungal epoxide hydrolase sequences is, on average, 34 amino acids longer than that of the other group members.

In group 4, both epoxide hydrolases and haloalkane dehalogenases appear to be present (Fig. 3; Table 2). These structurally related enzymes, which also have similar mechanisms (2), can be distinguished at the amino acid level by the presence of the ring-opening tyrosines, which are missing in the haloalkane dehalogenases. Since these tyrosines do not always align perfectly, the H-G-X-P motif, which is located between sheet H9251 and helix H9252 (Fig. 1), is more suitable for distinguishing epoxide hydrolases from haloalkane dehalogenases. The main chain amide of the X residue is part of the oxyanion hole in both enzyme classes (Fig. 1B). The side chain is part of the wall of the active site and can therefore interact with the substrate (31, 51). In epoxide hydrolases, the X residue is always an aromatic amino acid, whereas in haloalkane dehalogenases it is a more hydrophilic or even charged residue such as asparagine or glutamate (Fig. 5A). Although the exact function of the motif is unknown, it can possibly be used to distinguish between epoxide hydrolases and haloalkane dehalogenases. The fungal enzymes present in this group consist predominantly of homologous epoxide hydrolase sequences of various Saccharomyces spp. (Fig. 3).

More than half of the 42 known epoxide hydrolases that were included in the multiple sequence alignment clustered in group 5 (Fig. 3; Table 2). This group includes the well-known mammalian microsomal epoxide hydrolases (22) and the juvenile hormone epoxide hydrolases that are present in many insects (18). Group 5 is the best-defined group in the phylogenetic tree (Fig. 3). Its members are, on average, 100 (bacterial), 120 (fungal), and 150 (higher eukaryotes) amino acids longer than most other epoxide hydrolases (Table 2), which is mainly due to a large N-terminal extension. The mammalian and insect epoxide hydrolases of this group are membrane bound by an N-terminal membrane anchor (20). The fungal and bacterial epoxide hydrolases of this group do not contain this membrane anchor and are therefore probably soluble. The N-terminal extension is important for the phylogenetic distinction of the group.

![Figure 4](image-url)  
**FIG. 4.** Occurrence of putative epoxide hydrolases in prokaryotes and archaea in relation to genome size. The number of organisms with a given genome size is indicated above each percentage bar, with the size of the collective genomes in this category shown in parentheses (in Mb). The percentages of organisms that have at least one putative epoxide hydrolase are indicated in black. The horizontal line indicates the average number of epoxide hydrolase-containing organisms (17.5%).
5 epoxide hydrolases from all others, but the differences are also clearly present in the α/β-hydrolase main domain. In contrast to the case for most epoxide hydrolases, the charge relay acid in this group of epoxide hydrolases is a glutamate in 75% of the cases instead of an aspartate (Table 2).

Group 6 contains both fluoroacetate dehalogenases and epoxide hydrolases. For all members of this group, the charge relay aspartate is located at position a instead of the more common position b (Fig. 1; Table 2). The putative fluoroacetate dehalogenase sequences all grouped together with the fluoroacetate dehalogenase from *Moraxella* sp. (24) (Fig. 3). These sequences contain three consecutive arginines distal from the nucleophilic aspartate in a DRXXRXXXR motif, whereas epoxide hydrolases usually have a conserved aromatic residue distally flanking the catalytic nucleophile (Fig. 5B). It was proposed that one of these arginines plays a role in the binding of the acid part of the fluoroacetate substrate (25).

Furthermore, all of these sequences have the fluoroacetate dehalogenase from *Moraxella* sp. (24) as their highest scoring BLAST hit for proteins that have known activity. These putative haloacetate dehalogenases have often erroneously been annotated as epoxide hydrolases, including in a recent phylogenetic study of epoxide hydrolases (9). The presence of a conserved tyrosine in the cap domains of haloacetate dehalogenases that aligns perfectly with one of the ring-opening tyrosines of epoxide hydrolases contributes to the confusion. The epoxide hydrolase sequences in group 7 are similar to those in group 6. The first conserved ring-opening tyrosine and the charge relay aspartate, however, are located at different positions (Fig. 1; Table 2).

Group 8 contains a large number of known epoxide hydrolases from plants and mammals, including the well-known mammalian cytosolic epoxide hydrolases (Fig. 3; Table 2). There are no sequences originating from other, lower eukaryotic organisms, such as fungi, in this group. The plant epoxide hydrolases occur only in this group. The cap domain of this group of epoxide hydrolases has a 30-amino-acid excursion between helices α6 and α7 (Fig. 1), resulting in a larger size of the epoxide hydrolase segment (Table 2).

**Overexpression and purification of putative epoxide hydrolases.** Ten putative epoxide hydrolase-encoding genes from different phylogenetic groups and the bacterial epoxide hydrolases from *A. radiobacter* (35) (AraEchA) and *Corynebacterium* sp. strain C12 (29) (Coreh) were selected to be cloned and overexpressed in *E. coli* (Table 3). Since initial attempts to overexpress the epoxide hydrolases from *Mycobacterium tuberculosis* H37v (MtuEphF) and *Pseudomonas putida* KT2440 (Ppueh) in *E. coli* using the pGEF system (38) resulted in inclusion bodies (data not shown), it was decided to overexpress the enzymes as maltose binding protein-epoxide hydrolase fusion proteins (MalE-EH). The MalE tag was expected to facilitate proper folding of the epoxide hydrolase in *E. coli* (37). Moreover, it serves as a convenient purification tag. The MalE-EH fusion proteins were produced in *E. coli* at levels of 10 to 60% of the total soluble protein in CFEs, including the enzymes for which expression yielded inclusion bodies when no MalE was fused to the epoxide hydrolase (Table 3). The CFE prepared from an expression culture of the MalE-Bsueh (originating from *Bacillus subtilis*) construct in *E. coli* TOP10 cells resulted in rather low specific activities (0.03 μmol min⁻¹

---

**FIG. 5.** Multiple sequence alignments of conserved regions in functionally different α/β-hydrolase fold enzymes, represented as bit scores for each position, with the total height indicating the degree of conservation (maximum level, 4.32) and the relative heights of different symbols at the same position indicating the frequencies of the amino acids at that position. (A) H-G-X-P region in epoxide hydrolases and haloalkane dehalogenases. (B) Region around the nucleophilic aspartate in epoxide hydrolases and fluoroacetate dehalogenases.
were obtained at $10^5$ mg$^{-1}$, possibly due to poor expression levels caused by the large number of rare codons present in the Bsueh gene. Expression of the MalE-Bsueh protein in *E. coli* BL21(DE3)/pLysS Rosetta, an *E. coli* variant that has increased levels of rare tRNAs, resulted in a twofold improvement in specific activity in the cell extract.

Amylose resin purification was first tested with the MalE-Coreh construct, resulting in $10^6$ mg of 85% pure fusion protein from a 1-liter culture at an OD$_{600}$ of $1.5$, with a total activity yield of 84%. The MalE-EH fusion proteins of the three epoxide hydrolases from phylogenetic group 2 (Fig. 3; Table 2) were also produced at a preparative scale and conveniently purified from cell extracts using the MalE tag. The enzymes from *Deinococcus radiodurans* (Draeh) and *Nostoc punctiforme* (Npueh) were obtained at $>95$% purity with good yields. The *B. subtilis* enzyme could only be obtained at 50 to 70% purity with a single round of amylose resin purification. The enzyme was unstable under the purification conditions used and was therefore not purified further (see the supplemental material for details on enzyme purifications).

**Activity of MalE-epoxide hydrolase fusion proteins.** CFES and/or purified samples of the MalE-EH fusion proteins were tested for activity towards a range of epoxides, using either a colorimetric assay (substrates 5 and 10) (Fig. 2) or an indirect assay that measures periodate depletion by its reaction with a diol (substrates 1, 2, 4, 7, 8, 11 to 14, 16, 17, and 19 to 25). In the latter assay, the diol that is formed by epoxide hydrolase activity reacts with periodate, and the amount of remaining periodate is measured by incubating it with adrenaline to yield a colored product (46). Using these assays, it appeared that of the 12 epoxide hydrolases that were tested, 8 were active with one or more epoxides (Table 4). The substrate ranges of the enzymes were generally very broad and included styrene oxide derivatives, phenyl glycidyl ethers, and most terminal aliphatic epoxides. Only a few enzymes showed activity with the sterically more demanding $\alpha,\beta$-disubstituted epoxides (compounds 6 to 8, 20, 23, and 24). Among the proteins for which no activity was detected were MtuEphF and Ppueh. It is likely that the inactive proteins were correctly folded since they were expressed as soluble proteins. Activity was detected with proteins from different groups, and in combination with data in the literature, it was concluded that epoxide hydrolases occur in at least seven of the phylogenetic groups described above (Fig. 3; Table 2). The fact that Npueh2 is active toward pNPGE (compound 10) (Tables 3 and 4) shows that group 4 contains both haloketone dehalogenases and epoxide hydrolases (Fig. 3). The nature of residue X in the H-G-X-P motif (motif 5 in Fig. 1), as described above, could indeed be used for distinction between the two functional classes (Fig. 5A).

The epoxide hydrolase from *Rhodopsuedomonas palustris* (Rpach2), which belongs to group 8 (Table 2), was active toward trans-stilbene oxide (compound 7) but not toward cis-stilbene oxide (compound 6) (Table 4). Both the mammalian and the plant soluble epoxide hydrolases, which also belong to phylogenetic group 8, have the same characteristic (11, 19, 30). Thus, this particular cis versus trans specificity is clearly linked to phylogenetic group 8. The three-dimensional structure of mammalian sEH, a group 8 member, revealed an L-shaped substrate access tunnel with the catalytic nucleophile sitting right at the kink (7). Compared to the active-site geometry of
other EH structures, where the catalytic nucleophile sits at the very end of the substrate access tunnel, the particular active-site architecture of sEH, and possibly of the other group 8 members, does indeed favor the turnover of trans-1,2-disubstituted epoxides.

An analysis of the genes carried in the vicinity of the six putative epoxide hydrolase genes for which activity was found did not clearly reveal a function for the enzymes. However, the genes encoding the epoxide hydrolases from \( B. subtilis \), \( Burkholderia fungorum \) (Bfuehl), and \( R. palustris \) are located near genes encoding putative enzymes that may be involved in the detoxification of xenobiotics or antibiotics. The genes surrounding the epoxide hydrolase genes from \( D. radiodurans \) and \( N. punctiforme \) are of unknown or unrelated function. Thus, the biological function of the epoxide hydrolases identified here is uncertain.

**Enantioselectivity and unusual enantiopreference of group 2 enzymes.** The enantioselectivity of the active epoxide hydrolase fusion proteins was tested with various chiral substrates. The enantioselectivities with styrene epoxides and phenyl glycidyl ethers were generally low (\( E < 10 \)) to modest (\( 10 < E < 50 \)) (Tables 3 and 5). This could possibly be influenced by the fact that the tested enzymes were fusion proteins, but the MalE-AraEchA enzyme showed no reduction in stereospecificity compared to the native enzyme (43), so in this case the MalE protein did not influence the catalytic properties of the enzyme. The highest \( E \) values were found for \( para \)-nitrostyrene oxide, using the \( D. radiodurans \) enzyme (Draeh), and for \( para \)-nitrophenyl glycidyl ether, using the \( B. subtilis \) epoxide hydrolase (Bfuehl).

Most well-described epoxide hydrolases, such as the ones from \( A. radioacter AD1 \) (43), \( A. niger \) (16, 32), and \( Rhodothorula glutinis \) (47) and the mammalian microsomal epoxide hydrolase (13), are (\( R \)) specific toward styrene oxides and (\( S \)) specific toward phenyl glycidyl ethers. In contrast, Bfuehl, Draeh, and Npuehl (from \( N. punctiforme \)) had an opposite enantiopreference towards these compounds. The enantioselectivities varied from low (\( E < 10 \)) to intermediate (\( 10 < E < 50 \)) and were generally higher for the aromatic substrates (epoxides 1 to 5, 9, and 10). All three of these epoxide hydrolases belong to phylogenetic group 2 (Fig. 3; Table 2). Of these three proteins, the \( B. subtilis \) enzyme has the most restricted substrate range. The enzyme had reasonable activities toward the terminal epoxides pNPGE (compound 10) and 1,2-epoxy-

### TABLE 5. Activities and enantioselectivities of novel group 2 epoxide hydrolases for terminal epoxides

<table>
<thead>
<tr>
<th>Enzyme and parameter</th>
<th>Activity (( \mu )mol min (^{-1} ) mg (^{-1} )) or enantioselectivity ( ^{a} ) with indicated epoxide substrate number ( ^{b} ):</th>
<th>Aromatic substrates</th>
<th>Aliphatic substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td>11 15 16 17 18 19</td>
<td></td>
</tr>
<tr>
<td>Bfuehl Activity</td>
<td>&lt;0.003 &lt;0.003 &lt;0.003 &lt;0.003 0.01 0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enantioselectivity</td>
<td>( ^{a} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draeh Activity</td>
<td>0.42 0.06 0.68 0.19 0.11 0.85 0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enantioselectivity</td>
<td>4.2 (( S )) 5.2 (( S )) 2.3 (( S )) 6.4 (( S )) 21 (( S )) 1.4 (( R )) 1.9 (( R ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Npuehl Activity</td>
<td>0.37 0.42 2.1 0.74 1.3 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enantioselectivity</td>
<td>2.3 (( S )) 1.9 (( S )) 4.6 (( S )) 1 17 (( S )) 3.7 (( R )) 16 (( R ))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^{a} \) The preferentially hydrolyzed enantiomer is indicated in parentheses.

\( ^{b} \) —, activity was too low to determine an accurate \( E \) value.

\( ^{c} \) See Fig. 2 for epoxide structures.
TABLE 6. Steady-state kinetic parameters for Draeh and Npueh1 acting on each enantiomer of pNSO (substrate 5) and pNPGE (substrate 10).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Value for indicated enantiomer</th>
<th>(R)-pNSO</th>
<th>(S)-pNSO</th>
<th>(R)-pNPGE</th>
<th>(S)-pNPGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draeh</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>&gt;0.007</td>
<td>&gt;0.17</td>
<td>&gt;0.7</td>
<td>&gt;0.5</td>
</tr>
<tr>
<td></td>
<td>$K_m$ (mM)</td>
<td>&gt;0.5</td>
<td>&gt;0.5</td>
<td>&gt;0.3</td>
<td>&gt;0.3</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}/K_m$ (mM$^{-1}$ s$^{-1}$)</td>
<td>0.020</td>
<td>0.42</td>
<td>2.8</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>E value</td>
<td>21 (21)</td>
<td>21 (21)</td>
<td>1.3 (1.8)</td>
<td>1.3 (1.8)</td>
</tr>
<tr>
<td>Npueh1</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>&gt;0.2</td>
<td>3.4</td>
<td>&gt;2.6</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>$K_m$ (mM)</td>
<td>&gt;0.5</td>
<td>0.20</td>
<td>&gt;0.025</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}/K_m$ (mM$^{-1}$ s$^{-1}$)</td>
<td>0.60</td>
<td>17</td>
<td>174</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>E value</td>
<td>28 (17)</td>
<td>28 (17)</td>
<td>12 (15)</td>
<td>12 (15)</td>
</tr>
</tbody>
</table>

a Since for some enzymes the $K_m$ value exceeded substrate solubility, only lower limits could be determined for $k_{cat}$ and $K_m$, and $k_{cat}/K_m$ values were determined as first-order rate constants.

b Calculated from the $k_{cat}/K_m$ ratios of the two enantiomers. The E value calculated from a kinetic resolution experiment is given in parentheses.

c $k_{cat}/K_m$ could be determined only as a first-order rate constant at a low substrate concentration.

d The E value was calculated from the ee $p$ according to equation 1.

TABLE 7. Activities of different fusion proteins for meso-epoxides.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Parameter value for indicated substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cis-2,3-Epoxybutane (substrate 20)</td>
</tr>
<tr>
<td></td>
<td>Activity (μmol min$^{-1}$ mg$^{-1}$)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>AraEchA</td>
<td>0.003</td>
</tr>
<tr>
<td>Bfueh1</td>
<td>&lt;0.002</td>
</tr>
<tr>
<td>Coreh1</td>
<td>&lt;0.002</td>
</tr>
<tr>
<td>Rpaeh2</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Initial activities were determined at a 2 mM substrate concentration with cell extracts.

b Activities were measured with (partially) purified protein.

c —, activity was too low to determine the enantiomeric excess of the product (ee $p$).

d The E value was calculated from the ee $p$ according to equation 1.
reported for an epoxide hydrolase (BD10090) obtained from an environmental gene library (49). Epoxide hydrolases with excellent enantioselectivity toward meso-epoxides could thus be obtained from genomic databases.

**General conclusions.** The results reported here indicate that epoxide hydrolase genes that encode active proteins are widely present in microbial genomes. In about 20% of the organisms for which the whole genome sequence is known, putative epoxide hydrolase genes are present, and the expression of a number of these genes and testing with model substrates identified activity for 60% of them. Since only a restricted number of substrates was used, the percentage of organisms that carry active enzymes is probably larger. Thus far, the functions of most of these epoxide hydrolases remain unclear.

The various phylogenetic groups of α/β-hydrolase epoxide hydrolases harbor orthologs from phylogenetically very different organisms. Since the level of sequence identity within the groups is low, this is probably not caused by recent lateral gene transfer. A more likely explanation is that the common ancestor of epoxide hydrolases was present early in evolution and widespread among the various species. As a result of speciation, enzymes that cluster in a multiple sequence alignment are now present in phylogenetically unrelated organisms.

Active recombinant epoxide hydrolases could rapidly be obtained by cloning the genes as fusions to the C-terminal part of the maltose binding protein (MalE). This facilitated expression, prevented the formation of insoluble inclusion bodies, and served as a convenient purification tag. Preliminary experiments showed that the production of soluble protein failed for only 2 of the 12 enzymes that were synthesized as fusion proteins. This approach for obtaining new enzymes for biocatalysis is attractive since the gene of a positive hit is immediately available for further improvement by site-directed mutagenesis or directed evolution.

The cloning and characterization of the putative epoxide hydrolases led to several new active enzymes that can be evaluated as biocatalysts for enantioselective conversion. Three enzymes were (S) selective toward aromatic substrates, which is rather uncommon and has been observed only for a few other epoxide hydrolases (21, 27, 42, 50). Furthermore, the enzymes from *B. fungorum* (Bluhe1) and *N. punctiforme* (Npue1) converted the prochiral substrate cyclohexene oxide to optically enriched (1S,2S)-1,2-cyclohexanediol with much higher enantioselectivities than previously reported (Table 7) (49). Thus, new epoxide hydrolases with interesting biocatalytic activities can be obtained using genomic databases, which defines genome analysis as a promising strategy to widen the scope of biocatalysts, as previously shown for carbohydrate-converting enzymes (17).

**ACKNOWLEDGMENTS**

The various bacterial strains and/or their DNA material was kindly provided by James Tiedje (Michigan State University; *B. fungorum*), David Leak (Imperial College London, United Kingdom; *Corynebacterium sp.*), Theo Sonke (DSM Research, Geleen, The Netherlands; *D. radiodurans* R1), John C. Meeks (University of California; *N. punctiforme*), Stuart Levy (Tufts University; *Pseudomonas fluorescens Pto-1*), Juan L. Ramos (Consejo Superior de Investigaciones Científicas, Granada, Spain; *P. putida* KT2440), Caroline S. Harwood (University of Iowa; *R. palustris* CGA0097), Stewart Cole (Institut Pasteur, Paris, France; *M. tuberculosis*), and Diana Irwin (Cornell University; *Thermobifida fusca*). We thank Birgit Schulze and Theo Sonke at DSM Research for valuable discussions.

The work of B.V.L. was supported by a grant from DSM to the University of Groningen.

**REFERENCES**


