This is a pre- or post-print of an article published in
Standfuß-Gabisch, C., Al-Halbouni, D., Hofer, B.
Characterization of biphenyl dioxygenase sequences and
activities encoded by the metagenomes of highly
polychlorobiphenyl-contaminated soils
(2012) Applied and Environmental Microbiology, 78 (8),
pp. 2706-2715.
Characterization of Biphenyl Dioxygenase Sequences and Activities Encoded by the Metagenomes of Highly Polychlorobiphenyl-Contaminated Soils

RUNNING TITLE: Dioxygenase activities in PCB-contaminated soil

Christine Standfuß-Gabisch,1†, Djamila Al-Halbouni,1‡, and Bernd Hofer1,2§*

1 Division of Microbiology, Helmholtz-Zentrum für Infektionsforschung, Braunschweig, Germany.

2 Department of Chemical Biology, Helmholtz-Zentrum für Infektionsforschung, Braunschweig, Germany.

† Present address: Research Group Viral Immune Modulation, Helmholtz-Zentrum für Infektionsforschung, Braunschweig, Germany.

‡ Present address: Institute of Biology I, RWTH Aachen University, Aachen, Germany.

§ Present address: Department of Chemical Biology, Helmholtz-Zentrum für Infektionsforschung, Braunschweig, Germany.

* Corresponding author. Mailing address: Helmholtz-Zentrum für Infektionsforschung, Abteilung Chemische Biologie, Inhoffenstraße 7, D-38124 Braunschweig, Germany. Phone: (49-531) 61814200. Fax: (49-531) 61813499. E-mail: bernd.hofer@helmholtz-hzi.de.
SUMMARY

Total extracted DNA of two heavily polychlorobiphenyl-contaminated soils has been analysed with respect to biphenyl dioxygenase sequences and activities. This was done by PCR amplification and cloning of a DNA segment encoding the active site of the enzyme. The obtained translated sequences fell into three similarity clusters (I - III). Sequence identities were high within, but moderate or low between clusters. Members of clusters I and II showed high sequence similarities with well-known biphenyl dioxygenases. Cluster III showed low (43 %) sequence identity with a biphenyl dioxygenase from *Rhodococcus jostii* RHA1. Amplicons from the three clusters were used to reconstitute and express complete biphenyl dioxygenase operons. In most cases, the resulting hybrid dioxygenases were detected in cell extracts of the recombinant hosts. At least 83 % of these enzymes were catalytically active. Several amino acid exchanges were identified that critically affected activity. Chlorobiphenyl turnover by the enzymes containing the prototype sequences of clusters I and II was characterized with 10 congeners that were major, minor or no constituents of the contaminated soils. No direct correlations were observed between on-site concentrations and rates of productive dioxygenations of these chlorobiphenyls. The prototype enzymes displayed markedly different substrate and product ranges. The cluster II dioxygenase possessed a broader substrate spectrum towards the assayed congeners, whereas the cluster I enzyme was superior in the attack of *ortho*-chlorinated aromatic rings. These results demonstrate the feasibility of the applied approach to functionally characterize dioxygenase activities of soil metagenomes via amplification of incomplete genes.
INTRODUCTION

Environmental pollutions by polychlorobiphenyls (PCBs) pose a specific problem to bioremediation, as they typically consist of industrial mixtures of dozens of different congeners. Even if broad in substrate range, no single pathway is able to metabolize all PCBs in such mixtures. Thus, the recruitment of novel biocatalysts that may support their removal is of considerable interest. A key enzyme in the aerobic catabolism of PCBs is biphenyl dioxygenase (BphA), which carries out the initial attack of the inert aromatic nucleus. It belongs to class II of aryl-hydroxylating dioxygenases (ARHDOs) that typically hydroxylate substituted benzenes like toluenes and biphenyls (7). This enzyme represents a catabolic bottleneck, as its substrate range is typically narrower than that of subsequent pathway enzymes (9, 13, 43). Moreover, its regiospecificity is a crucial parameter, as it co-determines whether the initial dioxygenation products become dead-end metabolites or can be further transformed.

A number of enzyme engineering projects have been carried out to obtain BphAs with altered or broadened substrate ranges (6, 14, 19, 43). Another approach is the detection and isolation of naturally occurring, but so far inaccessible enzymatic activities by metagenomic methods (24, 31). Such techniques seem promising to discover novel biocatalysts, as only a tiny fraction of existing microorganisms is apparently culturable under laboratory conditions (3). Metagenomic approaches have been applied to characterize the diversity of dioxygenase sequences at a number of PCB-polluted sites (1, 10, 17). Sequences with high as well as with low similarities to those known from culturable organisms have been detected, depending on the examined site and probably also on the PCR primers used. So far, however, these investigations were limited to the determination of (incomplete) gene sequences. Therefore, it remained unclear in how far detected sequences belonged to active enzymes. Moreover, it
was impossible to deduce detailed substrate and product specificities from the translated DNA sequences.

Previously we developed a sequence-based strategy that permits the characterization of enzymatic properties of BphA and other class II ARHDO activities, whose genes are only fragmentarily amplified (9, 18). It should be applicable to DNA from any source, including metagenomes. In this approach, a "donor" segment is amplified which encodes the catalytic center. This is fused with sequences of a "recipient" bphA gene cluster that is efficiently expressed in an appropriate host. It was confirmed that the substrate ranges of the resulting hybrid dioxygenases are dependent on the nature of the donor segment (9, 18).

Here we report the use of this system for a first characterization of dioxygenase activities encoded by the metagenomes of two soil samples from a heavily contaminated site near the city of Wittenberg, Germany. This site has previously been characterized with respect to PCB profile and bacterial community structure (25, 26).

**MATERIALS AND METHODS**

**Chemicals.** Chlorobiphenyl (CB) congeners (99% purity) were obtained from Lancaster Synthesis (White Lund, Morecombe, England), Promochem (Wesel, Germany), or Restek (Sulzbach, Germany).

**Isolation of DNA from soil samples.** Soil sampling and storage have previously been described (25, 26). DNA was isolated using the "FastDNA SPIN Kit for Soil" from Qbiogene BIO 101 (MP Biomedicals, Heidelberg, Germany) according to the protocol of the supplier. Briefly, up to 500 mg of soil were mixed with 978 µl of sodium phosphate buffer and 122 µl of MT buffer, and cells were disrupted for 30 s in a Qbiogene "FastPrep" Instrument (MP
Biomedicals, Heidelberg, Germany) at a rate of 5.5 m/s. After centrifugation (12000 g, 30 s), the supernatant was mixed with 250µl of PPS (Protein Precipitation Solution). Precipitated proteins were removed by centrifugation (12000 g, 5 min). The supernatant was mixed with 1 ml of Binding Matrix Suspension. After settling of the matrix, 500 µl of the supernatant were discarded, and the re-suspended matrix was placed in two subsequent batches onto a "SpinFilter" and centrifuged (12000 g, 1 min). After addition of 500 µl of SEWS-M (salt/ethanol wash solution) to the filter and centrifugation (12000 g, 1 min), it was air-dried, and the DNA was eluted with 50 µl of DES (ultra-pure water). After another centrifugation (12000 g, 1 min) of the filter, the eluate was collected and supplemented with 0.1 vol. of 10 mM Tris-HCl, pH 8, and stored at -20°C.

PCR with DNA from soil samples. Reactions were carried out in PCR buffer containing 1.5 mM MgCl$_2$, (QIAGEN, Hilden, Germany) with about 50 ng of template DNA, 0.5 µM primers HDO2AF and HDO2AR (18), 0.25 mM dNTPs, 0.2 mg/ml BSA, 1.6 µl of DMSO and 1 unit of recombinant Taq DNA polymerase (Fermentas, St. Leon-Rot, Germany) in a total volume of 20 µl. The thermocycler program was as follows: 1 cycle of 30 s at 94 °C; 30 cycles of 30 s at 60 °C, followed by 90 s at 72 °C with an increment of 3 s per cycle; 1 cycle of 600 s at 72 °C.

Molecular cloning techniques. Restriction, ligation, dephosphorylation, agarose gel electrophoresis and bacterial transformations were carried out following standard protocols (30).

Plasmid constructions. pAIA6099 is a derivative of pAIA6100 (18) that harbors a deletion of 12 codons within the *MluI*/AflIII fragment of *bphA1*, which inactivates the gene. It was constructed as follows: Two segments of *bphA1*-LB400 were PCR-amplified, using pAIA111 (18) as template and BPH1917 (CGCTCCAGGCACGCGTGCC, *MluI*\(^\text{+}\) (underlined)) and BPH-2420MUT
(GGGTGCCAGATCCGGAAGATCGTCATATGCTGGCCGACC, NdeI\(^+\) (underlined)) or BPH2454MUT (ATGACGATCTTCCGGATCTGGCAACCCTCGAGGTCCCAATG, XhoI\(^+\) (underlined)) and BPH-2711 (AATCAGGGTGACCGTCTGC, AgeI\(^+\) (underlined)), respectively, as primers. Both products were fused in an overlap-extension PCR with BPH1917 and BPH-2711 as primers. The amplicon was cleaved with MluI and AgeI and was used to replace the corresponding fragment in pAIA50 (43). This introduced the deletion as well as NdeI and XhoI sites and yielded pAIA500. A part of its inactivated \textit{bphA1} gene was PCR-amplified with primers HDO2AF and HDO2AR (18). The latter introduced an AflII site. The product was cleaved with MluI and AflII and used to replace the MluI/AflII fragment of pAIA6100.

**Cloning of PCR products in TOPO vectors.** Taq-DNA-polymerase-generated PCR products were inserted into the T-overhang topoisomerase vector pCR-XL-TOPO (Invitrogen, Karlsruhe, Germany) according to the protocol of the supplier. \textit{E. coli} strain Top10 (Invitrogen, Karlsruhe, Germany) was transformed with the ligation reactions. Plasmid preparations from transformants were analysed by restriction and agarose gel electrophoresis.

**Subcloning of PCR products in pAIA6099.** The MluI/AflII fragments of the inserts of the TOPO clones were excised with these enzymes and were ligated into MluI/AflII-cleaved and dephosphorylated pAIA6099. \textit{E. coli} strains Top10 or XL10-Gold (Stratagene, Amsterdam, The Netherlands) were transformed with the ligation reactions. Plasmid preparations from transformants were analysed by restriction and agarose gel electrophoresis. Correct plasmids were used to transform \textit{E. coli} BL21[DE3](pLysS). The resulting clones were analysed for correct plasmid size.

**Preparation of resting cells.** Preparation of resting cells was carried out as previously described (33) with some modifications. Cells of \textit{E. coli} BL21(DE3)[pLysS], harboring the
respective plasmid, were grown in LB medium at 30 °C. At an optical density at 600 nm (OD\textsubscript{600}) of about 1.0, IPTG was added to 0.4 mM, and the incubation was continued for another 60 min. Cells were harvested, washed with 1 vol. of 50 mM sodium phosphate buffer (pH 7.5) and resuspended in the same buffer to give the concentrations specified below.

**Determination of specific ARHDO activity with biphenyl.** Biphenyl was placed into an Erlenmeyer flask at a final nominal concentration of 125 µM, and the solvent was evaporated. Five ml of resting cells (see above) were added to a final OD\textsubscript{600} of 1, and the flask was shaken at room temp with 120 RPM. At appropriate times, samples of 640 µl were withdrawn, mixed with 160 µl of 5 N NaOH, and centrifuged for 3 min at 12000 g. UV/Vis spectra of the supernatants were recorded, using the resting cell medium as baseline. The absorption at 600 nm was set to zero. Formation rates of extradiol- or meta-cleavage products (MCPs), expressed in mAbs/min, were determined from the linear parts of the resulting plots. Concentrations of wild-type (WT) and variant BphA1 subunits were determined by evaluation of digitalized images of SDS gels of cell extracts stained with Sypro-Ruby (5), using the AIDA 4.15 software (raytest, Straubenhardt, Germany) and bovine serum albumin as standard. The extracts were prepared with the Relay 96 Protein Screen (Invitrogen) according to the manufacturer’s instructions.

**Determination of CB catabolism by prototype hybrid dioxygenases.** Resting cell suspensions containing 0.5 % (w/v) of glucose and 2 OD\textsubscript{600} of *E. coli* BL21[DE3](pLysS) harbouring pAIA6B15 or pAIA6C18 were shaken in Erlenmeyer flasks at 30 °C with nominal concentrations of single CBs of 125 µM. At various times, aliquots were withdrawn and centrifuged for 5 min at 12000 g. UV/Vis spectra of the supernatants were recorded, and rates of MCP formation were determined as described above.

**Sequence determination and analysis.** DNA sequencing was carried out as previously described (4). DNA and protein sequence alignments and calculations of dendrograms were
performed with Clustal W2 (15, 21) at the EBI website (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Dendrograms were displayed with the iTOL tool (22, 23) at the same website (http://itol.embl.de/). Sequence database searches were done with the blastn and blastp programs (2) at the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

**Database accession numbers.** Newly determined sequences have been deposited in the GenBank/EMBL/DDBJ database under accession numbers FR877587 - FR877632 and HE577113 - HE577117.

**RESULTS AND DISCUSSION**

**Amplification of ARHDO gene segments from metagenomic DNA.** DNA was isolated from an uncontaminated and from two increasingly PCB-contaminated soil samples (A, B, C) from a moorland in the vicinity of the city of Wittenberg, Germany (25). The polluted samples contained average PCB concentrations of approximately 1 g/kg or 10 g/kg, respectively. Using this DNA as potential template, PCR amplifications targeting segments which encode the substrate-range-determining cores of the alpha subunits of class II ARHDOs, here collectively termed BphAs, were attempted, using the previously established consensus primers HDO2AF and HDO2AR (18), which amplify fragments of about 720 bp.

It is clear that such an approach will probably not detect PCB-attacking dioxygenase sequences not belonging to class II and that it can only be estimated which fraction of all available class II sequences will be amplified. It has been pointed out that theoretical considerations as well as experimental results suggest that the used oligonucleotides are able to amplify more than 80 % of the cores of known sequences encoding alpha subunits of class
II ARHDOs (18).

For high efficiency and restriction-independent cloning, the amplicons were inserted into a T-overhang TOPO cloning vector. With soil A only trace amounts of PCR products were observed, which were not further processed. The heavily contaminated soils B and C, however, yielded significant amounts of amplicons of the expected length, suggesting that $bphA$ sequences are enriched and thus are likely to be functionally relevant in these soils.

**Sequences of the alpha subunit core gene segments.** A total of 51 TOPO clones, 26 from soil B and 25 from soil C, were sequenced. Translation of the sequences showed that one contained a frameshift and four contained nonsense mutations. DNA and protein sequence alignments revealed that the sequences formed three similarity clusters, named I to III (Fig. 1A). They comprised 37, 12 or 2 sequences, respectively. Nucleotide (NT) and amino acid (AA) sequence identities between these clusters were 85 - 88 %, 38 - 42 % or 37 - 38 %, respectively (Table 1). Within a given cluster, NT and AA sequences were 97 - 100 % identical (Table 1). It can, of course, not be ruled out that minor sequence differences were due to PCR errors. However, the detection of similar micro-heterogeneities and of comparable frequencies of nonsense and frameshift mutations in metagenomic studies not involving PCR, but direct cloning of environmental DNA (36, 38), suggests that all or a major fraction of the apparent sequence diversity was of natural origin. Unequivocal consensus sequences could be determined for clusters I and II, as the bias towards one specific NT or AA, respectively, was always very strong. For cluster I, the consensus sequence itself was found in 7 of 37 clones at the NT level and in 10 of 37 clones at the AA level. For cluster II, the respective values were 1 and 3 for a total of 12 clones. Clones WB15 of cluster I and WC18 of cluster II, which contained the consensus sequences, are in the following referred to as prototypes. Cluster I sequences were predominant in both soils, even more so in soil C (Table 2). In contrast, the percentage of cluster II sequences was higher in
soil B. Cluster III sequences were only found in soil C. This might indicate that enzymes or organisms, respectively (see below), belonging to clusters I and III can more readily cope with the more heavily contaminated site.

A database search revealed that the sequence most closely related to the translated sequences of cluster I was that of the BphA alpha subunit (BphA1) of strain LB400, showing 96% AA sequence identity with the WB15 prototype (Table 1). Of dioxygenases experimentally shown to possess catalytic activity, BphA1 of *Pseudomonas pseudoalcaligenes* KF707, showed the highest degree of AA sequence identity (93%) with the sequences of cluster II (Table 1). Interestingly, the NT sequence of the WC18 prototype was 100% identical with the sequences of putative *bphA1* gene fragments from two strains isolated from the Wittenberg site, *Burkholderia* sp. WBF3 and WBF4 (Table 1). The two sequences of cluster III were most similar (56%) to a putative ring-hydroxylating dioxygenase alpha subunit from *Burkholderia ambifaria* IOP40-10 (Table 1). Of dioxygenases experimentally shown to be catalytically active, BphA1 of *Rhodococcus jostii* RHA1 possessed the highest (43%) AA sequence identity (Table 1). Thus, the enzymes of clusters I and II belong to known ARHDO subfamilies, whereas the dioxygenases of cluster III appear to be part of a novel subfamily.

A number of previous studies also characterized PCB-polluted soils by using PCRs that target parts of the ARHDO class II alpha subunit gene. Capodicasa *et al.* (10) investigated bioreactors containing soils contaminated with 0.89 g PCB/kg. They found highly similar sequences that, in translated form, showed 92 - 99% AA identity to known sequences, mostly from cultivated organisms like strains LB400, KF707 or *Pseudomonas* sp. Cam-1. Aguirrre de Cárcer *et al.* (1) examined soil with a PCB contamination of 0.18 g per kg of dry material. They discovered large sequence diversities before as well as after introduction of willow trees for rhizomediation. Sequencing of 28 clones revealed that the translated
sequences all showed similarities to characterized ARHDO sequences. However, they showed great heterogeneity among each other, displaying between 10 % and 100 % AA sequence identity. Iwai et al. (17) investigated PCB-contaminated soil with the comparatively low concentration of 0.015 g/kg. They directly subjected PCR products to pyro-sequencing, obtaining about 2600 sequences of 175 or 200 NT, depending on the primer used. In their analysis, the authors obtained 40 sequence clusters that contained newly determined as well as database sequences, and 25 clusters that contained only novel sequences, indicating a wide variety of primary structures. In summary, these studies, including the present one, identified abundances of either highly similar or of fairly diverse alpha subunit segment sequences. Diversity appears to decrease with increasing PCB contaminations of the soil samples.

Witzig et al. (40) investigated the sequence diversity of alpha subunit segments of diterpenoid dioxygenase (DitA), which also belongs to the ARHDO family. This work is of interest here, because it also examined PCB-contaminated soil from the Wittenberg site. The authors determined 77 sequences of PCR products encoding the very same region of the alpha subunit as our amplicons. Their template DNAs originated from bacterial isolates as well as from the metagenome. The latter sequences were obtained either after cloning or after electrophoretic separation of the amplicons. Their results resemble ours in several respects. For DitA1 as well as for BphA1, the large majority of sequences fall into two similarity clusters, I and II in Fig. 1. The relative sizes of the clusters are comparable. While inter-cluster distances differ, intra-cluster sequence identities are similarly high at 96 % or above. The assumption that both minor clusters represent the same group of organisms is corroborated by the finding that the \textit{bphA1} sequences of the Wittenberg isolates WBF3 and WBF4 are completely identical with those of clones WB25, WB62 and WC18, belonging to BphA1 cluster II, and that the \textit{ditA1} sequences of the latter strains (100 % identity; accession nos. DQ789336 and DQ789337) belong to DitA1 cluster II. As mentioned above, our PCR
amplifications suggest an enrichment of bphA1 genes in the polluted soils. Thus it appears likely that the ability to utilize certain CBs selected for certain bph operons, and, as long as horizontal gene transfer plays no major role, thereby for certain taxa, and that this selection is also reflected by the ditA1 genes. This interpretation agrees with results of Witzig et al. (40) for isolates from the Wittenberg site, which suggest that the observed clustering of ditA1 sequences is paralleled by taxonomic clustering, as determined by gyrB sequencing. The widespread occurrence of dit genes in CB and aromatic hydrocarbon degraders in general may originate from the utilization of ubiquitous resin acids prior to introduction of the pollutants (40).

**Reconstitution of complete BphA gene clusters and determination of gene expression.** In order to assess whether or not the environmental DNA fragments harbor the potential to encode parts of active dioxygenases and to establish sequence-function and sequence-specificity correlations, 21 of the different alpha subunit core sequences were used to reconstitute complete bphA gene clusters. This was done by supplementing them with the missing flanking sequences of the alpha subunit gene as well as with the other three genes required for BphA systems. These encode the beta subunit, a ferredoxin, and a ferredoxin reductase. In principle, this was done as previously described, by exchanging the respective core segment of the cloned bphA-LB400 gene cluster against an amplified core fragment (18). The procedure was modified in that a newly constructed plasmid, pAIA6099, harboring an inactive bphA1 core segment that lacks 36 bp (for details see MATERIALS AND METHODS) was used for the replacement. This recipient plasmid also harbored genes bphBC from strain LB400, encoding the two subsequent catabolic pathway enzymes. Their presence enables verification if the products of the initial dioxygenation are further metabolized. Many of the resulting MCPs possess characteristic electronic spectra which not only facilitate assessment of dioxygenase activity, but also permit to some extent assignments
of the regiospecificity of the initial dioxygenation (see below).

The primary transformants of this reconstitutive subcloning were checked for genetic correctness, and the expression strain *E. coli* BL21[DE3](pLysS) was transformed by the respective plasmids. The resulting clones, 14 of them belonging to cluster I, six to cluster II and one to cluster III, were used for further analysis.

Cellular concentrations of dissolved alpha and beta subunits were determined by quantitation of SDS-PAGE band intensities. In most, but not all cases, the concentrations of hybrid alpha subunits were similar to that of the parental BphA1-LB400 (data not shown). The WT beta subunits were generally found in some excess, compared to the concentrations of the alpha subunits. Three hybrid subunits were not detected. In accordance with this observation, no catalytic activity was observed (below).

**Catalytic activity of the hybrid enzymes and its correlation with AA substitutions.** Catalytic activity of the hybrid enzymes was quantitated with resting cells and biphenyl as substrate (Table 3). It has been shown with this and a wide range of other substrates that the activities of BphB and BphC in these strains are not rate-limiting (41, 43; C. S.-G. and B. H., unpublished results). Of 14 hybrid BphAs belonging to cluster I, 10 were active, 3 were inactive, and in one case no alpha subunit was found. Of 6 BphAs of cluster II, 5 were active; again in one case a large subunit was not detected. Also the alpha subunit of cluster III hybrid WC27 was not observed. In accordance with this, no activity was detected, neither with biphenyl, nor with a range of other aromatic compounds, such as toluene, isopentylbenzene, diphenylmethane and dibenzofuran. A possible reason for the absence of some hybrid subunits among the dissolved proteins is an incompatibility between recipient and donor protein segments with respect to proper folding, leading to rapid proteolysis and/or precipitation of the hybrid. Of all 18 detected hybrids, 83 % were definitely active, whereas 7 % showed no activity under assay conditions. As hybrid formation, even with core segments
from active ARHDOs, will not in all cases yield active enzymes, the percentage of "active" donor segments may actually be higher.

When AA deviations relative to the prototypes resulted in significant changes of catalytic activity, decreases rather than increases were observed, a behaviour expected for the introduction of random AA substitutions into the prototype sequence. In the following, AA deviations that strongly affected activity (Table 3) are discussed with respect to structure-function relationships. In cluster I, this was the case for the apparently inactive hybrids WB70, WC10 and WC65 as well as for hybrid WC23, which displayed a 40-fold decreased activity.

BphA-WB70h (h denotes hybrid) harbors only a single substitution relative to its WB15 prototype, Ser274Pro (AA numbering from BphA1-LB400). In order to examine which other residues are tolerated at this position in related ARHDOs, we neglected sequence data without a positive correlation to enzymatic activity and restricted our sequence alignment to 23 alpha subunit sequences of active class II ARHDOs. This showed that also Thr and Ala occur at this position. In the crystal structure of the closely related LB400 enzyme (20; PDB ID 2XRX), Ser274 has no direct contact to the biphenyl substrate. Its replacements by Ala or Thr do not appear to lead to steric clashes. A Pro residue, however, would shift Gly273. This displacement could, via Met324, well be transmitted to His323, resulting in steric interference with the substrate.

Similarly, the two replacements in hybrid WC10, Gln322Arg, and Ala354Thr should not directly affect interactions with the substrate. Ser354 as well as Pro354, but not Thr354 are found in our compilation of active enzymes. In the LB400 structure, neither of these 3 changes appears to cause steric interference. The only other residue found in position 322 is Glu, which is sterically similar to Gln. Although BphA-LB400 seems to be able to accommodate Glu as well as Arg at this position, it seems likely that the Gln322Arg exchange
is mainly responsible for the inactivity of WC10, as it probably affects the positions of its direct main chain neighbors, Gly321 and His323, which, according to the LB400 structure, make van der Waals contacts with the substrate.

BphA-WB65h also harbors two changes, Gly271Arg, and Tyr370His. In active enzymes Phe, but no His occurs at position 370. Gly 271, however, is invariant. The crystal structure of the LB400 dioxygenase clearly shows that both residues are remote from the substrate-binding site and that His370, but not the large Arg271 side chain can be accommodated without major rearrangements of the protein structure. This suggests that probably the latter exchange triggers structural changes that result in inactivity.

Hybrid WC23 contains two replacements in close proximity, Ile375Leu and Asn377Ser. Leu375 and Thr377, but no Ser377, are found in active enzymes. The smaller Ser would generate a cavity within the fold, unless this is prevented by shifts of Ser itself and of adjacent residues. This may well affect the active site, for example the substrate-lining Phe378, which has been shown to be critical for dioxygenation (42).

Only one of the six cluster II hybrids, WB14, showed a remarkable (about 35-fold) reduction in activity. It contains only a single change, Val352Met. The LB400 structure indicates that this Val is well remote from the active site and that a Met residue could be accommodated at position 352. Thus a mechanism that triggers the observed drastic decrease in activity is not obvious. A crucial role of this Val residue is in agreement with its invariance in the compilation of active enzymes. We note, however, that changes of other invariant residues such as Lys291, His343, Val358 and Asp361 did not lead to drastic losses of activity.

**Assay of BphA prototypes for productive dioxygenation of selected CBs.** The strains producing the prototype enzymes BphA-WB15h and WC18 were assayed for dioxygenation of 10 CBs that were major, minor or no constituents of the PCB mixture found
at the Wittenberg site. It had previously been shown that the BphB and BphC enzymes, which were also synthesized by the recombinant strains, were able to convert ortho,meta-dioxygenated products of all of these CBs into MCPs (33, 34, 43). Initial dioxygenations that formed this type of further degradable catabolites, were termed “productive”. The finding that BphC of strain LB400 is unable to convert meta,para-dihydroxylated biphenyl (11a) indicates that productive dioxygenations in the pathway examined here are directed to ortho and meta carbons. The 10 congeners used were di- or trisubstituted and possessed no unchlorinated ring. They contained all three types of monochlorinated rings. Six of them (2,2'-, 2,4'-, 4,4', 3,4,2'-, 2,4,3'- and 3,4,4'-CB) were present at the contaminated site in different amounts (Table 4), while the other four (3,3'-, 3,5,2'-, 2,3,3'- and 3,5,4'-CB) were not detected (25).

An overview on the results is shown in Table 4, where congeners are listed according to the type of their monochlorinated ring. With a single exception (below), productive dioxygenation was always directed towards the monochlorinated ring. The position of the chlorine at this ring largely determined the rate of dioxygenation. There was no obvious correlation with the aqueous solubilities of the congeners (Table 4).

Some simple rules can be deduced from experimental data for correlations between absorption maxima and substituent patterns of chlorinated MCPs (Table 5), which allow some assignments of the sites of the initial dioxygenations. (A) Substitutions at carbons 5, 4 or ortho (for numbering see footnote of Table 5) shift absorption maxima from values above 430 nm to increasingly lower values. (B) In cases of multiple substitutions, these effects dominate in the order ortho > 4 > 5. Values based on these rules are given for the different MCPs in the "Expected" column of Table 5. In the subsequent two columns, they are compared with other experimental data. As can be seen, the observed values agree in many, but not all cases.
CBs with an ortho-monochlorinated ring were most readily turned over by both enzymes (Table 4). There was a fundamental difference, however, regarding the site of attack, as reflected (with the exception of 2,2'-CB) by the different absorption maxima of the resulting MCPs (Table 5). These indicate, as shown in Fig. 2, that BphA-WC18h dioxygenated 2,4', 3,4,2'- and 3,5,2'-CB at unchlorinated carbons (positions 5,6 or 5',6', respectively), whereas BphA-WB15h attacked them at the semichlorinated side of the ring (positions 2,3 or 2',3', respectively). It appears very likely that the same scheme also applies to 2,2'-CB, where the resulting MCPs are neither expected nor found to possess a significant difference in their absorption maxima. Generally, the rates of attack of the ortho-monochlorinated ring were higher with BphA-WB15h than with BphA-WC18h. The two of these four CBs that were found in higher concentrations at the Wittenberg site (2,4' and 3,4,2'-CB) were the best substrates for the latter enzyme. Such a correlation was less clear for BphA-WB15h, as also 2,2'-CB was an excellent substrate for this dioxygenase.

CBs with a meta-monochlorinated ring were much "slower" substrates with both enzymes (Table 4). Independent of the substitution pattern of the non-oxidized ring, BphA-WC18h turned all three congeners over at similar rates. 3,3'-CB was dioxygenated at the unchlorinated side (Table 5). In analogy with this, we assigned the same regiospecificity to the dioxygenations of 2,3,3'- and 2,4,3'-CB (Fig. 2), which also agrees with the finding that an attack involving meta-dechlorination has very rarely been observed (37). In contrast to BphA-WC18h, the WB15 enzyme attacked the meta-monochlorinated ring only in 2,3,3'-CB. It also slowly dioxygenated 2,4,3'-CB; here, however, as deduced from Table 5, the attack was not directed against the monochlorinated ring, but against carbons 2 and 3 (Fig. 2), in agreement with the described preference of this enzyme for chlorinated ortho carbons. Only BphA-WB15h showed a somewhat faster productive dioxygenation of 2,4,3'-CB, which is the only of these three congeners that was found at the contaminated site.
For CBs possessing a *para*-chlorinated ring, almost no turnover was observed with both enzymes (Table 4). Only BphA-WC18h slowly dioxygenated of 4,4'-CB, which, like 3,4,4'-CB, belongs to the CBs that are predominant at the Wittenberg site.

In our assays, BphA-WC18h showed a broader CB range than BphA-WB15h. On the other hand, the latter enzyme clearly was a better catalyst for the dioxygenation of CBs with *ortho*-chlorinated rings. A correlation between the concentrations of the selected CBs at the polluted site and their rates of productive dioxygenation by the two prototype enzymes was not generally apparent. The expectation of such a correlation may well be based on an oversimplified view. Rapid dioxygenation of a given CB must not necessarily result in an evolutionary advantage. It may indeed result in a disadvantage, if accumulating catabolites exert toxic effects (8, 11, 12, 16, 27, 29). Moreover, the evolution of catalytic activity in the presence of substrate mixtures will necessarily lead to compromises, so that any given enzyme will only be able to efficiently transform a fraction of substrates. Thus evolution towards the efficient utilization of a substrate other than applied in our assays may obscure correlations with the congeners used here.

**Concluding remarks.** The present work demonstrated the feasibility of the applied approach in not only retrieving ARHDO sequence information from metagenomic DNA, but also experimental data on enzymatic properties such as activity, substrate and product ranges. In this context, it will be of interest to investigate in detail in how far the sequence diversity within a given sequence cluster affects the substrate spectrum. Moreover, it appears intriguing to obtain active enzymes from the donor segments of similarity cluster III and, generally speaking, of other classes of ARHDOs by constructing alternative recipient gene clusters, based on genes from strains such as RHA1, PAH degraders and others.
REFERENCES

bacterial populations and in biphenyl dioxygenase gene diversity in a polychlorinated
biphenyl-polluted soil after introduction of willow trees for rhizoremediation. Appl.


59:143-169.

Occurrence and expression of glutathione S-transferase-encoding bphK genes in
Burkholderia sp. strain LB400 and other biphenyl-utilizing bacteria. Microbiology
145:2821-2834.

5. Berggren, K., T. H. Steinberg, W. M. Lauber, J. A. Carroll, M. F. Lopez, E.
luminescent ruthenium complex for ultrasensitive detection of proteins immobilized on


FIGURE LEGENDS

FIG. 1. Dendrograms of metagenomic BphA1 (A) and DitA1 (B) sequences from the Wittenberg site. Dendrograms were derived from sequence alignments. Similarity clusters are indicated by brackets and are designated by roman numerals. Scale bars give distances in amino acid substitutions per site. DitA1 sequences (40) are identified by database accession numbers.

FIG. 2. Regiospecificity of CB dioxygenation by cluster I and cluster II prototype hybrid enzymes. Sites of productive attack were deduced as given in Table 5 and in the text.
Fig. 2

<table>
<thead>
<tr>
<th></th>
<th>BphA-B15h</th>
<th>BphA-C18h</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,2'-CB</td>
<td><img src="image" alt="2,2'-CB" /></td>
<td><img src="image" alt="2,2'-CB" /></td>
</tr>
<tr>
<td>2,4'-CB</td>
<td><img src="image" alt="2,4'-CB" /></td>
<td><img src="image" alt="2,4'-CB" /></td>
</tr>
<tr>
<td>3,4,2'-CB</td>
<td><img src="image" alt="3,4,2'-CB" /></td>
<td><img src="image" alt="3,4,2'-CB" /></td>
</tr>
<tr>
<td>3,5,2'-CB</td>
<td><img src="image" alt="3,5,2'-CB" /></td>
<td><img src="image" alt="3,5,2'-CB" /></td>
</tr>
<tr>
<td>3,3'-CB</td>
<td><img src="image" alt="3,3'-CB" /></td>
<td><img src="image" alt="3,3'-CB" /></td>
</tr>
<tr>
<td>2,3,3'-CB</td>
<td><img src="image" alt="2,3,3'-CB" /></td>
<td><img src="image" alt="2,3,3'-CB" /></td>
</tr>
<tr>
<td>2,4,3'-CB</td>
<td><img src="image" alt="2,4,3'-CB" /></td>
<td><img src="image" alt="2,4,3'-CB" /></td>
</tr>
<tr>
<td>4,4'-CB</td>
<td><img src="image" alt="4,4'-CB" /></td>
<td><img src="image" alt="4,4'-CB" /></td>
</tr>
</tbody>
</table>
Table 1. Similarities of amino acid sequences encoded by the Wittenberg clones.

<table>
<thead>
<tr>
<th>Sequence cluster</th>
<th>No. of clones sequenced</th>
<th>Amino acid sequence identity (%)</th>
<th>With the most similar sequence in the data base&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Within the cluster</td>
<td>With cluster II</td>
<td>With cluster III</td>
</tr>
<tr>
<td>I</td>
<td>37</td>
<td>97-100</td>
<td>85-88</td>
</tr>
<tr>
<td>II</td>
<td>12</td>
<td>97-100</td>
<td>-</td>
</tr>
<tr>
<td>III</td>
<td>2</td>
<td>99</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup> If the most similar sequence belongs to a putative enzyme, a second value is given, which refers to the most similar sequence of a non-putative enzyme.

<sup>b</sup> Biphenyl dioxygenase alpha subunit from *Burkholderia xenovorans* LB400 (NCBI Protein Database accession no. ABE37059).

<sup>c</sup> Putative ring-hydroxylating dioxygenase alpha subunit from *Burkholderia* sp. WBF3 (accession no. ABG75584) and WBF4 (accession no. ABG75585). Biphenyl dioxygenase alpha subunit from *Pseudomonas pseudoalcaligenes* KF707 (accession no. Q52028).

<sup>d</sup> Putative ring-hydroxylating dioxygenase alpha subunit from *Burkholderia ambifaria* IOP40-10 (accession no. EDT02834). Biphenyl dioxygenase alpha subunit from *Rhodococcus jostii* RHA1 (accession no. BAA06868).
Table 2. Distribution of DNA sequences of soils B and C between sequence clusters.

<table>
<thead>
<tr>
<th>Sequence Cluster</th>
<th>Sequences obtained from soil</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number</td>
<td>%</td>
<td>Number</td>
</tr>
<tr>
<td>I</td>
<td>17</td>
<td>65</td>
<td>20</td>
</tr>
<tr>
<td>II</td>
<td>9</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>III</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Sum</td>
<td>26</td>
<td>100</td>
<td>25</td>
</tr>
</tbody>
</table>
Table 3. Activity of hybrid dioxygenases with biphenyl as substrate, and correlation with AA substitutions in BphA1.

<table>
<thead>
<tr>
<th>Clone name</th>
<th>Sequence cluster</th>
<th>Specific activity [(pmol/min)/mg BphA1]</th>
<th>AA substitution relative to cluster prototype</th>
<th>Comparison of and comments on AA substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>WB6</td>
<td>I</td>
<td>21.6</td>
<td>D279G</td>
<td>G, N also found at this position.</td>
</tr>
<tr>
<td>WB11</td>
<td>I</td>
<td>30.0</td>
<td>K291E</td>
<td>K invariant.</td>
</tr>
<tr>
<td><strong>WB15</strong></td>
<td>I</td>
<td><strong>41.9</strong></td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>WB40</td>
<td>I</td>
<td>49.1</td>
<td>L309P</td>
<td>V also found at this position.</td>
</tr>
<tr>
<td>WB41</td>
<td>I</td>
<td>59.9</td>
<td>S283P</td>
<td>SBSR, I, T, M, L also found at this position.</td>
</tr>
<tr>
<td>WB70</td>
<td>I</td>
<td>nad</td>
<td>S274P</td>
<td>Probable steric clash between P274 and G273.</td>
</tr>
<tr>
<td>WB72</td>
<td>I</td>
<td>25.9</td>
<td>T356A</td>
<td>V, W, I also found at this position.</td>
</tr>
<tr>
<td>WB74</td>
<td>I</td>
<td>nad, npd</td>
<td>I247V</td>
<td>L, M also found at this position.</td>
</tr>
<tr>
<td>WB42</td>
<td>I</td>
<td>654</td>
<td>Y370H</td>
<td>F also found at this position.</td>
</tr>
<tr>
<td>WC5</td>
<td>I</td>
<td>75.9</td>
<td>L304H</td>
<td>K, R also found at this position.</td>
</tr>
<tr>
<td>WC10</td>
<td>I</td>
<td>nad</td>
<td>Q322R</td>
<td>SBSR, E also found at this position.</td>
</tr>
<tr>
<td>WC15</td>
<td>I</td>
<td>80.7</td>
<td>F265L</td>
<td>Y also found at this position.</td>
</tr>
<tr>
<td>WC23</td>
<td>I</td>
<td>0.904</td>
<td>I375L</td>
<td>L, V also found at this position.</td>
</tr>
<tr>
<td>WC42</td>
<td>I</td>
<td>94.9</td>
<td>I339V</td>
<td>V also found at this position.</td>
</tr>
<tr>
<td>WC65</td>
<td>I</td>
<td>nad</td>
<td>G271R</td>
<td>G invariant.</td>
</tr>
<tr>
<td>WC47</td>
<td>II</td>
<td>14.2</td>
<td>V352M</td>
<td>V invariant.</td>
</tr>
<tr>
<td>WB14</td>
<td>II</td>
<td>654</td>
<td>E250G</td>
<td>G, D, N also found at this position.</td>
</tr>
<tr>
<td>WC18</td>
<td>II</td>
<td>500</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>WC27</td>
<td>III</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
</tbody>
</table>

*Prototype lines are in bold.*

*SDs were ± 30%.*
Specific activity was calculated using a molar extinction coefficient of 33200 for the MCP (35).

**AA numbering from BphA1-LB400.**

With 23 alpha subunit sequences of active benzene-type (class II) ARHDOs. Their NCBI protein database accession nos. are: CAA56346, AAB07750, BAA06868, AAP74038, AAA26005, ADI95397, CAA06970, Q07944, AAC43632, AAC46390, ABE37059, AAB88813, Q52028, AAK14781, 1WQL_A, AAD12763, AAB36666, AAC03436, CAB99196, AAC44526, CAA08985, BAJ72245, BAC01052. We changed the BAC01052 sequence in positions 351-359 to VWAFVVVDA, because in this segment the database sequence obviously switched to a wrong reading frame.

**nad, no activity detected.**

**npd, no protein (BphA1) detected.**

**na, not applicable**

**SBSR, substrate binding site residue, i.e., AA is located within a distance of 6 Å from the biphenyl molecule in the BphA-LB400 structure 2XRX (20).**
Table 4. Productive dioxygenation of various CBs by prototype hybrid BphAs.

<table>
<thead>
<tr>
<th>Chlorinated carbons</th>
<th>Contamination of site&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Aqueous solubility [µM]</th>
<th>Rate of product formation [mAbs&lt;sub&gt;max&lt;/sub&gt;/h]</th>
<th>BphA-WB15h</th>
<th>BphA-WC18h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>BphA-WB15h mean SD</td>
<td>BphA-WC18h mean SD</td>
<td></td>
</tr>
<tr>
<td>2,2′</td>
<td>+</td>
<td>1.91&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1370 320</td>
<td>87 51</td>
<td></td>
</tr>
<tr>
<td>2,4′</td>
<td>++</td>
<td>3.47&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1170 270</td>
<td>626 223</td>
<td></td>
</tr>
<tr>
<td>3,4,2′</td>
<td>+++</td>
<td>0.616&lt;sup&gt;b&lt;/sup&gt;</td>
<td>507 11</td>
<td>121 16</td>
<td></td>
</tr>
<tr>
<td>3,5,2′</td>
<td>-</td>
<td>0.501&lt;sup&gt;b&lt;/sup&gt;</td>
<td>133 41</td>
<td>46 10</td>
<td></td>
</tr>
<tr>
<td>3,3′</td>
<td>-</td>
<td>0.354&lt;sup&gt;b&lt;/sup&gt;</td>
<td>npo&lt;sup&gt;d&lt;/sup&gt; npo</td>
<td>37 13</td>
<td></td>
</tr>
<tr>
<td>2,3,3′</td>
<td>-</td>
<td>0.426&lt;sup&gt;c&lt;/sup&gt;</td>
<td>31 5</td>
<td>54 7</td>
<td></td>
</tr>
<tr>
<td>2,4,3′</td>
<td>+</td>
<td>0.776&lt;sup&gt;b&lt;/sup&gt;</td>
<td>59 9</td>
<td>46 11</td>
<td></td>
</tr>
<tr>
<td>4,4′</td>
<td>+++</td>
<td>0.426&lt;sup&gt;b&lt;/sup&gt;</td>
<td>npo npo</td>
<td>7 1</td>
<td></td>
</tr>
<tr>
<td>3,4,4′</td>
<td>+++</td>
<td>0.301&lt;sup&gt;c&lt;/sup&gt;</td>
<td>npo npo npo npo</td>
<td>npo npo npo npo</td>
<td></td>
</tr>
<tr>
<td>3,5,4′</td>
<td>-</td>
<td>0.194&lt;sup&gt;c&lt;/sup&gt;</td>
<td>npo npo npo npo</td>
<td>npo npo npo npo</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Gross classification of CB contaminations, based on gas chromatography data (25). -, no; +, minor; ++, medium; ++++, major constituent of the Wittenberg site.

<sup>b</sup> Experimental (28).

<sup>c</sup> Calculated (40).

<sup>d</sup> npo, no product observed.
Table 5. Absorption maxima\(^a\) of MCPs formed via dioxygenation of CBs by prototype hybrid BphAs, and tentative assignments of initially oxidized carbons.

<table>
<thead>
<tr>
<th>CB</th>
<th>Potentially oxidized carbons</th>
<th>Chlorinated carbons</th>
<th>Expected(^{cd}) (\lambda_{\text{max}}) [nm]</th>
<th>Previous experimental data (\lambda_{\text{max}}) [nm]</th>
<th>Oxidized carbons</th>
<th>BphA-WB15h (\lambda_{\text{max}}) [nm]</th>
<th>BphA-WC18h (\lambda_{\text{max}}) [nm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,2'</td>
<td>2,3</td>
<td>8</td>
<td>(\approx 393)</td>
<td>392(^c)</td>
<td>2,3(^h)</td>
<td>393</td>
<td>393</td>
</tr>
<tr>
<td>2,4'</td>
<td>2,3</td>
<td>10</td>
<td>&gt; 430</td>
<td>438(^c)</td>
<td>2,3(^h)</td>
<td>433 (399)</td>
<td>400</td>
</tr>
<tr>
<td>3,4,2'</td>
<td>2,3</td>
<td>3,8</td>
<td>(\approx 393)</td>
<td>395</td>
<td>5,6(^e)</td>
<td>435 (401)</td>
<td>401</td>
</tr>
<tr>
<td>3,5,2'</td>
<td>2,3</td>
<td>4,8</td>
<td>(\approx 393)</td>
<td>395</td>
<td>5,6,10</td>
<td>435 (400)</td>
<td>400</td>
</tr>
<tr>
<td>3,3'</td>
<td>2,3</td>
<td>9</td>
<td>&gt; 430</td>
<td>430 (410)</td>
<td>5,6(^e,h)</td>
<td>npo</td>
<td>395</td>
</tr>
<tr>
<td>2,3,3'</td>
<td>5,6</td>
<td>4,5,9</td>
<td>(\approx 402)</td>
<td>400</td>
<td>5',6(^h)</td>
<td>385 ± 3</td>
<td>388(^j)</td>
</tr>
<tr>
<td>2,3'</td>
<td>5,6</td>
<td>8,9</td>
<td>(\approx 393)</td>
<td>395</td>
<td>5',6(^h)</td>
<td>385 ± 3</td>
<td>388(^j)</td>
</tr>
<tr>
<td>2,4,3'</td>
<td>2,3</td>
<td>3,9</td>
<td>&gt; 430</td>
<td>433(^f)</td>
<td>2,3(^f)</td>
<td>420 (438)</td>
<td></td>
</tr>
<tr>
<td>2,3'</td>
<td>5,6</td>
<td>3,5,9</td>
<td>(\approx 402)</td>
<td>437(^f)</td>
<td>5,6(^h)</td>
<td>420 (438)</td>
<td></td>
</tr>
<tr>
<td>5,6'</td>
<td>5,6</td>
<td>4,8,10</td>
<td>(\approx 393)</td>
<td>395</td>
<td>5,6,10</td>
<td>388(^j)</td>
<td></td>
</tr>
</tbody>
</table>

\(\lambda_{\text{max}}\) values ± 2 nm, unless otherwise indicated. Numbers in parentheses indicate final values in case of a shift of the absorption maximum during incubations.

Carbon numbering in MCP is as shown here:

```
10 11
 9 8
7 6
5 4
3 2
 1
```

\(\lambda_{\text{max}}\) values ± 2 nm, unless otherwise indicated. Numbers in parentheses indicate final values in case of a shift of the absorption maximum during incubations.

\(\lambda_{\text{max}}\) values ± 2 nm, unless otherwise indicated. Numbers in parentheses indicate final values in case of a shift of the absorption maximum during incubations.

\(\lambda_{\text{max}}\) values ± 2 nm, unless otherwise indicated. Numbers in parentheses indicate final values in case of a shift of the absorption maximum during incubations.
i npo, no product observed.

j Absorption unstable.