One-Component Styrene Monooxygenases: An Evolutionary View on a Rare Class of Flavoproteins

**Dirk Tischler · Janosch A. D. Gröning ·
Stefan R. Kaschabek · Michael Schlömann**

Abstract Styrene monooxygenases (SMOs) are catalysts for the enantioselective epoxidation of terminal alkenes. Most representatives comprise a reductase and a monooxygenase which are encoded by separate genes (*styA*, *styB*). Only six presumed self-sufficient one-component SMOs (*styA2B*) have previously been submitted to databases, and one has so far been characterized. StyA2B can be supported by another epoxidase (StyA1) encoded by *styA1*, a gene in direct neighborhood of *styA2B*. The present report describes the identification of a further *styA1/styA2B*-like SMO, which was detected in *Rhodococcus opacus* MR11. Based on the initially available sequences of *styA2B*-type SMOs, primers directed at conserved sequences were designed and a 7,012-bp genomic fragment from strain MR11 was obtained after PCRs and subsequent genome walking. Six open reading frames (ORFs) were detected and compared to genomic fragments of strains comprising either two- or one-component SMOs. Among the proteins encoded by the ORFs, the monooxygenase StyA1/StyA2B showed the highest divergence on amino acid level when comparing proteins from different sources. That finding, a rare distribution of *styA2B* genes among bacteria, and the general observation of evolution from simple to complex systems indicate that one-component SMOs evolved from two-component ancestors. Analysis of gene products from *styA/styB*- and *styA1/styA2B*-like SMOs revealed that a fusion of *styA/styB* to *styA2B* might have happened at least twice among microorganisms. This points to a convergent evolution of one-component SMOs.

Keywords One-component styrene monooxygenase · *Rhodococcus opacus* · Self-sufficient fusion protein · Covalent evolution · Enantioselective epoxidation · Flavin monooxygenase

D. Tischler (✉) · J. A. D. Gröning · S. R. Kaschabek · M. Schlömann
Interdisciplinary Ecological Center, TU Bergakademie Freiberg,
Leipziger Str. 29, 09599 Freiberg, Germany
e-mail: dirk-tischler@email.de

Introduction

Styrene monooxygenases (SMOs) perform a selective oxygenation of the vinyl side chain of styrene to styrene oxide and thus initiate its degradation process [1]. Most SMOs described so far are two-component, flavoprotein monooxygenases and originate from *Pseudomonas* species [2, 3]. The two components, a single epoxidase and a single FAD reductase, are encoded by two separate genes, *styA* and *styB*, respectively. The reductase (StyB) solely uses NADH to reduce the FAD cofactor which is then utilized by the epoxidase (StyA) to activate molecular oxygen and to catalyze the epoxidation of styrene to (*S*)-styrene oxide [4, 5]. The high specific oxygenation activity, excellent enantioselectivity, and a broad substrate range favor SMOs for biotechnological application [2, 6, 7]. However, the two-component character may imply some disadvantages related to recombinant expression, inefficient cofactor transfer possibly resulting in uncoupling, or complex purification strategies [4, 7–10].

Similar limitations were reported for multi-component P450 monooxygenases [11–13]. For these types of enzymes, evolutionary fusion processes leading to self-sufficient proteins like the cytochrome P450 BM3 from *Bacillus megaterium* [13] turned out to change biochemical properties in such a way, favoring biotechnological applicability.

Natural diversity among P450 and other monooxygenase systems shows how differently such enzymes can be composed or organized [3, 11, 13, 14]. Thus, in case of SMOs, also other than just two-component systems are likely to be functional in an effective way. StyA2B from *Rhodococcus opacus* 1CP has been recently described as first example of a biochemically characterized fused type of SMO [9]. Unfortunately, the oxygenating activity was found to be rather low, when compared to the two-component SMO (StyA and StyB) from *Pseudomonas* sp. strain VLB120 [5], making StyA2B—at first sight—not suited for biocatalysis. However, recent investigations revealed that the fusion protein StyA2B can be supported by another styrene monooxygenase (StyA1) to provide more efficiency as well as higher productivity [10]. The system obtained is encoded by two separate genes (*styA1* and *styA2B*) located in direct neighborhood on the genome [9]. The higher epoxidation efficiency, expressed as yield of styrene oxide per mole of reducing equivalent, brings the new fusion type of SMO into focus for biotransformation. But the moderate epoxidation rate hampers its applicability and asks for the identification of further representatives.

Since only a few genes of such naturally fused *styA2B*-type SMOs were available from databases [1], a sequence-based screening method seemed to be favorable in order to identify further *styA2B*-like genes in strain collections or in clone libraries. Approaches performed for other monooxygenase systems demonstrated the general applicability of such a procedure [12, 15, 16].

The one-component SMOs are interesting from an evolutionary point of view as well as for application in biotransformation [9]. Thus, questions can be raised as to the natural abundance of this SMO type or whether the corresponding genes are part of conserved operons. It is also a question from which type of ancestral protein and how often the one-component SMOs arose. Finally, the biocatalytic potential of these monooxygenases is of interest. In order to approach such questions, it was considered appropriate to search for additional representatives by a sequence-based screening.

Materials and Methods

Bacterial Strains, Plasmids, and Culture Conditions

Rhodococcus strains listed below were freshly streaked and kept on mineral medium plates adapted from Dorn et al. [17], but containing 54 mM sodium–potassium phosphate (pH 7.3) with phenylacetic acid as sole source of carbon and energy in a final concentration of 2 mM. The plates were incubated first at 30 °C (up to 24 h) and later at room temperature (up to 96 h). For strain preservation, the plates were kept at 4 °C. In addition, *R. opacus* strains 1CP and MR11 were cultivated in liquid minimal medium as described earlier with phenylacetic acid as sole carbon source [9].

The following *Rhodococcus* strains were used in this study: *Rhodococcus erythropolis* L88 [18], *Rhodococcus* sp. US-B1 (Fraunhofer-Gesellschaft, Stuttgart), *R. erythropolis* SQ1 [19], *R. erythropolis* B7g (this institute), *Rhodococcus zopfii* DSM44108, *R. erythropolis* HL24-1 [20], wild-type and mutants of *R. opacus* 1CP ([21], and this institute), *R. erythropolis* DSM43066T, *R. erythropolis* BD2.101 [22], *Rhodococcus ruber* DSM8425, *R. opacus* MR2822 (Institute for Microbiology, Göttingen University), *R. opacus* MR11 (DSM427; 23), *R. opacus* MR2226 [23]; Institute for Microbiology, Göttingen University), *R. opacus* DSM43943, *Rhodococcus* sp. PM1 and PM2 (this institute), *Rhodococcus rhodochrous* S5 (DSM6697), *Rhodococcus jostii* RHA1 [24], *R. rhodochrous* 172*, *Rhodococcus* sp. R2(2)*, *Rhodococcus* sp. 2 L*, *Rhodococcus* sp. 3 L*, *Rhodococcus* sp. 4 L*, *Rhodococcus* sp. 5 L*, *Rhodococcus* sp. 6 L*, *Rhodococcus rhodnii* 135*, *R. rhodochrous* 89*, *R. opacus* 13A*, and *R. opacus* 557*. The rhodococci marked with an asterisk (*) have previously been obtained from the strain collection of L. A. Golovleva (G. K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino).

Arthrobacter aurescens TC1 was provided by the lab of Prof. M. Sadowsky (University of Minnesota, USA) and cultivated on modified R minimal medium with atrazine as sole source of carbon and nitrogen as described elsewhere [25].

Escherichia coli DH5α from Gibco-BRL for cloning purposes and clones obtained in this study were cultivated as described previously on complex media with appropriate antibiotics [26].

DNA Techniques and PCR

Standard protocols were used for DNA cloning, transformation, and nucleic acid purification [26]. Primers designed and applied in this study as well as plasmids used or constructed are listed in the Supplementary Material section.

Primers (also degenerated ones) for general cloning, screening, or genome-walking approaches were designed on the basis of an already published alignment ([9]; see Fig. 1). So far, most *styA2B*-homologous genes were detected in Actinobacteria with a relatively high GC content. Here, the recent progress in the field of PCR techniques for GC-rich template DNA as obtained from Actinobacteria provided helpful information [12, 27, 28].

For PCRs, a general protocol with Taq or DreamTaq polymerase (Fermentas) was used according to the manufacturer's instructions. The reactions included additional DMSO (up to 5 % v/v). In a few cases, the $MgCl_2$ concentration was varied between 1.5 and 3 mM. For screening PCRs to identify *styA2B*-homologous genes from whole cells or purified DNA samples, conditions were further optimized by varying the annealing temperature or by performing touch-down PCRs. Products obtained from PCRs were agarose gel purified by
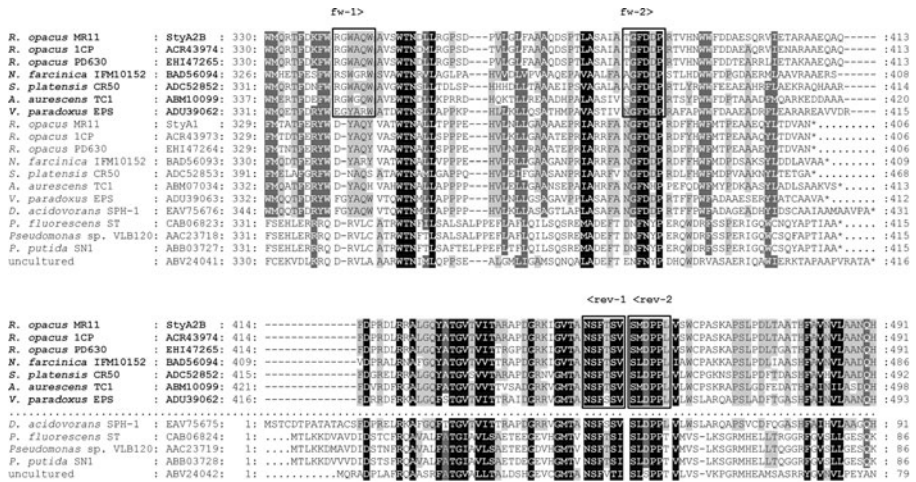
Fig. 1 Multiple-sequence alignment of SMOs. The self-sufficient one-component SMOs (StyA2B-like; *bold* designations) were aligned with corresponding single oxygenases (StyA1-like) as well as with conventional two-component SMOs (StyA/StyB-like). Relevant locus for screening primer design was the fusion region of StyA2B comprising parts of both the monooxygenase and the flavin reductase domain. The corresponding sequences are *boxed*. Two forward primers (fw-1 and fw-2) and two reverse primers (rev-1 and rev-2) were designed. Amino acid positions and if present the C-terminal end (*) of all proteins are indicated. *Shading* indicates different degrees of conservation of similar amino acids among the sequences compared (*black*, 100 %; *dark gray*, >80 %; *gray*, >60 %). An alignment of fewer representatives presented earlier showed the complete protein sequences and a similar degree of conservation was obtained [9]

means of 1 % to 3 % agarose gels, depending on the product size. Products were cloned into pJET1.2/blunt (Fermentas) or pBluescript II SK + (Stratagene). Representative clones were controlled after plasmid preparation by restriction digestion and PCR.

In order to prepare template DNA for the optimization of the screening PCR, genomic DNA of *R. opacus* 1CP and *A. aurescens* TC1 was isolated. Both strains were cultivated on minimal medium as described above. Biomass of actively growing cultures was harvested by centrifugation for subsequent DNA isolation. From 3-ml samples of such a growing culture, the DNA was isolated by means of a commercial kit (Dneasy Blood & Tissue Kit, Qiagen) according to the manufacturer's instructions. From larger samples, the genomic DNA was isolated by a phenol-based extraction protocol as described previously [29].

Plasmids harboring DNA fragments from *Rhodococcus* strains obtained by PCR or restriction digestion were subjected to DNA sequencing (Eurofins MWG Operon). Sequences were analyzed with BLAST [30, 31] and the Staden sequence analysis package [32]. Alignments and distance trees were generated with the software programs ClustalX (version 1.8) [33, 34], GeneDoc (version 2.6.003), and the PHYLIP 3.66 package [35]. From the latter, the following tools were chosen to create unrooted distance trees: PROTDIST (Jones–Taylor–Thornton matrix) and FITCH (Fitch–Margoliash method). The corresponding bootstrap values were calculated applying the same algorithms to the tools SEQBOOT (1,000 replicates, treated as individual data sets, random number seed=345), PROTDIST, FITCH, and CONSENSE. The values obtained were shifted to the original tree calculated previously. For comparison, distance trees were generated by applying other tools of the PHYLIP package and other algorithms such as maximum parsimony, maximum likelihood, and neighbor joining (not shown).

Screening PCR

To identify *styA2B*-like genes from isolated strains or clones via a PCR approach, specific screening primers were designed. The primer combination was intended to amplify the fusion part of *styA2B* genes comprising a sequence of the monooxygenase domain as well as of the FAD reductase domain (Fig. 1). Furthermore, the product obtained from a PCR applying one of these primer pairs was intended to allow to clearly differentiate between one-component (*styA2B*) and two-component (*styA* and *styB*) SMOs. The following amino acid motifs were identified (see Fig. 1) and used to design degenerated screening primers: RGWAQW (fw-01=5′-CGSRGSTGGGSNCRNTGG-3′), NSFTSV (rev-01=5′-NACNGWNGTRAANGWMTT-3′), TGFDDP (fw-02=5′-DCSGGNTTYGAYGAYCCB-3′), and SMDPPL (rev-02=5′-SAGSGGSGGRTCSAKSGA-3′).

The set of screening primers was first applied to test PCRs in various combinations to amplify corresponding fragments of known *styA2B*-homologous genes from pSRoA2B_P01 [9], *R. opacus* 1CP (genomic DNA), *A. aurescens* TC1 (genomic DNA), and pKNL011_E17 (*Nocardia farcinica* IFM10152, gene on plasmid; [36]). To increase product formation during PCR and to ensure product formation from different templates tested, while minimizing the unspecific formation of by-products, the PCR protocol was optimized. The PCR mix had the following composition in a final volume of 10 μl: 1-fold buffer of respective DNA polymerase, 6.25 pmol of forward and reverse primer, 2 nmol dNTPs (each), 0.2 μl template (of an appropriate concentration), and DNA polymerase (0.5 U; Fermentas). The concentrations of $MgCl_2$ (1.5 to 3 mM) and DMSO (0 to 5 %) were varied among assays. Besides the assay composition, also the PCR program was investigated and a touch-down protocol with a decreasing annealing temperature was found to be suitable (Table 1). Finally, the PCR mix as described above comprising 3 mM $MgCl_2$ and 2 % DMSO in combination with the optimized touch-down PCR program allowed to detect the *styA2B* fragments in all four DNA sources used as positive controls. This protocol was then used also for other strains. Previously, those strains were cultivated on mineral medium as described above. Then the screening PCR contained whole cells, transferred from agar plates, which served as template and products obtained were analyzed by agarose gel electrophoresis.

Nucleotide Sequences

The 7,012-bp genomic sequence derived from *R. opacus* MR11 was deposited at GenBank (accession no. JQ074239). The sequences of fragments derived from *R. opacus* 1CP which

Table 1 A screening PCR program for *styA2B* fragments

Step (cycles)	Designation	Temperature (°C)	Time (s)	Remark
1 (1)	Hot start	95	Up to 300	Optional for whole-cell PCR
2–4 (25)	Denaturation	95	Steps 2–7:	
	Annealing	71 to 65.5	Each 30	dT gradient: −0.2 °C each cycle
	Elongation	72		
5–7 (10)	Denaturation	95		
	Annealing	65.5		
	Elongation	72		
8 (1)	Final elongation	72	Up to 300	Optional for T/A-cloning procedures

were found to be in direct neighborhood of the already published one (accession no. FJ403049.1) were also uploaded to GenBank (accession no. FJ403049.2).

Results

Establishing a Screening PCR for *styA2B* Genes

In addition to the previously characterized self-sufficient one-component SMO of *R. opacus* 1CP (ACR43974), only two sequences of genes coding for such a one-component SMO have been in the database since several years (BAD56094 of *N. farcinica* IFM10152; and ABM10099 of *A. aurescens* TC1) [7, 9]. Three more such sequences originating from *R. opacus* PD630 (EHI47265), *Streptomyces platensis* CR50 (ADC52852), and *Variovorax paradoxus* EPS (ADU39062) have been submitted after the design of the screening primers. Most of the *styA2B*-type sequences were identified from Actinobacteria with a relatively high GC content. Thus, we considered especially these high-GC genes for the design of the degenerated screening primers. An alignment presented earlier [9] was used to identify conserved regions within StyA2B representatives (see "Materials and Methods"). However, the sequences submitted later (EHI47265, ADC52852, and ADU39062) and that of the StyA2B homolog from *R. opacus* MR11 found here were used to generate a multiple-sequence alignment (Fig. 1). The relevant part used for primer design was still found to be conserved. Thus, the screening primers designed here should allow the amplification of most *styA2B* homologs from Actinobacteria. The different codon usage in the *styA2B*-homologous gene encoded by the Gram-negative β-proteobacterium *V. paradoxus* EPS might hamper a successful amplification of the desired product via a screening PCR using the designed primers. Therefore, the solution can either be another set of screening primers or a higher degree of degeneracy of existing primers, which was not investigated further.

The set of screening primers was in various combinations applied to PCRs for the amplification of corresponding fragments of the *styA2B*-homologous genes from the positive controls (*A. aurescens* TC1, *N. farcinica* IFM10152, and *R. opacus* 1CP). In these experiments, each primer combination was expected to lead to a characteristic product size from the different template sources: (1) fw-01/rev-01 approx. 339 bp, (2) fw-01/rev-02 approx. 357 bp, (3) fw-02/rev-01 approx. 219 bp, and (4) fw-02/rev-02 approx. 237 bp. The best result was achieved with the primer pair fw-01/rev-02 in combination with the established touch-down PCR protocol (Table 1) since products of expected length (about 360 bp) were obtained from positive controls and less by-products were formed than from the PCRs applying other primer combinations.

Screening a Strain Collection for One-Component SMO Genes

Thirty *Rhodococcus* strains grown on phenylacetate agar were subjected to the optimized PCR protocol for a specific detection of *styA2B*-type SMOs. *R. opacus* 1CP (and its mutants) served as positive control while *R. jostii* RHA1, a genome-sequenced strain without *styA2B*, served as negative control. For *R. opacus* MR11, *R. rhodochrous* 172, and *R. rhodnii* 135, a product of expected size (about 360 bp) was determined, while for *Rhodococcus* sp. US-B1, *R. erythropolis* SQ1, *R. rhodochrous* S5, and *R. erythropolis* B7g, the reaction yielded no defined product. These strains and the positive as well as the negative control were subjected to another round of screening (Fig. 2). The previously gained results were proven. PCR products were cloned into pJET1.2/blunt and constructs obtained (pScreen_MR11, pScreen_172, and
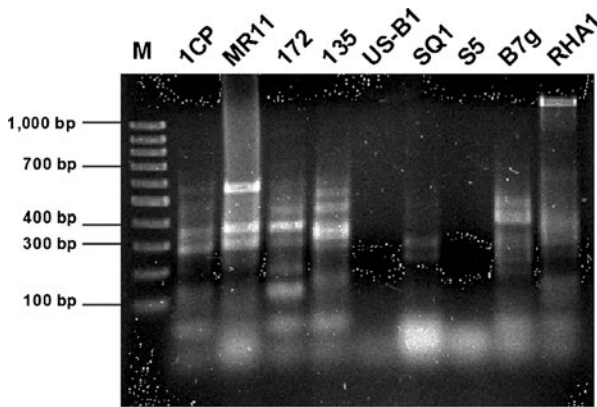
Fig. 2 Screening for one-component SMOs. *Rhodococcus* strains which in a pre-screening were either positively tested (*R. opacus* MR11, *R. rhodochrous* 172, *R. rhodnii* 135) or yielded poorly defined products (*Rhodococcus* sp. US-B1, *R. erythropolis* SQ1, *R. rhodochrous* S5, *R. erythropolis* B7g) were investigated with the screening protocol for *styA2B* homologs. *R. opacus* 1CP and *R. jostii* RHA1 were applied as a positive and a negative control, respectively. Products obtained from PCR were analyzed by agarose gel. The 100-bp DNA ladder (*M*) was used as size standard. Products of expected size (about 360 bp) were observed for strains 1CP, MR11, 172, and 135

pScreen_135) were analyzed by restriction digests and DNA sequencing. One of the constructs, pScreen_MR11, harbored an insert with a considerable similarity to StyA2B of *R. opacus* 1CP. Thus, the 359-bp fragment derived from *R. opacus* MR11 showed about 88 % identical positions on amino acid level to StyA2B and covered exactly the expected region comprising the fusion part as shown in Fig. 1 (also Fig. 3a). The other plasmid inserts sequenced, however, did not code for a product with similarity to an SMO. This means that only a single *styA2B*-homologous gene was found among 30 *Rhodococcus* strains investigated.

The Genomic Region of *R. opacus* MR11 around the Identified SMO Fragment

In order to obtain the genomic region of *R. opacus* MR11 neighboring the *styA2B*-fragment and thus comprising the complete ORFs of the putative SMO StyA1/StyA2B of the strain, different techniques were applied.

Assuming that the adjacent ORFs are highly conserved, since among the first strains all *styA2B* homologs were surrounded by the same ORFs encoding a short-chain dehydrogenase (SDR), a single monooxygenase (StyA1), and a dienelactone hydrolase (DLH) (Fig. 3b, c), the upstream region was cloned using consecutively PCR approaches with degenerated primer pairs binding to conserved sites within the mentioned putative genes (Fig. 3a) (see Supplementary Material). These attempts yielded three products of expected length: (1) A1ABS_B02 A (~2.1 kb), (2) SDRA1_B03 (~0.9 kb), and (3) PaaXSDR_B04 (~2.9 kb) as indicated in Fig. 3a. The approach aiming at the downstream region did not yield a product. The products obtained were propagated and directly or in form of subclones subjected to DNA sequencing.

For the downstream region, a genome-walking method was applied. For that, the protocol of Tan and coworkers [37] had previously been adapted to GC-rich samples [27]. In a first approach, the genome-walking technique was now used to amplify an already known fragment coding for the C-terminal part of StyA2B as well as for the N-terminal region of the DLH of *R. opacus* 1CP (Fig. 3b). After it had successfully been established, the same

Fig. 3 Genetic organization of regions coding (putative) SMOs. **a** A 359-bp fragment obtained from the *styA2B* screening was used to explore the surrounding genomic region of *R. opacus* MR11. By means of PCRs, three fragments A1ABS_B02 A, SDRA1_B03, and PaaXSDR_B04 and based on genome-walking approaches two further fragments (MR11_01 and _02) were amplified, then cloned and sequenced. Subclones obtained via the indicated restriction sites (DraII, NotI, XhoI) were analyzed later. The genome-walking directions (*black arrows*) are indicated. **b** The fragment of *R. opacus* 1CP, which was also increased in length from 8.965 kb to 10.664 kb by genome-walking approaches, is shown. Comparison of the *styA1* and *styA2B* encoding regions of the rhodococci with **c** chromosomal regions of *R. opacus* PD630, *A. aurescens* TC1, *N. farcinica* IFM10152, *S. platensis* CR50, *V. paradoxus* EPS, *D. acidovorans* SPH-1, *A. calcoaceticus* ADP1, *Rhodococcus* sp. ST-10, and of two metagenomic fragments [1, 7], as well as to **d** the styrene catabolic gene clusters of *Rhodococcus* sp. ST-5 [41] and of *Pseudomonas* sp. strain VLB120 [1, 5] are shown. The corresponding gene identification numbers (GI) or protein accession numbers are given. Identical hatching of ORFs indicates a similar protein function. *SDR* short-chain dehydrogenase, *DLH* dienelactone hydrolase (carboxymethylenebutenolidase), *StySc/StyR* two-component regulatory system, *StyA* styrene epoxidase, *StyB* FAD reductase, *StyC* styrene oxide isomerase, *StyD* phenylacetaldehyde dehydrogenase, *put.* putative. Fragments encoding a monooxygenase for which activity data have been reported are marked with a *cross* (X)

procedure was performed for strain MR11 and indeed a product with the so far missing sequence information encoding the C-terminal part of the one-component monooxygenase and the first amino acids of the N terminus of the dienelactone hydrolase was obtained (Fig. 3a). The same technique with other primers was also successfully applied to regions encoding the PaaX-like regulator of both *R. opacus* strains and in addition to the still incomplete DLH ORF of strain 1CP (Fig. 3a, b). The products derived were ligated into a pJET1.2/blunt and subjected to DNA sequencing.

The achieved results allowed to assemble a 7,012-bp genomic fragment from *R. opacus* MR11 (Fig. 3a). That fragment obtained from strain MR11 showed a high similarity to that of strain 1CP, with five complete ORFs and one incomplete ORF being assigned. Unfortunately, the last ORF which encodes a fragment of a DLH is lacking a major part and no sequence information beyond that was gained in the study. Since the organization including the ORFs and their orientation was found to be highly conserved for strain MR11 (Fig. 3a, b), similar assignments were made for this novel fragment. In congruence to strain 1CP, several ORFs encode hypothetical proteins involved either in the upper (ORF4 and ORF5) or lower styrene (ORF1) degradation pathway [9]. Interestingly, ORFs 3 to 6 (which encode for an SDR, a StyA, a StyA2B or StyB, and a DLH, respectively) were found to be highly conserved among microorganisms comprising styrene monooxygenase genes (Fig. 3).

Additional Sequence Information for Strain 1CP

The original fragment from *R. opacus* 1CP was enlarged from an 8,965-bp to a 10,664-bp one (Fig. 3b), and that revealed some interesting facts. First, ORF1 was completed and the high similarity of its product to regulatory proteins (PaaX) of the phenylacetic acid metabolism was proved [38]. Second, the sequence of ORF9 was completed; its polypeptide enlarged from 26 amino acids to 270 amino acids together with that of the incomplete ORF6 could be assigned to a DLH. That DLH type was found to be frequently associated with a SDR (Fig. 3). Third, a 108-amino-acid fragment of a transposase coded by ORF10 was discovered, which is an interesting finding with respect to the mobility discussed for styrene catabolic genes [1]. The transposase fragment shows considerable similarities to related proteins [accession nos. AAP74048 (97 % identical amino acids) and BAH47175 (94 % identical amino acids)] which are encoded on linear plasmids. This indicates that strain 1CP may carry the *styA1/styA2B*-comprising fragment on a plasmid, perhaps on its megaplasmid [39], or the found fragment may have been integrated from another source into the chromosome. It is remarkable that only in this actinobacterium these putative transposases (encoded by ORF7, ORF8, and ORF10) were found to be close to SMO genes and that only for *Pseudomonas fluorescens* ST mobile elements have also been reported to be located next to styrene catabolic genes [1, 38, 40].

Discussion

Flavin-dependent styrene monooxygenases are versatile biocatalysts [2] for the enantiose-letive epoxidation of styrene and chemically analogous compounds as well as for the sulfoxidation of methylphenylsulfide and related compounds to chiral sulfoxides [5–7, 9, 10, 41]. However, only a few representatives have been studied in respect to their biochemical properties (reviewed by [2, 3]), and of these most originated from *Pseudomonas* species [1]. Furthermore, SMOs are considered to be relatively rare within the microbial flora [3]. At

start of the project, only three representatives of self-sufficient one-component SMOs were known, of which only one was biochemically characterized [9].

Identification of *styA2B* Genes

A new *in silico* database mining still revealed only six representatives of such fused SMOs (StyA2B) and five of these are encoded by Actinobacteria. That demonstrates how rare such fused SMOs are among bacteria. In contrast, several two-component SMOs (StyA plus StyB) have been characterized and described in detail [2], and many more have been found by *in silico* analysis. However, just eight of such conventional two-component SMOs have so far been found to be part of a styrene catabolic operon comprising also *styC* and *styD* (Fig. 3d) [1], whereas the majority is either not part of such a defined larger operon or associated somehow with ORFs encoding an AraC-like regulator, a SDR, and a DLH (Fig. 3).

Besides database mining, activity-based screening may provide access to *styA2B* genes. However, the low activity of the first characterized fused SMO (StyA2B from *R. opacus* 1CP; [9]) suggested that an activity-based screening as reported for conventional two-component SMOs by van Hellemond and co-workers [7] is not promising.

An alternative approach is a sequence-based screening protocol [12, 15, 16], which is described here for the first time for SMOs. For that approach, a specific touch-down PCR (Table 1) comprising a suitable primer pair (fw-1 and rev-2) was successfully established and allowed the identification of *styA2B* homologs from various sources. Even various GC contents of strains tested (*A. aurescens* TC1=63 %; *R. opacus* 1CP=66 %; *N. farcinica* IFM10152=72 %) did not interfere with the protocol. This suggests that the method may be applicable to a rather broad range of (actino-)bacteria, thereby increasing chances to detect further one-component SMO genes.

The established protocol was applied to pure strains and of about 30 *Rhodococcus* strains investigated only a single strain, namely *R. opacus* MR11, was found to harbor such *styA2B*-homologous gene. The results obtained from the screening approach performed here as well as from the *in silico* database mining confirmed a low abundance of self-sufficient one-component monooxygenases (StyA2B) among bacteria. Nevertheless, the approach developed in this study using a specific sequence-based screening method for one-component SMOs seems promising to identify further representatives and thus to provide access to more such enzymes especially since the chosen primer binding regions were found to be conserved in all StyA2B proteins so far (Fig. 1).

The Genomic Region Surrounding *styA2B* Fragments

The identified 359-bp fragment of *R. opacus* MR11 obtained in the screening (Fig. 2) was the initiation point to characterize the adjacent genomic context. ORFs coding for StyA1 and StyA2B as well as the adjacent ORFs were conserved to those of other bacteria and overall a high similarity to the corresponding genomic fragment of *R. opacus* 1CP [9] was detected (Fig. 3b).

Interestingly, ORFs encoding a short-chain dehydrogenase (SDR), a single epoxidase (StyA or StyA1), a single or fused flavin reductase (StyB or StyA2B), and a dienelactone hydrolase (DLH; carboxymethylenebutenolidase) were often found to be clustered (Fig. 3a–c). This indicates a functional coherence of the respective proteins, possibly within the same metabolic pathway. The genes may even form an operon, possibly regulated by AraC-like regulators with a high sequence similarity among each other for which genes were detected in

direct neighborhood to the presumed operons (Fig. 3). The genes of StyA1 and StyA2B were so far never found in a conventional styrene catabolic operon comprising also *styC* and *styD* [1, 2, 38, 41]. In a few cases, such as for *Acinetobacter calcoaceticus* ADP1, that gene cluster was found to be located next to genes coding for proteins involved in anthranilate degradation (Fig. 3c) [42]. Also, other metabolic routes for heteroaromatic compounds as for indolepyruvate (*N. farcinica* IFM 10152; *Ralstonia solanacearum* GMI1000) or carbazole (*Pseudomonas resinovorans* CA10) were detected to be encoded in the neighborhood of proteins of the presumed *styA1/sty(A2)B* comprising operon (not shown). Thus, the gene cluster investigated here of strain MR11 might have a role in the degradation of heteroaromatic compounds.

## Convergent Evolution of StyA1 and StyA2B

Among the (putative) SMO genes available from databases, those which encode one-component-type SMOs (*styA2B*) are rare and always a single-like epoxidase gene (*styA1*) is located next to it. The corresponding proteins, StyA1 and StyA2B, share about 50 % similarities on amino acid level to each other (48 % for strain MR11 and 51 % for strain 1CP) (Fig. 4). Interestingly, the SMO components of strain MR11 share about 94 % identical positions with their counterparts of strain 1CP, while the other proteins encoded adjacent to *styA1/styA2B* share more than 97 % to the corresponding proteins in the other strain. Thus, the two monooxygenase components StyA1 and StyA2B of *R. opacus* MR11 among the six proteins encoded on both genomic fragments of the two rhodococci show the highest sequence divergence on amino acid level (Fig. 3a, b). Possibly, the obviously increased mutation rate of the SMO genes points to a functional change or to selective pressures resulting from the interaction of the subunits.

Representatives of both proteins from Actinobacteria, StyA1 and StyA2B, each respectively form a clade in a dendrogram presented in Fig. 4, and the two types can clearly be differentiated. Furthermore, it becomes obvious that StyA-like monooxygenases, originating from Proteobacteria such as *Delftia*, *Variovorax*, or *Ralstonia*, and encoded in a similar operon as StyA1 and StyA2B (Fig. 3) also form a distinct branch. Experimentally verified epoxidases (StyA) from styrene catabolic operons of *Pseudomonas* species [2, 3], in contrast, and also those of *Rhodococcus* spp. ST-5 and ST-10 [41] also form their own branches and are more distinct from the others mentioned above. The latter observation was also made for the corresponding flavin reductases (StyB and StyA2B) (Fig. 4). This clustering is in context with the genomic context presented in Fig. 3.

It can be assumed that the type of SMOs encoded by *styA1* and *styA2B* has evolved from conventional two-component SMOs (*styA* and *styB*) by some sort of recombination [1, 9]. This scenario is plausible, first since evolution usually moves from simple to more complex systems [13], and second since the two-component SMO genes are far more distinct to each other and occur in more different types of gene cluster than the fused SMOs.

Recently, the genome of *V. paradoxus* EPS has been announced (unpublished) and, as mentioned above, the strain encodes the novel StyA1/StyA2B-type monooxygenase (Fig. 3c). Thus, the known representatives of this type of monooxygenases are not anymore restricted to Actinobacteria. In spite of the distinct origin, the monooxygenase components of strain EPS show a considerable similarity to those of Actinobacteria (Figs. 1 and 4). They share between 51 % and 61 % similarity on amino acid level with their counterparts, but StyA1 and StyA2B of *V. paradoxus* EPS together form a branch in the dendrogram distinct from that of the actinobacterial enzymes. The distance tree for oxygenase components shown in Fig. 4 (algorithm: Fitch–Margoliash) and thus the unique position of StyA1 and StyA2B from *V. paradoxus* EPS was confirmed by other algorithms (maximum parsimony,
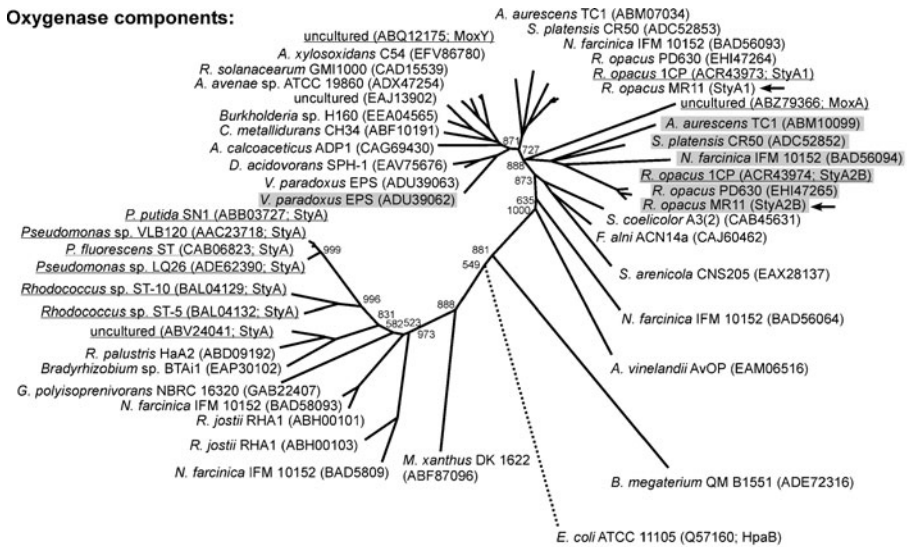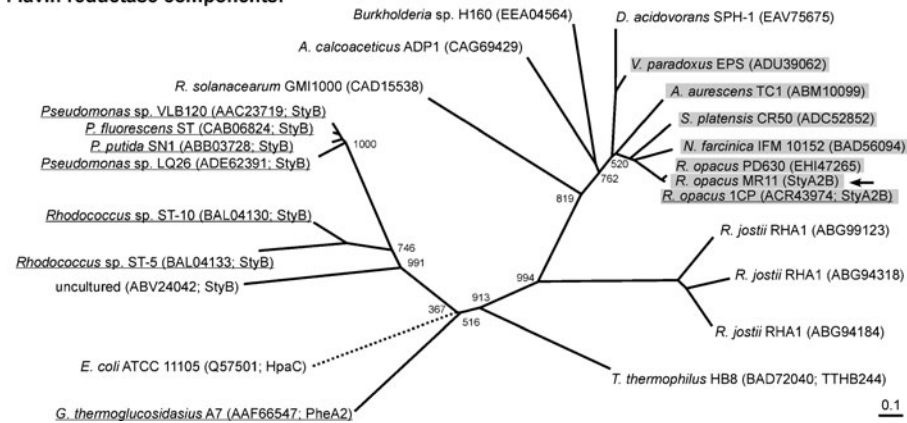
Fig. 4 Dendrograms showing the relatedness of (putative) one- and two-component (styrene) monooxygenase subunits. The distance trees were created according to the Fitch–Margoliash algorithm by means of the software programs ClustalX (version 1.8), GeneDoc (version 2.6.003), and the PHYLIP 3.66 package (PROTDIST and FITCH), and bootstraps of 1,000 replicates were calculated from the corresponding alignment by means of the PHYLIP 3.66 package (SEQBOOT, PROTDIST, FITCH, and CONSENSE). Corresponding values are shown at nodes of the major branches. The proteins StyA1 and StyA2B of *R. opacus* MR11 are highlighted by *black arrows*, those with reported activity are *underlined*, and one-component SMOs are marked with *gray* background. Protein identification numbers are given in *parentheses* and the outgroups are presented by a *dotted line* (one third of calculated length)

maximum likelihood, neighbor joining; not shown). Interestingly, the components StyA1 and StyA2B also share among each other about 74 % similarity, which is quite unique among StyA1/StyA2B subunits from same organism. This finding and the distinct position of the StyA2B homolog in the distance tree lead to the suggestion that the complex monooxygenase system of StyA1/StyA2B type may have evolved at least twice independently from conventional two-component SMOs by functionally convergent evolution. A similar case of "sequence divergence and functional convergence" between Actinobacteria

and Proteobacteria has previously been reported for chlorocatechol catabolic enzymes [43, 44]. With respect to the presented alignment (Fig. 1), it is remarkable that in all cases the fused region of StyA2B proteins between the epoxidase domain (A2) and the reductase domain (B) seems to be highly conserved. Thus, there probably are steric or electronic restrictions for the functionality of the fusion protein. It will be interesting to obtain deeper insights into such restrictions and into the function of the StyA1/StyA2B type.

References

1. Tischler, D., & Kaschabek, S. R. (2012). In S. N. Singh (Ed.), *Microbial degradation of xenobiotics. Environmental Science and Engineering* (pp. 67–99). Heidelberg: Springer.
2. Montersino, S., Tischler, D., Gassner, G. T., & van Berkel, W. J. H. (2011). *Advanced Synthesis and Catalysis, 353*, 2301–2319.
3. van Berkel, W. J. H., Kamerbeek, N. M., & Fraaije, M. W. (2006). *Journal of Biotechnology, 124*, 670–689.
4. Kantz, A., Chin, F., Nallamothu, N., Nguyen, T., & Gassner, G. T. (2005). *Archives of Biochemistry and Biophysics, 442*, 102–116.
5. Otto, K., Hofstetter, K., Roethlisberger, M., Witholt, B., & Schmid, A. (2004). *Journal of Bacteriology, 186*, 5292–5302.
6. Hollmann, F., Lin, P.-C., Witholt, B., & Schmid, A. (2003). *Journal of the American Chemical Society, 125*, 8209–8217.
7. van Hellemond, E. W., Janssen, D. B., & Fraaije, M. W. (2007). *Applied and Environmental Microbiology, 73*, 5832–5839.
8. Nikodinovic-Runic, J., Flanagan, M., Hume, A. R., Cagney, G., & O'Connor, K. E. (2009). *Microbiology, 155*, 3348–3361.
9. Tischler, D., Eulberg, D., Lakner, S., Kaschabek, S. R., van Berkel, W. J. H., & Schlömann, M. (2009). *Journal of Bacteriology, 191*, 4996–5009.
10. Tischler, D., Kermer, R., Gröning, J. A. D., Kaschabek, S. R., van Berkel, W. J. H., & Schlömann, M. (2010). *Journal of Bacteriology, 192*, 5220–5227.
11. De Mot, R., & Parret, A. H. A. (2002). *Trends in Microbiology, 10*, 502–508.
12. Liu, L., Schmid, R. D., & Urlacher, V. B. (2006). *Applied Microbiology and Biotechnology, 72*, 876–882.
13. Munro, A. W., Girvan, H. M., & McLean, K. J. (2007). *Biochimica et Biophysica Acta, 1770*, 345–359.
14. Torres Pazmiño, D. E., Winkler, M., Glieder, A., & Fraaije, M. W. (2010). *Journal of Biotechnology, 146*, 9–24.
15. Kim, B. S., Kim, S. Y., Park, J., Park, W., Hwang, K. Y., Yoon, Y. J., et al. (2007). *Journal of Applied Microbiology, 102*, 1392–1400.
16. Roberts, G. A., Grogan, G., Greter, A., Flitsch, S. L., & Turner, N. J. (2002). *Journal of Bacteriology, 184*, 3898–3908.
17. Dorn, E., Hellwig, M., Reineke, W., & Knackmuss, H.-J. (1974). *Archives of Microbiology, 99*, 61–70.
18. Mitani, Y., Meng, X. Y., Kamagata, Y., & Tamura, T. (2005). *Journal of Bacteriology, 187*, 2582–2591.
19. Quan, S., & Dabbs, E. R. (1993). *Plasmid, 29*, 74–79.
20. Lenke, H., Pieper, D. H., Bruhn, C., & Knackmuss, H.-J. (1992). *Applied and Environmental Microbiology, 58*, 2928–2932.
21. Gorlatov, S. N., Maltseva, O. V., Shevchenko, V. I., & Golovleva, L. A. (1989). *Microbiology (Mikrobiologiya), 58*, 647–651.
22. Dabrock, B., Keßler, M., Averhoff, B., & Gottschalk, G. (1994). *Applied and Environmental Microbiology, 60*, 853–860.
23. Sensfuss, C., Reh, M., & Schlegel, H. G. (1986). *Journal of General Microbiology, 132*, 997–1007.
24. Seto, M., Kimbara, K., Shimura, M., Hatta, T., Fukuda, M., & Yano, K. (1995). *Applied and Environmental Microbiology, 61*, 3353–3358.

25. Mongodin, E. F., Shapir, N., Daugherty, S. C., DeBoy, R. T., Emerson, J. B., Shvartzbeyn, A., et al. (2006). *PLoS Genetics, 2,* e214.
26. Sambrook, J., Fritsch, E., & Maniatis, T. (2001). *Molecular cloning: a laboratory manual* (3rd ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
27. Gröning, J. A. D., Tischler, D., Kaschabek, S. R., & Schlömann, M. (2010). *Journal of Basic Microbiology, 50,* 499–502.
28. Wei, M., Deng, J., Feng, K., Yu, B., & Chen, Y. (2010). *Analytical Chemistry, 82,* 6303–6307.
29. Moiseeva, O. V., Solyanikova, I. P., Kaschabek, S. R., Gröning, J., Thiel, M., Golovleva, L. A., et al. (2002). *Journal of Bacteriology, 184,* 5282–5292.
30. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). *Nucleic Acids Research, 25,* 3389–3402.
31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). *Journal of Molecular Biology, 215,* 403–410.
32. Staden, R. (1996). *Molecular Biotechnology, 5,* 233–241.
33. Higgins, D. G., & Sharp, P. M. (1988). *Gene, 73,* 237–244.
34. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). *Nucleic Acids Research, 25,* 4876–4882.
35. Felsenstein, J. (2005). *PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author.* Seattle: Department of Genome Sciences, University of Washington.
36. Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., et al. (2004). *Proceedings of the National Academy of Sciences of the United States of America, 101,* 14925–14930.
37. Tan, G., Gao, Y., Shi, M., Zhang, X., He, S., Chen, Z., et al. (2005). *Nucleic Acids Research, 33,* e122.
38. Velasco, A., Alonso, S., Garcia, J. L., Perera, J., & Diaz, E. (1998). *Journal of Bacteriology, 180,* 1063–1071.
39. König, C., Eulberg, D., Gröning, J., Lakner, S., Seibert, V., Kaschabek, S. R., et al. (2004). *Microbiology, 150,* 3075–3087.
40. Bestetti, G., Galli, E., Ruzzi, M., Baldacci, G., Zennaro, E., & Frontali, L. (1984). *Plasmid, 12,* 181–188.
41. Toda, H., & Itoh, N. (2012). *Journal of Bioscience and Bioengineering, 113,* 12–19.
42. Bundy, B. M., Campbell, A. L., & Neidle, E. L. (1998). *Journal of Bacteriology, 180,* 4466–4474.
43. Eulberg, D., Kourbatova, E. M., Golovleva, L. A., & Schlömann, M. (1998). *Journal of Bacteriology, 180,* 1082–1094.
44. Solyanikova, I. P., Maltseva, O. V., Volmer, M. D., Golovleva, L. A., & Schlömann, M. (1995). *Journal of Bacteriology, 177,* 2821–2826.