Identification of a complete dibenzothiophene biodesulfurization operon and its regulator by functional metagenomics.

Guadalupe Martín-Cabello^{1§}, Laura Terrón-González^{1§}, Manuel Ferrer² and Eduardo Santero^{1*}

¹Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide/Consejo Superior de Investigaciones Científicas/Junta de Andalucía, and Departamento de Biología Molecular e Ingeniería Bioquímica, Universidad Pablo de Olavide^a, Spain;

²Instituto de Catálisis y Petroleoquímica, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain^b.

§ These authors contributed equally to this work and should both be considered first authors.

* Address correspondence to Eduardo Santero, esansan@upo.es.

Running Head: **DBT biodesulfurization operon from a metagenome**

Originality/Significance Statement

Dibenzothiophene biodesulfurization is an important process for removing sulfur from fossil fuels. Functional metagenomics identified a complete operon for biodesulfurization (*dszEABC*) with a different gene arrangement and containing a gene coding for the FMN-dependent reductase that is missing in all other *dsz* operons.

Regulation of the operon is fully characterized *in vivo* and *in vitro* and shown to be σ_N -dependent, unlike other operons of cultured bacteria. The σ_N -dependent promoter, the activator gene and the reductase-coding gene have been apparently recruited from a different pathway.

An insulation phenomenon that could be relevant in nature to prevent spurious expression of “mobile” operons when laterally transferred to different replicons is also reported here and its mechanism elucidated: an insulator sequence in the promoter region binds to the ribosome binding site of the first gene *dszE*, thus preventing its translation, combined with translational coupling of the downstream *dszABC* genes.

Abstract

Functional screening for aromatic ring oxygenases of an oil contaminated soil metagenome identified 25 different clones bearing monooxygenases coding genes. One fosmid bore an operon containing 4 tightly linked genes coding for a complete dibenzothiophene biodesulfurization pathway, which included the predicted monooxygenases DszC and DszA, the desulfinase DszB, and an FMN-oxidoreductase designated DszE. The *dszEABC* operon provided *E. coli* with the ability to use dibenzothiophene as the only sulfur source. Transcription of the operon is driven from a σ_N -dependent promoter and regulated by an activator that was designated *dszR*. DszR has been purified and characterized *in vitro*, and shown to be a constitutively active σ_N -dependent activator of the group IV, which binds to two contiguous sequences located upstream of the promoter. The *dsz* promoter, and *dszE* and *dszR* genes have apparently been recruited from an aliphatic sulfonate biodegradation pathway. If transcribed from a heterologous upstream promoter, the σ_N -dependent promoter region functions as an “insulator” that prevents translation of *dszE*, by binding with its ribosome binding site. Translational coupling, in turn, prevents translation of the downstream *dszABC* genes. The silencer combined with translational coupling thus represent an effective way of preventing expression of operons when spuriously transcribed from upstream promoters.

Introduction

Combustion of fossil fuel containing sulfur inevitably releases sulfur dioxide (SO_2) to the atmosphere, thus contributing to environmental pollution and acid rain. Because of that, environmental regulations have set strict decreasingly limits to the sulfur content of the fuel (Swaty, 2005) while there is an increasing social pressure to transition toward ultra-low sulfur fuel. This, together with the need of access to lower quality oils, in order to satisfy the increasing transportation fuel demands, makes it essential to develop ever improved desulfurization processes (Kilbane, 2006; Kilbane and Stark, 2016).

Hydrodesulfurization is the dominant desulfurization process used by the petroleum industry (Wang et al., 2017). However, the low availability of hydrogen and the high costs associated to the process, mainly due to the requirements of high temperature and pressure, makes it imperative the discovery of alternative ways of desulfurization (Kilbane II, 2016). In addition, some organic sulfur forms such as benzothiophenes, dibenzothiophene (DBT) and, above all, their alkylated derivatives, are refractory to chemical desulfurization (Wang et al., 2017), which will require the development of alternative or complementary desulfurization processes for sulfur rich oils.

Biodesulfurization, the sulfur removal by enzymatic cleavage of carbon–sulfur bonds from organic compounds, is an attractive complement to hydrodesulfurization and has received much attention in the last two decades, above all biodesulfurization of DBT, which has been used as the model compound for biodesulfurization studies. A variety of microorganisms including Gram-negative bacteria, Gram-positive bacteria and yeasts have been shown to be capable of cleaving carbon–sulfur bonds to remove sulfur from crude oil and petroleum compounds, in order to use the resulting product as the sulfur source (Kilbane II, 2016).

All microorganisms share a single biodesulfurization pathway, also known as the 4S pathway, which involves three sequential mono-oxygenation reactions to progressively transform DBT to DBT sulfoxide, to DBT sulfone and to 2-hydroxybiphenyl-2-sulfinate, from which bisulfite (HSO_3^-) is finally released, thus yielding 2-hydroxybiphenyl (Piddington et al.,

1995). The two first reactions are catalyzed by the same sulfide-sulfoxide mono-oxygenase DszC, while the third mono-oxygenation is catalyzed by the dibenzothiophene-5,5-dioxide mono-oxygenase DszA. Bisulfite ion is finally released by hydrolysis in a reaction catalyzed by the hydroxybiphenyl-2-sulfinate desulfinate DszB.

The genes encoding for the biodesulfurization enzymes were first discovered in *Rhodococcus erythropolis* strain IGTS8 (Denome et al., 1994) and found to be clustered in an operon finally called *dszABC* (Piddington et al., 1995). Homologous genes have been subsequently found in plasmids of different *Rhodococcus* strains (Denis-Larose et al., 1997), in the genomes of different Gram-positive bacterial genera such as *Gordonia* (Park et al., 2003), *Mycobacterium* (denoted *mds* operon) (Nomura et al., 2005), thermophilic strains of *Bacillus* (*bdzABC* operon) (Kirimura et al., 2004), and *Paenibacillus* (*tdsABC* operon) (Ishii et al., 2000b), as well as in metagenomic sequences from uncultured bacteria (Abbasian et al., 2016). In all cases, the three genes are highly conserved, with proteins identities usually higher than 80%, and arranged in the same order in very compact operons almost lacking intergenic regions, and in many instances with overlapped coding regions. It is also usual to find insertion elements or sequences coding for transposases next to the biodesulfurization operon (Denis-Larose et al., 1997; Ishii et al., 2000b), which suggest that these genes may be horizontally transferred to different replicons by these transposable elements (Chalmers and Blot, 1999).

The monooxygenation reactions catalyzed by DszC and DszA require electrons, which are provided by an NAD(P)H-FMN oxidoreductase DszD (Gray et al., 1996). In all described bacteria the genes encoding for the oxidoreductase are not linked to the biodesulfurization operons and in *R. erythropolis* strain XP, the oxidoreductase coding gene and the *dszABC* operon are located even in different replicons, being the *dszD* located in the chromosome, whilst the *dszABC* operon on a plasmid (Ishii et al., 2000a; Furuya et al., 2004; Furuya et al., 2005).. Flavin reductases involved in DBT biodesulfurization are less conserved and do not seem to be very specific since they provide electrons to both DszC and DszA, while their

function can be complemented by other reductases from different bacteria and biodesulfurization systems (Galan et al., 2000; Reichmuth et al., 2000).

Very little is known about the transcriptional regulation of DBT desulfurization genes. By assaying expression of *dsz::lacZ* gene fusions and probing *dsz* mRNA it was shown that *dsz* genes of *Rhodococci* are strongly repressed by sulfate, cysteine, methionine and casaminoacids (Li et al., 1996; Denis-Larose et al., 1997). Similar sulfur repression has also been reported for *Gordonia* (Rhee et al., 1998; Alves et al., 2007; [Feng, 2016](#)) and *Mycobacterium* (Takada et al., 2005) *dsz* genes, which has prompted the construction of engineered strains expressing *dsz* genes from heterologous promoters (Gallardo et al., 1997; Takada et al., 2005; Alves et al., 2007). Partially constitutive expression of *dsz* genes in two *R. erythropolis* KA2-5-1 mutants, bearing the insertions in a gene coding for cystathionine β -synthase, indicate that sulfur repression requires sulfur assimilation (Tanaka et al., 2002). Mutational analyses of the IGTS8 *dsz* promoter region revealed mutations upstream of the promoter that increased expression and partially released this repression, thus suggesting a negative control of *dsz* gene expression (Li et al., 1996). Very recently, a regulatory protein of the TetR family from *Gordonia* sp. IITR100 has been found to bind to upstream sequences of the *dsz* promoter region and to activate expression of the *dszABC* operon in different genomic backgrounds (Murarka et al., 2019).

Most of the analyses on DBT desulfurization have been performed with cultivated bacteria and only one publication has reported on identification of DBT biodesulfurization genes by sequence analysis of a metagenome (Abbasian et al., 2016). However, cultivated bacteria represent a very small fraction of the actual bacterial diversity, thus providing an extremely limited picture of the biodesulfurization potential of bacterial populations (Vilchez-Vargas et al., 2010). Metagenomic functional analyses of contaminated sites screening for biodegradation activities such as extradiol dioxygenases has substantially increased the known diversity of these enzymes (Brennerova et al., 2009; Terron-Gonzalez et al., 2016). In this report we describe a metagenomic functional screening for oxygenase activities that has

led to the identification of a fosmid clone containing a new, more complete, *dsz* operon, which also comprises a gene coding for an FMN oxidoreductase. Expression analysis indicates that transcription of the new operon is dependent on the alternative σ_N factor, and regulated by a DszR activator also present in the fosmid.

Results

Identification of DBT biodesulfurization genes by metagenomic functional analysis screening for indigo production.

A metagenomic library comprising 185,000 independent clones constructed from a soil contaminated with crude oil (Terron-Gonzalez et al., 2016) was screened for indigo production by the specialized host strain *E. coli* MPO554 grown on rich medium. The medium also contained arabinose to induce the copy number of the fosmid vector and salicylate to promote transcription of the metagenomic DNA from a heterologous *P_{sal}* promoter, whose transcription is subjected to antitermination by the N protein of the lambda phage (Terron-Gonzalez et al., 2013). Indigo production from indole that comes from tryptophan catabolism by monooxygenases and dioxygenases of aromatic compounds has been well documented in a number of publications referenced in (van Hellemond et al., 2007). Among the identified blue colonies a total of 25 different clones were selected and sequenced, and revealed the presence of genes potentially coding for monooxygenases of different compounds. One of the clones, UPO21, which showed the blue colony color only in the presence of arabinose plus salicylate, harboured 34,226 bp of environmental DNA encompassing 31 complete *orfs*. Functional annotation of each *orf* together with the best hits, E-values and possible origin of the DNA sequences are shown in Supplementary Table S1. The predicted products of three overlapped genes showed high similarity to those of known *dszABC* genes from different bacteria, reported to be involved in dibenzothiophene biodesulfurization *via* the 4S desulfurization pathway (Figure 1). Similarity of these gene products was very high, showing 82%, 76% and 86% identity, respectively, to the closest characterized sequences in the non-redundant databases (Fig. 1B). Upstream of these

genes, and separated 4 bp from *dszA*, an additional *orf* potentially coding for an FMN oxidoreductase was found. In this case, similarity to the best matches in the databases was lower (52% identity with the best match, E-value 2e-48, accession number WP_039549499.1). A phylogenetic analysis of this *orf* revealed it to be more related to FMN reductases that are partners of aliphatic sulfonates monooxygenases, particularly to SsuE reductases, than to the known FMN reductases involved in DBT biodesulfurization (Fig.1C). The four genes are assembled in a very compact way. Actually, all of them overlap or are separated by just 4 nucleotides, which suggest that translation of all four genes is coupled. Because of its tight linkage to the *dsz* genes we considered this *orf* a *dsz* gene but due to its lower similarity to the reductases involved in DBT biodesulfurization we designated this *orf* *dszE* instead of *dszD*.

Phylogenetic analyses of each *orf* (Supplementary Table S1) suggested two possible origins for the sequences present in UPO21, being the *dszABC* genes more related to *Actinobacteria*, whilst the remaining sequence, including *dszE*, having a probable proteobacterial origin. However, no sequences related to mobile DNA, which would suggest possible horizontal transfer, were found within the DNA insert in UPO21.

The UPO21 clone provided *E. coli* with the ability to grow on DBT as the only sulfur source. However, this ability was dependent on the addition to the culture medium of either arabinose to increase fosmid copy-number or salicylate to promote transcription from the external *P_{sal}* promoter, obtaining the best growth on DBT when both were added (Fig. 2A). Growth on DBT was concomitant to 2-hydroxybiphenyl accumulation in culture, which was detected from day 5, up to 0.15 mM at day 7 (not shown), thus confirming that DBT was metabolized via the 4S pathway. This, together with the fact that the colonies in rich LB medium were blue only if arabinose and salicylate were added, indicated that expression of the *dsz* operon in *E. coli* was very low but the clone could be detected in the functional screening because of heterologous transcription of the metagenomic DNA from the vector

P_{sal} promoter. However, intriguingly, the *dsz* genes of the operon are oriented opposite to the *P_{sal}* promoter (Fig. 1A).

Subcloning of the dsz operon and identification of the dszR activator gene.

The whole *dsz* operon comprising the 4 genes was subcloned into the fosmid vector used for the library construction in such a way that they were located 1250 bp downstream of the *P_{sal}* promoter in the appropriate orientation. Surprisingly, this subclone, pMPO587, could not provide the host strain with the ability to grow on DBT as the sulfur source (Fig. 2B) or to produce blue colonies in rich medium (not shown), independently from the addition of arabinose and/or salicylate. Considering that the *dsz* operon expression in the original UPO21 was dependent on salicylate although the *dsz* genes were in the opposite orientation, we considered that heterologous transcription was needed to express a regulatory gene, whose product, in turn, would activate the *dsz* operon. A search for the open reading frames of UPO21 in the orientation of the *P_{sal}* promoter revealed one gene potentially coding for a σ_N -dependent activator and located at approximately 10.3 kb downstream of *P_{sal}*. This new gene was subcloned under control of the *lacI/P_{tac}* regulatory couple in the pIZ1016 expression vector to produce pMPO588, which was introduced into the MPO554 strain bearing pMPO587. The *E. coli* strain harboring both subclones was then able to grow on DBT as the only source of sulfur when transcription of the regulatory gene was induced by addition of IPTG (Fig. 2C), and grew even faster than when bearing the original fosmid,. In this case, maximal growth rate did not require increasing the *dsz* operon copy-number with arabinose, while addition of salicylate was actually deleterious for using DBT as the sulfur source. These data together indicated that growth of the host strain on DBT as the only sulfur source requires expression of a regulatory gene also located in the fosmid UPO21, which was designated *dszR*.

Expression of the dsz operon from its own promoter and from heterologous promoters upstream.

However, one intriguing question remained: Why the Dsz gene products could not be produced when the operon was transcribed from the heterologous *P_{sal}* promoter even when this transcription is subjected to antitermination by the N protein? There should be an unidentified element in the fragment cloned in pMPO587 that “insulated” *dszEABC* operon expression, or particular genes within it, by preventing either transcription or translation. In order to answer this question, and to know whether all the genes were coordinately controlled, transcriptional and translational *lacZ* gene fusions to *dszE*, *dszA* and *dszC* were constructed (Figure 3A) and their expression tested when transcription proceeded from their own promoter (activated by DszR) or from the *P_{sal}* promoter. For this, the EPI300-T1_R strain, which increases fosmid vector copy-number in response to arabinose but cannot provide heterologous transcription from *P_{sal}* since it lacks the NahR regulator, and its specialized derivative strain MPO554, which can both increase the copy number of the vector and drive transcription from *P_{sal}*, were transformed with the fosmid vectors bearing each of gene fusions and with pMPO588, which produces DszR, or its parental vector pIZ1016 lacking *dszR*. The two strains with different plasmid combinations were grown in minimal M9 medium in the presence of arabinose, salicylate, IPTG and DBT, and *dsz* expression levels estimated after 6 hours of growth.

Levels of β -galactosidase activity showed slightly variation depending on the *dsz* gene where the fusion was constructed, but the fusions in all 3 genes responded similarly. In general, β -galactosidase activity was much higher in the case of transcriptional fusions than in the translational ones, thus indicating that *dsz* genes are translated less efficiently than *lacZ* in the transcriptional fusions. As shown in Figure 3B and 3C, both transcriptional and translational fusions to *lacZ* in the EPI300-T1_R strain could only be expressed (from their own promoter) if DszR was produced. On the other hand, all 3 transcriptional fusions could be expressed in MPO554, either in the presence of DszR (from its own promoter or from *P_{sal}*) or in its absence (from *P_{sal}*) (Fig. 3D), thus indicating that the *dsz* operon could be actively transcribed from the *P_{sal}* promoter. However, translational fusions could only be expressed in

MPO554 if they were transcribed from its own promoter (i.e., in the presence of DszR). The fact that the transcriptional fusions could be expressed without DszR in this specialized strain but translational fusions could not, suggests that transcription of *dszEABC* from the heterologous upstream P_{sal} promoter is not productive since the transcribed genes could not be translated. In fact, expression levels of the translational fusions in MPO554 (Fig. 3E) were even lower than those in EPI300-T1_R (Fig. 3C), which suggests that in MPO554 there is an interference of the non-productive transcription from the upstream P_{sal} promoter with the productive transcription activated by DszR that initiates at the *dsz* promoter.

dsz genes can be translated in the absence of DszR if transcription of the operon starts at its native site.

Two important features are different between the translatable mRNA produced from its own promoter and the non-translatable mRNA produced from the heterologous P_{sal} promoter: (i) transcription starts at different sites depending on the promoter, therefore, the 5' leader sequence is different, and (ii) transcription is activated by DszR in the case of its own promoter. In order to determine which of these two factors was behind the observed differences in translatability of the genes, we first defined the transcription start site of the *dszEABC* operon and then generated a deletion of the sequence upstream of the +1 native site, removing the *dsz* promoter region and upstream sequences in such a way that transcription from P_{sal} had to start at the position of native transcription.

Sequence analyses upstream of the *dszE* coding region revealed a putative σ_N -dependent promoter that, if functional, should give rise to an mRNA with a 34-36 nucleotides long untranslated leader. The primer PEX-DBT1 was designed for 5' mapping of the transcripts of the *dszEA::lacZ* transcriptional and translational fusions when transcription was activated by DszR. As shown in Supplementary Figure S1, the transcription start site was the same for both gene fusions, and located at a G, 13 nucleotides downstream of the putative GG-N10-GC σ_N -dependent promoter, thus confirming that this sequence was the actual *dsz* promoter.

A deletion that substituted the σ_N -dependent promoter with P_{sal} to produce the same transcript but in a DszR independent manner was constructed in the plasmids bearing the *dszE* or *dszEA* translational gene fusions (Figure 4A) and their expression tested in the presence or absence of DszR in EPI300-T1_R or in MPO554 (Figure 4B and 4C). Expression of both genes was now strictly dependent of transcription from P_{sal} (compare expression in EPI300-T1_R to that in MPO554) since it was the only promoter present on the new constructions. Translation of the transcripts originated from P_{sal} but at the native start site was now possible, and the presence of DszR did not affect their expression levels.

Two mRNA structural analyses were performed at the RNAfold Web Server (Lorenz et al., 2011): (i) the leader and the first 30 codons of *dszE* and (ii) the same sequence with the addition of a 35 nt upstream sequence that includes the sigma-N dependent promoter. As shown in Figure 5, the *dszE* initiation codon in the native mRNA sequence is essentially single stranded and its putative Shine-Dalgarno sequence only partially occluded in a weak secondary structure. However, if the mRNA were transcribed from an upstream promoter, both the *dszE* initiation codon and its SD sequence would be occluded within strong secondary structures that would prevent its translation. Consistently with this, the estimated translation rate of *dszE* from its initiation codon at the RBS calculator Web Server (Salis et al., 2009) dropped 47-fold (from 1446 au to 31 au) when the mRNA sequence contained the additional 35 nt upstream. Further structural analyses going up to 500 nt upstream from the native +1 showed that in all cases the relevant features were conserved: (i) the *dszE* SD sequence was partially occluded in the secondary structure in which the CTTTATG sequence at the -12 to -6 promoter region was involved (See Figure 5B) and (ii) the ATG codon was also fully occluded within the same 6-bp stem shown in Fig. 5B.

DszR is constitutively active.

BLAST protein sequence comparison of the 375 residue DszR protein to Uniprot/Swissprot database unambiguously showed that DszR is a transcriptional activator of σ_N -dependent promoters, showing highest similarity to SfnR from *Pseudomonas fluorescens* (Q3K999; 58%

identity, E-value $2e-136$). As SfnR, DszR is shorter than most activators of this kind and consists of the conserved central AAA⁺ activation domain and the C-terminal helix–turn–helix DNA-binding domain, thus lacking the N-terminal regulatory domain that controls its function. DszR (and SfnR) would then be members of the group IV of σ_N -dependent activators (Bush and Dixon, 2012), which should be constitutively active. A phylogenetic tree of group IV activators lacking the N-terminal domain that shows clustering of DszR with the members of the SfnR family is presented in Supplementary Figure S2.

To check whether DszR produced in *E. coli* was constitutively active, expression of the *dszEA* transcriptional fusion was tested in EPI300-T1_R carrying the plasmid pMPO588 (DszR) or the empty vector pIZ1016 in the presence or absence of DBT using high (+ IPTG) or low (no IPTG) levels of DszR. Results shown in Figure 6 indicated that expression of both fusions was strictly dependent on DszR, which could activate transcription even in the absence of IPTG, although maximal levels of expression were obtained when DszR production was induced by IPTG. In the presence of IPTG, a slight difference of up to 1.5-fold was observed depending on the conditions, expression levels being lower in the absence of sulfate, when cells have grown at the expenses of their own sulfur. Thus, this result indicates that once produced in *E. coli*, DszR is constitutively active.

In vitro characterization of DszR activity.

The *dszR* gene was cloned in the expression vector pTYB21 to yield plasmid pMPO1554, which overproduce an intein-tagged DszR protein. The plasmid was transformed into the specialized overproducing strain and, after induction with IPTG, DszR in its native form was purified (see Methods). A 98% pure protein preparation at a concentration of 5.1 μ M (dimer) was obtained.

Binding of DszR to the *dszE* promoter regulatory region was assessed by DNase I footprinting assays using a DNA fragment comprising the promoter and sequences upstream from -150 to +7. As shown in Supplementary Figure S3, two DNA regions located between -130 to -112 and -98 to -81 from the transcription start site were protected by DszR against

DNase I digestion, which indicate direct interaction of the activator DszR with the regulatory region *in vitro*. Sequence comparison of the two regions revealed the presence of a conserved palindromic TGC-N₁₀-GCA sequence in each region, which presumably represents the DszR binding site. The putative binding sequences are separated by 31 nucleotides; therefore, they are located in the same face of the helix, which would facilitate binding of DszR (Supplementary Fig. S3B).

To test if DszR can activate transcription from the *dszE* promoter *in vitro* in an isolated environment using purified components of the σ_N -RNA polymerase holoenzyme, we selected a DNA fragment from coordinates -160 to +40 with respect to the transcriptional start site, which covers the whole *dszE* promoter regulatory region. This sequence was cloned into the vector pTE103 for *in vitro* transcription. As shown in Fig. 7, no transcription was detected in the absence of DszR or the *dszE* promoter. On the other hand, a transcription band was detected in presence of both, whose signal intensified with the increasing concentrations of DszR, thus confirming that DszR activated transcription from *dszE* promoter. Transcription was 10 to 20-fold higher in the presence of the DNA-bending protein IHF, thus indicating that IHF also plays a direct role as co-activator of this promoter, as it has been shown for multiple σ_N -dependent promoters.

Discussion

Functional metagenomic analysis for indigo production aimed at finding aromatic oxygenases resulted in the identification of 25 independent fosmids, all bearing genes coding for putative monooxygenases. Although indigo production from indole oxidation with aromatic ring oxygenases was first described for naphthalene dioxygenases (Ensley et al., 1983) a number of later reports have assigned this capacity to different monooxygenases (van Hellemond et al., 2007 and references therein). Our data support the view that the ability to oxidize indole to produce indigo is actually more abundant among ring-hydroxylating monooxygenases than among dioxygenases. One of the fosmids providing blue color harbored a tightly linked cluster of genes that would putatively be involved in DBT biodesulfurization. Subcloning and

expressing these genes in *E. coli* conferred the recipient strain the ability to use DBT as the only source of sulfur (Fig. 2), thus confirming that these genes really were *dsz* genes. It appears that the DBT monooxygenase activity of DszC, involved in the first two oxidation steps, and not the activity of DszA, is responsible for oxidizing indole to produce indigo since the clone bearing the complete *dszEABC* operon produces indigo if the operon transcription is activated by DszR, but the one bearing a *dszEABC::lacZ* fusion did not (not shown), which is consistent with the reported ability to oxidize indole of two DBT monooxygenases (Coco et al., 2001; Ohshiro et al., 2009).

Regardless of their location (plasmid or chromosome) genes encoding for the two monooxygenases and the desulfurase involved in DBT biodesulfurization *via* the 4S pathway are found very tightly linked in an operon and in the same gene arrangement *dszABC* in all the bacteria where they have been described. The similarity in the gene arrangement and sequence conservation suggest that these genes may have been spread in nature by horizontal transfer to many different bacteria (Martínez, 2016; Nour Shafik El-Gendy, 2018). Another common feature is that the operons seem incomplete since they do not code for the additional activity required for DBT desulfurization, the NAD(P)H-dependent FMN reductase to provide electrons to both monooxygenases. In all reported cases, the gene encoding for the reductase (*dszD*) is present elsewhere in the genome, frequently in different replicons (plasmid and chromosome) and in several cases, this reductase gene has not been identified to date. This has hampered the readily transfer of the DBT biodesulfurization capability to recombinant *E. coli* strains (Denome et al., 1993), and represents one of the bottlenecks to overcome in engineering DBT biodesulfurization *via* the 4S pathway (Martínez, 2016).

The *dsz* operon reported here conserves the *dszABC* gene arrangement and their sequences, which are very similar to the others previously described in *Actinobacteria*, while it differs in that the *dszABC* genes are preceded by a flavin reductase gene, thus representing the first reported example of a complete *dsz* operon that includes the electron transfer element.

By sequence analysis we show that the FMN reductase described is the Dsz protein that shows the lowest level of similarity to its related proteins in the database (Fig. 1C). Interestingly, DszE was not most similar to identified FMN reductases involved in DBT biodesulfurization, and instead showed highest sequence similarity to SsuE FMN reductases from *Proteobacteria* involved in biodesulfurization of aliphatic sulfonates. This would suggest that this enzyme could have been recruited from other biodesulfurization pathways. In most instances, a FMN reductase has to be provided when *dsz* genes are cloned in *E. coli* because it apparently lacks a reductase able to donate electrons to the monooxygenases involved in DBT biodesulfurization. In fact, *E. coli* has been used as a host system to identify the flavin reductases that are partners of DBT monooxygenases (Ishii et al., 2000a; Furuya et al., 2004; Furuya et al., 2005). Therefore, although no functional assay specific for DszE has been performed, the fact that the subcloned *dszEABC* operon led to growth of *E. coli* on DBT as the only sulfur source suggests that DszE is fully functional. None of the *dsz* operon sequences in the databases has any linked FMN reductase-coding gene, which suggest that this particular arrangement may be not be frequent in nature. However, the functional screening may have selected for this configuration because the ones lacking the reductase would not exhibit oxygenase activity in *E. coli*.

Intriguing results on salicylate-dependent growth on DBT of clone UPO21 and inability of pMPO587 to support growth on DBT led to the discovery of the transcriptional activator DszR. Although very little is known about regulation of *dsz*, or equivalent genes in the different bacterial strains, the *Rhodococcus dsz* promoter appears to be σ_{70} -dependent and subjected to repression by sulfate and sulfur-containing amino acids (Li et al., 1996). This has represented an important bottleneck in devising DBT biodegrading biocatalysts that had to be alleviated by expressing the *dsz* operon using heterologous expression systems (Gallardo et al., 1997). Similar promoter sequences and sulfate repression of *dsz* genes have been reported in *Gordonia alkanivorans* strain 1B (Alves et al., 2007). The *dsz* promoter regions from *Gordonia alkanivorans* RIPI90A and *R. erythropolis* IGST8 showed much more

differences than the coding regions (Shavandi et al., 2010). However, they were almost identical in the -10 and -35 regions defined for strain IGS8, thus suggesting that the *dsz* promoter in *Gordonia* is also σ_{70} -dependent.

DszR is the first regulatory protein of *dsz* genes reported so far, and its sequence, and its functional role in transcribing the identified *dsz* promoter characterized *in vitro* unambiguously showed that it is an activator of σ_N -dependent promoters, which binds to two contiguous palindromic binding sites upstream of the σ_N -dependent promoter in the same face of the helix, in an arrangement very common for this type of activators. *In vitro* assays also showed that DszR-mediated activation of P_{dsz} is facilitated by IHF, a frequent co-activator of σ_N -dependent promoters (Garcia-Maurino et al., 2013 and references therein). Thus regulation of the *dszEABC* operon described here seems to be completely different from that reported in culturable bacteria. DszR is most similar to SfnR, an activator involved in dimethyl sulfide utilization as the sulfur source (Endoh et al., 2003a; Endoh et al., 2005). DszR and the SfnR sequences annotated in GenBank appear to be σ_N -dependent activators of the group IV (Bush and Dixon, 2012), which lack the N-terminal regulatory domain, and appear to be constitutively active, as recently shown for SfnR2 *in vivo* (Lundgren et al., 2019), and for DszR *in vitro* (Supplementary Fig. S3) and *in vivo* (Fig. 6) in this report. Given the constitutive activity of DszR, this system could be an appropriate part of a framework for DBT biodesulfurization (Martínez, 2016; Martinez et al., 2017) without requiring a heterologous expression system to control the *dsz* operon since its expression would not be limited by sulfur assimilation.

An *ssuEADCBF* operon, whose first gene codes for an FMN-reductase, has been reported as essential for dimethyl sulfide utilization in *P. putida* strain DS1, and the function of SsuE in this pathway has been compared to that of DszD in DBT desulfurization (Endoh et al., 2003b). SfnR has also been reported as an activator essential for utilization of dimethyl sulfone in *P. putida* DS1 (Endoh et al., 2003a) and methane sulfonate in *P. aeruginosa* PAO1 (Lundgren et al., 2019) as sulfur sources, and involved in transcription activation of

sfnFG (in DS1) and *msuED* (in PAO1), coding both for an FMN-reductase and an FMN-dependent monooxygenase in turn involved in dimethyl sulfone (Endoh et al., 2005) or methane sulfonate desulfonation (Kertesz et al., 1999), respectively. Therefore, an attractive hypothesis is that the DBT desulfurization pathway reported here has recruited not only the FMN-reductase from an aliphatic sulfonate biodesulfurization pathway but also a regulated σ_N -dependent promoter and its activator.

Expression analyses of *lacZ* fusions to 3 genes of the operon indicate that DszR is strictly required not just for *dsz* operon transcription but also for proper translation of the Dsz proteins, since they are not produced if the operon is transcribed from an upstream heterologous promoter. However, the idea that DszR is also required for translation is not supported by its sequence, which indicates it is just a transcriptional activator. When the operon is transcribed from its own promoter, production of Dsz proteins is diminished if the heterologous *P_{sal}* promoter is also actively promoting *dsz* transcription (Fig. 3C and 3E). The most plausible explanation for this interference is that *dsz* reading frames cannot be translated from the *dsz* mRNA produced from *P_{sal}*, and that this transcription is interfering with *dsz* transcription from its own promoter, which would yield the translationally functional mRNA.

A precise replacement of the *dsz* promoter region with *P_{sal}* showed that the *dsz* genes could then be translated in the absence of DszR (Fig. 4). Sequence analysis of the leader region and sequences upstream revealed that a stable secondary structure between the leader region and the *dsz* promoter sequences could be formed if the region is transcribed from an upstream heterologous promoter. This structure apparently occludes the SD and the initiation codon of *dszE*, preventing its translation. This would represent a very efficient system for insulating expression from an operon to prevent spurious gene expression, which might be particularly useful to preserve expression of “mobile” operons that are laterally transferred and inserted in different replicons. A similar phenomenon of insulation was previously reported at the *E. coli glnK* σ_N promoter and thoroughly characterized (Levy et al., 2017).

This study revealed the presence of short CT-rich sequences that associate to the ribosome binding site, thus preventing translation of the downstream gene, which were preferentially located within a subset of σ_N promoters of *E. coli*. In the *dsz* promoter, the GCTTT sequence encompassing the -12 region (Supplementary Fig. S2B), which occludes the *dszE* Shine-Dalgarno sequence (Fig. 5B), might be the “silencer” sequence. Thus, our results appear to extend this insulation phenomenon to other bacteria. In the case of the *dsz* operon, expression of all 4 genes of the operon appeared insulated, indicating that translation of the downstream genes is also prevented when *dszE* is not translated.

Given the tight proximity or even overlap of the reading frames, this effect is presumably achieved by translational coupling. Therefore, this insulation mechanism, when combined with translational coupling, is an effective way to prevent expression of all genes of an operon when spuriously transcribed from upstream sequences.

Experimental procedures

Bacterial strains and growth conditions

Supplementary Table S4 shows the bacterial strains used in this work and their genotypes. Cells were grown in M9 minimal medium (Sambrook, 2001) containing 2 g l⁻¹ glucose as carbon source, magnesium sulfate 0.2 to 2 mM or DBT 0.2 mM (in NN-Dimethylformamide) as the sulfur source, and leucine 0.5 mM and isoleucine 0.3 mM. The rich medium used in this work (when required) was, Lysogeny Broth (LB) (Sambrook, 2000). Cultures were incubated in tubes or flasks with shaking (180 r.p.m.) at 30°C (M9) or 37°C (LB) (depending on the experiment). The concentrations of the antibiotics used (when required) were (μg ml⁻¹): ampicillin (Ap) 100; chloramphenicol (Cm) 12.5; gentamicin (Gm) 10; nalidixic acid (Nal) 15. Other additions (when required) were: arabinose 1 mM, salicylate 1 mM and isopropyl β-D-thiogalactopyranoside (IPTG) 1 mM. The reagents were purchased from Sigma-Aldrich.

Plasmids construction

Plasmids and oligonucleotides used in this work are summarized in Supplementary Tables S5 and S6, respectively. All DNA manipulations were made following standard protocols (Sambrook, 2000). Plasmid DNA preparation kit was purchased from Macherey-Nagel, total DNA kit purification was acquired from Promega and DNA clean kit from General Electric Healthcare. In all cases they were used following the manufacturer's specifications. In all cloning procedures involving PCR amplification, the polymerase used was Velocity DNA polymerase (Bioline), and the amplifications were carried out following manufacturer's specific instructions. Sequences of the amplified fragments were determined by commercial sequencing (Secugen, Madrid, Spain).

For the pMPO587 construction, a 5.7 kb fragment containing *dszE*, *dszA*, *dszB* and *dszC* genes was obtained by digesting UPO21 with *Ce*III. This region includes 816 pb upstream from *dszEABC* genes, containing the last third of *orf18*, coding for a hypothetical protein, and 597 pb downstream of *dszEABC* genes harboring the whole *orf13*, encoding for a hypothetical protein, which starts 121 nucleotides downstream *dszC* (see supplementary Table S1). The fragment was blunt-ended with Klenow, and cloned into pMPO579 digested with *P*mI. These four genes were located downstream of the heterologous promoters, in the same orientation.

To build pMPO588, a 1.4 kb fragment containing *dszR* gene was obtained by digesting UPO21 clone with *N*coI and *K*pnl. The fragment was blunt-ended with T4 DNA polymerase, and cloned into pIZ1016 (Martínez-Pérez *et al.*, 2004) digested with *H*indIII and treated with Klenow to place the *dszR* downstream of the *P_{tac}* promoter, in the same orientation.

The *dszE::lacZ* transcriptional (pMPO1395) and translational (pMPO1394) fusions were constructed at nucleotide 262 of *dszE*. They were constructed by digesting pMPO587 with *B*amHI and *S*tul (filled in with Klenow). The 1.13 kb insert was ligated into pMPO200 (for the transcriptional fusion) or into pMPO234 (for the translational fusion), digested both of them with *X*maI and their ends filled-in with Klenow. In both cases, the new plasmids were then

digested with *EcoRI-SalI*, their ends filled-in with Klenow, and the 4.8 kb inserts were ligated into pMPO579-*SphI* (filled in with T4 DNA polymerase).

The *dszEA::lacZ* transcriptional (pMPO1390) and translational (pMPO597) fusions were constructed at nucleotide 464 of *dszA* by digesting pMPO200 (for the transcriptional fusion) or pMPO234 (for the translational fusion) with *EcoRI* and *SalI*, and their ends filled-in with Klenow. The resulting 3.5 kb inserts were ligated into pMPO587-*SphI* (filled in with T4 DNA polymerase).

The *dszEABC::lacZ* transcriptional (pMPO1388) and translational (pMPO593) fusions were constructed at nucleotide 1241 of *dszC*. pMPO1388 and pMPO593 were constructed by pMPO587 PCR amplification with EcoKpndszCfw and BamKpndszCrv primers (Annealing temperature: 60°C. Extension time: 1 minute). The 0.65 kb fragment was cloned into pMPO200 (for the transcriptional fusion) or into pMPO234 (for the translational fusion) digested with *EcoRI* and *BamHI*. In both cases, the new plasmids were then digested with *KpnI* and *SalI*, filling in the ends with T4 DNA polymerase. The 4.1 kb inserts bearing the fusions were ligated into pMPO587 digested with *SphI*.

The *P_{dsz}* deletion in the *dszE::lacZ* and *dszEA::lacZ* translational fusions were constructed by PCR amplification of pMPO579 with del pdszE-Fw and del pdszE-Rv primers (Annealing temperature: 54°C. Extension time: 1 minute). The 1.77 kb fragment was ligated with an 11.4 kb fragment from pMPO1394 (in the case of *dszE* fusion) resulting in pMPO1396 (fusion at nucleotide 262 of *dszE*) or an 11.4 kb fragment from pMPO597 (in the case of *dszEA* fusion), getting pMPO1541 (fusion at nucleotide 464 of *dszA*), digested both of them with *Apal* and *NsiI*.

Plasmid for DszR overproduction was constructed by *dszR* amplification from UPO21, with primers Purif dszR Fw and Purif dszR Rv (Annealing temperature: 60°C. Extension time: 1 minute). The 1.1 kb insert was digested with *SapI* and *EcoRI* and ligated into pTYB21 digested with the same enzymes.

Plasmid pMPO1547 for DNase footprinting assay was constructed by PCR amplification of a 0.167 kb fragment (containing the possible binding sites for DszR, and the P_{dsz} promoter) from pMPO1394, with primers dszE retFW and dszE retRV (Annealing temperature: 64°C. Extension time: 30 seconds). The insert was cloned into pBSKII+ digested with *HindIII* and *Sall*.

The *in vitro* transcription template plasmid pMPO1548 harbouring P_{dsz} was constructed by PCR amplification of the *dsz* promoter region with primers dszE retFW and dszE IVTRv (Annealing temperature: 64°C. Extension time: 30 seconds). The 0.212 kb fragment was ligated into pTE103 digested with *HindIII*.

Identification of clones conferring indigo production

For the functional screening for indigo production we used a metagenomic library previously constructed in our laboratory. The library was constructed with DNA extracted from a soil sample from a petrochemical industry site contaminated with crude oil, using the pMPO579 fosmid (Terron-Gonzalez et al., 2013). The library comprises approximately 6.5 Gb of metagenomic DNA distributed across approximately 185,000 different clones, with an average size of about 35 kb, and is maintained in the strain EPI300-T1_R (Epicentre) (Terron-Gonzalez et al., 2016).

The metagenomic library was transferred by triparental conjugation (Figurski and Helinski, 1979) to MPO554NaI_R strain, with DH5 α /pRK2013 as the helper strain. Conjugative matings were carried out on LB-agar without antibiotics, and incubated overnight at 37 °C. Then, the mating mixtures were plated on LB-agar with antibiotics for transconjugants selection (chloramphenicol 12.5 μ g ml⁻¹ and nalidixic acid 15 μ g ml⁻¹), salicylate 1 mM to increase the transcription levels and arabinose 1 mM to increase fosmid copy-number. The plates were incubated at 30 °C for 48 h, and subsequently 3 more days at room temperature. Positive clones were identified by the production of a blue colony color due to the indigo yield. The substrate of this reaction is indol, spontaneously produced from the tryptophan contained in LB.

DNA sequencing and data analysis

Metagenomic clones were sequenced with a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT, USA) at Lifesequencing S.L. (Valencia, Spain), and the sequences obtained were assembled using a Newbler GS v.2.3 *de novo* assembler (Roche).

Sequences corresponding to each clone were compared to those at the NCBI database utilizing the BLASTX and BLASTN programs (Altschul et al., 1997) using default parameters and annotated according to their similarities.

MOLE-BLAST tool (<https://blast.ncbi.nlm.nih.gov/moleblast/moleblast.cgi>) was used to try to identify the origin of the UPO21 sequence. The genomic insert was initially divided in 5 kb fragments and parts of these fragments subsequently separated into smaller fragments. Database of Nucleotide collection (nr/nt) was selected.

Multiple-sequence alignment was performed with MEGAX software (Kumar et al., 2018), using MUSCLE as the alignment method. Phylogenetic trees were estimated using the Maximum Likelihood method, choosing the best model indicated by MEGAX in each case (WAG+G+F for DszE tree and LG+G for DszR tree) and selecting in the Gap/Missing Data Treatment parameter the Partial Deletion option. A bootstrap analysis was run, selecting a No. of Bootstrap Replicated of 100.

Structural analyses of mRNA were carried out with RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>).

Estimation of the *dszE* translation rate was performed on the RBS calculator Web Server (<https://www.denovodna.com/software/reverse>) using the reverse Engineer RBSs method.

Determination of 2-hydroxybiphenyl production

Formation of 2-hydroxybiphenyl, as the final product of the DBT desulfurization pathway, was determined by the Gibbs assay. MPO554 harbouring UPO21 was grown in M9 medium with 0.2 mM DBT as the only sulfur source. Samples of 1 ml of the culture were taken at different times and the assay was performed as previously described (Murarka et al., 2019).

β-galactosidase activity assays

Steady-state β-galactosidase activity assays were performed to determine the expression of the *dszE::lacZ*, *dszEA::lacZ*, and *dszEABC::lacZ* translational and transcriptional fusions and their derivatives, in MPO554Nal^R or EPI300-T1^R. Preinocula of bacterial strains containing the relevant plasmids were grown to saturation in LB at 37 °C. Cells were then diluted to 0.05 of A_{600} in minimal medium with MgSO₄ at 2 mM as the sulfur source and grown during 24 h at 30 °C. Then were diluted to 0.1 of A_{600} in minimal medium (with DBT at 0.2 mM as the sulfur source when required) (with or without arabinose, salicylate or IPTG), and grown for 6 hours (depending on the experiment) at 30 °C. Cells were permeabilized with SDS and chloroform. β-galactosidase activity was determined as previously described (Miller, 1972). All the results are the average of at least three independent replicates. The error bars represent standard errors.

DszR purification

Cultures of *E. coli* NCM631 harboring the plasmid pIZ227 and the *dszR* expression plasmid pMPO1554 were grown overnight in LB medium containing ampicillin and chloramphenicol. They were used to inoculate 1 l of LB plus the appropriate antibiotics and incubated at 37°C until A_{600} reached 0.3. At this point, expression from the $\phi 10$ promoter from the T7 phage was induced by adding 0.4 mM isopropyl β-D-thiogalactopyranoside (IPTG). The incubation was continued for 21.5 h at 16°C. Cells were harvested by centrifugation at 4000 *g* for 20 minutes at 4°C, and the cell pellet frozen at -80°C until further use. DszR was purified by IMPACT™ Protein Purification System (NEB E6901) with some modifications. All the following procedures were carried out at 4°C, unless specifically mentioned. Cell pellets were resuspended in column buffer (20 mM Tris-HCl buffer, pH 8, containing 100 mM NaCl, 1 mM EDTA) with protease inhibitor cocktail (1183617001 Roche) and disrupted by sonication. Broken cells were centrifuged 30 minutes at 15000 *g*, and the clarified cell extract was loaded onto a chitin column pre-equilibrated with column buffer. The column was washed extensively with the same buffer and the on-column cleavage was induced with DTT 50mM.

Cleavage was performed during 72 h at 18°C. DszR was eluted, dialyzed (to 20 mM tris HCl pH8, 0.2 M NaCl, 0.2 M EDTA, 5 mM DTT) and stored at -80°C. Protein quantification was calculated using RC/DC kit Bio-Rad (5000121) and purity was estimated visually by SDS-PAGE to be about 98%.

RNA preparation and primer extension

Total RNA was obtained from EPI300-T1_R/pMPO597 and EPI300-T1_R/pMPO1390 with pMPO588 or pIZ1016 grown to mid-exponential phase in minimal medium with magnesium sulfate 2 mM as the sulfur source, arabinose and IPTG. Primer extension reactions were performed using 10 µg of RNA as template, [γ -³²P]-end-labelled primer PEX-DBT1, and Superscript III reverse transcriptase (Invitrogen, Carlsbad, California), as previously described (Govantes et al., 2000). DNA from pMPO587 was used for sequencing reactions using the Thermo Sequenase Cycle Sequencing Kit (USB, Cleveland, Ohio) according to manufacturer's instructions. 6% polyacrylamide-urea sequencing gels were run in Tris-borate-EDTA buffer to solve the samples. Gels were then dried, exposed to radiosensitive screens and these processed using a Typhoon 9410 scanner (GE Healthcare).

DNase I Footprinting

DszR DNase I footprinting was performed essentially as previously described (Porrua et al., 2007). DNA fragment (167 pb) for DNase I footprinting containing both putative DszR binding sites, was obtained as Sall-EcoRV digestion of pMPO1547. The probe was [α -³²P]-dCTP labelled by Klenow filling of the 3' recessed ends. The binding reactions were performed as previously described (Rivas-Marin et al., 2016) with modifications in the footprint buffer (10 mM Tris HCl pH 8, 50 mM NaCl, 0.4 mM DTT, 2% glycerol, 2 mM CaCl₂, 5 mM MgCl₂). As a size marker, a sequencing reaction carried out with the Sequenase 2.0 kit (USB) using a specific primer (SeqFpDszE1) for the labelled strand was used. The gels were processed and analyzed as described above.

In vitro transcription (IVT) assays

Multi-round IVT reactions were carried out in a final volume of 20 μ l containing 40 mM Tris HCl (pH 7.5), 150 mM KCl, 10 mM MgCl₂, 2% glycerol, 1 mM DTT, 5 mg BSA, 4 mM ATP and 0.5 μ g of a supercoiled plasmid template that contained *P_{dsz}* (pMPO1548) or the empty vector (pTE103). The *E. coli* core RNA polymerase (Epicentre Biotechnologies, Madison, WI, USA) (100 nM), *P. putida* σ_N (80 nM), and *E. coli* IHF (75 nM) (when required) were pre-incubated at 30°C during 10 min. Then, increasing concentrations of DszR were added to the mixture, and incubated at 30°C for 10 min. *E. coli* IHF was a gift from Fernando Govantes and Ana Platero (Platero et al., 2016). *P. putida* σ_N was a gift from Victoria Shingler. After incubation, a mixture of ATP, GTP, CTP (final concentration 0.4 mM each), UTP (0.07 mM) and [α -³²P]-UTP (0.033 mM, Perkin Elmer) was added to initiate multi-round IVT. After 5 min incubation at 30°C, re-initiation was prevented by the addition of heparin (0.1 mg ml⁻¹). The reactions were further incubated for 5 min at 30°C, and stopped by adding 5 μ l of stop/loading buffer (0.125% w/v bromophenol blue, 0.125% w/v xylene cyanol, 20 mM EDTA, 95% v/v formamide). Samples were run through 4 % polyacrylamide-urea gels in Tris-borate-EDTA buffer, gels dried, and exposed to radiosensitive screens. These were processed using a Typhoon 9410 scanner (GE healthcare), and analyzed using ImageQuant software (GE Healthcare).

Accession number

The insert sequence in the UPO21 fosmid is available under the EMBL/GenBank/DBBJ Accession No. MK510930.

Acknowledgements

We are grateful to Elena Rivas for her advise on the *in vitro* transcription assays and all members of the laboratory for their insights and helpful suggestions. We thank Fernando Govantes, Ana Platero and Victoria Shingler for their generous gift of IHF and σ_N . This project has received funding from the European Union's Horizon 2020 research and innovation program (Blue Growth: Unlocking the potential of Seas and Oceans) through the

Project 'INMARE' under grant agreement No. 634486 and ERA-IB 5 'METACAT'. This work was further funded by grants PCIN-2014-107 (within ERA NET IB2 grant nr. ERA-IB-14-030—MetaCat), PCIN-2017-078 (within the Marine Biotechnology ERA-NET (ERA-MBT) funded under the European Commission's Seventh Framework Programme, 2013-2017, Grant agreement 604814), BIO2014-54494-R and BIO2017-85522-R from the Ministerio de Ciencia, Innovación y Universidades, formerly Ministerio de Economía, Industria y Competitividad.

Declaration of conflict of interest

The authors declare no conflict interest of any kind.

References

Abbasian, F., Lockington, R., Megharaj, M., and Naidu, R. (2016) Identification of a new operon involved in desulfurization of dibenzothiophenes using a metagenomic study and cloning and functional analysis of the genes. *Enzyme Microb Technol* **87-88**: 24-28.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Alves, L., Melo, M., Mendonca, D., Simoes, F., Matos, J., Tenreiro, R., and Girio, F. (2007) Sequencing, cloning and expression of the dsz genes required for dibenzothiophene sulfone desulfurization from *Gordonia alkanivorans* strain 1B. *Enzyme Microb Technol* **40**: 1598–1603.

Brennerova, M.V., Josefiova, J., Brenner, V., Pieper, D.H., and Junca, H. (2009) Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environ Microbiol* **11**: 2216-2227.

Bush, M., and Dixon, R. (2012) The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. *Microbiol Mol Biol Rev* **76**: 497-529.

Chalmers, R., and Blot, M. (1999) Insertion sequences and transposons. In *Organization of the Prokaryotic Genome*. Charlebois, R.L. (ed). Washington, D. C.: American Society for Microbiology, pp. 151-169.

Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T. et al. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol* **19**: 354-359.

Denis-Larose, C., Labbe, D., Bergeron, H., Jones, A.M., Greer, C.W., al-Hawari, J. et al. (1997) Conservation of plasmid-encoded dibenzothiophene desulfurization genes in several rhodococci. *Appl Environ Microbiol* **63**: 2915-2919.

Denome, S.A., Olson, E.S., and Young, K.D. (1993) Identification and Cloning of Genes Involved in Specific Desulfurization of Dibenzothiophene by *Rhodococcus* sp. Strain IGTS8. *Appl Environ Microbiol* **59**: 2837-2843.

Denome, S.A., Oldfield, C., Nash, N.J., and Young, K.D. (1994) Characterization of the Desulfurization Genes from *Rhodococcus* sp. Strain IGTS8. *Journal of Bacteriology* **176**: 6707-6716.

Endoh, T., Habe, H., Yoshida, T., Nojiri, H., and Omori, T. (2003a) A CysB-regulated and sigma54-dependent regulator, SfnR, is essential for dimethyl sulfone metabolism of *Pseudomonas putida* strain DS1. *Microbiology* **149**: 991-1000.

Endoh, T., Habe, H., Nojiri, H., Yamane, H., and Omori, T. (2005) The sigma54-dependent transcriptional activator SfnR regulates the expression of the *Pseudomonas putida* sfnFG operon responsible for dimethyl sulphone utilization. *Mol Microbiol* **55**: 897-911.

Endoh, T., Kasuga, K., Horinouchi, M., Yoshida, T., Habe, H., Nojiri, H., and Omori, T. (2003b) Characterization and identification of genes essential for dimethyl sulfide utilization in *Pseudomonas putida* strain DS1. *Appl Microbiol Biotechnol* **62**: 83-91.

Ensley, B.D., Ratzkin, B.J., Osslund, T.D., Simon, M.J., Wackett, L.P., and Gibson, D.T. (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* **222**: 167-169.

Feng, S.Y., H.; Zhan, X.; Wang, W (2016) Enhancement of dibenzothiophene biodesulfurization via weakening feedback inhibition effects based on system-understanding of biodesulfurization mechanism by *Gordonia* sp. through potential “4S” pathway *Royal Society Advances* **6**: 8271-82881.

Figurski, D.H., and Helinski, D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**: 1648-1652.

Furuya, T., Takahashi, S., Ishii, Y., Kino, K., and Kirimura, K. (2004) Cloning of a gene encoding flavin reductase coupling with dibenzothiophene monooxygenase through coexpression screening using indigo production as selective indication. *Biochem Biophys Res Commun* **313**: 570-575.

Furuya, T., Takahashi, S., Iwasaki, Y., Ishii, Y., Kino, K., and Kirimura, K. (2005) Gene cloning and characterization of *Mycobacterium phlei* flavin reductase involved in dibenzothiophene desulfurization. *J Biosci Bioeng* **99**: 577-585.

Galan, B., Diaz, E., and Garcia, J.L. (2000) Enhancing desulphurization by engineering a flavin reductase-encoding gene cassette in recombinant biocatalysts. *Environ Microbiol* **2**: 687-694.

- Gallardo, M.E., Ferrandez, A., De Lorenzo, V., Garcia, J.L., and Diaz, E. (1997) Designing recombinant *Pseudomonas* strains to enhance biodesulfurization. *J Bacteriol* **179**: 7156-7160.
- Garcia-Maurino, S.M., Perez-Martinez, I., Amador, C.I., Canosa, I., and Santero, E. (2013) Transcriptional activation of the CrcZ and CrcY regulatory RNAs by the CbrB response regulator in *Pseudomonas putida*. *Mol Microbiol* **89**: 189-205.
- Govantes, F., Albrecht, J.A., and Gunsalus, R.P. (2000) Oxygen regulation of the *Escherichia coli* cytochrome d oxidase (cydAB) operon: roles of multiple promoters and the Fnr-1 and Fnr-2 binding sites. *Mol Microbiol* **37**: 1456-1469.
- Gray, K.A., Pogrebinsky, O.S., Mrachko, G.T., Xi, L., Monticello, D.J., and Squires, C.H. (1996) Molecular mechanisms of biocatalytic desulfurization of fossil fuels. *Nat Biotechnol* **14**: 1705-1709.
- Ishii, Y., Konishi, J., Suzuki, M., and Maruhashi, K. (2000a) Cloning and expression of the gene encoding the thermophilic NAD(P)H-FMN oxidoreductase coupling with the desulfurization enzymes from *Paenibacillus* sp. A11-2. *J Biosci Bioeng* **90**: 591-599.
- Ishii, Y., Konishi, J., Okada, H., Hirasawa, K., Onaka, T., and Suzuki, M. (2000b) Operon structure and functional analysis of the genes encoding thermophilic desulfurizing enzymes of *Paenibacillus* sp. A11-2. *Biochem Biophys Res Commun* **270**: 81-88.
- Kertesz, M.A., Schmidt-Larbig, K., and Wuest, T. (1999) A novel reduced flavin mononucleotide-dependent methanesulfonate sulfonate encoded by the sulfur-regulated msu operon of *Pseudomonas aeruginosa*. *J Bacteriol* **181**: 1464-1473.
- Kilbane II, J.J. (2016) Biodesulfurization: How to Make it Work? *Arab J Sci Eng*.

Kilbane, J.J., 2nd (2006) Microbial biocatalyst developments to upgrade fossil fuels. *Curr Opin Biotechnol* **17**: 305-314.

Kilbane, J.J., 2nd, and Stark, B. (2016) Biodesulfurization: a model system for microbial physiology research. *World J Microbiol Biotechnol* **32**: 137.

Kirimura, K., Harada, K., Iwasawa, H., Tanaka, T., Iwasaki, Y., Furuya, T. et al. (2004) Identification and functional analysis of the genes encoding dibenzothiophene-desulfurizing enzymes from thermophilic bacteria. *Appl Microbiol Biotechnol* **65**: 703-713.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**: 1547-1549.

Levy, L., Anavy, L., Solomon, O., Cohen, R., Brunwasser-Meirom, M., Ohayon, S. et al. (2017) A Synthetic Oligo Library and Sequencing Approach Reveals an Insulation Mechanism Encoded within Bacterial sigma(54) Promoters. *Cell Rep* **21**: 845-858.

Li, M.Z., Squires, C.H., Monticello, D.J., and Childs, J.D. (1996) Genetic analysis of the dsz promoter and associated regulatory regions of *Rhodococcus erythropolis* IGTS8. *J Bacteriol* **178**: 6409-6418.

Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.

Lundgren, B.R., Sarwar, Z., Feldman, K.S., Shoytush, J.M., and Nomura, C.T. (2019) SfnR2 Regulates Dimethyl Sulfide-Related Utilization in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **201**.

Martinez, I., El-Said Mohamed, M., Santos, V.E., Garcia, J.L., Garcia-Ochoa, F., and Diaz, E. (2017) Metabolic and process engineering for biodesulfurization in Gram-negative bacteria. *J Biotechnol* **262**: 47-55.

- Martínez, I.G., J.L.; Díaz, E. (2016) Genetic Engineering for Removal of Sulfur from Fuel Aromatic Heterocycles. In *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals, Chemicals, Handbook of Hydrocarbon and Lipid Microbiology*. Lee, S.Y. (ed): Springer International Publishing AG 2016, pp. 1-23.
- Miller, J.H. (1972) *Experiments in Molecular Genetics*: Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Murarka, P., Bagga, T., Singh, P., Rangra, S., and Srivastava, P. (2019) Isolation and identification of a TetR family protein that regulates the biodesulfurization operon. *AMB Express* **9**: 71.
- Nomura, N., Takada, M., Okada, H., Shinohara, Y., Nakajima-Kambe, T., Nakahara, T., and Uchiyama, H. (2005) Identification and functional analysis of genes required for desulfurization of alkyl dibenzothiophenes of Mycobacterium sp. G3. *J Biosci Bioeng* **100**: 398-402.
- Nour Shafik El-Gendy, H.N.N. (2018) Enzymology and Genetics of Biodesulfurization Process. In *Biodesulfurization in Petroleum Refining*: Scrivener Publishing LLC. John Wiley & Sons, Inc.
- Ohshiro, T., Nakura, S., Ishii, Y., Kino, K., Kirimura, K., and Izumi, Y. (2009) Novel reactivity of dibenzothiophene monooxygenase from Bacillus subtilis WU-S2B. *Biosci Biotechnol Biochem* **73**: 2128-2130.
- Park, S.J., Lee, I.S., Chang, Y.K., and Lee, S.Y. (2003) Desulfurization of dibenzothiophene and diesel oil by metabolically engineered Escherichia coli. *J Microbiol Biotechnol* **13**: 578-583.
- Piddington, C.S., Kovacevich, B.R., and Rambosek, J. (1995) Sequence and molecular characterization of a DNA region encoding the dibenzothiophene

desulfurization operon of *Rhodococcus* sp. strain IGTS8. *Appl Environ Microbiol* **61**: 468-475.

Platero, A.I., Lopez-Sanchez, A., Tomas-Gallardo, L., Santero, E., and Govantes, F. (2016) Mechanism of Antiactivation at the *Pseudomonas* sp. Strain ADP sigmaN-Dependent PatzT Promoter. *Appl Environ Microbiol* **82**: 4350-4362.

Porrua, O., Garcia-Jaramillo, M., Santero, E., and Govantes, F. (2007) The LysR-type regulator AtzR binding site: DNA sequences involved in activation, repression and cyanuric acid-dependent repositioning. *Mol Microbiol* **66**: 410-427.

Reichmuth, D.S., Hittle, J.L., Blanch, H.W., and Keasling, J.D. (2000) Biodesulfurization of dibenzothiophene in *Escherichia coli* is enhanced by expression of a *Vibrio harveyi* oxidoreductase gene. *Biotechnol Bioeng* **67**: 72-79.

Rhee, S.K., Chang, J.H., Chang, Y.K., and Chang, H.N. (1998) Desulfurization of dibenzothiophene and diesel oils by a newly isolated gordona strain, CYKS1. *Appl Environ Microbiol* **64**: 2327-2331.

Rivas-Marin, E., Floriano, B., and Santero, E. (2016) Genetic dissection of independent and cooperative transcriptional activation by the LysR-type activator ThnR at close divergent promoters. *Sci Rep* **6**: 24538.

Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**: 946-950.

Sambrook, J.F., E. F.; Maniatis, T. (2000) *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sambrook, J.R., D. W. (2001) *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Shavandi, M., Sadeghizadeh, M., Khajeh, K., Mohebbali, G., and Zomorodipour, A. (2010) Genomic structure and promoter analysis of the dsz operon for

dibenzothiophene biodesulfurization from *Gordonia alkanivorans* RIPI90A. *Appl Microbiol Biotechnol* **87**: 1455-1461.

Swaty, T.E. (2005) Global refining industry trends: the present and future. *Hydrocarbon Processing* **84**: 35-46.

Takada, M., Nomura, N., Okada, H., Nakajima-Kambe, T., Nakahara, T., and Uchiyama, H. (2005) De-repression and comparison of oil-water separation activity of the dibenzothiophene desulfurizing bacterium, *Mycobacterium* sp. G3. *Biotechnol Lett* **27**: 871-874.

Tanaka, Y., Yoshikawa, O., Maruhashi, K., and Kurane, R. (2002) The cbs mutant strain of *Rhodococcus erythropolis* KA2-5-1 expresses high levels of Dsz enzymes in the presence of sulfate. *Arch Microbiol* **178**: 351-357.

Terron-Gonzalez, L., Medina, C., Limon-Mortes, M.C., and Santero, E. (2013) Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. *Sci Rep* **3**: 1107.

Terron-Gonzalez, L., Martin-Cabello, G., Ferrer, M., and Santero, E. (2016) Functional Metagenomics of a Biostimulated Petroleum-Contaminated Soil Reveals an Extraordinary Diversity of Extradial Dioxygenases. *Appl Environ Microbiol* **82**: 2467-2478.

van Hellemond, E.W., Janssen, D.B., and Fraaije, M.W. (2007) Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl Environ Microbiol* **73**: 5832-5839.

Vilchez-Vargas, R., Junca, H., and Pieper, D.H. (2010) Metabolic networks, microbial ecology and 'omics' technologies: towards understanding in situ biodegradation processes. *Environ Microbiol* **12**: 3089-3104.

Wang, J., Butler, R.R., 3rd, Wu, F., Pombert, J.F., Kilbane, J.J., 2nd, and Stark, B.C. (2017) Enhancement of Microbial Biodesulfurization via Genetic Engineering and Adaptive Evolution. *PLoS One* **12**: e0168833.

Figure Legends

Figure 1. A. Schematic diagram showing the UPO21 clone. Putative *orfs* and their orientations are shown. *orfs* involved in dibenzothiophene desulfurization are expanded. P_{sal} , inducible by salicylate, and P_{T7} , dependent on the T7 RNA-polymerase, are heterologous promoters present in the fosmid vector. B. Identity between DszA, DszB and DszC and the most similar characterized sequences in non-redundant protein sequences database using Blastp (protein-protein BLAST). C. Phylogenetic tree of DszE and other FMN oxidoreductases. The scale indicates the amino acid substitutions per position. FMN reductases involved in DBT biodegradation and in aliphatic sulfonates biodegradation appear separated into two groups. GenBank accession numbers of the previously sequenced oxidoreductases included in the phylogenetic tree, and the bacterial strains from which they come are shown in Supplementary Table S1.

Figure 2. A. Growth of MPO554/UPO21 in minimal M9 medium with 0.2 mM DBT as the only sulfur source. B. Growth of MPO554/pMPO587 in minimal M9 medium with 0.2 mM DBT as the only sulfur source. C. Growth of MPO554/pMPO587, pMPO588 in minimal M9 medium with DBT as the only sulfur source. In all panels, values are the average of at least three independent replicates; error bars represent standard errors. Residual growth up to $OD_{600}=0.2$ due to sulfur previously accumulated by the cells was also observed in the control strain bearing the empty vector pMPO579 but it is not shown for clarity.

Figure 3. A. Scheme of the transcriptional and translational *lacZ* gene fusions to *dszE*, *dszEA* and *dszEABC*. B, C, D, E. Expression levels of *dszE::lacZ* (pMPO1395, transcriptional; pMPO1394, translational), *dszEA::lacZ* (pMPO1930, transcriptional; pMPO597, translational)

and *dszEABC::lacZ* (pMPO1388, transcriptional; pMPO593, translational) with or without *dszR*, after 6 hours of growth in minimal M9 medium with DBT and SO₄ 0.2 mM, arabinose, salicylate and IPTG. Graphs show transcriptional fusions (B and D) and translational fusions (C and E), only transcribed from the *dsz* promoter (EPI300-T1_R strain, B and C) or also transcribed from *P_{sal}* (MPO554 strain, D and E).

Figure 4. A. Scheme of the deletion of the sequence upstream the native transcription start site. The resulting translational fusions allow the study of the expression from *P_{sal}* starting from the same position as the native transcription. B, C. Expression levels of ΔP_{dsz} *dszE::lacZ* and ΔP_{dsz} *dszEA::lacZ* translational fusions as compared to the fusions bearing the *P_{dsz}* and *P_{sal}* promoters upstream, in EPI300-T1_R (B) and in MPO554 (C) with (grey) or without (white) *dszR*, after 6 hours of growth in minimal M9 medium with DBT and SO₄ 0.2 mM, arabinose, salicylate and IPTG.

Figure 5. *dszE* ribosome binding site and start codon accessibility. Predicted structures of the *dsz* leader region when transcribed from its native start site (A) or when starting 35 nucleotides upstream (B). The *dszE* initiation codon and its putative Shine-Dalgarno sequence are marked with a line. The color intensity indicates the probability that the nucleotide is in the proposed form, either paired or unpaired (the darker a nucleotide is, the more likely its predicted structure is correct).

Figure 6. DszR-mediated activation in different sulfur sources. *dszEA::lacZ* transcriptional fusion activity, in *E. coli* EPI300-T1_R with pZ1016 (empty vector) or pMPO588 (*dszR*), after 24 hours of growth in minimal medium M9 with different S sources: without S but with N,N-dimethylformamide (NNDMF), the DBT solvent, with SO₄ 2 mM, DBT 0.2 mM, and DBT+SO₄ 0.2 mM each, at low (left) or high (right) levels of DszR.

Figure 7. *In vitro* transcription from *P_{dsz}*. A. Primary result of one representative experiment of the *in vitro* transcription activated by DszR. The discontinuous line between right and left in the image indicates that they come from two different parts of the same gel. The same experiment was performed without addition of σ_N and no signal was detected (data not

shown) B. Average value of three independent experiments. Maximum value at the highest concentration of DszR (150 nM) was considered 100%, and the transcription levels are expressed relative to this value. Filled circles, IHF addition; open circles, no IHF addition. The error bars represent the standard errors.

A

B

	Most similar catacterized protein and accession numbers	Identity	E-value
DszA	DszA, dibenzothiophene desulfurization protein (Rhodococcus sp. DS-3) ABE26644.1	82 %	0.0
DszB	hydroxybiphenyl-2-sulfinate desulfinase (uncultured bacterium) ANC50862.1	76 %	1e-177
DszC	dibenzothiophene monooxygenase (uncultured bacterium) ANC50861.1	86 %	0.0

C

A

A

B

C

A

B

A

B