Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the *alk* genes

Jan B. van Beilen, Sven Panke,[†] Sacha Lucchini,[‡] Alessandro G. Franchini, Martina Röthlisberger and Bernard Witholt

Author for correspondence: Jan B. van Beilen. Tel: +41 1 6333444. Fax: +41 1 6331051.
e-mail: vanbeilen@biotech.biol.ethz.ch

Institute of Biotechnology,
Swiss Federal Institute
of Technology (ETH),
ETH-Hönggerberg,
CH-8093 Zürich,
Switzerland

The *Pseudomonas putida* GPo1 (commonly known as *Pseudomonas oleovorans* GPo1) *alkBFGHJKL* and *alkST* gene clusters, which encode proteins involved in the conversion of n-alkanes to fatty acids, are located end to end on the OCT plasmid, separated by 9.7 kb of DNA. This DNA segment encodes, amongst others, a methyl-accepting transducer protein (AlkN) that may be involved in chemotaxis to alkanes. In *P. putida* P1, the *alkBFGHJKL* and *alkST* gene clusters are flanked by almost identical copies of the insertion sequence *IS_{Ppu4}*, constituting a class 1 transposon. Other insertion sequences flank and interrupt the *alk* genes in both strains. Apart from the coding regions of the GPo1 and P1 *alk* genes (80–92 % sequence identity), only the *alkB* and *alkS* promoter regions are conserved. Competition experiments suggest that highly conserved inverted repeats in the *alkB* and *alkS* promoter regions bind AlkS.

Keywords: alkane degradation, insertion sequence, chemotaxis, regulation

INTRODUCTION

Even though the ability to grow on alkanes is a very common phenomenon, most biochemical and genetic studies have focused on a limited number of prototype systems. An example is the *Pseudomonas putida* GPo1 (commonly known as *Pseudomonas oleovorans* GPo1 = TF4-1L = ATCC 29347) alkane hydroxylase system (Baptist *et al.*, 1963), which can be used to carry out a wide range of stereo- and regioselective oxidation reactions (Witholt *et al.*, 1990). In the alkane degrader *Pseudomonas putida* P1, closely related genes were found. Recent PCR studies (Smits *et al.*, 1999) and genome sequencing have shown that many strains contain homologues of the GPo1 alkane hydroxylase, usually encoded on the chromosome. In contrast, the

GPo1 alkane degradation (*alk*) genes are grouped in two clusters located on a large plasmid, named OCT (Chakrabarty *et al.*, 1973). Mapping experiments had indicated that the two loci (Benson *et al.*, 1977; Fennewald & Shapiro, 1977) are transcribed towards each other and are separated by about 40 kb (Fennewald *et al.*, 1979; Owen, 1986). The G + C content of the *alk* genes is much lower than that of the host strain and the OCT plasmid (Fennewald *et al.*, 1978; Kok *et al.*, 1989b; Panke *et al.*, 1999b), suggesting that these genes are part of a 55 kb mobile element, which integrated in the OCT plasmid. This appears to be a common finding for catabolic genes, and many catabolic transposons have now been described (Tan, 1999).

Witholt and co-workers cloned the two *alk* clusters in pLAFR1 as 18 kb and 16.9 kb *EcoRI* fragments to create plasmid pGEc47 (Eggink *et al.*, 1987a). Subsequent minicell experiments and partial sequence analysis showed that the 18 kb fragment contains an operon named *alkBFGHJKL*, which encodes two components of the alkane hydroxylase system and enzymes involved in further metabolic steps (Eggink *et al.*, 1987b; Kok *et al.*, 1989a, b; van Beilen *et al.*, 1992a). The 16.9 kb fragment was found to encode the third component of the alkane hydroxylase system: rubredoxin re-

[†] **Present address:** DSM Research, Postbus 18, 6160 MD Geleen, The Netherlands.

[‡] **Present address:** Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK.

The EMBL accession numbers for the sequences reported in this paper are AJ245436 [*P. putida* (*oleovorans*) GPo1 *alk* gene clusters and flanking DNA], AJ233397 [*P. putida* P1 *alk* gene clusters and flanking DNA], AJ249793 [*P. putida* P1 *nahKJ* genes], AJ249825 [*P. putida* (*oleovorans*) GPo1 16S RNA gene] and AJ271219 [*P. putida* P1 16S RNA gene].

ductase, and *AlkS*, which regulates expression of the *alkBFGHJKL* operon (Kok *et al.*, 1989b; Eggink *et al.*, 1990; Panke *et al.*, 1999b; Canosa *et al.*, 2000), as well as the *alkST* genes (Canosa *et al.*, 1999, 2000) (see Fig. 1). Only *alkB*, *alkG* and *alkS* (Kok *et al.*, 1989a; van Beilen, 1994; van Beilen *et al.*, 1992a) cannot be complemented by genes present on the chromosome of GPo1.

To obtain more evidence for the existence of an *alk* transposon, and to investigate the possibility that additional non-essential functions involved in alkane degradation might be encoded outside the regions that had been sequenced to date, we completed the sequence of the two *EcoRI* fragments in pGEc47 and cloned additional DNA that links the two *EcoRI* fragments. Comparison of the *P. putida* P1 *alkBFGHJ* genes (Smits *et al.*, 1999) with the GPo1 *alkBFGHJKL* genes showed that the homology between the *alkB* promoter regions ends with an inverted repeat. To better understand the significance of this and other sequence features, we also cloned and sequenced the remaining *P. putida* P1 *alk* genes and flanking DNA. Analysis of sequences flanking the *alk* gene clusters in both strains revealed the presence of several complete and incomplete insertion sequences (IS), which are likely to have played a role in mobilizing the *alk* gene clusters.

METHODS

Bacterial strains and growth conditions. *P. putida* GPo1 [commonly known as *P. oleovorans* GPo1 = TF4-1L = ATCC 29347 (Schwartz, 1973)] is a prototroph strain that carries the OCT-plasmid. *P. putida* P1 is a prototroph strain isolated from a pentane enrichment culture (Smits *et al.*, 1999). *E. coli* DH10B (Gibco-BRL) was used for cloning and the production of plasmid DNA for sequencing. *E. coli* W3110 [$F^- \lambda^-$ IN(*rrnD-rrnE*)1] is a prototroph K-12 strain (Bachmann, 1987) used for chemotaxis studies. *E. coli* JM101(*alkS** *alkBp-xylE*) (Panke *et al.*, 1999a) was used for regulation studies. All strains were grown in E2 medium (Lageveen *et al.*, 1988) supplemented with the appropriate carbon source, or Luria-Bertani broth (Luria *et al.*, 1960). *E. coli* DH10B was transformed by electroporation (Dower *et al.*, 1988). To select *E. coli* transformants, antibiotics were used at the following concentrations ($\mu\text{g ml}^{-1}$): ampicillin, 200; tetracycline, 12.5; kanamycin, 50.

DNA methods. Restriction enzymes, T4-DNA ligase and dideoxynucleotides were obtained from Roche Diagnostics. PCR reactions were carried out as described by Innis *et al.* (1990), using a Perkin Elmer GeneAmp PCR System 9600, and oligonucleotides were synthesized by Microsynth. *Taq* DNA polymerase and the Expand Long Template PCR system were obtained from Promega and Roche Diagnostics, respectively. For cloning, PCR products were digested with appropriate restriction enzymes, purified over a 1% agarose gel in TBE buffer and isolated by electroelution (Sambrook *et al.*, 1989). PCR fragments were cloned between the same or compatible sites of pGEM-7Zf(+) (Promega) or pZERO-2 (Invitrogen). Plasmid DNA was isolated with the Roche Diagnostics High Pure plasmid isolation kit or, for large plasmids, as described by Kado & Liu (1981). PFGE (Bustamante *et al.*, 1993) was carried out using the Bio-Rad Chef-DR II Electrophoresis Cell. Chromosomal DNA was isolated with the Qiagen Genomic DNA kit.

The Roche Diagnostics DIG kit was used for Southern blots, using CSPD [disodium 3-(4-meth-oxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo (3.3.1.1^{3,7}) decan-4-yl}phenyl phosphate)] as the chemiluminescent substrate. To prepare probes, appropriate DNA fragments were purified using a 1% agarose gel, electroeluted and labelled (DIG labelling kit; Roche Diagnostics). Hybridizations were carried out using standard hybridization buffer without formamide at 64 °C.

Cloning and sequencing of *alk* genes and flanking DNA. To construct enriched gene banks, suitable restriction fragments were identified by Southern blotting using previously cloned GPo1 *alk* genes or flanking DNA as probes. Restriction fragments around the target size were cut from a preparative agarose gel, isolated by electroelution, ligated between the appropriate sites of pGEM-7Zf(+) or pZERO-2 and transformed into *E. coli* DH10B. Transformants containing the desired DNA fragment were identified by colony blotting using Roche Diagnostics nylon membrane for colony and plaque hybridization and the Roche Diagnostics DIG kit.

To clone DNA downstream of *alkL* in GPo1, a 0.8 kb *SalI* fragment located between *alkL* and the end of the *EcoRI* fragment was DIG-labelled and used as probe to identify an overlapping 3.0 kb *NheI*–*HindIII* fragment. A DIG-labelled GPo1 *alkST* fragment was used as probe to clone the P1 *alkST* genes as an *EcoRI*–*HindIII* and a *HindIII* fragment. Sequencing showed that an additional 400 bp *HindIII* fragment was missing from *alkS*, which was obtained by PCR using primers P1-*alkST*Bam (GCTCGCGGTGGATCCTC-AGGTC) and P1-*alkST*Sac (GCGGCGGAGCTCGCGGTGGATGCTCAGGTC), and sequenced. The previously cloned *EcoRI*–*HindIII* fragment containing the P1 *alkBFGHJ'* genes (Smits *et al.*, 1999) forms a contiguous sequence with the *EcoRI*–*HindIII* fragment containing most of P1 *alkS* (the *EcoRI* sites upstream of *alkS* and *alkB* are the same). The Expand Long Template system was used with primers P1-J-Xba (GGTCTAGAGTGGCGCTGCGTGCCCCGTAAGC) and P1-STBamSac (GCGGCGGAGCTCGCGGTGGATCCTCAGGTC) to obtain DNA segments downstream of *alkJ* and *alkT*. (This PCR was carried out based on the erroneous assumption that the P1 *alk* genes are organized as in GPo1, where *alkST* is located downstream of *alkL*. A 7.2 kb PCR fragment was obtained because almost identical copies of insertion element *IS_{Ppu4}* flank the P1 *alk* genes, and cause a crossover reaction in the PCR.) In addition, a 3.1 kb *EcoRI* fragment containing *IS_{Ppu4.2}*, and a 7.5 kb *EcoRI*–*SalI* fragment containing the P1 *alkL* gene, *IS_{Ppu4.1}*, and DNA further downstream, were cloned.

A combination of techniques was used to sequence the cloned DNA fragments. Subclones for sequencing of the 8.0 kb DNA segment downstream of *alkT* were prepared by making nested deletions with Exonuclease III (Henikoff, 1984). The other DNA fragments were sequenced by subcloning selected restriction fragments, by making directed deletions and by primer walking. Cycle sequencing reactions were carried out using the Amersham Thermosequenase kit with deaza-GTP and IRD-800-labelled primers obtained from MWG-Biotech. Sequencing was carried out on a Li-Cor 4000L sequencer. DNA and protein sequences were analysed using Lasergene Navigator (DNASTAR). BLAST sequence comparisons were carried out at NCBI (Altschul *et al.*, 1990), using standard settings or match value, mismatch value and gap extension penalty $-r$ 5, $-q$ -4 and $-E$ 4, respectively, to detect low levels of sequence homology. 16S rRNA genes were amplified with primers 16F27 and 16R1525, and sequenced with 16F355, 16R519 and 16R1488 (Hauben *et al.*, 1997).

Construction of plasmids and strains for *alkB* and *alkS* promoter studies. The three inverted repeats (IR-B, IR-S and IR-syn) were synthesized as linkers with *Eco*RI and *Hind*III sticky ends (see Fig. 3c) and ligated to plasmid pGEM-7Zf(+) (Promega) digested with *Eco*RI and *Hind*III. Plasmid DNA isolated from *E. coli* DH10B transformants, selected by standard blue/white screening, was sequenced to check the inserts. *E. coli* JM101(*alkS** *alkBp*-*xylE*) was transformed with the three IR-vectors, pGEM-7Zf(+) and pBG201. M9 medium (100 ml), containing 0.2% glucose and 0.001% thiamine, was inoculated with 2 ml overnight LB-medium cultures. At an OD₄₅₀ of 0.5 the cultures were induced with 0.02% dicyclopropylketone (Grund *et al.*, 1975). Catechol-2,3-dioxygenase (XylE) activity was assayed (Nozaki, 1970) 2 h after induction.

Accession numbers and nomenclature. Sequences determined in this study were deposited in the EMBL database and are available under accession numbers AJ245436 [*P. putida* (*oleovorans*) GPo1 *alk* gene clusters and flanking DNA], AJ233397 [*P. putida* P1 *alk* gene clusters and flanking DNA], AJ249793 [*P. putida* P1 *nabKJ* genes], AJ249825 [*P. putida* (*oleovorans*) GPo1 16S RNA gene] and AJ271219 [*P. putida* P1 16S RNA gene]. The insertion sequences were named as proposed by the curators of the insertion sequence database (http://www-is.biotoul.fr/is/is_about.html, 18 October 2000). In the absence of a standardized nomenclature for transposons, the *P. putida* P1 *alk* transposon was named Tn*Ppu-alk1*.

RESULTS AND DISCUSSION

P. putida (*oleovorans*) GPo1 utilizes medium-chain-length alkanes as its sole carbon and energy source and has been studied in detail with respect to the genetics and enzymology of alkane oxidation (reviewed by van Beilen *et al.*, 1994) (Fig. 1). We cloned and sequenced DNA segments flanking the *alk* genes of GPo1 to identify additional genes involved in alkane degradation and to investigate the possibility that the *alk* genes of this strain are organized in a transposon-like structure. To be able to judge the significance of sequence features and gene organization of the GPo1 *alk* genes, we also cloned and sequenced the related *alk* genes of *P. putida* P1.

Taxonomical analysis of *Pseudomonas* strains GPo1 and P1

The strain commonly known as *P. oleovorans* GPo1 (= TF4-1L = ATCC 29347) was first described in 1963 (Baptist *et al.*, 1963) and tentatively identified as a strain of *P. oleovorans*. After strain improvement for the production of epoxides (Schwartz & McCoy, 1973) it was submitted to ATCC as *P. oleovorans* TF4-1L, and is available as ATCC 29347. However, in the extensive phylogeny paper of Stanier *et al.* (1966), the original isolate was listed as strain 266 and already classified as a *P. putida* biotype A (Chakrabarty *et al.*, 1973). Strain 266 was also submitted to ATCC and is available but unlisted (ATCC 17633). As expected, ATCC 17633 is able to grow on n-octane and the 16S sequences of GPo1 and ATCC 17633 are identical. Sequencing of the complete GPo1 16S sequence showed that it is closely related to the 16S sequences of *P. putida* F1 (gb:L37365)

Fig. 1. Metabolic pathway of alkane degradation, role and cellular localization of Alk proteins and regulation of the *alk* genes. The *alkBFGHJKL* operon encodes the alkane hydroxylase (AlkB), two rubredoxins (AlkF and AlkG), an alcohol and an aldehyde dehydrogenase (AlkJ and AlkH, respectively), an acyl-CoA synthetase (AlkK) and an outer-membrane protein of unknown function (AlkL) (Kok *et al.*, 1989a, b; van Beilen *et al.*, 1992a). The *alkST* locus encodes the third component of the alkane hydroxylase system, rubredoxin reductase (AlkT), and AlkS, which positively regulates expression of the *alkBFGHJKL* operon (Eggink *et al.*, 1990; Panke *et al.*, 1999b) and the *alkST* genes (Canosa *et al.*, 2000). AlkN is a putative chemotactic transducer for alkanes.

and the *P. putida* biotype A type strain DSM 291^T (5 and 12 differences, or 99.7 and 99.2% sequence identity, respectively). It is more distantly related to the 16S sequence of other *Pseudomonas* species (>30 differences, or <98% sequence identity), including the type strain of *P. oleovorans*, DSM 1045^T, which was first described in 1941 (Lee & Chandler, 1941). This confirms the earlier classification of GPo1 by Stanier *et al.* (1966). In this paper we refer to GPo1 as *P. putida* (*oleovorans*) GPo1.

P. putida P1 was originally isolated from a pentane enrichment culture inoculated with soil from a gas station in Groningen (Bioclear). It was classified as a *P. putida* biotype A strain by NCIMB. Sequencing of the complete P1 16S sequence showed that it is closely related to the GPo1 16S sequence (three differences, or 99.8% sequence identity).

Cloning and sequencing of DNA flanking the *P. putida* (*oleovorans*) GPo1 *alk* gene clusters

Witholt and co-workers cloned, and partially sequenced, an 18 kb *Eco*RI fragment containing the *alkBFGHJKL* genes (9.1 kb sequenced), and a 16.9 kb *Eco*RI fragment containing the *alkST* genes (4.7 kb sequenced) (Eggink *et al.*, 1987a, 1990; Kok *et al.*, 1989a, b; Panke *et al.*,

Fig. 2. Sequence analysis of the *P. putida* (*oleovorans*) GPo1 and the *P. putida* P1 *alk* genes and flanking DNA. The scale is in bp. Arrows represent potential coding regions. Black arrows represent genes involved, or presumably involved, in alkane degradation and vertically hatched arrows represent complete or partial transposase genes. Horizontally hatched arrows (also in combination with vertically hatched) indicate that the ORFs are incomplete or interrupted by stop codons or frameshifts. Black bars correspond to (incomplete) insertion sequences. ISPPu1–5 are new IS elements named in this study. The other black bars represent incomplete IS elements, with the name of the most closely related IS element shown between parentheses. The orientation of the IS elements is marked by L and R for left and right end. Bars linking the GPo1 and P1 DNA segments indicate homologous regions. DR, direct repeats formed by ISPPu4.1 and ISPPu4.2. See Table 1 and Fig. 1 for additional information on coding regions.

1999b; van Beilen *et al.*, 1992a). We completed the sequence of both *Eco*RI fragments (Fig. 2). The region downstream of *alkL* in the 18 kb *Eco*RI fragment was found to encode a putative transposase of which a few N-terminal codons were missing due to the *Eco*RI cloning site. A 3.0 kb *Nhe*I–*Hind*III fragment which was cloned to complete the putative transposase (*tnpA1*) sequence was found to overlap with the sequence downstream of *alkT* on the 16.9 kb *Eco*RI fragment: a 651 bp *Eco*RI fragment sufficed to join the 16.9 and 18 kb *Eco*RI fragments. The resulting restriction map is in agreement with all restriction sites mapped in Southern blots (data not shown). This confirmed the hypothesis that the *alk* gene clusters are located back to back on the OCT plasmid. However, the two clusters are separated by only 9.7 kb (Fig. 2), instead of the 40 kb suggested by crossover experiments (Fennewald *et al.*, 1979).

A chemotactic transducer for alkanes?

Sequence analysis of the 9.7 kb DNA segment separating the GPo1 *alk* gene clusters showed that a protein encoded 2 kb downstream of *alkL* shows up to 30% amino acid sequence identity with methyl-accepting chemotactic transducers, recognizing, for example, naphthalene (Grimm & Harwood, 1999) or amino acids (Taguchi *et al.*, 1997). Like the *alk* genes, the chemoreceptor gene has a much lower G + C content than the flanking DNA and the OCT plasmid as a whole (Table 1). In addition, the 230 bp region directly upstream of the ATG start codon has 56% DNA sequence identity with the GPo1 *alkB* promoter region and the first 76 bases of the *alkB* gene, which suggests that the chemoreceptor gene (*alkN*) plays a role in alkane degradation (Fig. 3). Unfortunately, GPo1 is not very motile and

could not be used for chemotaxis experiments, whilst P1 only contains a remnant of the *alkN* gene (see below).

Cloning and sequencing of the *P. putida* P1 *alk* gene clusters

Previously, a 7.5 kb *Eco*RI–*Hind*III fragment encoding the P1 *alkBFGHJ* genes was cloned, sequenced and shown to complement an *E. coli* recombinant containing all genes necessary for growth on n-octane except *alkB* (Smits *et al.*, 1999) or *alkG* (rubredoxin 2) (J. B. van Beilen, unpublished data). The remaining *alk* genes and flanking DNA were cloned as overlapping DNA fragments using DIG-labelled probes (the GPo1 *alkST* or *alkJKL* genes) or PCR. Attempts to localize the P1 *alk* genes were not conclusive. P1 contains a 114 ± 7 kb plasmid, as demonstrated by PFGE experiments (data not shown), but subsequent Southern blots proved that the *alk* genes are not located on this plasmid. We were unable to detect mobilization of the P1 *alk* gene clusters to other *P. putida* strains.

Sequence analysis and comparison of the *P. putida* (*oleovorans*) GPo1 and *P. putida* P1 DNA segments

The GPo1 and P1 sequences (34902 and 23312 bp, respectively) were analysed for coding regions, homology to each other and homology with database sequences. Table 1 lists all ORFs that have significant homology with protein database sequences and other ORFs over 100 aa. For each of the ORFs, length, position, G + C content of coding DNA, putative ribosome-binding site and best score in BLASTX searches are given. The data are graphically presented in Fig. 2. In some cases ORFs were extended or shortened to alternative start codons if BLASTX searches indicated

Table 1. Proteins encoded by the *alk* gene clusters of *P. putida* (*oleovorans*) GPo1 and *P. putida* P1, and flanking DNA

Gene	Length (aa)	Position	G + C (mol %)	RBS	BLASTP score	Identity (%)	Overlap/length of best score	Best scores*	Comments
<i>P. putida</i> (<i>oleovorans</i>) GPo1 OCT plasmid									
<i>orf1</i>	179	212–751	44	GAG(8)ATG	—	—	—	—	No hit
<i>orf2</i>	307	748–1671	44	GGTG(5)ATG	—	—	—	—	No hit
<i>orf3</i>	130	2396–2755	54	GGAG(10)ATG	299	92	109/118	gi:8037784	Hypothetical 13 kDa protein in <i>P. syringae</i> . Similarity to IS ORFs
<i>orf4</i>	510	2774–4306	59	GGTA(11)ATG	808	76	506/507	gi:8037785	Hypothetical 60 kDa protein in <i>P. syringae</i> . Similarity to IS ORFs
<i>orf5</i>	192	5612–5034	54	GAGG(11)ATG	145	41	172/194	sp:P50361	Hypothetical 21.7 kDa protein y4hQ, <i>Rhizobium</i> sp. NGR234
<i>orf6</i>	96	6230–5940	61	GAG(10)ATG	82	60	74/332	sp:P55636	Putative integrase/recombinase y4rC, <i>Rhizobium</i> sp. NGR234
<i>orf7</i>	274	6284–7108	61	AGGAG(8)GTG	289	73	193/276	pir:S27641	Putative transposase in <i>P. syringae</i> IS53 (fragment)
<i>alkB</i>	401	7520–8722	47	GGAGA(7)ATG	—	—	—	sp:P12691	Alkane hydroxylase (Kok <i>et al.</i> , 1989b)
<i>alkF</i>	132	8942–9337	46	AGGAGA(8)ATG	—	—	—	sp:P12692	Rubredoxin 1 (Kok <i>et al.</i> , 1989a)
<i>alkG</i>	172	9388–9906	46	GGTGA(6)ATG	—	—	—	sp:P00272	Rubredoxin 2 (Kok <i>et al.</i> , 1989a)
<i>alkH</i>	483	9956–11404	46	AGGA(10)ATG	—	—	—	sp:P12693	Aldehyde dehydrogenase (Kok <i>et al.</i> , 1989a)
<i>alkJ</i>	558	11446–13119	46	GAGAA(6)ATG	—	—	—	sp:Q00593	Alcohol dehydrogenase (van Beilen <i>et al.</i> , 1992a)
<i>alkK</i>	546	13182–14819	46	GAGG(8)ATG	—	—	—	sp:Q00594	Acyl-CoA synthetase (van Beilen <i>et al.</i> , 1992a)
<i>alkL</i>	230	14924–15613	43	GAGG(8)ATG	—	—	—	sp:Q00595	Outer-membrane protein (van Beilen <i>et al.</i> , 1992a)
<i>tnpA1</i>	355	16753–15686	55	GGAG(7)ATG	502	70	353/361	gi:5759102	Putative transposase in <i>P. syringae</i> IS870 (see TnpA4,5)
<i>alkN</i>	492	17714–19192	40	AGAGG(7)ATG	161	30	406/629	gi:BAA23413	Chemotactic transducer of <i>P. aeruginosa</i>
<i>orf8</i>	219	19973–19314	55	GGAGG(6)GTG	—	—	—	—	No hit
<i>orf9</i>	809	22377–19948	55	GAAGA(8)ATG	—	—	—	—	No hit
<i>orf10</i>	483	23987–22536	56	GGG(8)ATG	35	21	215/320	gi:3095165	Site-specific recombinase Int4 of <i>Vibrio cholerae</i>
<i>orf11</i>	92	24712–24434	53	?ATG	123	65	79/245	sp:P55617	Putative transposase subunit <i>Rhizobium</i> sp. NGR234 (fragment)
<i>orf12</i>	112	25148–24810	57	GGG(7)ATG	54	35	78/114	gi:2995634	Hypothetical protein in <i>P. putida</i> H Tn5502
<i>alkT</i>	385	26654–25500	47	GGAGAG(6)ATG	—	—	—	sp:P17052	Rubredoxin reductase (Eggink <i>et al.</i> , 1990)
<i>alkS</i>	882	29352–26707	45	GAGAA(7)ATG	—	—	—	sp:P17051	Regulator of <i>alkBFGHJKL</i> expression (Panke <i>et al.</i> , 1999b)
<i>orf13</i>	150	30067–29612	62	—	268	88	151/326	gi:2995640	Transposase of <i>P. putida</i> IS1384 (fragment)
<i>orf14</i>	199	31050–30451	60	GAG(6)ATG	158	50	155/197	pir:JC2320	Hypothetical 22 kDa protein in <i>Acetobacter</i> sp. plasmid pAH4
<i>recA</i>	348	32466–31420	51	?ATG	522	75	346/353	sp:P03017	RecA of <i>E. coli</i>
<i>recA'</i>	174	33088–32564	64	GGAGA(6)ATG	310	90	173/346	pir:JN0304	RecA of <i>P. aeruginosa</i> (fragment)
<i>orf15</i>	333	34453–33491	56	?GTG	—	—	—	—	No hit
<i>P. putida</i> P1									
<i>tnpA2</i>	332	2583–3633	54	GGAGA(8)ATG	443	63	332/334	sp:Q56897	Transposase of <i>Y. enterocolitica</i> IS1328 (frameshifts)
<i>tnpA3</i>	150	3908–4357	55	—	99	41	134/334	sp:Q56897	Transposase of <i>Y. enterocolitica</i> IS1328 (fragment)
<i>alkT</i>	385	6085–4931	48	GAGAG(6)ATG	696	91	385/385	sp:P17052	Rubredoxin reductase AlkT of GPo1
<i>alkS</i>	883	8788–6137	45	GAGA(8)ATG	1479	83	882/883	emb:CAA37105	Positive regulator AlkS of GPo1
<i>tnpA4</i>	363	9015–10108	56	GGCG(8)ATG	310	46	350/363	pir:B36919	Transposase of <i>A. tumefaciens</i> IS870 (frameshifts)
<i>tnpA5</i>	363	11229–10138	56	GGATTGG(8)ATG	309	45	345/363	pir:B36919	Transposase of <i>A. tumefaciens</i> IS870
<i>alkB</i>	402	11581–12789	48	GGAGA(8)ATG	761	90	398/401	sp:P12691	Alkane hydroxylase AlkB of GPo1
<i>alkF</i>	134	13072–13476	43	GGAGA(7)ATG	132	47	134/132	sp:P12692	Rubredoxin 1 (AlkF) of GPo1
<i>alkG</i>	175	13517–14044	46	GGTAA(5)ATG	249	63	173/173	sp:P00272	Rubredoxin 2 (AlkG) of GPo1
<i>alkH</i>	483	14085–15536	44	AGGA(11)ATG	799	80	483/483	sp:P12693	Aldehyde dehydrogenase AlkH of GPo1
<i>alkJ</i>	522	15575–17233	44	AGGA(7)ATG	978	85	551/558	sp:Q00593	Alcohol dehydrogenase AlkJ of GPo1
<i>alkK</i>	546	17318–18958	46	GGAG(7)ATG	908	80	546/546	sp:Q00594	Acyl-CoA synthetase AlkK of GPo1
<i>alkL</i>	230	19064–19756	41	GGAG(7)ATG	381	79	224/230	sp:Q00595	Outer-membrane protein AlkL of GPo1
<i>alkN'</i>	(292)	19898–20777	37	—	—	—	—	—	87 % DNA homology to <i>alkN</i> of GPo1 (frameshifts)
<i>orf16</i>	154	20854–21318	58	GAGAG(5)ATG	180	64	139/153	gi:1100232	Hypothetical protein in <i>Pseudomonas syringae</i> IS1240
<i>tnpA6</i>	305	21488–22401	58	GACG(6)ATG	516	71	342/343	gi:1100231	Transposase in <i>Pseudomonas syringae</i> IS1240 (frameshift)
<i>orf17</i>	205	22700–23312	56	—	180	46	203/454	emb:AAF10600	<i>Deinococcus radiodurans</i> oxidoreductase (fragment)
Opposite end of <i>P. putida</i> P1 EcoRI–SalI fragment									
<i>'nahK</i>	85	1–255	68	—	168	100	85/264	gb:AAD13220.1	4-Oxalocrotonate decarboxylase of <i>P. putida</i> (truncated)
<i>nahJ</i>	63	313–501	58	GGAGG(7)ATG	108	85	63/63	gb:AAD13221.1	4-Oxalocrotonate tautomerase of <i>P. putida</i>
<i>orf17</i>	125	564–928	63	—	171	67	122/140	gb:AAD13222.1	Hypothetical protein encoded by <i>P. putida nah</i> genes (frameshifts)

* emb, EMBL; gi, GenBank; pir, PIR protein database; sp, SWISS-PROT.

that the coding region should be longer or shorter and/or if no ribosome-binding site could be detected.

Although the *alkBFGHJKL* and *alkST* gene clusters are organized similarly in both strains (Fig. 2), an interesting difference between GPo1 and P1 is the relative position of the gene clusters: in P1, *alkST* is located upstream of

alkBFGHJKL. On average, the level of DNA sequence identity between the GPo1 and P1 *alk* genes is about 80 % for the *alkBFGHJKL* genes, and 92 % for the *alkST* genes, but only within the ORFs. The intergenic regions could not be aligned.

Most of the GPo1 and P1 Alk proteins show similar

Fig. 3. (a) Alignment of the GpO1 and P1 *alkB* promoter regions and the GpO1 *alkN* promoter region. The underlined sequences marked -35 and -10, and the arrow labelled +1, refer to the promoter and mRNA start site identified previously (Canosa *et al.*, 2000; Kok *et al.*, 1989b). The translation start sites of AlkB and AlkN (predicted) and the stop codon of the short AlkB' peptide encoded by the DNA segment upstream of the *alkN* coding region are also underlined. Box A indicates that this and DNA further upstream are homologous to an internal fragment of IS1384. Box B indicates the extended homology between the P1 *alkB* and GpO1 *alkN* promoter regions, compared to the GpO1 *alkB* promoter region. The horizontal arrows (C) mark the 20 bp inverted repeat proposed as AlkS binding site. Box D marks the end of the homology between the sequence upstream of the *alkN* coding region and the *alkB* promoter regions, about 80 bases within the *alkB* gene and 25 bases upstream of the *alkN* start codon. RBS, Ribosome-binding site. (b) Alignment of the GpO1 and P1 *alkS* promoter regions. The underlined sequences marked -35 and -10, and the arrows labelled +1, refer to the σ^S - and AlkS-dependent promoters identified previously (Canosa *et al.*, 1999, 2000). Box A indicates that this DNA and sequences further upstream are homologous to the end of IS1384. Box B marks the start of the homology between the GpO1 and P1 *alkS* promoter regions. The horizontal arrows (C) mark the 18 bp perfect inverted repeat proposed as AlkS-binding site. The start codon of *alkS* is underlined. (c) Sequence of inverted repeat primers used to construct pGEM-7Zf(+) IR-B, IR-S and IR-syn. The horizontal arrows mark the inverted repeats.

levels of sequence identity (Table 1). However, specific regions within the proteins, and the corresponding DNA sequences are not (well) conserved. The C-terminal 50 bp of the GpO1 and P1 *alkB* genes could not be aligned. In the case of the GpO1 *alkB* gene, this segment was shown previously to be non-essential to mono-oxygenase function; it could be replaced by the *lacZ* gene without losing AlkB activity (van Beilen *et al.*, 1992b).

The GpO1 rubredoxins AlkF and AlkG are unusual compared to rubredoxins isolated from anaerobic organisms. AlkF consists of an N-terminal rubredoxin-like domain and a C-terminal 80 bp extension that has no homology with database sequences. Similarly, AlkG consists of an N-terminal and a C-terminal rubredoxin-like domain, linked by a 70 bp sequence that has no homology with database sequences. Only the C-terminal rubredoxin-like domain of AlkG is essential for the alkane hydroxylase system (Kok *et al.*, 1989a). In P1, the C-terminal extension of AlkF and the AlkG linker

are not well conserved. However, the N-terminal rubredoxin-like domains of AlkF and AlkG (AlkF1 and AlkG1), which cannot replace the C-terminal domain of AlkG (AlkG2) in alkane oxidation (Kok *et al.*, 1989a), are almost as highly conserved as AlkG2. This suggests that these domains do play a role in alkane oxidation, perhaps under conditions that are different from those used in the complementation experiments.

In P1, a DNA segment which shows 87% sequence identity to the first half of the GpO1 *alkN* gene is truncated by the insertion sequence IS*Ppu4.1* (see below). The gene remnant contains three frameshift mutations relative to the GpO1 *alkN* sequence and an insert composed of a perfect sevenfold AAATGT repeat, while the ATG-start codon has changed to ATC. The *alkN* gene remnant is located 135 bp downstream of *alkL*, which is only slightly more than the distance between the coding regions of the *alkBFGHJKL* operon, suggesting that in P1 *alkN* was part of the *alkBFGHJKL* operon. The gene organization in GpO1 suggests that the

IS element *ISPpu1* has interrupted a previously existing *alkBFGHJKLN* operon, leaving the *alkN* gene without a promoter. In a later evolutionary step, the *alkN* gene acquired an *alkB* promoter sequence, which also includes 76 bases of the *alkB* coding sequence. Interestingly, this sequence has significantly higher sequence identity to the P1 *alkB* promoter (83 % identity) than to the GPo1 *alkB* promoter (56 % identity), suggesting that the source of this DNA may be a different *alk* operon, more closely related to the P1 *alk* genes than to the GPo1 *alk* genes.

Analysis of the *alkB*, *alkS* and *alkN* promoter regions

Apart from the *alk* gene-coding regions, only the *alkB* and *alkS* promoter regions show clear homology. An alignment of the *alkB* promoter regions of GPo1 and P1 (Fig. 3a) shows that 154 bases upstream of the *alkB* start codon are conserved with 56 % identity. The homology ends with a conserved 20 bp imperfect inverted repeat (IR-B), directly upstream of the -35 region of the *alkB* promoter (*alkBp*) characterized previously (Canosa *et al.*, 2000; Kok *et al.*, 1989b). As discussed above, the 230 bp DNA segment upstream of the GPo1 *alkN* gene is more closely related to the P1 *alkB* promoter than to the GPo1 *alkB* promoter. In addition, the homology between the P1 *alkB* and GPo1 *alkN* promoter regions extends 15 bases upstream of the inverted repeat (Fig. 3a).

The 103 bases upstream of the *alkS* ATG start codon in GPo1 and P1 are well conserved at 93 % sequence identity, including by 3 bases the σ^S -dependent *alkS* promoter (*alkSp1*) (Canosa *et al.*, 1999). Here, an 18 bp perfect inverted repeat (IR-S), similar to IR-B, is positioned directly upstream of the -35 region of *alkSp2*, the AlkS-dependent promoter located downstream of *alkSp1* (Canosa *et al.*, 2000). IR-S overlaps with the -10 region of *alkSp1* (Fig. 3b). Based on their positions relative to the promoters, IR-B and IR-S might be control site locations that bind AlkS (Canosa *et al.*, 1999, 2000; Collado-Vides *et al.*, 1991; Kok *et al.*, 1989b). If the inverted repeats bind AlkS, the presence of these sequences *in trans* on a high-copy-number plasmid should influence expression of a reporter gene controlled by *alkBp*. To test this hypothesis, we used *E. coli* JM101 containing a single copy of a cassette composed of *xylE* under control of *alkBp*, and *alkS** (*NdeI* site removed), integrated in the chromosome (Panke *et al.*, 1999a). Three versions of the inverted repeats (Fig. 3c), the first two identical to IR-B and IR-S, and a third in which IR-B and IR-S are combined to a perfect 20 bp inverted repeat (IR-syn), were cloned in the high-copy-number vector pGEM-7Zf(+) and transferred to the reporter strain. As control plasmids pGEM-7Zf(+) and pBG201, a pGEM-7Zf(+)-derived vector containing the *alkB* gene and the *alkBp* promoter, were used. The recombinants were cultured in minimal medium with glucose as carbon source to repress the *lac* promoter of pGEM-7Zf(+), as transcriptional activity of this promoter might replace AlkS from the IR sequences. The recombinant strains were assayed for XylE activity 2 h after

induction with the non-metabolizable inducer dicyclopropylketone (Grund *et al.*, 1975).

The presence of pGEM-7Zf(+) did not affect XylE activity. Recombinants containing IR-B, IR-S or IR-syn, however, show a strongly reduced XylE activity (to 7–8 % for IR-B, 5–6 % for IR-S, and 1 % for IR-syn), which suggests that AlkS indeed binds to the inverted repeats. Interestingly, pBG201 reduces the XylE activity by only 30–40 %, perhaps because transcriptional activity of *alkBp* in this construct replaces AlkS from its binding site. The competition experiments confirm that AlkS binds to the *alkS* as well as the *alkB* promoter (Canosa *et al.*, 2000), at positions that are appropriate for control site locations. The fact that the inverted repeats are well conserved between GPo1 and P1 further supports this notion.

As discussed above, the homology between the GPo1 and P1 *alk* promoter regions ends almost exactly at the inverted repeat of *alkBp* or the -35 region of *alkSp1*. In GPo1, fragments of insertion sequences related to IS1384 begin only 15 and 3 bp from the conserved promoter regions (Fig. 3). This indicates that DNA upstream of the conserved regions is not involved in regulation.

DNA flanking the GPo1 and P1 *alk* gene clusters

Sequence analysis of DNA segments flanking the GPo1 *alk* genes reveals a mosaic structure of (incomplete) insertion sequences (listed in Table 2) and ORFs (Fig. 2, Table 1). The 5.5 kb DNA segment (55 mol % G + C) directly upstream of the GPo1 *alkB* gene, the 1.8 kb segment between *alkL* and *alkN*, and shorter segments up- and downstream of *alkST* are entirely composed of insertion sequences or their remnants, and encode proteins with clear homology to transposases or other proteins involved in DNA transposition events. The remaining DNA segments encode proteins that have no homologues in the databases (ORFs 1, 2, 8, 9 and 15), or only have distant relatives that do not allow the assignment of a function (ORFs 10, 12 and 14). The only exceptions are a complete and an incomplete *recA* gene upstream of *alkS*, the latter showing 90 % DNA sequence identity with *P. aeruginosa* PAO1 *recA*.

DNA flanking the P1 *alk* genes is almost exclusively composed of insertion sequences (6.2 out of 7.4 kb) with a G + C content close to 56 mol % (P1 *alk* genes: 45 %). Only the far end of the *EcoRI*–*SalI* fragment containing P1 *alkLN'* and *ISPpu4.1* is not related to insertion sequence elements, but is highly homologous to genes in the *P. putida* G7 naphthalene lower pathway (Grimm & Harwood, 1999).

Based on three criteria (encoded transposase, terminal inverted repeats and flanking direct repeats), six full-length insertion sequences were assigned registration numbers (http://www-is.biotoul.fr/is/is_about.html, 18 October 2000) (Table 2). *ISPpu1* is located immediately downstream of *alkL* in GPo1, and is related to the *Agrobacterium tumefaciens* and *Agrobacterium vitis* IS870 copies (Fournier *et al.*, 1993). The IS element

Table 2. Complete and partial insertion sequences flanking the *alk* gene clusters

IS element*	Location	Related to:†	Family‡	Left and right inverted repeats§
<i>P. putida</i> (oleovorans) GPo1				
Fragment (right end) IS53K'	2440–4351R 6525–7214L	IS1131 (1900/2700)	IS66	R : GATA.GTAAGCGCTCCA–TAAACCCACCTAACCAAGATGGGCAGCGCACCTAGC (R : ATAC.GTAAGC–CTCCGGTGAAGGCCACAGGTCAGGCAGCTTGAGCGTTGACGGGT IS 66) L : CTTG.TGTTGCCGCTGACCGAAAAGTACCCAGTAGAGGCTGTTCTGCCGACTGC (L : GCAA.TGTTGCCGCTGACCGAAAAGTACCCAGTAGAGGCTGTTCTGCCGACTGA IS 53)
Fragment ISPpu1	7266–7353 R15661–16824L	IS1384/IS1051 IS870	IS5 IS630	– (internal) L : CTAA.TGTCCTGCCTCCGAAATAAGTTTCTAGCCTGAGCTATGCTTCTGTGGTCA R : TTAG.TGTGCCGCTCACAAATAATGCTCCTCATCCCTGTGAGACCATTGCTCTCA L : GGAC.ACCTGTTCACGATCATTCGGACAGGCAGTCGGAGCCGACGACGACAGAG (L : GGCC.ACCCATCCACGCGCATCCGGCCACTGATTCCACGGCCATCCGGCCACTC IS 1491) R : CTAG.GGAAGGTCTGAAGTAGGCCGCTATTCTGGCCACTTCCGGCCTTCGCCG (R : TTAAG.GGAAGGTCTGAAGTAGACTTGATTCTGGCCACTTCCGGCCCTCCCCA IS 1384)
<i>P. putida</i> P1				
Fragment ISPpu4.2	1–875 L880–2521R	IS401 IS1240/IS1141	IS3	– (internal) L : CGTA.CCCCGG–TGTCAGGCTAGTTGTCGCTGAGCTGATTATCAAAACGATA R : GGTC.CGGCGTTTTCGAAGG–TAGTTGTCGCGTCAGCTGTGTTTTATGCGGCTA L : AAAACGCCGACCCCGGCACCTTGTGAGGCTACGCTTTCACAGGGTGTTTTA (L : TTGAG.ATGGTCACCCCGG–ACCTGTGAGCGGTAAGCTG–ACAGGGTTTCTT IS 1328)
ISPpu5	12522–3860R	IS1328	IS110	R : ATCT.ATGGTCACCCCGGTTTTTGAATACTGATTAGCGGGTGAAGTGGTTGGCT L : TATA.CTGTGCATCCCTCCTTTTCGGCGCTCTGTAGCATGGCTCGACCCAGCGGC R : CGGA.CTGTGTTGTTAGTTCAATAACTTATCTTGTATTGCGGTGAGCTTTCGCGCA L : TATA.CTGTGCATCCAATGTCATGGA–TTGGTTATACGCATGGCCCGACCGACTTC R : TGGC.CTGAGCGGTTAAT–TCATTAACCTATTTTTATCAGCTCAGCTTCGCTTT L : TTTC.CCCCGG–TGTCAGGCTAGTTGTCGCGTGAAGTATGATTATCAAAACGATAG R : TACG.CGGCGTTTTCGAAGG–TAGTTGTCGCGTCAGCTGTGTTTTATGCGGCTAT L : ATTG.TGAACCGCCCGSSTATCGCGGAGGCTCCACCTTTTAGAAACTGGGGCC (L : CGAC.TGAACCGCCCGGTTTTCGCGGAGGCTCCACCTTTAGAGAACTGAGGCCIS 401) CGTA – ISPpu4.2 – <i>alkS</i> – <i>alkBFGHJKLM</i> – ISPpu4.1 – CGTA TATA – ISPpu2 – TCCGCTGGC – ISPpu3 – TATA
Tn <i>Ppu-alk1</i>	880–22442	ISPpu4		
ISPpu2-ISPpu3	L8983–11262L	IS870		

* Only complete insertion sequences are named.

† The most closely related IS elements in the GenBank and EMBL databases are listed. If the most closely IS element is not in the IS database (http://www-is.biotoul.fr/is/is_about.html, 18 October 2000), the closest relative in the IS database is listed as well.

‡ The IS elements were assigned to the families proposed by Mahillon & Chandler (1998).

§ The left (L) and right (R) ends of the insertion sequences are shown, with the most likely end indicated by the dot. Terminal inverted repeats are underlined. If only one end of an insertion sequence was found, the same end of the most closely related IS element is shown for comparison, with the terminal inverted repeat underlined. The box at the left IR of ISPpu5 marks bases that belong to the right end of ISPpu4, shown directly above (reverse complement).

includes terminal inverted repeats that are quite similar to the IS870 inverted repeats, and is bordered by a 4 bp direct repeat (CTAA). ISPpu2 and ISPpu3 are located between *alkS* and *alkB*. They are quite similar to each other (70% DNA sequence identity) and more distantly related to GPo1 ISPpu1 and IS870. In particular, the right terminal inverted repeats (adjacent to each other) of these two IS elements are not well conserved, whilst flanking direct repeats were not present. The two left (outside) inverted repeats are more conserved and ISPpu2 and ISPpu3 together are flanked by the direct repeat TATA, which suggests that they transposed as a single IS element (Table 2). ISPpu5 is located downstream of *alkT* and is related to the *Yersinia enterocolitica* IS1328 (64% DNA sequence identity) (Rakin & Heesemann, 1995). Alignment of the terminal inverted repeats of ISPpu5 with those of IS1328 shows that the last 6 bases of the left inverted repeat are replaced by the left end of ISPpu4.2, suggesting that ISPpu4.2 inserted itself in the left end of ISPpu5, which removed the direct repeat in the process.

ISPpu4.1 and ISPpu4.2 are almost identical IS elements that show 62% DNA sequence identity to IS1240 (Hanekamp *et al.*, 1997) and are located on either side of the P1 *alk* genes. As discussed above, ISPpu4.1 has

truncated the P1 *alkN* sequence, while ISPpu4.2 is located downstream of *alkT*. Analysis of the left and right ends of ISPpu4.1, ISPpu4.2 and IS1240 shows that the homology between the left and right ends extends beyond the imperfect inverted repeats (Table 2), which suggests that these extensions are part of the insertion sequence. ISPpu4.1 and ISPpu4.2 individually are not flanked by direct repeats. However, the combination of both insertion sequences (including the *alk* genes) is flanked by a 4 bp direct repeat (CGTA). In addition, further analysis showed that the entire cassette has interrupted an IS element related to IS401. Although the TnpA ORFs in both ISPpu4 copies are disrupted by frameshift mutations, including a 23 bp deletion in ISPpu4.2, which most likely has inactivated both insertion sequences, the above features are evidence for a class I transposon (Kleckner, 1981), which we have named Tn*Ppu-alk1*.

Analysis of the GPo1 sequence did not reveal a similar structure. However, a 638 bp DNA segment starting only 6 bases upstream of the *alkS* promoter has 77% sequence identity to the right end of IS1384 (accession no. AF052751), whilst an 88 bp region ending only 15 bp from the *alkB* promoter shows 86% identity to an internal fragment of the same insertion sequence (the

ends of these DNA segments are marked in Fig. 3). These fragments do not overlap, which makes it impossible to ascertain whether the IS1384-related sequences were identical. Nevertheless, the arrangement suggests that the GPo1 *alk* genes were part of a class I transposon as well, before large segments of the flanking insertion sequences were lost. This analysis also demonstrates that class I transposons are not very stable; they are formed by the fortuitous insertion of the same IS element on both sides of a given set of genes, and become inactive again when one of the constituting IS elements is destroyed, for example by the insertion of other IS elements, deletions or frameshift mutations. Eventually, mosaics of (incomplete) insertion sequences flank the catabolic genes. Such sections are, as they no longer contain functional genes or genes necessary for the survival of the host organism, selected targets for new transposition events, which destroy previously inserted IS elements, but also provide new opportunities for the mobilization of catabolic genes, such as the *alk* gene clusters. Analysis of the *alk* genes of two *P. putida* strains provides a clear example of this evolutionary process in action and at the same time has helped us to identify new functions and aspects of alkane degradation.

ACKNOWLEDGEMENTS

This research was supported by the Swiss National Science Foundation through the Swiss Priority Program in Biotechnology, grant no. 5002-037023.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Bachmann, B. (1987). Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 1190–1219. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.
- Baptist, J. N., Gholson, R. K. & Coon, M. J. (1963). Hydrocarbon oxidation by a bacterial enzyme system: I. Products of octane oxidation. *Biochim Biophys Acta* **69**, 40–47.
- van Beilen, J. B. (1994). *Alkane oxidation by Pseudomonas oleovorans: genes and proteins*. PhD thesis, University of Groningen.
- van Beilen, J. B., Eggink, G., Enequist, H., Bos, R. & Witholt, B. (1992a). DNA sequence determination and functional characterization of the OCT-plasmid-encoded *alkJ* genes of *Pseudomonas oleovorans*. *Mol Microbiol* **6**, 3121–3136.
- van Beilen, J. B., Penninga, D. & Witholt, B. (1992b). Topology of the membrane-bound alkane hydroxylase of *Pseudomonas oleovorans*. *J Biol Chem* **267**, 9194–9201.
- van Beilen, J. B., Wubbolts, M. G. & Witholt, B. (1994). Genetics of alkane oxidation by *Pseudomonas oleovorans*. *Biodegradation* **5**, 161–174.
- Benson, S., Fennewald, M., Shapiro, J. & Huettner, C. (1977). Fractionation of inducible alkane hydroxylase activity in *Pseudomonas putida* and characterization of hydroxylase-negative plasmid mutations. *J Bacteriol* **132**, 614–621.
- Bustamante, C., Gurrieri, S. & Smith, S. B. (1993). Towards a molecular description of pulsed-field gel electrophoresis. *Trends Biotechnol* **11**, 23–30.
- Canosa, I., Yuste, L. & Rojo, F. (1999). Role of the alternative sigma factor σ^s in the expression of the AlkS regulator of the *Pseudomonas oleovorans* alkane degradation pathway. *J Bacteriol* **181**, 1748–1754.
- Canosa, I., Sanchez-Romero, J. M., Yuste, L. & Rojo, F. (2000). A positive feedback mechanism controls expression of AlkS, the transcriptional regulator of the *Pseudomonas oleovorans* alkane degradation pathway. *Mol Microbiol* **35**, 791–799.
- Chakrabarty, A. M., Chou, G. & Gunsalus, I. C. (1973). Genetic regulation of octane dissimilation plasmid in *Pseudomonas*. *Proc Natl Acad Sci USA* **70**, 1137–1140.
- Collado-Vides, J., Magasanik, B. & Gralla, J. D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol Rev* **55**, 371–394.
- Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127.
- Eggink, G., Lageveen, R. G., Altenburg, B. & Witholt, B. (1987a). Controlled and functional expression of *Pseudomonas oleovorans* alkane utilizing system in *Pseudomonas putida* and *Escherichia coli*. *J Biol Chem* **262**, 17712–17718.
- Eggink, G., van Lelyveld, P. H., Arnberg, A., Arfman, N., Witteveen, C. & Witholt, B. (1987b). Structure of the *Pseudomonas putida* *alkBAC* operon. Identification of transcription and translation products. *J Biol Chem* **262**, 6400–6406.
- Eggink, G., Engel, H., Vriend, G., Terpstra, P. & Witholt, B. (1990). Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J Mol Biol* **212**, 135–142.
- Fennewald, M. & Shapiro, J. (1977). Regulatory mutations of the *Pseudomonas* plasmid *alk* regulon. *J Bacteriol* **132**, 622–627.
- Fennewald, M., Prevatt, W., Meyer, R. & Shapiro, J. (1978). Isolation of Inc P-2 plasmid DNA from *Pseudomonas aeruginosa*. *Plasmid* **1**, 164–173.
- Fennewald, M., Benson, S., Oppici, M. & Shapiro, J. (1979). Insertion element analysis and mapping of the *Pseudomonas* plasmid *alk* regulon. *J Bacteriol* **139**, 940–952.
- Fournier, P., Paulus, F. & Otten, L. (1993). IS870 requires a 5'-CTAG-3' target sequence to generate the stop-codon for its large ORF1. *J Bacteriol* **175**, 3151–3160.
- Grimm, A. C. & Harwood, C. S. (1999). NahY, a catabolic plasmid-encoded receptor required for chemotaxis of *Pseudomonas putida* to the aromatic hydrocarbon naphthalene. *J Bacteriol* **181**, 3310–3316.
- Grund, A., Shapiro, J., Fennewald, M., Bacha, P., Leahy, J., Markbreiter, K., Nieder, M. & Toepfer, M. (1975). Regulation of alkane oxidation in *Pseudomonas putida*. *J Bacteriol* **123**, 546–556.
- Hanekamp, T., Kobayashi, D., Hayes, S. & Stayton, M. M. (1997). Avirulence gene D of *Pseudomonas syringae* pv. *tomato* may have undergone horizontal gene transfer. *FEBS Lett* **415**, 40–44.
- Hauben, L., Vauterin, L., Swings, J. & Moore, E. R. B. (1997). Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int J Syst Bacteriol* **47**, 328–335.
- Henikoff, S. (1984). Unidirectional digestion with exonuclease-III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.
- Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (1990).

PCR Protocols. A Guide to Methods and Applications. San Diego: Academic Press.

Kado, C. I. & Liu, S.-T. (1981). Rapid procedure for detection and isolation of large and small plasmids. *J Bacteriol* **145**, 1365–1373.

Kleckner, N. (1981). Transposable elements in prokaryotes. *Annu Rev Genet* **15**, 341–404.

Kok, M., Oldenhuis, R., van der Linden, M. P. G., Meulenberg, C. H. C., Kingma, J. & Witholt, B. (1989a). The *Pseudomonas oleovorans* *alkBAC* operon encodes two structurally related rubredoxins and an aldehyde dehydrogenase. *J Biol Chem* **264**, 5442–5451.

Kok, M., Oldenhuis, R., van der Linden, M. P. G., Raatjes, P., Kingma, J., van Lelyveld, P. H. & Witholt, B. (1989b). The *Pseudomonas oleovorans* alkane hydroxylase gene. Sequence and expression. *J Biol Chem* **264**, 5435–5441.

Lageveen, R. G., Huisman, G. W., Preusting, H., Ketelaar, P. E. F., Eggink, G. & Witholt, B. (1988). Formation of polyester by *Pseudomonas oleovorans*: the effect of substrate on the formation and composition of poly-(R)-3-hydroxyalkanoates and poly-(R)-3-hydroxyalkenoates. *Appl Environ Microbiol* **54**, 2924–2932.

Lee, M. & Chandler, A. C. (1941). A study of the nature, growth and control of bacteria in cutting compounds. *J Bacteriol* **41**, 373–386.

Luria, S. E., Adam, J. N. & Teng, R. C. (1960). Transduction of lactose utilizing ability among strains of *Escherichia coli* and *Shigella dysenteriae* and the properties of the transducing phage particle. *Virology* **12**, 348–390.

Mahillon, J. & Chandler, M. (1998). Insertion sequences. *Microbiol Mol Biol Rev* **62**, 725–774.

Nozaki, M. (1970). Metapyrocatechase. *Methods Enzymol* **17**, 522–525.

Owen, D. J. (1986). Molecular cloning and characterization of sequences from the regulatory cluster of the *Pseudomonas* plasmid *alk* system. *Mol Gen Genet* **203**, 64–72.

Panke, S., de Lorenzo, V., Kaiser, A., Witholt, B. & Wubbolts, M. G. (1999a). Engineering of a stable whole-cell biocatalyst

capable of (S)-styrene oxide formation for continuous two-liquid phase applications. *Appl Environ Microbiol* **65**, 5619–5623.

Panke, S., Meyer, A., Huber, C. M., Witholt, B. & Wubbolts, M. G. (1999b). An alkane-responsive expression system for the production of fine chemicals. *Appl Environ Microbiol* **65**, 2324–2332.

Rakin, A. & Heesemann, J. (1995). Virulence-associated *fyuA/irp2* gene cluster of *Yersinia enterocolitica* biotype 1B carries a novel insertion sequence IS1328. *FEMS Microbiol Lett* **129**, 287–292.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schwartz, R. D. (1973). Octene epoxidation by a cold-stable alkane-oxidizing isolate of *Pseudomonas oleovorans*. *Appl Microbiol* **25**, 574–577.

Schwartz, R. D. & McCoy, C. J. (1973). *Pseudomonas oleovorans* hydroxylation-epoxidation system: additional strain improvements. *Appl Microbiol* **26**, 217–218.

Smits, T. H. M., Röthlisberger, M., Witholt, B. & van Beilen, J. B. (1999). Molecular screening for alkane hydroxylase genes in Gram-negative and Gram-positive strains. *Environ Microbiol* **1**, 307–318.

Stanier, R. Y., Palleroni, N. J. & Doudoroff, M. (1966). The aerobic pseudomonads: a taxonomic study. *J Gen Microbiol* **43**, 159–271.

Taguchi, K., Fukutomi, H., Kuroda, A., Kato, J. & Ohtake, H. (1997). Genetic identification of chemotactic transducers for amino acids in *Pseudomonas aeruginosa*. *Microbiology* **143**, 3223–3229.

Tan, H.-M. (1999). Bacterial catabolic transposons. *Appl Microbiol Biotechnol* **51**, 1–12.

Witholt, B., de Smet, M. J., Kingma, J., van Beilen, J. B., Kok, M., Lageveen, R. G. & Eggink, G. (1990). Bioconversions of aliphatic compounds by *Pseudomonas oleovorans* in multiphase bioreactors: background and economic potential. *Trends Biotechnol* **8**, 46–52.

Received 25 October 2000; revised 2 February 2001; accepted 13 February 2001.