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 Rothlisberger and Bernard Witholt

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

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 ISPpu4, 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; Fennwwald & 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-
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) (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 prototrophic 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 DNA methods.

DNA methods. Restriction enzymes, T4-DNA ligase and dideoxynucleotides were obtained from Roche Diagnostics. The Roche Diagnostics DIG kit was used for Southern blots, using CSPD [disodium 3-(4-methyl-oxyspiro {1,2-dioxetane-

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 Nhel–HindIII fragment. A DIG-labelled GPo1 alkST fragment was used as probe to clone the P1 alkST genes as an additional 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-alkSTBam (GCTCGCGGTGGATCTCTCAGGTC) and P1-alkSTSac (GGCGCGAGCTCGCGGT-AGATCAGGTC) 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 (GGTCTAGAGTGCCGCTGGCTGCGTGCGTGACG) and P1-STBamSac (GGCGCGAGCTCGCGGTAGATCAGGTC) 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 ISPpuw flank the P1 alk genes, and cause a crossover reaction in the PCR.) In addition, a 3.1 kb EcoRI fragment containing ISPpuw-2, and a 7.5 kb EcoRI–SalI fragment containing the P1 alkL, gene, ISPpuw-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 16L1525, 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 EcoRI and HindIII sticky ends (see Fig. 3c) and ligated to plasmid pGEM-7Zf (+) (Promega) digested with EcoRI and HindIII. 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 OD600 of 0.5 the cultures were induced with 0.02% dicyclopentylketone (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 nalkKJ 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 TnPpu-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.

Taxonomic 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) and the P. putida biotype A type strain DSM 291T (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 1045T, 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 EcoRI fragment containing the alkBFGHJKL genes (9.1 kb sequenced), and a 16.9 kb EcoRI fragment containing the alkST genes (4.7 kb sequenced) (Eggink et al., 1987a, 1990; Kok et al., 1989a, b; Panke et al., 1990).
1999b; van Beilen et al., 1992a). We completed the sequence of both EcoRI fragments (Fig. 2). The region downstream of alkT in the 18 kb EcoRI fragment was found to encode a putative transposase of which a few N-terminal codons were missing due to the EcoRI cloning site. A 30 kb Nhel–HindIII fragment which was cloned to complete the putative transposase (tnpAI) sequence was found to overlap with the sequence downstream of alkT on the 16-9 kb EcoRI fragment: a 651 bp EcoRI fragment sufficed to join the 16-9 and 18 kb EcoRI 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, respectively, for naphthalene (Grimm & Harwood, 1999) or amino acids (Taguchi et al., 1997). Like the alk genes, the chemo-receptor 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 chemo-receptor 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 EcoRI–HindIII 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...
that the coding region should be longer or shorter. Similarity to IS ORFs.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.

Most of the GPO1 and P1 Alk proteins show similar identity between the GPo1 and P1 gene clusters, although the GPo1 and P1 Alk proteins could not be aligned. The differences between the GPo1 and P1 gene clusters are mainly due to the presence of ORFs in the GPo1 gene clusters. The intergenic regions could not be aligned.
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 ISpPu4 (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 alkBFGHJL operon. The gene organization in GPo1 suggests that the

---

**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− 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.
IS element IS\textsubscript{Ppu1} has interrupted a previously existing alk\textsubscript{BFGHJKLN} operon, leaving the alk\textsubscript{N} gene without a promoter. In a later evolutionary step, the alk\textsubscript{N} gene acquired an alk\textsubscript{B} promoter sequence, which also includes 76 bases of the alk\textsubscript{B} coding sequence. Interestingly, this sequence has significantly higher sequence identity to the P1 alk\textsubscript{B} promoter (83\% identity) than to the GPo1 alk\textsubscript{B} 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 alk\textsubscript{B}, alk\textsubscript{S} and alk\textsubscript{N} promoter regions**

Apart from the alk gene-coding regions, only the alk\textsubscript{B} and alk\textsubscript{S} promoter regions show clear homology. An alignment of the alk\textsubscript{B} promoter regions of GPo1 and P1 (Fig. 3a) shows that 154 bases upstream of the alk\textsubscript{B} 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 alk\textsubscript{B} promoter (alk\textsubscript{Bp}) characterized previously (Canosa \textit{et al.}, 2000; Kok \textit{et al.}, 1989b). As discussed above, the 230 bp DNA segment upstream of the GPo1 alk\textsubscript{N} gene is more closely related to the P1 alk\textsubscript{B} promoter than to the GPo1 alk\textsubscript{B} promoter. In addition, the homology between the P1 alk\textsubscript{B} and GPo1 alk\textsubscript{N} promoter regions extends 15 bases upstream of the inverted repeat (Fig. 3a).

The 103 bases upstream of the alk\textsubscript{S} ATG start codon in GPo1 and P1 are well conserved at 93\% sequence identity, including by 3 bases the \(\sigma^\text{alkS}\)-dependent alk\textsubscript{S} promoter (alk\textsubscript{Sp1}) (Canosa \textit{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 alk\textsubscript{Sp2}, the alk\textsubscript{S}-dependent promoter located downstream of alk\textsubscript{Sp1} (Canosa \textit{et al.}, 2000). IR-S overlaps with the −10 region of alk\textsubscript{Sp1} (Fig. 3b). Based on their positions relative to the promoters, IR-B and IR-S might be control site locations that bind AlkS (Canosa \textit{et al.}, 1999, 2000; Collado-Vides \textit{et al.}, 1991; Kok \textit{et al.}, 1989b). If the inverted repeats bind AlkS, the presence of these sequences \textit{in trans} on a high-copy-number plasmid should influence expression of a reporter gene controlled by alk\textsubscript{Bp}. To test this hypothesis, we used E. \textit{coli} JM101 containing a single copy of a cassette composed of xyl\textsubscript{E} under control of alk\textsubscript{Bp}, and alk\textsubscript{S}\textsuperscript{N} (Ndel site removed), integrated in the chromosome (Panke \textit{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 in 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 alk\textsubscript{B} gene and the alk\textsubscript{Bp} promoter, were used. The recombinants were cultured in minimal medium with glucose as carbon source to repress the lac promoter of pGEM-7Zf(+) and assayed for Xyl\textsubscript{E} activity 2 h after induction with the non-metabolizable inducer dicyclopropylketone (Grund \textit{et al.}, 1975).

The presence of pGEM-7Zf(+) did not affect Xyl\textsubscript{E} activity. Recombinants containing IR-B, IR-S or IR-syn, however, show a strongly reduced Xyl\textsubscript{E} activity (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 Xyl\textsubscript{E} activity by only 30–40\%, perhaps because transcriptional activity of alk\textsubscript{Bp} in this construct replaces AlkS from its binding site. The competition experiments confirm that AlkS binds to the alk\textsubscript{S} as well as the alk\textsubscript{B} promoter (Canosa \textit{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 alk\textsubscript{Bp} or the −35 region of alk\textsubscript{Sp1}. In GPo1, fragments of insertion sequences related to IS\textsubscript{1384} 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 alk\textsubscript{B} gene, the 1·8 kb segment between alk\textsubscript{L} and alk\textsubscript{N}, and shorter segments up- and downstream of alk\textsubscript{ST} 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 \(\text{recA}\) gene upstream of alk\textsubscript{S}, the latter showing 90\% DNA sequence identity with \(P. \text{aeruginosa}\) PA\textsubscript{O}1 \(\text{recA}\).

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 alk\textsubscript{LN}, and IS\textsubscript{Ppu1}, and IS\textsubscript{Ppu4.1} is not related to insertion sequence elements, but is highly homologous to genes in the \(P. \text{putida}\) 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). IS\textsubscript{Ppu1} is located immediately downstream of alk\textsubscript{L} in GPo1, and is related to the \(\text{Agrobacterium tumefaciens}\) and \(\text{Agrobacterium vitis}\) IS\textsubscript{870} copies (Fournier \textit{et al.}, 1993). The IS element
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 in-
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


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