JOURNAL OF BACTERIOLOGY, Oct. 1988, p. 4924–4930
0021-9193/88/104924-07$02.00/0

Vol. 170, No. 10

Cloning and Sequencing of *Pseudomonas* Genes Encoding Vanillate Demethylase

FRANÇOISE BRUNEL* AND JOHN DAVISON

Unit of Molecular Biology, International Institute of Cellular and Molecular Pathology, 75 Avenue Hippocrate, 1200 Brussels, Belgium

Received 4 April 1988/Accepted 1 July 1988

A 2,598-base-pair (bp) *Sal*I-*Hin*cII DNA fragment has been cloned which codes for vanillate demethylase, the enzyme responsible for the demethylation of vanillate (3-methoxy-4-hydroxybenzoate) to protocatechuate (3,4-dihydroxybenzoate). Complementation and insertional inactivation experiments have shown that this fragment carries two genes (*vanA* and *vanB*) which are predominantly cotranscribed from a promoter upstream of *vanA*. Nucleotide sequencing of the *Sal*I-*Hin*cII fragment confirmed the genetic data: two open reading frames of 987 and 942 bp were present in the transcribed orientation. These had a very high G+C content in the third base of each codon, which is characteristic of *Pseudomonas* chromosomal genes. Expression of the genes in *Escherichia coli* with the T7 RNA polymerase-promoter system gave rise to two polypeptides of 36 and 33 kilodaltons which could be identified by deletion analysis as the products of *vanA* and *vanB*, respectively. A search of the protein sequence data bank indicated that the *vanB* gene product was related to the ferredoxin family.

Strains of *Pseudomonas* spp. can be quite readily isolated that are capable of degrading aromatic hydrocarbons despite the intrinsic stability of these compounds. With rare exceptions, they do not open the benzene ring unless two hydroxyl groups have been introduced in *cis* into the benzene nucleus (8). Thus, the reactions which prepare the ring for fission tend to be specialized. For example, dealkylation of phenolic ethers is effected by monooxygenases. In these two- or three-component enzyme systems, the electron flow from NADH to the oxygenase is mediated via short electron transport chains (8). The oxygenases themselves fall into two groups: iron sulfur proteins, as in 4-methoxy-benzoate demethylase (2), and cytochrome P450-like enzymes, as in the demethylation of guiaicol (9) and isovanillate (6). It has also been proposed that two proteins may be involved in the demethylation of vanillate (3-methoxy-4-hydroxybenzoate) to protocatechuate (3,4-dihydroxybenzoate) (Fig. 1). Yet purification of the relevant enzyme components has not been successful (5, 24), and their properties and mode of action remain to be characterized.

Vanillate is a key intermediate in the biodegradation of lignin. In chlorinated form, it constitutes one of the major waste products of the paper and pulp industry. To gain further knowledge on the proteins involved in vanillate demethylation, we have cloned the corresponding genes from a *Pseudomonas* sp. This paper describes their isolation and characterization by complementation tests, insertional inactivation, nucleotide sequencing, and expression in *Escherichia coli*.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* S17-1 (*pro hsdR hsdM*$^+$) contains a modified chromosomally integrated RP4 plasmid which retains its ability to mobilize RSF1010-based vectors, such as pJRD215, to other hosts (26). *Pseudomonas* sp. strain ATCC 19151 was isolated for its ability to utilize sodium dodecyl sulfate (SDS) as a carbon source (15). It is also able to utilize ferulate and vanillate as sole carbon sources. Plasmid vectors pJRD203 and pJRD215 have been described previously (10, 11). Vector pJRD205 is an improved version of pJRD204 (10) containing the *Sal*I-*Xho*I segment of restriction site bank vector pJRD158 (11) inserted into the *Sal*I site of pJRD203. Interposon pHP45Ω was used to interrupt gene activity and transcription on the cloned DNA (22).

Media and growth conditions. All *E. coli* and *Pseudomonas* strains were routinely grown in L broth or on M9 minimal medium at 37°C. When SDS, vanillate, or other phenolic compounds were used as a carbon source, they were added to the growth medium at a final concentration of 0.2%. Kanamycin and streptomycin were added to the selective medium at final concentrations of 300 μg/ml and 1 mg/ml, respectively, for the *Pseudomonas* sp.

Enzymes and reagents. All restriction enzymes, T4 DNA ligase, T4 DNA polymerase, and *E. coli* Klenow fragment were obtained from Boehringer Mannheim or New England BioLabs and used according to the manufacturer's instructions. The DNA packaging kit and [^{32}P]dATP (specific activity, >3,000 Ci/mol) were purchased from Amersham Ltd. Phenolic compounds were obtained from Janssen Pharmaceutica, hydrazine was from Kodak, piperidine was from Merck, diethylpyrocarbonate was from Sigma, and dimethyl sulfate was from Aldrich Chemical.

Nitrosoguanidine and mutagenesis. The nitrosoguanidine mutagenesis of ATCC 19151 was performed as described by Miller (19). The final concentration of nitrosoguanidine was 200 μg/ml. The mutants were screened for loss of the ability to use vanillate as a carbon source while retaining the ability to grow on protocatechuate.

Cloning and nucleotide sequencing. All of the recombinant DNA methods used to construct the plasmids or to study the cloned fragments have been described previously (10, 11). The shotgun cloning of the genes involved in vanillate catabolism proceeded in two steps. DNA fragments from an *Mbo*I partial restriction digest of ATCC 19151 chromosomal DNA were subjected to sucrose gradient (10 to 30%) centrifugation. Fragments of about 30 kilobases (kb) were

* Corresponding author.

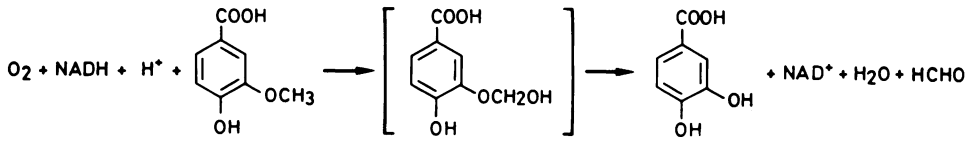

FIG. 1. Oxidative demethylation of vanillate by a *Pseudomonas* sp. (23). The product of the reaction is probably the hemiacetal shown, which spontaneously decomposes in formaldehyde and protocatechuate (8).

recovered and ligated to *Bam*HI-digested pJRD203 DNA which had been dephosphorylated in order to avoid reforming parental molecules (17). The gene bank thus constructed was introduced into *E. coli* S17-1 by λ DNA packaging, selecting for the Km^r marker carried by the plasmid. Individual *E. coli* transformants in microtiter wells were conjugated to a *Pseudomonas van* mutant by using a multiprong replicating tool. The exconjugants were then screened on M9 minimal medium plates containing vanillate as the carbon source and kanamycin, on which neither parent is able to grow. Colonies growing on these selective plates are complemented by the recombinant cosmids.

Nucleotide sequencing was performed by the method of Maxam and Gilbert (18), and sequences were analyzed with the Beckman Microgenie sequence analysis program. Protein homology searches against the National Biomedical Research Foundation Protein Identification Resource were performed by P. Terpstra (University of Groningen, Groningen, The Netherlands) by using the FASTP program designed by W. R. Pearson and D. J. Lipman.

Complementation tests. A 0.1-ml amount of an overnight culture of each *Pseudomonas van* mutant was spread onto a fresh L plate together with 0.1 ml of the *E. coli* S17-1 derivative containing the pJRD215 recombinant plasmid to be tested. After 6 h, the mixture of bacteria was scraped off the plate surface and washed twice in M9 minimal salts. The bacterial suspension was then plated on M9 minimal medium plates supplemented with kanamycin and vanillate (0.2%) or SDS (0.2%) as a carbon source. The complementation frequency was expressed as the ratio of the number of colonies capable of growing on this medium to the total number of *Pseudomonas* exconjugants (i.e., colonies capable of growing on SDS-kanamycin).

Gene expression in *E. coli*. Expression of *Pseudomonas* genes in *E. coli* was achieved by using the T7 polymerase-promoter system of Tabor and Richardson (27). This method uses two compatible plasmids: pGP1-2, which expresses the T7 RNA polymerase after temperature inactivation of the phage λ repressor, and pT7-5, which contains the T7 φ10 promoter and a polylinker which facilitates insertion of foreign DNA fragments. The [^{35}S]methionine-labeled polypeptides obtained were separated by electrophoresis on a 10% polyacrylamide gel containing 0.1% SDS for 3 h at 300 V.

RESULTS

Isolation of *Pseudomonas* sp. strain ATCC 19151 mutants altered in the vanillate demethylation pathway. The first step in the degradation of vanillate by *Pseudomonas* most probably involves the demethylation of this compound to protocatechuate (1, 16, 23) (Fig. 1). Thus, mutants defective in the vanillate demethylase reaction should be unable to use vanillate as a carbon source while retaining the ability to use protocatechuate. Six mutants of this type were retained for study following nitrosoguanidine mutagenesis. Each of them had a reversion frequency (RF) of less than 2×10^{-5}. The *vanB5* derivative ($RF = 10^{-6}$) was used as a host strain for the initial screening of recombinant DNA carrying the genes involved in vanillate demethylation.

Cloning of the genes involved in vanillate demethylation. A genomic library of *Pseudomonas* sp. strain ATCC 19151 DNA was constructed by inserting 30-kb *Mbo*I fragments of the chromosomal DNA into the *Bam*HI site of plasmid pJRD203 (10). The recombinant DNA molecules were recovered in *E. coli* S17-1 by DNA packaging in λ particles and then transferred by conjugation to the *vanB5* mutant. Three exconjugants were found to be capable of growing on vanillate as a carbon source. The cosmid DNAs extracted from them were all of much higher molecular weight than the pJRD203 parent. When retransformed into S17-1 and again conjugated to the *vanB5* mutant, they could still complement the mutant for growth on vanillate. However, the percentage of individual S17-1 transformants capable of transferring the van^+ genes to the *vanB5* mutant varied between 60 and 100%. The S17-1 transformants which had lost the property of conjugating these markers were shown to contain cosmid DNAs of lower molecular weight than the initial cosmid DNAs. Thus, it is likely that the large size of the recombinant DNA molecules makes them unstable in *E. coli* S17-1 and that, in some cases, the deletions remove the genes involved in vanillate catabolism.

Subcloning of the genes involved in vanillate demethylation. To improve stability and facilitate genetic analysis, the *van*-complementing region was subcloned on a 4.7-kb *Sal*I-*Asu*II fragment (Fig. 2A) into plasmid pJRD205 (10). When this newly formed recombinant cosmid was conjugated to the six independent *van* mutants, it was found to be capable of complementing all of them for growth on vanillate. Therefore, the genetic information necessary for the catabolism of vanillate is within the 4.7-kb *Pseudomonas* DNA insert.

Genetic analysis of the 4.7-kb DNA fragment. In order to correlate the physical and genetic maps of the 4.7-kb fragment, deletion derivatives were constructed and the complementing properties of the corresponding regions were investigated. The results obtained indicate that the mutants were distributed into two complementation groups, *vanA* and *vanB* (Fig. 2C). The *vanA* group was complemented by clones containing the *Sal*I-*Pvu*II fragment, and the *vanB* group was complemented by those containing the *Bgl*II-*Hin*cII fragment (Fig. 2A).

The complementation pattern of the *vanB* group with subclones lacking the *Sal*I-*Bgl*II region was peculiar, as it was delayed (±4 days instead of 2 days) and had a different phenotype from the positive control. This suggests that the deletion removed the normal *van* promoter but not the *vanB* structural gene and that *vanB* complementation is due to reduced *vanB* transcription from a secondary promoter. In

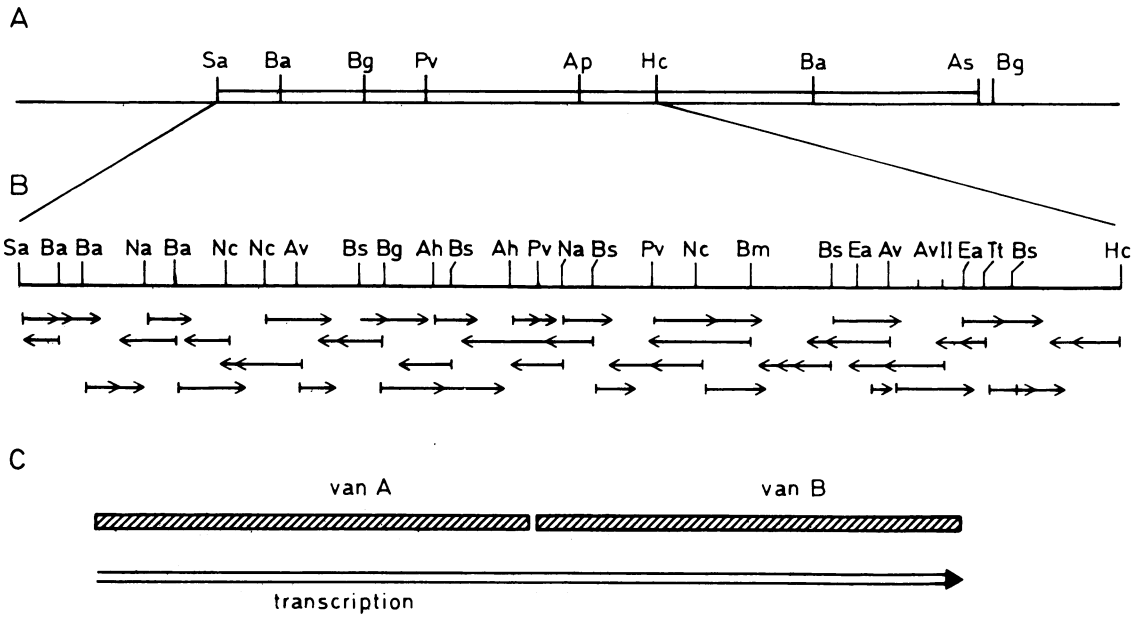

FIG. 2. Restriction map and sequencing strategy for the *Sal*I-*Hin*cII fragment containing the *vanA* and *vanB* genes. (A) Simplified restriction map of the 4.7-kb *Sal*I-*Asu*II fragment carrying the *vanAB* genes. The sites shown are those used for subcloning and deletion analysis. The double line indicates cloned *Pseudomonas* DNA, and the single line shows plasmid DNA. (B) Restriction sites used for sequence determination are shown as vertical lines. Horizontal arrows indicate the extent and direction of sequencing. Abbreviations for restriction sites: Sa, *Sal*I; Ba, *Bam*HI; Na, *Nar*I; Nc, *Nci*I; Av, *Ava*I; Bs, *Bst*EII; Bg, *Bgl*II; Ah, *Aha*I; Pv, *Pvu*II; Bm, *Bsp*MI; Ea, *Eag*I; Av, *Ava*II; Tt, *Tth*111I; Hc, *Hin*cII; Ap, *Apa*I. (C) Positions of the *vanA* and *vanB* genes relative to the physical map of the *Sal*I-*Hin*cII fragment.

agreement with this was the observation that the *Bam*HI-*Bam*HI (Fig. 2A) subclone showed the same phenotype irrespective of orientation.

The direction of transcription within *Sal*I-*Asu*II (Fig. 2A) was determined by in vitro insertional mutagenesis with the 2-kb Ω DNA segment from pHP45Ω (22) (to perform this experiment, a variant of pJRD205 carrying a deletion of the vector Sm^r gene was used). This DNA carries an Sm^r gene flanked in both orientations by transcription and translation stop signals. When inserted in a coding region, it terminates RNA and protein synthesis prematurely and thus permits characterization of the functional units. The presence of Ω in the plasmid-borne *Sal*I or *Bgl*II site (Fig. 2A) did not affect complementation, indicating that *van* region expression depends on a promoter situated within the cloned fragment. Insertion of Ω at the *Bgl*II site in the insert eliminated complementation of both *vanA* and *vanB* mutants. In contrast, insertion at the *Apa*I site eliminated complementation only for *vanB* mutants (Fig. 2A). These results show that the cloned DNA contained the promoter necessary for the expression of the *van* genes. *Bgl*II is situated within gene *A*, and *Apa*I is within gene *B*. Transcription proceeds from *Sal*I towards *Hin*cII and is polycistronic, since both *vanA* and *vanB* mutant complementation was affected by the presence of Ω at the *Bgl*II site.

Sequencing of the genes for vanillate biodegradation. The nucleotide sequence between the *Sal*I and *Hin*cII restriction sites was determined by using the strategy indicated in Fig. 2B. When possible, both strands were sequenced. Otherwise, at least two different strategies were used on one strand. As with most *Pseudomonas* chromosomal DNA, the sequence of the *van* region was very G+C rich (69.9%). Computer analysis of the nucleotide sequence (Fig. 3) showed two open reading frames in the transcribed orientation (as determined in the previous and following sections). The first (ORF1) extended from coordinates 259 (ATG) to 1246 (TGA) and had a coding capacity of 329 amino acids. It should be noted that there was a second ATG codon at coordinate 277 (in the same phase). The use of this second ATG, instead of the first, would give a coding capacity of 323 amino acids. The second open reading frame (ORF2), in the transcribed direction, immediately followed ORF1 with an ATG at position 1248 (overlapping the TGA of ORF1) and a TAG stop at position 2190, to give a coding capacity of 314 amino acids. Overlapping or closely adjacent TGA and ATG codons are frequently observed for neighboring cistrons on a polycistronic mRNA (20, 25). Computer analysis (Table 1) showed a very high G+C content in the third base in these open reading frames (ORF1, 90%; ORF2, 92%). This high G+C content is characteristic of bona fide *Pseudomonas* coding sequences (4). Thus, it is most likely that ORF1 and ORF2 are used by the bacterium as genetic information for protein synthesis. This interpretation is supported by the finding that there were sequences resembling ribosome-binding sites upstream of the putative ATG initiator codons (GGGAGGCAAG situated 4 base pairs [bp] from the ATG in position 1248, GACGGCAAG at 11 bp from ATG 259, and GCAACGAA at 4 bp from ATG 277). Both *E. coli* and *P. aeruginosa* 16S rRNAs have similar 3′ termini (13), and it is assumed here that *Pseudomonas* sp. strain ATCC 19151 is not greatly different. Table 1 shows the codon usage for the *vanA* and *vanB* genes. They were remarkable for the scarce utilization of codons containing only A and T or having these as the third base.

Open reading frames are common in G+C-rich DNA, and inspection of the nontranscribed strand showed two open reading frames, from base 1655 to 914 and from base 869 to 119, which could encode proteins of reasonable sizes. In contrast with ORF1 and ORF2, these two reading frames did

```
        10         20         30         40         50         60         70         80         90        100        110        120
GGATCCAATT TCACGAAGGC CCGTACTCTA GCCAGCCGCC CAGCCCGCTG CCAGGGCCAC GACGCAGGCG TTCCTTGTCT CAGCGTTCGC TGATCGCACG GTTATTGGAT CCAATCAGCC

       130        140        150        160        170        180        190        200        210        220        230        240
GTAAATCATG GATTCTACCC GCCCGAAGCG CCAGCATTGG ATCCAATACA ACAACAACGA GGACCGCTCC ATGTTCCCGA AGAATGCCTG GTACGTCGCC TGCACCCCGG ATGAAATCGA

       250        259                                   288                                   318                                   348
CGGCAAGCCC TTGGGCGC ATG ATC TGC AAC GAA CGG ATG GTG ATC TAC CGG GGC GCC GGC CAA CGT GTC GCA GCC CTG GAG GAC TTC TGC CCT CAT CGC GGT GCG CCG
                    Met Ile Cys Asn Glu Arg Met Val Ile Tyr Arg Gly Ala Gly Gln Arg Val Ala Ala Leu Glu Asp Phe Cys Pro His Arg Gly Ala Pro

                                        378                                   408                                   438
CTG TCC CTT GGA TCC ATT CAG GAC GGC AAG CTG GTG TGC GGC TAC CAC GGC CTG GTG ATG GAC TGC GAC GGC CGT ACC GCC TCC ATG CCT GCC CAG CGG GTG CAG
Leu Ser Leu Gly Ser Ile Gln Asp Gly Lys Leu Val Cys Gly Tyr His Gly Leu Val Met Asp Cys Asp Gly Arg Thr Ala Ser Met Pro Ala Gln Arg Val Gln

                   468                                   498                                   528                                   558
GCG TTT CCC TGC ATC CGC GCG TTC CCG GCC CAG GAG CGC CAC GGC TTC ATC TGG GTC TGG CCC GGT GAC GCC GCC CTC GCC GAC CCG GCC CTG ATC CCC CAC CTG
Ala Phe Pro Cys Ile Arg Ala Phe Pro Ala Gln Glu Arg His Gly Phe Ile Trp Val Trp Pro Gly Asp Ala Ala Leu Ala Asp Pro Ala Leu Ile Pro His Leu

                                        588                                   618                                   648
GAG TGG GCC GAG AAC CCG GCC TGG GCC TAC GGT GGC GGG CTC TAC CAC ATC GCC TGC GAC TAC CGG CTG ATG ATC GAC AAC CTC ATG GAC CTG ACC CAC GAG ACC
Glu Trp Ala Glu Asn Pro Ala Trp Ala Tyr Gly Gly Gly Leu Tyr His Ile Ala Cys Asp Tyr Arg Leu Met Ile Asp Asn Leu Met Asp Leu Thr His Glu Thr

                   678                                   708                                   738                                   768
TAC GTC CAC GCC TCC AGC ATC GGC CAG AAG GAA ATC GAC GAG GCC CCC GTG TCC ACC CGC GTG GAA GGC GAC AGG CTC ATC ACC GGC CGC TTC ATG GAA GGC ATC
Tyr Val His Ala Ser Ser Ile Gly Gln Lys Glu Ile Asp Glu Ala Pro Val Ser Thr Arg Val Glu Gly Asp Arg Leu Ile Thr Gly Arg Phe Met Glu Gly Ile

                                        798                                   828                                   858
CTG GCC CCG CCC TTC TGG CGC GCC GCC CTG CGC GGG AAT GGC CTG GCC GAC GAC GTG CCG GTG GAC CGC TGG CAG ATC TGC CGC TTC ACC CCG CCC AGC CAT GTG
Leu Ala Pro Pro Phe Trp Arg Ala Ala Leu Arg Gly Asn Gly Leu Ala Asp Asp Val Pro Val Asp Arg Trp Gln Ile Cys Arg Phe Thr Pro Pro Ser His Val

                   888                                   918                                   948                                   978
CTC ATC GAA GTC GGC GTG GCC CAC GCC GGC CGA GGG GGC TAC GAC GCC CCC GCC GAC TGC AAG GCG AGC AGC ATC GTG GTG GAC TTC ATC ACC CCC GAG ACG GAC
Leu Ile Glu Val Gly Val Ala His Ala Gly Arg Gly Gly Tyr Asp Ala Pro Ala Asp Cys Lys Ala Ser Ser Ile Val Val Asp Phe Ile Thr Pro Glu Thr Asp

                                       1008                                  1038                                  1068
ACC TCC ATC TGG TAC TTC TGG GGC ATG GCG CGC AGC TTC CGC CCC GAG GAC AAC GAG CTG ACC GCA CGC ATC CGC GAG GGC CAG GGC ACG ATC TTC GCC GAG GAT
Thr Ser Ile Trp Tyr Phe Trp Gly Met Ala Arg Ser Phe Arg Pro Glu Asp Asn Glu Leu Thr Ala Arg Ile Arg Glu Gly Gln Gly Thr Ile Phe Ala Glu Asp

                  1098                                  1128                                  1158                                  1188
CTG GAG ATG CTC GAA CAG CAG CAG CGC AAC CTG CTG GCC TGG CCC GAG CGG CCC CTG CTC AAG CTC AAC ATC GAC GCC GGC GGC GTG CAG TCC CGG CGC ATC ATC
Leu Glu Met Leu Glu Gln Gln Gln Arg Asn Leu Leu Ala Trp Pro Glu Arg Pro Leu Leu Lys Leu Asn Ile Asp Ala Gly Gly Val Gln Ser Arg Arg Ile Ile

                                       1218                                  1248                                  1277
GAG CGC CTG GTC AGC GCC GAA CGC GCC GCC GAG GCC CAG CTG ATC GGG AGG CAA GCC TGATG  CTC GAC CTG ATG ATC CGT GGC CTG CGC CTG GAG GCG CCG GGC
Glu Arg Leu Val Ser Ala Glu Arg Ala Ala Glu Ala Gln Leu Ile Gly Arg Gln Ala ***
                                                                              Met  Leu Asp Leu Met Ile Arg Gly Leu Arg Leu Glu Ala Pro Gly

                  1307                                  1337                                  1367                                  1397
ATC CTC GGC CTG GAG CTG GTG GCC ACC GAC GGT TCG CCC CTG CCC ACC TTC GAG GCC GGC GCG CAC CTG GAC CTG CAC CTG CCC GGC GGC CTG GTG CGC CCC TAC
Ile Leu Gly Leu Glu Leu Val Ala Thr Asp Gly Ser Pro Leu Pro Thr Phe Glu Ala Gly Ala His Leu Asp Leu His Leu Pro Gly Gly Leu Val Arg Pro Tyr

                                       1427                                  1457                                  1487
AGC CTG TGC AAC GCC CCC GGC GAG ACG CAC CGC TAC TGC CTC GCC GTG CTG CTG GAC CCG GCG TCC CGG GGC GGC TCC CGC GCC GTG CAC GAA CAG CTG CGG GTC
Ser Leu Cys Asn Ala Pro Gly Glu Thr His Arg Tyr Cys Leu Ala Val Leu Leu Asp Pro Ala Ser Arg Gly Gly Ser Arg Ala Val His Glu Gln Leu Arg Val

                  1517                                  1547                                  1577                                  1607
GGC CAG CAC CTG ACC ACC AGC GCA CCG CGC AAC CTC TTC CCC CTT GTG GCC GAA AGC AGC CGC AGC CTG CTG TTC GCC GGC GGC ATC GGC ATC ACC CCG ATC CTC
Gly Gln His Leu Thr Thr Ser Ala Pro Arg Asn Leu Phe Pro Leu Val Ala Glu Ser Ser Arg Ser Leu Leu Phe Ala Gly Gly Ile Gly Ile Thr Pro Ile Leu

                                       1637                                  1667                                  1697
GCC ATG GCC CAG GTG CTG GCG GCC CGG GGC GAC ACA TTC GAG CTG CAT TAC TGC GTG CGC TCG CGC AGG CTT GCG GCC TTC ATC GAC TGG CTG GAG GCC AGC ACC
Ala Met Ala Gln Val Leu Ala Ala Arg Gly Asp Thr Phe Glu Leu His Tyr Cys Val Arg Ser Arg Arg Leu Ala Ala Phe Ile Asp Trp Leu Glu Ala Ser Thr

                  1727                                  1757                                  1787                                  1817
TTC GCC GCG CAC GTC CAC CTG CAC GCC GAC GAC GGC CCC ACG CCC TTC GAC GCC ACC GCG CTG CTG CGG GAC GCC GGC GAC GCC CAC CTC TAC GTG TGC GGG CCG
Phe Ala Ala His Val His Leu His Ala Asp Asp Gly Pro Thr Pro Phe Asp Ala Thr Ala Leu Leu Arg Asp Ala Gly Asp Ala His Leu Tyr Val Cys Gly Pro

                                       1847                                  1877                                  1907
GGC GGC TTC ATG GAG CAC GTG CTG GGC TGT GCC CGC ACC GCC GGC TGG GAC GAA ACG CGC CTG CAC CGC GAA TAC TTC GCC GCC CCC GTG CAA CCG GCC GGG GAC
Gly Gly Phe Met Glu His Val Leu Gly Cys Ala Arg Thr Ala Gly Trp Asp Glu Thr Arg Leu His Arg Glu Tyr Phe Ala Ala Pro Val Gln Pro Ala Gly Asp

                  1937                                  1967                                  1997                                  2027
GCG CGC GCC TTC GAG GGT CGC CTG GCG CGC AGC GGC CTG ACC TTG CAG GTG CCG GCC GAG CGC AGC GTG GCC CAG GTG CTG GAC GAC GCC GGG GTG TGC ATC CCG
Ala Arg Ala Phe Glu Gly Arg Leu Ala Arg Ser Gly Leu Thr Leu Gln Val Pro Ala Glu Arg Ser Val Ala Gln Val Leu Asp Asp Ala Gly Val Cys Ile Pro

                                       2057                                  2087                                  2117
CTG GCC TGC GAA CAG GGC ATC TGC GGC ACC TGC CTG ACC CGG GTG CTG GAC GGC GAA CCC GAG CAC CGC GAC AGC TTC CTC ACC GAC GCG GAG CGG GCC CGC AAC
Leu Ala Cys Glu Gln Gly Ile Cys Gly Thr Cys Leu Thr Arg Val Leu Asp Gly Glu Pro Glu His Arg Asp Ser Phe Leu Thr Asp Ala Glu Arg Ala Arg Asn

                  2147                                  2177                                  2202       2212       2222       2232       2242
GAC CAG TTC ACC CCC TGC TGT TCG CGG GCC CGC AGT GCC TGC CTG GTG CTG GAC CTC TAG CAGGAGGAAC CGGGCGGCTT GGCGGTGCCC GGGTTAGTTA CCGGCCGACA
Asp Gln Phe Thr Pro Cys Cys Ser Arg Ala Arg Ser Ala Cys Leu Val Leu Asp Leu ***

      2252       2262       2272       2282       2292       2302       2312       2322       2332       2342       2352       2362
GCCCGCAGTC GATGACCAGG TCCGCCCCGG TGATGGAGCG CGCCAGGGGC TGGCAGAGGA AGTGGCACAG CTCGGCCACT TCCTCCGGCT GGATGAAGCG CGCCTGTTCG CCCTGGGGGT

      2372       2382       2392       2402       2412       2422       2432       2442       2452       2462       2472       2482
ACTTGGCCAG CAGCTCGCGG TAGTAGCCGT CGGGGTCGCC GTTGCCGTAG CGCTCGGCCT GGAAGCTCAG CGTCGGGGGG TCGATGTCCC CCGGGGACAC CGCGTTAGCA GCGCCGCCGT

      2492       2502       2512       2522       2532       2542       2552       2562       2572       2582       2592
CCGGCGCCAG GTTCAGGGCC AGGGTCTTGG TCAGCAGCGC CACCCCACCC TTGCTGGCGC AGTAGGCGGC GGCGTTACGG TTGCCCTGGC GCCCCGCATC GCTGGAGATG TTGGGA
```

FIG. 3. Nucleotide sequence of the *SalI-Hin*cII fragment containing the *vanA* and *vanB* genes. The ORF1 and ORF2 open reading frames, corresponding to *vanA* and *vanB*, respectively, have been translated into their predicted amino acid sequences. Only the 5′-3′ (sense) DNA strand is shown.

TABLE 1. Distribution of codons in the *vanA* and *vanB* genes

Codon	Amino acid	No. of codons		Codon	Amino acid	No. of codons		Codon	Amino acid	No. of codons		Codon	Amino acid	No. of codons	
		vanA	*vanB*			*vanA*	*vanB*			*vanA*	*vanB*			*vanA*	*vanB*
TTT	Phe	1	0	TCT	Ser	0	0	TAT	Tyr	0	0	TGT	Cys	0	2
TTC	Phe	10	12	TCC	Ser	7	2	TAC	Tyr	8	5	TGC	Cys	8	10
TTA	Leu	0	0	TCA	Ser	0	0	TAA	End	0	0	TGA	End	1	0
TTG	Leu	0	1	TCG	Ser	0	3	TAG	End	0	1	TGG	Trp	9	2
CTT	Leu	1	2	CCT	Pro	2	0	CAT	His	2	1	CGT	Arg	2	1
CTC	Leu	8	8	CCC	Pro	11	11	CAC	His	7	12	CGC	Arg	17	18
CTA	Leu	0	0	CCA	Pro	0	0	CAA	Gln	2	1	CGA	Arg	1	0
CTG	Leu	18	33	CCG	Pro	7	8	CAG	Gln	12	7	CGG	Arg	6	7
ATT	Ile	1	0	ACT	Thr	0	0	AAT	Asn	1	0	AGT	Ser	0	1
ATC	Ile	22	8	ACC	Thr	9	13	AAC	Asn	6	3	AGC	Ser	6	9
ATA	Ile	0	0	ACA	Thr	0	1	AAA	Lys	0	0	AGA	Arg	0	0
ATG	Met	9	4	ACG	Thr	2	3	AAG	Lys	4	0	AGG	Arg	2	1
GTT	Val	0	0	GCT	Ala	0	0	GAT	Asp	1	0	GGT	Gly	3	2
GTC	Val	5	2	GCC	Ala	31	30	GAC	Asp	20	20	GGC	Gly	21	24
GTA	Val	0	0	GCA	Ala	2	1	GAA	Glu	7	6	GGA	Gly	1	0
GTG	Val	13	16	GCG	Ala	5	10	GAG	Glu	15	11	GGG	Gly	4	3

not have a high G+C content in the third base of the codon and are probably not used to code proteins.

Expression of the vanillate biodegradative genes in *E. coli*. In order to characterize the polypeptide products of the *vanA* and *vanB* genes, expression in *E. coli* maxicells was attempted. No expression was seen, possibly because *Pseudomonas* promoters are generally not recognized by *E. coli* RNA polymerase. To avoid this problem and to specifically label the cloned *Pseudomonas* gene products, use was made of the T7 polymerase-promoter expression system in *E. coli* (27). In this system, T7 polymerase is expressed from the λ p_L promoter under the control of the temperature-sensitive repressor *c*I857. When induced, the T7 polymerase specifically initiates transcription of genes cloned downstream of the T7 ϕ10 promoter on a second compatible plasmid. These genes are the only ones to be transcribed when rifampin is added to block *E. coli* RNA polymerase (T7 RNA polymerase being insensitive to rifampin), and their products are easily identified on an SDS-polyacrylamide gel following labeling with [^{35}S]methionine.

Figure 4 shows that when the *Sal*I-*Hin*cII fragment was cloned and expressed by using the T7 RNA polymerase system, two polypeptides were made. Their sizes were in reasonable agreement with those deduced from the nucleotide sequences. The *vanA* and *vanB* gene products estimated by gel electrophoresis were 37.5 and 33 kilodaltons (kDa), respectively (Fig. 4, lanes 3 and 4), compared with 36.5 and 33.7 kDa estimated from the nucleotide sequence. The use of appropriate deletion derivatives permitted unambiguous identification of the coding regions for the two polypeptides. Thus, the *Sal*I-*Bgl*II fragment which lacks ORF2 and part of ORF1 did not make the *vanB* polypeptide and synthesized a truncated *vanA* polypeptide (Fig. 4, lane 5). Loss of the ORF1 5′ end in the *Bgl*II-*Asu*II fragment only permitted *vanB* polypeptide synthesis (Fig. 4, lane 6). No polypeptides were synthesized from the reverse strand (Fig. 4, lane 7). The difference in intensity of the two polypeptides is almost certainly a reflection of their different methionine content; the *vanA* gene product had nine methionine residues, and the *vanB* gene product only had four.

DISCUSSION

In this paper, the cloning and sequencing of two genes involved in the demethylation of a key intermediate in the biodegradation of lignin, 3-methoxy-4-hydroxybenzoate (vanillate), are described. The screening procedure for the

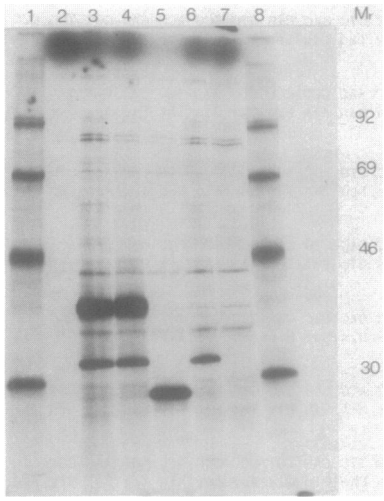

FIG. 4. Polypeptides synthesized in *E. coli* from the *Sal*I-*Asu*II fragment and deleted derivatives. *E. coli* MM294 cultures containing both the pGP1-2 plasmid (T7 RNA polymerase expression plasmid) and the pT7-5 T7 ϕ10 promoter plasmid (into which various fragments of the *van* region had been inserted) were induced and labeled with [^{35}S]methionine by the protocol of Tabor and Richardson (27). The ^{35}S-labeled polypeptides were separated by electrophoresis through a 10% polyacrylamide gel containing 0.1% SDS for 3 h at 300 V and visualized by autoradiography. The pT7-5 recombinant plasmids contained various parts of the *van* region. Lane 2, No *van* insert; lane 3, *Sal*I-*Asu*II; lane 4, *Sal*I-*Hin*cII; lane 5, *Sal*I-*Bgl*II; lane 6, *Bgl*II-*Asu*II. All of these plasmids contained the *van* insert oriented so that it would be transcribed from left to right, as drawn in Fig. 3. Inverting the orientation of the *Sal*I-*Asu*II fragment gave no expression (lane 7). Lanes 1 and 8 contained ^{14}C-labeled molecular weight markers (shown in thousands).

recombinant clones was based on the assumption that in *Pseudomonas* sp. strain ATCC 19151, as in other *Pseudomonas* strains (1, 16, 23), vanillate is degraded via protocatechuate and that expression of the cloned genes would enable *van* mutants (defined as capable of growth on protocatechuate but not on vanillate) to grow on vanillate. This assumption was strengthened by the fact that the wild-type strain ATCC 19151 used in the study does not degrade compounds such as guiaicol and catechol, which are intermediates of the vanillate biodegradation pathway of other organisms (21). Moreover, ATCC 19151 does not grow on isovanillate (3-hydroxy-4-methoxybenzoic acid), veratrate (3-4-methoxybenzoic acid), or *o-*, *m-*, or *p*-anisates (2-, 3-, and 4-methoxybenzoic acids), which seems to indicate that the demethylation is a very specific process.

A recombinant clone able to complement six independently isolated *van* mutants was isolated and subcloned on a 4.7-kb *Pseudomonas* DNA fragment. This DNA contains all of the genes necessary for vanillate demethylation, as expression in *Pseudomonas oleovorans* GPO1 (which naturally degrades protocatechuate) conveys to this strain the additional property of using vanillate as a carbon source (data not shown). The intact clone also confers the ability to utilize vanillate on *vtu* (vanillate utilization) mutants of *Pseudomonas aeruginosa* PAO2175 (from H. Matsumoto). Curiously, deletion clones containing only *vanA* or *vanB* did not complement the mutants. It is probable that interaction between the vanillate monooxygenase and its electron transport protein is highly specific and that one protein cannot be replaced by the equivalent protein from another species. Alternatively, the mechanisms of vanillate degradation may differ between *Pseudomonas* sp. strain ATCC 19151 and *P. aeruginosa*.

The genetic evidence, derived from deletion analysis and insertion inactivation, suggests that two different *van* genes, *vanA* and *vanB*, are located in the left half of the fragment and that these are cotranscribed from left to right. The presence of a secondary promoter for the *vanB* gene was suggested by the delayed complementation of the *vanB* mutants observed when the left-hand end of the cloned DNA was missing. A computer search for sequences similar to the *Pseudomonas* promoter consensus sequence (12) showed that one of them was in fact present 106 bp upstream of the *vanB* start codon (coordinates 1121 to 1142; C*TGGC*-11 nt-CT*GC*T). An in-depth study of the regulatory elements carried by the *Sal*I-*Asu*II fragment will be published elsewhere.

The characterization of the *vanA* and *vanB* genes, as well as their direction of transcription, was confirmed by DNA sequence analysis, which showed two open reading frames in positions corresponding to those suggested by deletion analysis. The high G+C content of the third base of the codons confirmed the functionality of these open reading frames (4).

The *vanA* and *vanB* genes were expressed in *E. coli* by using a strong promoter from phage T7. In agreement with the nucleotide sequencing results, two polypeptides were visualized, corresponding in size to those predicted by the *vanA* and *vanB* open reading frames. These polypeptides could be correlated with the *vanA* and *vanB* open reading frames by deletion analysis.

The identity of the gene products encoded by the *vanA* and *vanB* genes is not yet known. In *Nocardia* and *Moraxella* spp., the monooxygenases active in the demethylation of isovanillate and guaiacol, respectively, are cytochrome P450s (7, 9, 28). On the contrary, the demethylation of 4-methoxybenzoate by *Pseudomonas putida* is carried out by an iron sulfur protein (3, 8). Comparison of the *vanA* and *vanB* protein sequences with those of all other recorded proteins indicates that *vanB* has similarity to many members of the ferredoxin family. The involvement of a ferredoxin in the vanillate demethylase reaction is quite possible. For example, ferredoxin$_{TOL}$ functions in the electron transport system of the toluene dioxygenase reaction (14). However, it is worth mentioning that the *vanB* gene product (33 kDa) is considerably larger than most ferredoxins, which typically have molecular weights between 8,000 and 15,000. Comparison of the *vanA* gene product with the protein bank failed to give any convincing similarities. Additional biochemical experiments will obviously be needed to obtain a better understanding of the mode of action of the *vanA* and *vanB* proteins.

ACKNOWLEDGMENTS

This research was carried out under research contract GBI-3-016-B of the Biomolecular Engineering Programme and research contract BAP-OO48-B of the Biotechnology Action Program of the Commission of European Communities as well as under research grant 2.4333.86F of the Fonds de la Recherche Fondamentale Collective.

We are grateful for extensive discussions with A. M. Chakrabarty during the tenure of a NATO travel grant for international collaboration in research (no. 0177/87). We thank S. Tabor for providing the T7 polymerase expression system, M. Kok and H. Matsumoto for providing the *P. oleovorans* and the *P. aeruginosa vtu* mutants, respectively, and P. Terpstra for performing the protein homology search. We acknowledge N. Chevalier and M. F. Pilaete for expert technical assistance and P. Mulder for the DNA sequence analysis and typing the manuscript.

LITERATURE CITED

1. **Bernhardt, F. H., N. Erdin, H. Staudinger, and V. Ullrich.** 1973. Interactions of substrates with a purified 4-methoxybenzoate monooxygenase system (O-demethylating) from *Pseudomonas putida*. Eur. J. Biochem. **35**:126–134.
2. **Bernhardt, F. H., H. Pachowsky, and H. Staudinger.** 1975. A 4-methoxybenzoate O-demethylase from *Pseudomonas putida*. A new type of monooxygenase system. Eur. J. Biochem. **57**: 241–256.
3. **Bernhardt, F. H., and H. Kuthan.** 1983. Kinetics of reduction of putidamonooxin by NADH-putidamonooxin oxidoreductase, sodium dithionite and superoxide radicals. Eur. J. Biochem. **130**:99–103.
4. **Bibb, M. J., P. R. Findlay, and M. W. Johnson.** 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene **30**:157–166.
5. **Cartwright, N. J., and A. R. W. Smith.** 1967. Bacterial attack on phenolic ethers; an enzyme system demethylating vanillic acid. Biochem. J. **102**:826–841.
6. **Cartwright, N. J., and D. A. Broadbent.** 1974. Bacterial attack on phenolic ethers. Preliminary studies on systems transporting electrons to the substrate binding components in bacterial O-dealkylases. Microbios **10**:87–96.
7. **Crawford, R. L., E. McCoy, J. M. Harkin, T. K. Kirk, and J. R. Obst.** 1973. Degradation of methoxylated benzoic acids by a *Nocardia* sp. from a lignin-rich environment: significance to lignin degradation and effect of chloro substituents. Appl. Microbiol. **26**:176–184.
8. **Dagley, S.** 1986. Biochemistry of aromatic hydrocarbon degradation in Pseudomonads, p. 527–555. *In* I. C. Gunsalus, J. R. Sokatch, and L. N. Ornston (ed.), The bacteria, vol. X: the biology of *Pseudomonas*. Academic Press, Inc., Orlando, Fla.
9. **Dardas, A., D. Gal, M. Barrelle, G. Sauret-Ignazi, R. Sterjiades, and J. Pelmont.** 1985. The demethylation of guaiacol by a new bacterial cytochrome P-450. Arch. Biochem. Biophys. **236**:585–

592.

10. **Davison, J., M. Heusterspreute, N. Chevalier, V. Ha-Thi, and F. Brunel.** 1987. Vectors with restriction site banks. V. pJRD215, a wide host range cosmid vector with multiple cloning sites. Gene **51**:275–280.
11. **Davison, J., M. Heusterspreute, and F. Brunel.** 1987. Restriction site bank vectors for cloning in gram-negative bacteria and yeast. Methods Enzymol. **153**:34–55.
12. **Deretic, V., J. F. Gill, and A. M. Chakrabarty.** 1987. Alginate biosynthesis: a model system for gene regulation and function in *Pseudomonas*. Bio/Technology **5**:469–477.
13. **Frantz, B., and A. M. Chakrabarty.** 1986. Degradative plasmids in *Pseudomonas*, p. 295–325. *In* I. C. Gunsalus, J. R. Sokatch, and L. N. Ornston (ed.), The bacteria, vol. X: the biology of *Pseudomonas*. Academic Press, Inc., Orlando, Fla.
14. **Gibson, D. T., and V. Subramanian.** 1984. Microbial degradation of aromatic hydrocarbons, p. 181–252. *In* D. T. Gibson (ed.), Microbial degradation of organic compounds. Marcel Dekker, New York.
15. **Hsu, Y.** 1963. Detergent (sodium lauryl sulphate)-splitting enzyme from bacteria. Nature (London) **200**:1091–1092.
16. **Kersten, P., P. J. Chapman, and S. Dagley.** 1985. Enzymatic release of halogens or methanol from some substituted protocatechuic acids. J. Bacteriol. **162**:693–697.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
19. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 125–129. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon of *E. coli*. Genetics **95**:785–795.
21. **Pometto, A. L., J. B. Sutherland, and D. L. Crawford.** 1981. *Streptomyces setonii*: catabolism of vanillic acid via guaiacol and catechol. Can. J. Microbiol. **27**:636–638.
22. **Prentki, P., and H. M. Krisch.** 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene **29**:303–313.
23. **Ribbons, D. W.** 1970. Stoichiometry of O-demethylase activity in *Pseudomonas aeruginosa*. FEBS Lett. **8**:101–104.
24. **Ribbons, D. W.** 1971. Requirement of two protein fractions for O-demethylase activity in *Pseudomonas testosteroni*. FEBS Lett. **12**:161–165.
25. **Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Petersen.** 1982. Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. **162**:729–773.
26. **Simon, R., U. Priefer, and A. Pühler.** 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram-negative bacteria. Bio/Technology **1**:784–791.
27. **Tabor, S., and C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
28. **Unger, B. P., S. G. Sligar, and I. C. Gunsalus.** 1986. *Pseudomonas* cytochromes P-450, p. 557–590. *In* I. C. Gunsalus, J. R. Sokatch, and L. N. Ornston (ed.), The bacteria, vol. X: the biology of *Pseudomonas*. Academic Press, Inc., Orlando, Fla.