

Organization and nucleotide sequence determination of a gene cluster involved in 3-chlorocatechol degradation

(gene expression/maxicell analysis/nuclease S1 mapping/*Pseudomonas* promoters)

B. FRANTZ* AND A. M. CHAKRABARTY†

Department of Microbiology and Immunology, University of Illinois College of Medicine, Chicago, IL 60612

Communicated by I. C. Gunsalus, January 5, 1987

ABSTRACT Three critical enzymes, catechol oxygenase II (chlorocatechol dioxygenase), muconate cycloisomerase II, and dienelactone hydrolase, are involved in the degradation of chlorocatechols, which are obligatory intermediates in the catabolism of chlorinated aromatic compounds. The organization and complete nucleotide sequence of the genes for these enzymes have been determined on a 4.2-kilobase-pair (kbp) *Bgl* II fragment cloned from the plasmid pAC27, based on the agreement of open reading frame lengths with apparent mobilities of polypeptides expressed in *Escherichia coli* maxicells, predicted N-terminal amino acid sequences with those of the purified proteins, and predicted total amino acid compositions with those of the purified proteins. The 4.2-kbp fragment contains the three genes and ribosome binding sites for those genes but no promoter. When placed downstream of the *tac* promoter in the broad-host-range plasmid pMMB22, this fragment directs the synthesis of all three enzymes in both *E. coli* and *Pseudomonas putida* only on induction with isopropyl β -D-thiogalactopyranoside, suggesting that the gene cluster is regulated as a single unit under the control of a single promoter. Endogenous transcription initiation of the gene cluster on pAC27, however, occurs from a site present within a 386-bp *Bgl* II fragment upstream of the 4.2-kbp fragment, and sequences 5' to that site are similar to the sequences of other positively controlled *Pseudomonas* promoters occurring on the TOL and NAH plasmids.

A large number of chlorinated compounds have been manufactured by the chemical industry during the last several decades and released into the environment in the form of herbicides, pesticides, solvents, refrigerants, etc., resulting in massive pollution problems (1). Highly chlorinated compounds such as polychlorinated biphenyls (PCBs), 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), and dichlorodiphenyltrichloroethane (DDT) persist in the environment because of their recalcitrance to microbial attack. In contrast, simple chlorinated compounds such as 3-chlorobenzoic acid and 2,4-dichlorophenoxyacetic acid (2,4-D) are biodegradable, and the degradative genes are borne on conjugative plasmids pAC27 and pJP4, respectively (2, 3). In both instances, the degradative genes have been cloned from the respective plasmids as parts of a 4.2-kilobase-pair (kbp) *Bgl* II fragment of pAC27 and 15-kbp *Eco*RI fragment of plasmid pJP4 (4, 5). While the plasmid pJP4 encodes the complete 2,4-dichlorophenoxyacetic acid degradative pathway, the plasmid pAC27 encodes only a chlorocatechol degradative pathway. The two enzymes that are involved in the conversion of 3-chlorobenzoate to 3-chlorocatechol are chromosomally encoded in *Pseudomonas putida* and are assumed to be the same needed for growth on benzoate. Thus, transfer of pAC27 to *P. putida* allows total degradation of 3-chloro-

benzoate through both chromosomal and plasmid genes. Knackmuss and Reineke (6, 7) showed that there are three enzymes critical for chlorocatechol degradation: catechol oxygenase II (chlorocatechol dioxygenase, or pyrocatechase II), muconate cycloisomerase II, and dienelactone hydrolase (Fig. 1). Whereas the analogous chromosomally encoded enzymes involved in the oxidation of catechol (catechol oxygenase I, cycloisomerase I, and enol-lactone hydrolase) have high specificity for the nonchlorinated substrates with little activity toward chlorinated analogues, the plasmid-encoded catechol oxygenase II and cycloisomerase II are highly active toward chlorinated substrates but retain diminished activity toward the nonchlorinated substrates. Dienelactone hydrolase is, however, specific for dienelactone and has no activity toward enol-lactones. The plasmid pAC27 was demonstrated to complement *P. putida* chromosomal Ben⁻ mutant PRS2015 (8), which is deficient in cycloisomerase I activity, to Ben⁺ (allowing growth on benzoate as well as 3-chlorobenzoate) (9), suggesting that the plasmid-encoded enzyme can substitute for the defective chromosomal enzyme while retaining a different substrate specificity. These similarities and differences in substrate specificity raise the interesting question as to how genes encoding degradation of synthetic chlorinated compounds such as 3-chlorocatechol evolve in nature, how they are organized and regulated on a plasmid, and whether any clue to such a process can be obtained by comparing the nucleotide sequences of the newly evolved genes to the sequences for the corresponding chromosomal genes specifying degradation of nonchlorinated analogues. In this paper we report the organization and complete nucleotide sequence of the pAC27 gene cluster specifying 3-chlorocatechol degradation. We compare the promoter region of that cluster with other positively regulated *Pseudomonas* biodegradative gene functions from the TOL and NAH plasmids.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Phage. All strains used in this report have been described previously: *Escherichia coli* JM103 (10) for transformation of M13 recombinants, C600F⁺ (11) for analysis of recombinants and purification of template DNA for sequencing, and CSR603 (12) for maxicell analysis and *P. putida* PRS2015 (8) containing pAC27 (2) for isolation of RNA. The overproducing plasmid pDC100, containing the 4.2-kbp *Bgl* II fragment of pAC27, has been described (11), and subfragments of the 4.2-kbp *Bgl* II fragment were used to construct pKK223-3 (Pharmacia) derivatives pBF101 and pBF102 (Fig. 1) for maxicell analysis. M13 phage mp18 and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IPTG, isopropyl β -D-thiogalactopyranoside; ORF, open reading frame.

*Present address: Department of Chemistry, Northwestern University, Evanston, IL 60201.

†To whom requests for reprints should be addressed.

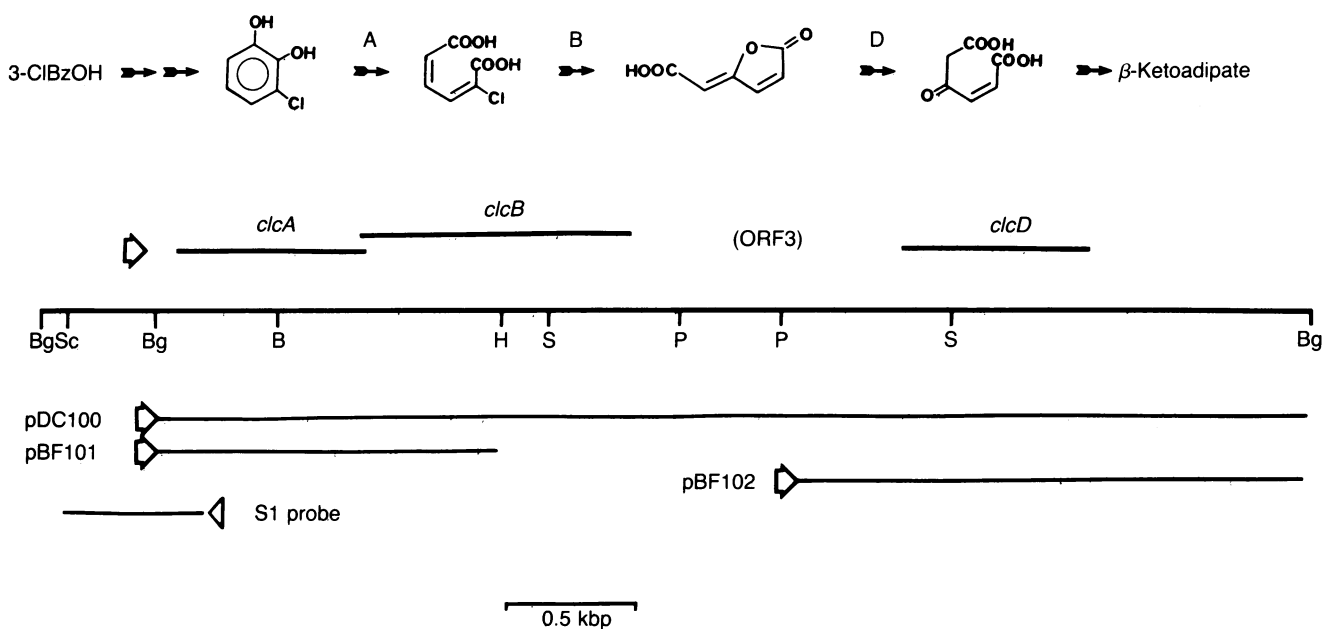


FIG. 1. Enzymatic pathway for the degradation of 3-chlorocatechol [derived from 3-chlorobenzoic acid (3-CIBzOH)] to maleylacetate, organization of chlorocatechol-degradative genes, restriction map, and plasmids used for expression analysis and nuclease S1 protection. The pathway shows the three key steps leading to the production of maleylacetate from 3-chlorocatechol. A chromosomally encoded maleylacetate reductase is believed to convert this product to β-ketoadipate for its ultimate utilization. A, catechol oxygenase II; B, muconate cycloisomerase II; D, dienelactone hydrolase. Step C, normally used for isomerization to the enol-lactone in the nonsubstituted catechol pathway, is nonenzymatic in the chlorocatechol pathway, with the spontaneous release of chloride. The locations of the corresponding genes (A, *clcA*; B, *clcB*; D, *clcD*) and of the unassigned open reading frame ORF3 are shown above the 4.2-kbp *Bgl* II fragment. A 385-bp *Bgl* II fragment immediately precedes the 4.2-kbp *Bgl* II fragment and contains operator/promoter sequences (open arrow) for expression of the *clc* gene cluster. B, *Bam*HI; Bg, *Bgl* II; H, *Hind*III; P, *Pst* I; S, *Sal* I; Sc, *Sac* II. pDC100, pBF101, and pBF102 represent the whole or part of the 4.2-kbp *Bgl* II fragment cloned under control of the *tac* promoter.

mp19 (13) were used for dideoxy sequencing and deletion cloning.

DNA Sequencing and Nuclease S1 Protection Analysis. All reagents and procedures used for sequencing and maxicell analysis have been described (11). Deletion cloning by the method of Dale *et al.* (14) was used to construct clones for sequencing and probes for S1 mapping. RNA was isolated from PRS2015 harboring pAC27, with or without induction by 5 mM 3-chlorobenzoate, by the method of Deretic *et al.* (15). The DNA probe for hybridization was prepared by primer extension of a deletion clone in the presence of [α -³²P]dCTP, by the method of Burke (16). Nuclease S1 was purchased from Boehringer Mannheim.

RESULTS AND DISCUSSION

Cloning and Expression of the Chlorocatechol (*clc*) Degradative Gene Cluster. The cloning of a 4.2-kbp *Bgl* II fragment (E fragment) from the degradative plasmid pAC27 that conferred slow growth on 3-chlorobenzoate was reported previously (4). This suggested that all the degradative enzymes were encoded by this fragment. In order to determine if such enzymes could be detected in a heterologous host such as *E. coli*, the 4.2-kbp fragment was placed (after addition of *Eco*RI linkers) downstream of the *tac* promoter on the broad-host-range plasmid pMMB22 to produce the recombinant plasmid pDC100 (Fig. 1). Introduction of pDC100 into *E. coli* maxicell strain CSR603 allowed the synthesis of three major polypeptides, of 40, 33, and 28 kDa, only upon induction with isopropyl β-D-thiogalactopyranoside (IPTG) (Fig. 2, lane 4). Similar introduction of pDC100 into other strains of *E. coli* and induction with IPTG allowed detection of all three chlorocatechol degradative enzyme activities (catechol oxygenase II, muconate cycloisomerase II, and dienelactone hydrolase) and their purification (ref. 11; K.-L. Ngai, B.F., D. K. Chatterjee, L. N. Ornston, and A.M.C.,

unpublished results). Subfragments of the 4.2-kbp fragment, when placed downstream of the *tac* promoter of pKK223-3 to give plasmids pBF101 and pBF102 (Fig. 1), similarly allowed the synthesis of only the 33-kDa polypeptide (Fig. 2, lane 6) and the 28-kDa polypeptide (Fig. 2, lane 8) in *E. coli* maxicells. These and the following sequencing results suggest that the catechol oxygenase II gene (*clcA*) is at the 5' end

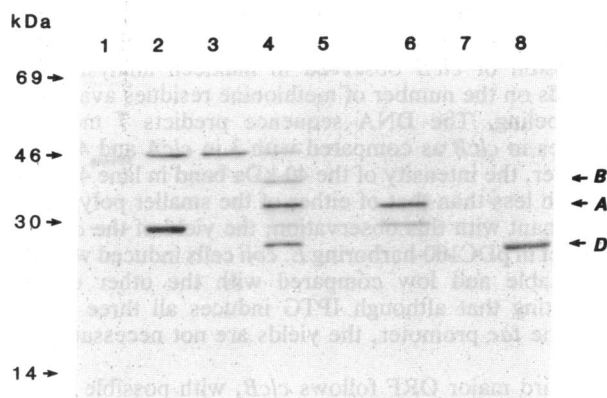


FIG. 2. Autoradiogram of L-[³⁵S]methionine-labeled proteins after UV irradiation of *E. coli* maxicells carrying various plasmids. Lanes: 1, ¹⁴C-labeled protein standards; 2, pMMB22 induced with IPTG; 3, pDC100 uninduced; 4, pDC100 induced with IPTG; 5 and 7, pKK223-3; 6, pBF101; 8, pBF102. Expression of polypeptides in pKK223-3 derivatives is constitutive in the maxicell strain CSR603 and thus independent of IPTG induction. The 30-kDa polypeptide present in pKK223-3 for ampicillin resistance is not detectable in pBF102 because that construct contains putative transcriptional termination sequences following the coding region for the 28-kDa polypeptide, which may prevent read-through from the *tac* promoter. Arrows with letter designations indicate associated *clc* genes responsible for expression of the polypeptides shown.

and the diene lactone hydrolase gene (*clcD*) is at the 3' end of the 4.2-kbp fragment.

Nucleotide Sequence Analysis of *clc* Genes. Plasmid pAC27 contains a 385-bp *Bgl* II fragment, immediately upstream of the 4.2-kbp *Bgl* II E fragment (Fig. 1), that was previously unrecorded. The recombinant plasmid (pDC25) carrying these fragments was reported to elicit slow growth on 3-chlorobenzoate (4) but occasionally gave rise to larger, faster-growing colonies that were shown to have undergone as much as an 8-fold amplification of the 4.2-kbp segment. Amplification was seen as a direct consequence of the severe selective pressure on cells carrying a recombinant plasmid devoid of necessary regulatory signals to grow on 3-chlorobenzoate as a sole source of carbon and energy and was suggested to allow growth on that substrate by a gene-dosage effect (4). To determine the presence (or absence) of regulatory and structural gene elements within the two fragments, we determined the nucleotide sequences of the two fragments and performed nuclease S1 protection analysis to determine the site of mRNA initiation.

The nucleotide sequence of the 385-bp and 4.2-kbp *Bgl* II fragments revealed four major open reading frames (ORFs), the first initiating 23 bp downstream from the 5' *Bgl* II end of the 4.2-kbp segment (Fig. 3, base 402). N-terminal amino acid sequence analysis of purified catechol oxygenase II from IPTG-induced pDC100 in *E. coli* agreed with the predicted amino acid sequence from the DNA, and the total amino acid composition of the purified protein agreed, within experimental error, with that predicted by the DNA (K.-L. Ngai, B.F., D. K. Chatterjee, L. N. Ornston, and A.M.C., unpublished data). Thus, we designated ORF1 as *clcA*. The molecular mass of the protein predicted by the DNA sequence is, however, 4 kDa lower than that consistently observed in NaDodSO₄/PAGE. Similar to other dioxygenases, catechol oxygenase II complexes non-heme Fe³⁺ ions as cofactors (17), and thus the anomalous migration may be due to incomplete denaturation of the protein (K.-L. Ngai, personal communication).

The termination codon for *clcA* at base 1182 overlaps the initiation codon of the second major ORF (Fig. 3), which by agreement with N-terminal amino acid sequence analysis and total amino acid composition of the purified protein, we have designated *clcB*, the gene for muconate cycloisomerase II. This overlap may play a role in negative interference of the expression of *clcB* observed in maxicell analysis, which depends on the number of methionine residues available for ³⁵S labeling. The DNA sequence predicts 7 methionine residues in *clcB* as compared with 3 in *clcA* and 4 in *clcD*; however, the intensity of the 40-kDa band in lane 4 of Fig. 2 is much less than that of either of the smaller polypeptides. Consonant with this observation, the yield of the *clcB* gene product in pDC100-harboring *E. coli* cells induced with IPTG is variable and low compared with the other enzymes, suggesting that although IPTG induces all three enzymes from the *tac* promoter, the yields are not necessarily coordinate.

A third major ORF follows *clcB*, with possible initiation codons at bases 2320 and 2464, although relatively poor accompanying ribosome-binding sequences appear to be present. No enzyme function has been identified with this coding region and the sequence is not translated in Fig. 3. There is no major corresponding polypeptide directed in *E. coli* maxicells, although a very light band is observed below the 30-kDa product of the ampicillin-resistance gene.

The fourth major ORF, initiating at base 3325, predicts a polypeptide of molecular mass 26.5 kDa, again slightly lower than that observed in NaDodSO₄/PAGE. The N-terminal coding sequence of ORF4 matched the N-terminal sequence of purified diene lactone hydrolase from *Pseudomonas* sp. B13 harboring the plasmid pWR1, a plasmid shown to be

extensively homologous to pAC27 (18), and the total amino acid composition of diene lactone hydrolase purified from pDC100-harboring, IPTG-induced cells matched the composition predicted by the DNA sequence. Thus, we designated ORF4 as *clcD* (11).

All three genes (*clcA*, *-B*, and *-D*) are preceded by potential ribosome-binding sequences with the conserved sequence GGAGA appearing in each case. The G+C content of each coding region is maintained at ≈60%, resembling that of the *P. putida* chromosome (19), and a pronounced preference for codons ending in guanosine or cytidine is universal. Base preference for G or C in the wobble position has been observed in a variety of organisms with high G+C content (11) and is useful in identifying reading frames within a given sequence.

Nuclease S1 Protection Analysis. Because the coding sequence of the *clcA* gene initiates close to the 5' end of the 4.2-kbp *Bgl* II fragment, sufficient space for promoter/operator sequence appeared to be absent within this fragment. Also, we observed that growth on 3-chlorobenzoate by *P. putida* PRS2015 harboring pDC100 was dependent on IPTG induction. This observation indicated that the 4.2-kbp *Bgl* II fragment lacks a functional promoter. That a functional promoter is present on the upstream, 385-bp *Bgl* II fragment was inferred from the fact that cloning of this fragment upstream of a promoterless streptomycin phosphotransferase gene in plasmid pKT240 led to enhanced streptomycin resistance (A. Berry and B.F., unpublished observations). In order to localize the site of mRNA transcription initiation, a 1.8-kbp *Bam*HI fragment carrying DNA overlapping the 385-bp *Bgl* II fragment and spanning the 4.2-kbp *Bgl* II fragment to base 845 (Fig. 3) was subcloned from pDC10 (4) into M13 vectors. Deletion clones from the *Bam*HI site (base 845) toward the 385-bp *Bgl* II fragment were generated by the method of Dale *et al.* (14). One of these deletion clones, with base 559 at the 3' end, was used to produce a single-stranded labeled DNA probe by primer extension (16) followed by digestion with *Sac* II (site at base 72). The labeled fragment was excised from a 5% polyacrylamide (60:1 acrylamide/*N,N'*-methylenebisacrylamide weight ratio) gel after electrophoresis under denaturing conditions and then was hybridized at 65°C with 60 μg of total RNA from *P. putida* cells harboring pAC27 grown with 3-chlorobenzoate. The results of digestion with 900 units of nuclease S1 are shown on the autoradiogram in Fig. 4. No protection was observed when the probe was hybridized with RNA isolated from cells grown with glucose (data not shown). Because the sequencing ladder bands used as size markers in the analysis contain the 17-base M13 primer as well as 18 nucleotides generated in cloning, a correction of 35 bases in the 3' direction of the RNA-complementary strand (SL) is necessary. Assignment of RNA initiation was correlated with the most intense (**) band after nuclease treatment, although a short series of minor bands around the central band is seen. This localized the major transcription initiation site at a deoxycytosine residue (base 356) on the 385-bp *Bgl* II fragment upstream of the 4.2-kbp *Bgl* II E fragment. There is some evidence that the *clc* gene cluster is under positive control, presumably by a trans-acting element (4). When the DNA sequence 5' to the transcription initiation site is compared to promoters of the *nah* (20), *sal* (20), and *xyl* (21) operons, sequences in the -10 region show considerable similarity (Fig. 5). Raibaud and Schwartz (22) demonstrated that the most highly conserved residues in positively regulated *E. coli* promoters are the adenine and thymine nucleotides in the -10 region and that these residues probably provide a minimal requirement for participation of RNA polymerase. This requirement appears to have been met also in the *Pseudomonas* promoters shown in Fig. 5, while very little similarity is seen in the -35 region.

Bgl II

AGATCTGGTAAAGTGGAGTGGCCCTGCTCACTTAAATGTGTACCCCGTCCCGTCAATTTGGAGCCGGGGTGGCTTCAACACGACCCAGCCGAGATCTCTTCCAGGG 120

CTGTATTTGGCGGATATGGCCGGCTGGAAATATCGACGGGCTCTTCCGGCAGCAGCGATATTTCCCTCTCTGGGACCGCGATGAATAGCAAGCTGGCGAAATCCATTAAGACC 240

CTCTGGTTTCCCTAGGGGTCACCCGCTAAGCCATCCGATCCCGTATTGCAAGGCTAAAAAGGATTGGACCGCATGACACGGGAATCTTAGATTTCATGTTTGAAGCACAAT 360

Bgl II

CATCGGTCTTCAACCATCAGATCTTGAAGAGGACGAGCTC ATG GAT AAA CGA GTT GCC GAG GTC GCA GGC GCG ATC CTC GAG GCA CTA CGC AAA ATT 458

Met Asp Lys Arg Val Ala Glu Val Ala Gly Ala Ile Val Glu Ala Val Arg Lys Ile

1 start c1cA 10

TTG CTG GAC AAG GCG GTC ACC GAA GCC GAA TAC GCG GCG GGT CTC GAC TAT CTC ACC GAG GTC CCA CAG ACC GCG GAA ACC GCG CTG CTT 548

Leu Leu Asp Lys Arg Val Thr Glu Ala Glu Tyr Arg Ala Gly Val Asp Tyr Leu Thr Glu Val Ala Gln Thr Arg Glu Thr Ala Leu Leu

20 30 40

CTG GAC GAT TTC CTG AAC ACC ACC ATC ATC GAA GGC AAG GCG CAG CCG TCG GCG ACC TCT GCG CCT GCG ATC CAG GCG GCG TAC TTC CTG 638

Leu Asp Val Phe Leu Asn Ser Thr Ile Ile Glu Gly Lys Ala Gln Arg Ser Arg Thr Ser Ala Pro Ala Ile Gln Gly Pro Tyr Phe Leu

50 60

GAA GGT GCT CCT GTA GTT GAA GGC GTC CTC AAG ACC TAC GAT ACC GAC GAC CAC AAA CCG CTG ATC ATT CCG GGT ACG GTG GCG TCG GAC 728

Glu Gly Ala Pro Val Val Glu Gly Val Leu Lys Thr Tyr Asp Thr Asp Asp His Lys Pro Leu Ile Ile Arg Gly Thr Val Arg Ser Asp

80 90

ACG GCG GAG TTG CTC GCT GCG GCT GTC ATC GAC CTG TGG CAC TCG ACC CCT GAT GCG TTG TAC ACG GCG ATC CAG GAC AAC ATC CCC GTG 818

Thr Gly Glu Val Ala Glu Ala Val Ile Asp Val Trp His Ser Thr Pro Asp Gly Leu Tyr Ser Gly Ile His Asp Asn Ile Pro Val

110 120 130

GAC TAC TAC CCG GGA AAA CTC GTG ACC GAT TCC CAG GCG AAC TAT CCG GTG GCG ACC ACG ATG CCA GTG CCA TAC CAG ATC CCC TAC GAG 908

ASP Tyr Thr Arg Gly Lys Leu Val Thr Asp Ser Gln Gly Asn Tyr Arg Val Arg Thr Thr Met Pro Val Pro Tyr Gln Ile Pro Tyr Glu

140 150 160

GGC CCG ACT GGG GGT CTG CTC GGC CAC CTG GCG ACC CAT ACC TGG CGT CCG GCG CAC GTG CAC TTC AAG GTG GCG AAG GAC GGT TTC GAA 998

Gly Pro Thr Gly Arg Leu Leu Gly His Leu Gly Ser His Thr Trp Arg Pro Ala His Val His Phe Lys Val Arg Lys Asp Gly Phe Glu

170 180 190

CCG TTG ACC ACG CAA TAC TAC TTC GAA GGG GCG AAA TGG CTG GAC GAT GAC TGC TGT CAC GGC GTC ACC CCC GAC CTG ATT ACG CCC GAG 1088

Pro Leu Thr Thr Gln Tyr Tyr Phe Glu Gly Gly Lys Trp Val Asp Asp Asp Cys Cys His Gly Val Thr Pro Asp Leu Ile Thr Pro Glu

200 210 220

ACC ATC GAG CAC GGG GTG CCG GTC ATG ACC CTG GAC TTC GTA ATC GAG CGT GAG CAG GCC GAG CAA CCG AAG TCG CCG ACT GAG ACA GTG 1178

Thr Ile Glu Asp Ala Glu Val Arg Val Ile Glu Arg Glu Gln Ala Glu Gln Arg Lys Ser Ala Thr Glu Thr Val

230 240 250

GCATG AAG ATC GAA GCG ATC GAT GTG ACG CTG GTG GAC GTC CCA CCT TCG CGT CCC ATC CAG ATG TCG TTT ACC ACG GTG CAG AAG CAG 1267

Ala * end c1cA

260 Met Lys Ile Glu Ala Ile Asp Val Thr Leu Val Asp Val Pro Ala Ser Arg Pro Ile Gln Met Ser Phe Thr Thr Val Gln Lys Gln 20

1 start c1cB

ACC TAT GCG ATC GTG CAC ATC GGT GCG GGC GGG CTT TCC GGC ATC GGC GAG GGC ACG ACC GTA GGT GGG CCG ACT TCG TCC GAA TGC 1357

Ser Tyr Ala Ile Val Gln Ile Ala Glu Gly Glu Cys Gly Ile Gly Glu Gly Ser Ser Val Gly Gly Pro Thr Trp Ser Ser Glu Cys

30 40

CGT GAA ACC ATC AAG GTC ATC ATC GAA ACG TAC TTG GCG CCG CTC CTC ATC GCG AAG GAC GCG ACA AAT CTC GGA GAG CTC CAG CAC TTA 1447

Ala Glu Thr Ile Lys Val Ile Ile Glu Thr Tyr Leu Ala Pro Leu Leu Ile Gly Lys Asp Ala Thr Asn Leu Arg Glu Leu Gln His Leu

60 70 80

ATG GAG GCG GCC GTA ACC CGA AAC TAT TCG GCC AAG GCG GCC ATC CAG GTT GCG CTG CAT GAT CTG AAG GCA CAC TCT CTG AAC CTG CCG 1537

Met Glu Arg Ala Val Thr Gly Asn Tyr Ser Ala Lys Ala Ala Ile Asp Val Ala Leu His Asp Leu Lys Ala His Ser Leu Asn Leu Pro

90 100 110

CTG ACC GAT TTG ATC GCG GCG CCG ATC CAG GCG ATC CCC ATT GCC TGG ACC CTG GCG ACC GCG GAC ACG CAG CCG GAC ATC GCA ATC 1627

Leu Ser Asp Leu Ile Gly Gly Ala Ile Gln Gln Gly Ile Pro Ile Ala Trp Thr Leu Ala Ser Gly Asp Thr Gln Arg Asp Ile Ala Ile

120 130 140

GCC GAG GAA ATG ATC GAG GCG GCG GCG CAC AAC CCG TTC AAG ATC AAC CTT GCC GTG CCG TCC CCG GCA GAT GAT TTG GCG CAT ATC GAG 1717

Ala Glu Glu Met Ile Glu Arg Arg Arg His Asn Arg Phe Lys Ile Lys Leu Gly Val Arg Ser Pro Ala Asp Asp Leu Arg His Ile Glu

150 160 170

AAG ATT ATC GAG CCG CTC GGT GAC CGT GCT GCG GTG GCG GTC GAT ATC AAC CAG GCC TGG GAT GAG AAC ACG GCA TCG GTG TGG ATT CCG 1807

Lys Ile Ile Glu Arg Val Gly Asp Arg Ala Ala Val Arg Val Asp Ile Asn Gln Ala Trp Asp Glu Asn Thr Ala Ser Val Trp Ile Pro

180 190 200

CGC CTG GAG GCG GCG GGT GTC GAA CTG GTC GAA CAG CCG GTG GCA GCG ACG AAC TTC GAT GCG CTT GCG GCG CTG TCG GCC AAC GCG 1897

Arg Leu Glu Ala Ala Gly Val Leu Val Glu Gln Pro Val Ala Arg Ser Asn Phe Asp Ala Leu Arg Arg Leu Ser Ala Asp Asn Gly

210 220 230

GTG GCC ATC CTG GCC GAT GAA ACC CTG ACC TCG CTG GCG TCC GCC TTC GAA CTG GCG CCG CAT CAT TCG GTC GAC GCG TTC TCG CTG AAG 1987

Val Ala Ile Leu Ala Asp Glu Ser Leu Ser Ser Leu Ala Ser Ala Phe Glu Leu Ala Arg His His Cys Val Asp Ala Phe Ser Leu Lys

240 250 260

CTG TGC AAC ATC GCG GCG GCG GCA AAT ACC CTC AAG CTC GCT GCG ATC GCG GAA GCG TCG GCG ATT GCG TCC TAT GCG GCG ACC ATG TTG 2077

Leu Cys Asn Met Gly Gly Val Ala Asn Thr Leu Lys Val Ala Ala Ile Ala Glu Ala Ser Gly Ile Ala Ser Tyr Gly Gly Thr Met Leu

270 280 290

GAT TCA TCA ATC GCG ACC GCT GCT CTC CAT GTG TAT GCC ACA TTG CCG ACC ATG CCC TTC GAA TGT GAA CTG CTA GCG CCC TGG GTG 2167

ASP Ser Ser Ile Gly Thr Ala Ile Leu His Val Tyr Ala Thr Leu Pro Thr Met Pro Phe Glu Cys Glu Leu Leu Gly Pro Trp Val

300 310 320

TTA GCC GAC ACG CTT ACG CAG ACC CAA CTC GAG ATC AAG GAC TTC GAC ATT CCG TTG CCC TCG GGT CCT GGG TTG GGT GTT GAT ATC GAT 2257

Leu Ala Asp Thr Leu Thr Gln Thr Gln Leu Glu Ile Lys Asp Phe Glu Ile Arg Leu Pro Ser Gly Pro Gly Leu Gly Val Asp Ile Asp

330 340 350

CCG GAC AAG CTG GCG CAC TTC ACC GCG GCG GGT TGA TTGAAGTCAGTTAGGGGAAAATCACCATGTTGCTCTGTTTATCAATTCGTTTGGTCTGCTGGCTAGGCTGC 2364

Pro Asp Lys Leu Arg His Phe Thr Arg Ala Gly * end c1cB

360 370

GTAGGCTTGAAGCTTCTGTCGAGTAAACGGGAGTCTGCTGAGACTACCCGAGCAAGCCCATTCGATGGATGCTTCCATTCGCGCGGGGGTGAATCGACATGGTTTCTGCTGCTTGGCC 2484

CAGCGGCTCTGGGAAAGCGTGGGCGACCGCTGACAGTGGAGATGCTCCGGGCTGGGGCGGATCAGCGGACGGGAGGTCGCGGGCTCCCGGCTCAGCGTTAAGCTGCTGCTGCTG 2604

AACTCGACAGGGCTCGAGTGGCAAGTGGTTTATTCGGAAGCTCCGCTAGATGCTCGCGAAAGCATTTGCCCGCTGGCGCTTGGGAACATCCCAAGCTTACTGCTGCTGCGCGCT 2724

GATTTCCCATACAAGATGCGCAGGAGCCTACTAGTGGTAGCGAAAGGCCAACCCCGGTAAGTGTCTATTCGCTCTGCGAGCATCCATCTGCGGCGCCCTCTGCTGCTG 2844

GCTCAGCGCGGGTTGACTGCTCCACGCTCCCTACAAGAGGAGCACCGCCAGCTGCGCGCCAGCTGATGCGGGGGAGCTCCATCCATGATGGTGGATTGATCAGCCGACAGAGTGGTTT 2964

ATCAAACTGGACGGTGGCTGGCTGGTGTCCACAGCTTCCAACCCGCGCCCTCATGCTGGGATTCCTCCCTGGCGCAAGCAGCCGACCAACCGAATTCGAGCTCTGACATGG 3084

TTTGACCTTACTCTGGGCTCACTCCGCGACATGTTCAAGTCTTCAACACAGCAATGAATGAAGCACTGAAGACCCCGGAGTCAAAAAGGCGCTGGCGGATATCCGCGCTTCA 3204

GCGGAGGGGGAACATCGCAAGCTCTGGGGTCTTATGGGACAAGAGATGCGACCGATGGGCGCATGATGCTCCACCATGACTCAACCAATGAATCGAGATGCGAAG 3324

ATG TTG ACT GAA GGG ATA TCG ATT CAA TCG TAT GAC GGG CAT ACA TTC GCG GCG CTC GTG GCG TCG CCG GCC AAA GCG CCC CCT CCC GTG 3414

Met Leu Thr Glu Gly Ile Ser Ile Gln Ser Tyr Asp Gly His Thr Phe Gly Ala Leu Val Gly Ser Pro Ala Lys Ala Pro Ala Pro Val

1 start c1cD 10 20 30

ATT GTG ATC GCT CAA GAA ATA TTT GGT GTG AAC GCG TTC ATG CGA GAA ACG GTG TCA TGG CTG CTC GAC CAG GGG TAT GCG GCA GTT TGC 3504

Ile Val Ile Ala Gln Glu Ile Phe Gly Val Asn Ala Phe Met Arg Glu Thr Val Ser Trp Leu Val Asp Gln Gly Tyr Ala Ala Val Cys

40 50 60

CCT GAT CTG TAC GCG CCG CAG GCG CCA GGT ACA GCA CTC GAT CCG CAG GAT GAG CCG CAG AGA GAG CAA GCG TAC AAG CTC TCG CAG GCC 3594

Pro Asp Leu Tyr Ala Arg Gln Ala Glu Tyr Thr Ala Leu Asp Pro Gln Asp Gly Thr Ala Leu Asp Pro Gln Asp Gly Ala Tyr Lys Leu Trp Gln Ala

70 80 90

TTC GAC ATG GAG GCG GCG GTG GCG GAT CTG GAG GCT GCT ATC CCG TAT GCG GCA CAC CAA CCC TAC AGC AAC GGC AAG GTG GGA TTG GTG 3684

Phe Asp Met Glu Ala Gly Val Gly Asp Leu Glu Ala Ala Ile Arg Tyr Ala Arg His Gln Pro Tyr Ser Asn Gly Lys Val Gly Leu Val

100 110 120

GGG TAT TCG GCG GGT GCG CTT GCG TTT CTA GTG GCG GCG AAA GGA TAC GTG GAT GCG GCG CTA GGC TAC TAC GGT CTT GGA CTG GAG 3774

Gly Tyr Cys Leu Gly Gly Ala Leu Ala Phe Leu Val Ala Ala Lys Gly Tyr Val Asp Arg Ala Val Gly Tyr Tyr Gly Val Gly Leu Glu

130 140 150

AAG CAG CTC AAG AAG CTC GCG GAA GTC AAG CAT CCG GCG TTG TTT CAC ATG GCG GCG CAA GAC CAC TTC GTG CCG GCG CCA ACG CCG CAG 3864

Lys Gln Leu Lys Val Pro Glu Val Lys His Pro Ala Lys Phe His Met Gly Gly Gln Asp His Phe Val Glu Gln Pro Ser Arg Gln

160 170 180

CTG ATT ACT GAA GCG TTC GCT GCG AAT CCA TTG CTG CAA GTG CAC TGG TAC GAA GAG CCG GGA CAC TCG TTC GCG AGG ACG ACC ACT TCG 3954

Leu Ile Thr Glu Gly Phe Gly Ala Asn Pro Leu Leu Gln Val His Trp Tyr Glu Glu Ala Gly His Ser Phe Ala Arg Thr Ser Ser Ser

190 200 210

GGC TAT GTG GCG AGT GCG GCG GCG TTG GGC AAC GAA GGT AGA CTG GAT TTC CTG GCG CCG TTG CAG ACG AAC CCA TGA ATTTTCATTAC 4046

Gly Tyr Val Ala Ser Ala Ala Ala Leu Ala Asn Leu Ala Ser Phe Leu Ala Pro Leu Gln Ser Lys Lys Pro * end c1cD

220 230

GACTATGTTGTCGGCGGTAATTTTGAACCGCACCGCTGGCCGATTCGCGAGAACTCGAAGCTTGGACATTTGACCGGGCGCTGCTGTTGACCGCGCTGACGAGGGCGCGCTGGG 4166

AGGCAAGTACGCGAGCGCGCTGATGGCCACCTGGCAGCTGCTTACGATGTTGGCACCATCGATGATACCGGCTTTGGTGGCTGAGGAGCGCTCAAGATTTGGCGCGCAGTGGAGCTAA 4286

TGGGCTGATGGATTTGGTGGGGATCTACTATGAGCTCGCTAAGATGCTGGCGCTCGGTAAGAGCTTGGCATGCTGCGTGGCCAGCAGATATGCGAGTGGAGATGAGCTGCA 4406

TCTTGGGTATCAGAGGGTGGCTCAAGAAACCGGGCGGAGCGCGGTGTCATGCGCCGTCGAGTTATATGACGACCGCGGCTGACCTTAGAGCTACCCCTTAGCATCAGCGTACCA 4526

GCGGATCAATGCAATGACCCATGCACTTGAAGCGCTTATGCGCCAGTGCACCGCGCTGCTTACCATGATGGCCGAGGAATGCGCAGCAGGTAAGCGGGATTTCCAGAAATGTA 4646

TCAATCAACAGAGATCT 4664

Bgl II

FIG. 3. Nucleotide sequence of chlorocatechol-degradative genes and DNA 5' to the initiation codon of the first gene. mRNA initiation site mapped by nuclease S1 protection is indicated at +1. Bgl II sites shown in Fig. 1 are indicated above the sequence. The first amino acid residue for each coding sequence identified by direct comparison with N-terminal amino acid sequence analysis of the purified protein is designated no. 1, and stop codons are indicated with asterisks. Ribosome-binding sequences for each gene are overlined. Possible start codons for ORF3 are indicated (>).

The complete nucleotide sequence of the *clc* genes revealed the homology of a pentapeptide sequence near a cysteine residue (Cys-60) of the dienelactone hydrolase (*clcD*

gene product) with the chromosomally encoded enol-lactone hydrolase, which otherwise shows no appreciable homology (11). Nucleotide sequence analysis of the chromosomal gene



FIG. 4. Nuclease S1 mapping of the transcription initiation site of the *clc* gene cluster. Lanes 1 and 5: DNA fragments protected by RNA from nuclease S1 digestion. Lanes G, A, T, and C: products of sequencing reactions of the same recombinant used to make the probe. The sequencing ladder (SL) is read 5' to 3' from bottom to top (shown enlarged) and represents the mRNA-complementary strand. Because the ladder bands each contain M13 primer nucleotides as well as intervening bases generated in deletion cloning, a correction of 35 bases in the 3' direction is necessary to align mRNA initiation with the sequence in the ladder. Double asterisk indicates the major protected band and is aligned after correction with the appropriate nucleotide on the DNA sense strand (+1 or base 356 in Fig. 3). To save space, the 487-base DNA probe, which migrates extremely high in the gel, is not shown.

catB (encoding cycloisomerase I) from *P. putida* demonstrated a high degree of homology (52%) with the *clcB* gene. Although the two enzymes encoded by these genes show divergent substrate specificities, long regions of amino acids are conserved throughout the two proteins, with the most homologous stretches localized to the middle portions of both (23). Similarly, catechol oxygenases from divergent backgrounds—namely, protocatechuate 3,4-dioxygenase and catechol oxygenase II—have been compared at the amino acid level, and homologous regions thought to be involved in the iron-binding domains of these proteins appear to be conserved (K.-L. Ngai, B.F., D. K. Chatterjee, L. N. Ornston, and A.M.C., unpublished results). Thus, comparison of nucleotide sequences of the plasmid-borne *clc* structural and regulatory genes with those of chromosomal genes specifying analogous reaction mechanisms or with other plasmid genes specifying catabolic activities against natural aromatic hydrocarbons is expected to throw considerable light on the mode of evolution of new degradative functions against new synthetic compounds.

	-35	-10
<u>nah</u>	ATTGACAAATAAAGCACGCTCACCATCATCGGGAATACA	
<u>sal</u>	TGTATTATCAATATTGTTGCTCCGTTATCGTTATTAACA	
<u>xylABC</u>	CGGTATAAGCAATGGCATGGCGGTTGCTAGCTATACGAGA	
<u>xylDEFG</u>	TGGCTATCTCTAGAAAGCCTACCCCTTAGGCTTTATGCA	
<u>clcABD</u>	ACCGCATGACACGCGAATCTTAGCAITTCATGTTTGAAGCACC	
<u>E. coli</u>	TTGACA (17±2)	TATAAT (5-8)
PRCS (<u>E. coli</u>)		A T
PRCS (<u>Pseudomonas</u>)		C T [*] A T A

FIG. 5. Comparison of five positively regulated promoters from *Pseudomonas* plasmids involved in metabolism of aromatic compounds. In each case the transcription initiation site was identified by either nuclease S1 or reverse transcriptase mapping [*nah* and *sal* (20); *xylABC* and *xylDEFG* (21)]. The most conserved residues, which are similar to those found in the *E. coli* consensus sequence, are clustered in the -10 region and are shown in boldface. PRCS (*E. coli*), positively regulated conserved sequence from *E. coli* promoters (discussed in ref. 22). PRCS (*Pseudomonas*), positively regulated conserved sequence derived by comparison of the promoters shown in the figure. Asterisk indicates deoxyadenosine residue conserved in all cases.

This research was supported by a grant from the National Science Foundation (DMB85-14671) and by a Public Health Service grant (ESO4050) from the National Institute of Environmental Health Sciences.

- Schneider, M. H. (1979) *Persistent Poisons: Chemical Pollutants in the Environment* (N. Y. Acad. Sci., New York).
- Chatterjee, D. K. & Chakrabarty, A. M. (1982) *Mol. Gen. Genet.* **188**, 279-285.
- Don, R. H. & Pemberton, J. M. (1981) *J. Bacteriol.* **145**, 681-686.
- Ghosal, D., You, I.-S., Chatterjee, D. K. & Chakrabarty, A. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1638-1642.
- Weightman, A. J., Don, R. H., Lehrback, P. R. & Timmis, K. N. (1984) in *Genetic Control of Environmental Pollutants*, eds. Omenn, G. S. & Hollaender, A. (Plenum, New York), pp. 47-80.
- Reineke, W. & Knackmuss, H.-J. (1980) *J. Bacteriol.* **142**, 467-473.
- Reineke, W. (1984) in *Microbial Degradation of Organic Compounds*, ed. Gibson, D. T. (Dekker, New York), pp. 379-421.
- Wheelis, M. L. & Ornston, L. N. (1972) *J. Bacteriol.* **109**, 790-795.
- Chakrabarty, A. M. (1982) in *Biodegradation and Detoxification of Environmental Pollutants*, ed. Chakrabarty, A. M. (CRC, Boca Raton, FL), pp. 127-139.
- Messing, J. (1979) *Recombinant DNA Tech. Bull.* **2**, 43-48.
- Frantz, B., Ngai, K.-L., Chatterjee, D. K., Ornston, L. N. & Chakrabarty, A. M. (1987) *J. Bacteriol.* **169**, 704-709.
- Sancar, A., Hack, A. M. & Rupp, W. D. (1979) *J. Bacteriol.* **137**, 692-693.
- Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101-106.
- Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31-40.
- Deretic, V., Gill, J. F. & Chakrabarty, A. M. (1987) *J. Bacteriol.* **169**, 351-358.
- Burke, F. B. (1984) *Gene* **30**, 63-68.
- Que, L. (1980) *Struct. Bonding (Berlin)* **40**, 38-72.
- Chatterjee, D. K. & Chakrabarty, A. M. (1983) *J. Bacteriol.* **153**, 532-534.
- Stanier, R. Y., Palleroni, N. J. & Doudoroff, M. (1966) *J. Gen. Microbiol.* **43**, 159-271.
- Schell, M. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 369-373.
- Inouye, S., Nakazawa, A. & Nakazawa, T. (1984) *Gene* **29**, 323-330.
- Raibaud, P. & Schwartz, M. (1984) *Annu. Rev. Genet.* **18**, 173-206.
- Aldrich, T. L., Frantz, B., Gill, J. F., Kilbane, J. & Chakrabarty, A. M. (1987) *Gene* **52**, 185-195.