1

PM-1 3 1/2" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Notice to the user

Notice à l'utilisateur

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

**THE BIOCHEMICAL PATHWAY AND CATABOLIC GENES FOR
CHLOROBENZOATE DEGRADATION ENCODED ON THE TRANSPOSON Tn5271
OF *ALCALIGENES* SP. STRAIN BR60**

**BY
CINDY HATSUYO NAKATSU**

**submitted to:
the OTTAWA-CARLETON INSTITUTE OF BIOLOGY
CARLETON UNIVERSITY**

**In partial fulfilment of the requirements
for the degree of Doctor of Philosophy**

**November 1992
©
Cindy H. Nakatsu**

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Number of copies ordered

Number of copies received

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315 84038 2

Canada

Carleton University
Ottawa, Canada K1S 5B6

Thesis contains black & white
photographs &/or explanatory tables
which when microfilmed may lose their
significance. The hardcopy of the
thesis is available upon request from
Carleton University Library.

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis,

**THE BIOCHEMICAL PATHWAY AND CATABOLIC GENES
FOR CHLOROBENZOATE DEGRADATION ENCODED ON
THE TRANSPOSON Tn5271 OF ALCALIGENES SP. STRAIN BR60**

submitted by

CINDY HATSUYO NAKATSU, B.Sc. M.Sc.

in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Chair, Department of Biology
John Webb, Ph.D.

Thesis Supervisor
Campbell Wyndham, Ph.D.

External Examiner
Charles Greer, Ph.D.

Carleton University

December, 1992

ABSTRACT

Alcaligenes sp. strain BR60 is a eubacterial isolate of the β -Proteobacteria from Bloody Run Creek, a chemically contaminated tributary of the Niagara River. This organism was selected for its ability to utilize 3-chlorobenzoate as a carbon source. Previous studies have shown that information vital for the degradation of 3-chlorobenzoate by this organism is carried on a transposon which is part of a conjugative, broad-host-range plasmid, pBRC60 (Wyndham and Straus 1988, Wyndham et al. 1988).

This study shows that *Alcaligenes* sp. strain BR60 initiates catabolism of 3-chlorobenzoate by the action of a dioxygenase enzyme that introduces molecular oxygen at the meta- and ortho-positions. This is the first report of such a pathway and its expression appears to be limited to members of the Γ and β Proteobacteria that cleave protocatechuate (3,4-dihydroxybenzoate), an intermediate of the pathway, using the meta-ring-fission route.

The description of the Tn5271 element in this thesis and related publications is the first for a transposon carrying information for chloroaromatic catabolic activities. Tn5271 has unique features that show relationships to both the class I and class II transposons. It fits the description of the proposed ancestral structure of the Tn3 group of transposons.

The genetic information for the catabolic activity of

the transposon is carried on a 3.7 kb *NotI*-*EcoRI* fragment within the non-repeated region of Tn5271. Expression of these genes in *Escherichia coli* revealed two unique polypeptides 42 and 51 kd in size. Nucleotide sequence analysis of this region revealed three potential genes designated *cbaA*, *cbaB* and *cbaC*. Comparisons to other dioxygenase genes showed that the deduced amino acid sequence of *cbaA* has conserved amino acids typical of a Rieske-type [2Fe-2S] centre, much like that found in the terminal oxygenase component of other dioxygenases. The comparison to the databases of the translated *cbaB* sequence revealed conserved cysteine residues indicative of a plant-type ferredoxin centre, and potential NAD- and FAD-binding sites found in the reductase component of some dioxygenases. The sequence of the deduced amino acids of *cbaC* had similarities to several oxidoreductases, and the results indicated that this gene may code for a dehydrogenase.

ACKNOWLEDGEMENTS

I came to Carleton University specifically to work with Cam Wyndham because I felt he was a person I would respect and a supervisor I could count on during the worst hours, days... of my Ph.D thesis. He did not let me down. I have truly enjoyed my stay in his laboratory and the interactions with all those that work in his laboratory and in ELBA, they are too numerous to mention here.

I also appreciated the molecular genetics advice and every day interactions with Bob Iyer and John Vierula. I would also like to thank Neil Straus, Bess Wong and the members of their laboratory at the University of Toronto for the use of their facilities and many helpful hints they were always willing to offer. I would also like to thank Clem Kazekoff for running my GC-MS samples and assisting in their interpretation. Financial assistance was provided by scholarships from NSERC and OGS. The superb photography was provided by Joan Mallett.

Finally, I would like to thank my family and friends from HCI, U of T, baseball and darts for their support and the confidence they have shown in me throughout this thesis and everything leading up to it. I would especially like to thank Greg for his love, support and for performing more than his share of the household duties.

TABLE OF CONTENTS

Description	Page
Thesis Acceptance Form	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
 CHAPTER 1: GENERAL INTRODUCTION	 1
INTRODUCTION	1
OBJECTIVES	16
 CHAPTER 2: PHYSICAL STRUCTURE OF pBRC60 AND Tn5271	 17
INTRODUCTION	17
MATERIALS AND METHODS	20
BACTERIAL STRAINS AND PLASMIDS	20
GROWTH CONDITIONS	20
DNA ISOLATION	22
DNA MANIPULATION FOR PLASMID MAPPING	22
FILTER BLOTTING AND HYBRIDIZATION	23
PLASMID MAPPING STRATEGY	24
DNA MANIPULATIONS FOR TRANSPOSON MAPPING AND CLONING	 24
DNA SEQUENCING	25
RESULTS	26
MAPPING OF THE PLASMID pBRC60	26
MAPPING OF TRANSPOSON Tn5271	27
DISCUSSION	39
 CHAPTER 3: BIOCHEMICAL PATHWAY FOR 3-CHLOROBENZOATE DEGRADATION	 48
INTRODUCTION	48
MATERIALS AND METHODS	52
BACTERIAL STRAINS AND GROWTH CONDITIONS	52
ANALYTICAL METHODS AND CHEMICALS	53
DNA MANIPULATIONS	54
RESULTS AND DISCUSSION	55
 CHAPTER 4: CLONING AND EXPRESSION OF 3-CHLOROBENZOATE DEGRADATION GENES ENCODED ON Tn5271	 67
INTRODUCTION	67
MATERIALS AND METHODS	73
BACTERIAL STRAINS AND GROWTH CONDITIONS	73
CONSTRUCTION OF VECTOR pBW13	73

	Page
DNA MANIPULATIONS	77
GENE EXPRESSION	78
RESULTS	80
ISOLATION OF <i>cba</i> GENES	80
PROTEIN EXPRESSION	83
METABOLITES FORMED <i>IN VIVO</i>	83
DISCUSSION	87
 CHAPTER 5: NUCLEOTIDE SEQUENCE OF <i>cbaABC</i> GENES	 96
INTRODUCTION	96
MATERIALS AND METHODS	99
STRAINS AND GROWTH CONDITIONS	99
DNA MANIPULATIONS	102
DNA SEQUENCING AND ANALYSIS	102
RESULTS	103
NUCLEOTIDE SEQUENCE ANALYSIS	103
CODON USAGE AND G+C CONTENT	110
SIMILARITIES TO OTHER OXYGENASES	110
MUTATIONS WITHIN <i>cbaB</i> and <i>cbaC</i>	121
DISCUSSION	124
 CHAPTER 6: A Summary	 135
Objective 1	136
Objective 2	136
Objective 3	137
Objective 4	138
FUTURE RESEARCH PROJECTS	139
 APPENDICES	 144
APPENDIX 1: Nucleotide Sequence of Tn5271	144
APPENDIX 2: GC-MS ANALYSES	160
 REFERENCES	 204

LIST OF TABLES

	Description	Page
Table 2.1:	Bacterial Strains and Plasmids	21
Table 2.2:	Restriction Fragment Sizes of pBRC60	30
Table 2.3:	Restriction Enzyme Fragment sizes of pBRC40	31
Table 3.1:	Oxygen Uptake Rates by <i>Alcaligenes</i> sp. BR60	59
Table 5.1:	Distribution of codons in the <i>cbaABC</i> genes	111

LIST OF FIGURES

Description	Page
Figure 1.1: Examples of Intermediates Formed in the Metabolism of Complex Aromatic Compounds	2
Figure 1.2: Biodegradation mechanisms	5
Figure 2.1: Restriction Enzyme Digest Map of Plasmid pBRC60	28
Figure 2.2: Restriction Enzyme Digest Map of Tn5271	32
Figure 2.3: Nucleotide sequence of IS1071	35
Figure 2.4: Comparison of Inverted Repeat Sequences	37
Figure 2.5: Alignment of TnpA Sequences	40
Figure 2.6: Evolution of class II transposable elements	46
Figure 3.1: Structural map and the metabolic pathway for 3-chlorobenzoate encoded by Tn5271.	56
Figure 3.2: The phylogenetic distribution of isolates expressing the Tn5271-encoded dioxygenase	63
Figure 4.1: Degradation Pathways of 3-Chlorobenzoate	69
Figure 4.2: Construction of pBW13	74
Figure 4.3: Cloning Map of Tn5271	81
Figure 4.4: 7.5% SDS-PAGE gel resolving unique polypeptides expressed by pBRCN13a	84
Figure 5.1: Location of <i>cbaABC</i> on Tn5271	100
Figure 5.2: Nucleotide and deduced Amino Acid Sequence of <i>cbaABC</i>	104
Figure 5.3: Alignment of CbaA to VanA and Component A of Sulphobenzoate dioxygenase	112
Figure 5.4: Alignment of oxygenase component of <i>cis</i> -diol forming dioxygenases	115
Figure 5.5: Alignment of Amino Acid Sequence of CbaB and VanB	117
Figure 5.6: Proposed conserved regions in the reductase component of dioxygenases	119
Figure 5.7: Alignment of CbaC to other reductases	122

LIST OF ABBREVIATIONS

3-Cba: 3-chlorobenzoate
2,4-D: 2,4-dichlorophenoxyacetic acid
2,4,5-T: 2,4,5-trichlorophenoxyacetic acid
bp: base pairs
d: daltons
GC-MS: gas chromatography-mass spectrometry
HPLC: high pressure liquid chromatography
IHF: integration host factor
IPTG: isopropylthio- β -D-galactoside
IR: inverted repeat sequence
IS: insertion sequence
kb: kilobase pairs
LB: Luria Bertani, bacterial growth medium
NAH: plasmid carrying naphthalene catabolic genes
nm: nanometre
ORF: open reading frame
PAGE: polyacrylamide gel electrophoresis
SDS: sodium dodecyl sulphate
TMS: trimethylsilyl
TOL: plasmid carrying toluene catabolic genes
TYE: tryptone-yeast extract, bacterial growth medium
UV-VIS: ultra violet-visible wavelengths

CHAPTER 1: GENERAL INTRODUCTION

INTRODUCTION

The surge in technology and industrialization since the second world war has resulted in the production of enormous quantities and varieties of synthetic organic chemicals. These chemicals have been produced for uses such as pesticides, herbicides, solvents and refrigerants. Chlorinated aromatics are one of the basic structures that have been utilized in the construction of these compounds. These anthropogenic chemical compounds have structures that are not normally a part of the natural assortment of chemicals found in the environment, earning them the name of xenobiotics. They tend to be more resistant to microbial degradation than their natural counterparts. This has led to problems in treating wastes contaminated with industrial chemicals and has resulted in their accumulation in the environment. As many of these chemicals are toxic and may accumulate in the tissues of organisms, they present a significant hazard.

In general, the greater the structural similarity between a xenobiotic compound and a naturally occurring compound, the greater the probability that an existing microbial enzyme system can be directly used or modified to enable either cometabolism or metabolism of that chemical. The degradation pathways of chloroaromatic pollutants often

**Figure 1.1: Examples of Intermediates Formed in the
Metabolism of Complex Aromatic Compounds**

Complex chloroaromatic pollutants can be metabolized or
cometabolized into chlorinated aromatic carbonic acids.

(from Reineke and Knackmuss 1988).

converge at simple halogenated benzene derivatives (Figure 1.1). The mechanisms that have been documented for biodegradation of haloaromatics have been reviewed by Reineke and Knackmuss (1988) and are summarized here. In prokaryotes, four possible mechanisms are described (Figure 1.2): a) displacement of halogen by hydrogen, b) displacement of halogen by a hydroxyl group, c) oxygenolytic halogen-carbon bond cleavage and d) chloride elimination from non-aromatic intermediates.

The displacement of the halogen through hydrogen is an anaerobic process that has been reviewed by Evans and Fuch (1988) and will not be discussed here.

In the second mechanism, halogen is replaced in the first step by a hydroxyl group that is donated from water, thus forming a hydroxybenzoate. Although this pathway may pertain to 3-chlorobenzoate degradation (Johnston et al. 1972), the pathway has been studied more extensively in the case of 4-chlorobenzoate degradation (Savard et al. 1986, Scholten et al. 1991).

Halogen substituents can also be removed fortuitously by the dioxygenation of haloaromatics (Reineke and Knackmuss 1988). This oxygenolytic attack generates catechols that may then be mineralized by existing aromatic pathways. This mechanism was used in the degradation of 2-chloro-, 2-bromo- and 2-fluorobenzoate by *Pseudomonas putida* CLB250 (Engesser and Schulte 1989).

Figure 1.2: Biodegradation mechanisms

1.2a. Displacement of halogen by hydrogen. Complete mineralization in this anaerobic process generally requires a consortium that includes bacteria able to degrade the product, benzoate, anaerobically.

1.2b. Displacement of halogen by hydroxyl, using water as a hydroxyl donor and not molecular oxygen.

1.2c. Displacement of halogen fortuitously by dioxygenases. During degradation of 2-fluorobenzoate both pathways occur but most of the substrate is defluorinated (upper pathway).

1.2d. Displacement of halogen from nonaromatic intermediates. The chlorides are removed spontaneously after the *ortho*-cleavage of chlorocatechols (from Reineke and Knackmuss 1988).

a.

b.

c.

d.

The final mechanism is the spontaneous removal of the halogen after the ortho-ring-fission of halocatechols by a process of cycloisomerization to form muconolactone intermediates (Reineke and Knackmuss 1978a, b).

The majority of xenobiotic-, and in general, aromatic-degrading bacteria that have been studied have been assigned to the genus *Pseudomonas* (Sayler et al. 1990). As the phylogenetic reclassification of bacteria proceeds, it is becoming increasingly apparent that this genus name has been an expedient choice for the grouping of bacteria for which insufficient taxonomic detail exists. The genus is now being fragmented according to both phenotypic and phylogenetic data, and many of the so-called pseudomonads are acquiring new family, genus and species names (De Ley 1992, Palleroni 1992). It is now recognized that there exists an enormous amount of diversity in the microbes that degrade xenobiotics, and in particular, chlorinated aromatics (Busse et al. 1992).

Most xenobiotic degrading microbes have been isolated using culture enrichment techniques. This involves growth of a consortium of microbes from a contaminated source in serial batch cultures or continuous cultures with gradual increases in concentration of the chemical of interest. Enrichment is continued until the chemical is the growth limiting source of an essential nutrient, for example carbon (Reineke and Knackmuss 1988). Using these approaches, three

bacterial strains capable of 3-chlorobenzoate degradation have been isolated and described extensively in the literature with regard to the biochemistry of aromatic ring-fission and their genetics. Strain B13 of *Pseudomonas* sp. was isolated from the Gottingen sewage plant in Germany by Dorn et al. (1974). The cultures were enriched with only 3-chlorobenzoate but the isolate was found to catabolize both benzoate and 3-chlorobenzoate. A strain of *Pseudomonas putida*, AC858 was also isolated from sewage, but it originated in Niskayuna, New York (Chatterjee et al. 1981). In Australia, six different strains of *Alcaligenes eutrophus* or *Alcaligenes paradoxus* were isolated from soil (Pemberton and Fisher 1977). Although 2,4-dichlorophenoxyacetic acid (2,4-D) was used as the only selection factor in the enrichment cultures, all the isolates had the ability to degrade 3-chlorobenzoate and 4-chloro-2-methylphenoxyacetic acid (MCPA). Some strains were also resistant to mercuric ions, phenylmercuric acetate and merbromin (Don and Pemberton 1981). From this group one strain of *Alcaligenes eutrophus*, JMP134, has been studied extensively. In addition to these isolates there have been reports of the degradation of 3-chlorobenzoate by *Pseudomonas alcaligenes* (Focht and Shelton 1987), *P. aeruginosa* JB2 (Hickey and Focht 1990), *P. putida* P111 (Hernandez et al. 1991), *Pseudomonas* sp. (Johnston et al. 1972), *Pseudomonas* sp. (Taeger et al. 1988), and *Acinetobacter calcoaceticus*

(Zaitsev and Karasevich 1985).

Microbes capable of complete mineralization of haloaromatic compounds must possess enzymes capable of removing the halogen substituent at some point in the metabolism. Although there are several potential pathways for 3-chlorobenzoate degradation, as discussed above, most of the genetic studies to date have focused on the pathways that eliminate chloride after the *ortho*-ring-fission of chlorocatechols (Reineke and Knackmuss 1978a, b; Schmidt and Knackmuss 1980). Analysis of the metabolites produced by *Pseudomonas* sp. strain B13 when grown with 3-chlorobenzoate as the sole carbon source indicated that *ortho*-ring-fission of chlorocatechols was the pathway utilized by this microbe to catabolize the compound. The metabolites and the enzymes involved in this pathway have been studied extensively (Reineke and Knackmuss 1978a, b; Dorn and Knackmuss 1978a, b; Schmidt et al. 1980). 3-Chlorobenzoate was converted to chlorocatechols via the intermediate chlorosubstituted 1,2-dihydro-1,2-dihydroxy-cyclohexa-3,5-diene by means of a benzoate 1,2-dioxygenase. The pathway utilized two routes, one forming 4-chlorocatechol and the other 3-chlorocatechol, the latter being the major product (Dorn et al. 1974). They found that the three next steps required specialized enzymes that were adapted to the metabolism of halogenated catechols. The 3- and 4-chlorocatechols were *ortho*-cleaved and converted into 2- and 3-chloro-*cis,cis*-muconates by the

addition of oxygen by a pyrocatechase II (catechol dioxygenase II). The next step was common for the two pathways. The enzyme muconate cycloisomerase II cycloisomerizes 2- and 3-chloro-*cis,cis*-muconate forming a dienelactone. During or after this process the halide is spontaneously removed. Finally, a dienelactone hydrolase converts the dienelactone to maleylacetate which can then enter the β -ketoadipate pathway.

The other extensively studied 3-chlorobenzoate degrading isolate, *Pseudomonas putida* AC858, utilizes the same pathway (Chatterjee et al. 1981). In *Alcaligenes eutrophus* JMP134, 2,4-D degradation parallels the 3-chlorobenzoate catabolic pathway (Don and Pemberton 1981). Detailed investigation of the relationship between degradation of 3-chlorobenzoate, 4-chlorobenzoate and 2,4-D in JMP134 indicated that the pathways converged at the formation of halocatechols. Until that point, existing enzymes that were normally produced for the degradation of non-halogenated aromatics were nonspecific and functional on halogenated compounds. Specialized enzymes similar to those isolated from *Pseudomonas* sp. strain B13 were required to continue the degradation. Although these enzymes are analogous to the existing enzymes for non-halogenated aromatics, they appear to be quite specialized since the two can not be interchanged (Dorn et al. 1974, Don et al. 1985). These findings illustrate that not all the enzymes used in

the 3-chlorobenzoate catabolic pathway are specialized for haloaromatic catabolism, but may be functional to some degree on non-halogenated compounds. The specialized enzymes are similar to existing enzymes in terms of catalytic function but their substrates are not interchangeable.

Biochemical studies have illustrated the unique nature of the enzymes involved in halocatechol degradation and that the presence of this pathway determines whether the microbe is capable of ultimately mineralizing a haloaromatic compound. Genetic analysis of 3-chlorobenzoate degrading *Pseudomonas* strains B13 and AC858 indicates that the genes specifying conversion of 3-chlorobenzoate to 3-chlorocatechol are encoded on their chromosome (Ghosal et al. 1985). The genes for chlorocatechol degradation are organized in operons on plasmids. These plasmids are pB13 in *Pseudomonas* sp. strain B13 (Chatterjee and Chakrabarty 1983a), pAC25 (or pAC27) in *P. putida* AC858 (Chatterjee et al. 1981, Chatterjee and Chakrabarty 1982, 1983b) and pJP4 in *Alcaligenes eutrophus* JMP134 (Don and Pemberton 1985).

Specific loci on these plasmids involved in 3-chlorocatechol degradation were first identified by cloning and transposon mutagenesis (Chatterjee and Chakrabarty 1983b, Don and Pemberton 1985, Weisshaar et al. 1987). Nucleotide sequences and deduced amino acid sequences have been determined for genes encoded on pAC25 (Frantz and

Chakrabarty 1987) and pJP4 (Don et al. 1985, Ghosal and You 1988, 1989, Perkins et al. 1990). There are three structural genes involved in the conversion of chlorocatechol to maleylacetate: pyrocatechase II or chlorocatechol-1,2-dioxygenase (*clcA*, *tfdC*), cycloisomerase II or chloromuconate cycloisomerase (*clcB*, *tfdD*), and hydrolase II or dienelactone hydrolase (*clcD*, *tfdE*). The *clc* genes are encoded on pAC25 and the *tfd* genes are on pJP4. Some features common to the two operons that are likely to be of evolutionary significance are: the homology of the nucleotide and amino acid sequences; the coding length of each gene; the four nucleotide basepair overlap between the pyrocatechase and cycloisomerase genes; and the TGA pyrocatechase stop codon (Perkins et al. 1990).

Although the genes in B13 have not been sequenced, preliminary comparisons of pAC25 and pB13 show high homology with DNA-DNA hybridizations and an identical restriction pattern, with the exception of a 6 kb deletion in pB13 (Chatterjee and Chakrabarty 1983a). Frantz et al. (1987) have also shown that the deduced amino acid sequence of *clcD* (encoded on pAC25) is identical to the first 34 residues of the dienelactone hydrolase purified from strain B13, leading them to speculate that the operons are identical.

Ghosal et al. (1985) compared pJP4 to pAC25 and found a high degree of homology in their DNA-DNA hybridizations. The subsequent sequencing of both operons has shown that

clcA has 63% nucleotide sequence homology and 60% amino acid sequence homology to *tfdC* (Ghosal and You 1988), *clcB* has 63% nucleotide sequence and 62% amino acid sequence homology to *tfdD* (Ghosal and You 1989) and *clcD* has 53% nucleotide sequence homology to *tfdE* (Perkins et al. 1990).

These observations have contributed to the hypothesis that the *clc* and *tfd* genes evolved from a common ancestor, possibly genes involved in the degradation of analogous non-halogenated aromatics. Comparisons of deduced amino acid sequences of genes involved in benzoate degradation have shown that *catA*, encoding catechol-1,2-dioxygenase from *Acinetobacter calcoaceticus*, has 31% homology to *clcA*, the chlorocatechol-1,2-dioxygenase gene of *P. putida* (Neidle et al. 1988). The *catB* gene encoding *cis,cis*-muconate lactonizing enzyme isolated from *P. putida* and the *clcB* gene for 3-chloro-*cis,cis*-muconate lactonizing enzyme show even greater homology. There is 52% nucleotide sequence homology and 41% homology of the deduced amino acid sequence (Aldrich et al. 1987). Thus, the evidence in the literature indicates that the chlorocatechol catabolic genes show significant genetic similarity to one another and, to a lesser but still significant extent, to analogous genes of nonhalogenated aromatic pathways.

The presence of xenobiotic chemicals in the environment and their recalcitrant nature has led to concerns about their accumulation. It is important to determine the extent

of biodegradation that can occur in nature and potentially remedy the problems of accumulation by treating the waste on-site with specially enriched or modified bacteria. One approach to this has been demonstrated in the laboratory by utilizing aromatic catabolic genes to expand the substrate range of microbes to enable biodegradation of novel isomers of halogenated and methylated aromatic compounds. Both *in vivo* and *in vitro* techniques have been successfully used to create novel catabolic pathways by recombining genes from existing aromatic pathways (Chapman 1987). One of the main restrictions of *in vitro* manipulations is the limited number of organisms available in which the genes for aromatic degradation have been identified.

Many aromatic chemicals have degradation pathways that are convergent. The most studied route, as discussed above, converges on chlorocatechols, and the genes encoding enzymes involved in 3-chlorobenzoate degradation via chlorocatechols have been well characterized. The literature for this particular pathway shows that the number of strains studied is limited and the genes that have been found are highly similar. However, there are a number of alternatives to catechol formation in benzoate degradation which have been known for some time (Stanier et al. 1966). With this knowledge in hand, research to explore the nature of chloroaromatic degradation in waters contaminated by an existing industrial landfill was undertaken.

Previous studies by Wyndham et al. (1988) demonstrated the existence of a 3-chlorobenzoate degrading bacterial strain, *Alcaligenes* sp. strain BR60. It was isolated from a 3-chlorobenzoate continuous enrichment culture inoculated with a sediment/water mixture from Bloody Run Creek. This creek, is a small tributary of the Niagara River, contained surface water and groundwater discharges from the Hyde Park industrial landfill. The predominant contaminants from Hyde Park include a mixture of phenols, chlorobenzoates, hydrocarbons and some metals.

The ability to degrade 3-chlorobenzoate was lost at a frequency of about $1 \times 10^{-3} \cdot \text{cell}^{-1} \cdot \text{generation}^{-1}$ when *Alcaligenes* sp. BR60 was grown without the selective carbon source, forming strains BR6007 (in pure culture) and BR40 (from the original continuous culture enrichment) (Wyndham et al. 1988). Strain BR60 harbours a plasmid pBRC60 (formerly pBR60) that was found to have a 14 kb deletion in strains BR40 and BR6007. This deletion plasmid has been designated pBRC40 (formerly pBR40). The deletion was found to be a spontaneous loss of a single fragment of unique DNA of approximately 14 kb from pBRC60. This DNA was flanked by direct repeats of 3.2 kb of DNA. The entire element, which apparently carries genes essential to 3-chlorobenzoate catabolism, has been designated transposon Tn5271. The lack of hybridization, even at low stringency, of the DNA of *Alcaligenes* sp. strain BR60 plasmid pBRC60 to the known *clc*

(chlorocatechol) catabolic genes of *Pseudomonas putida*, indicated the unique nature of these genes (Wyndham *et al.* 1988). The research presented in this Thesis was undertaken to characterize the genes encoded on the transposon Tn5271 that are involved in 3-chlorobenzoate degradation.

OBJECTIVES

The presence of haloaromatic compounds in the environment has resulted in the evolution of novel catabolic pathways. My hypothesis is that genes involved in one of these pathways are encoded on the transposon Tn5271 in *Alcaligenes* sp. strain BR60. Thus the objectives of this research are:

1. To determine the physical structure of the plasmid, pBRC60 and the transposon, Tn5271.
2. To determine the biochemical pathway utilized by *Alcaligenes* sp. BR60 to degrade 3-chlorobenzoate.
3. To determine the genes present on Tn5271 that are involved in 3-chlorobenzoate degradation.
4. To determine the relationship of these genes to genes for other aromatic catabolic pathways that have been characterized.

CHAPTER 2: PHYSICAL STRUCTURE OF pBRC60 AND Tn5271

INTRODUCTION

Two of the most effective pressures driving natural selection in bacterial populations are the widespread use of antibiotics and the provision of unusual organic carbon sources in the form of environmental contaminants. In response to these pressures, bacteria have evolved varied mechanisms of resistance to antibiotics and novel means of catabolizing unusual organic compounds. It is not surprising that the molecular mechanisms of adaptation to antibiotics and xenobiotics are similar. Plasmids have been implicated in the catabolism of complex organic compounds for two decades (Chakrabarty and Gunsalus 1971, Williams and Murray 1974, Saylor et al. 1990). There is interest in determining the incompatibility group and host range of these plasmids in order to determine the potential for transfer and stability of these catabolic traits in the environment (Fulthorpe and Wyndham 1991).

The rearrangement of catabolic genes along with changes in their expression have been implicated as mechanisms in the evolution of bacteria exposed to novel carbon sources (Bayley et al. 1977). The discovery of transposable elements specifying resistance to antibiotics (Hedges and Jacob 1974) was followed within a very few years by the first descriptions of transposition of the determinants for

toluene degradation carried on plasmid pWWO (TOL) in *Pseudomonas putida* mt-2 (Jacoby et al. 1978, Nakazawa et al. 1978, Chakrabarty et al. 1978). The presence of catabolic genes on transposons increases the potential of transferring them into hosts where the conjugative plasmid which acts as the vector for transfer can not be stably maintained.

Bacterial transposable elements fall into two well defined structural classes (Kleckner 1981). Class I elements include insertion sequences (IS), containing the genetic determinants for transposition only, and composite transposons formed when flanking IS elements mobilize an intervening sequence (Syvanen 1988). Class II elements, or the Tn3-family transposons, are related by inverted repeat (IR) similarities, transposase amino acid sequence homologies and transposition mechanism (Sherratt 1989). The class II catabolic transposons that have been characterized encode lactose metabolism (Tn951; Michiels et al. 1987) and toluene metabolism (Tn4653; Tsuda et al. 1989). Both are complex, nested class II elements. The naphthalene catabolic genes on plasmid NAH7 have recently been localized to a defective class II transposon, Tn4655, that lacks a transposase function but has an intact resolution system (Tsuda and Iino 1990).

Catabolic genes specifying the biodegradation of chlorinated aromatic compounds have been observed to undergo rearrangements, however most of these have not been shown to

transpose as a defined element. Tomasek et al. (1989, Haugland et al. 1990) mapped a 1,477 bp IS element flanking 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) catabolic genes of a *P. cepacia* strain, suggesting a possible class I composite transposon structure. During the course of this Thesis research there appeared a report of a functional composite transposon, Tn5280, encoding chlorobenzene catabolic genes, found on plasmid pP51 in *Pseudomonas* sp. strain P51 (van der Meer 1991). The flanking IS elements of Tn5280, IS1066 and IS1067, differ by a single base pair and show similarity to the *RhsA* elements of *Bradyrhizobium japonicum* (61.5% DNA sequence identity) and weak similarity to IS630 of *Shigella sonnei* (21.8% in 312 amino acid overlap).

In 1988 our laboratory published work on an *Alcaligenes* sp. strain BR60, isolated from a chlorobenzoate contaminated landfill, that undergoes deletion of chlorobenzoate catabolic genes from the indigenous plasmid pBRC60 (formerly pBR60), and exhibits recombination of plasmid sequences into the chromosome (Wyndham et al. 1988). This host-plasmid association was used to study gene transfer in the environment (Fulthorpe and Wyndham 1991). In the following chapter I describe the structure of the plasmid, pBRC60 and the chlorobenzoate catabolic transposon in *Alcaligenes* sp. BR60, the nucleotide sequence of the flanking direct repeats in this element, and their relationship to both class I and

class II transposable elements. This chapter will give the reader some background into the elements carrying the genes involved in 3-chlorobenzoate degradation. The transposition properties of Tn5271, and the mobility, host range and incompatibility functions of pBRC60 are being investigated by other researchers in our group.

MATERIALS AND METHODS

BACTERIAL STRAINS AND PLASMIDS

The various bacterial strains, plasmids and their sources used in this study are listed in Table 2.1.

GROWTH CONDITIONS

Alcaligenes strains were grown at 25°C, on Luria Bertani broth (LB) (Sambrook et al. 1989), TYE (1% Tryptone, 1% Yeast Extract, 0.5% NaCl) or defined minimal medium A (Wyndham 1986) with 4 mM 3Cba added as its carbon source for plasmid or total genomic DNA isolations. *E. coli* strains JM109, HB101 and DH5 α , with their plasmids pUC18, pUC19, pDPL13, or phage M13mp18 and M13mp19, were grown in LB at 37°C, with addition of required antibiotics as listed in Table 2.1. Filter sterilized ampicillin or kanamycin sulfate (Sigma Chem. Co.) were added at 50 $\mu\text{g}\cdot\text{mL}^{-1}$ concentration.

Table 2.1: Bacterial Strains and Plasmids

Strain	Plasmid & Phenotypes*	References
<i>E. coli</i>		
DH5 α	-, (Cm ^r , Km ^r , Tc ^r)	Hanahan 1983
HB101(pUC18)	pUC18, (β -gal, Ap ^r)	Yannisch-Perron et
HB101(pUC19)	pUC19, (β -gal, Ap ^r)	al. 1985
	phage M13mp18	Yannisch-Perron et
	phage M13mp19	al. 1985
	pDPL13, (Ap ^r)	Gendel et al. 1983
<i>Alcaligenes sp.</i>		
BR40	pBRC40, (3Cba ⁻)	Wyndham & Straus
BR60	pBRC60, (3Cba ⁺)	1988

* Ap, ampicillin; Cm, chloramphenicol; Km, kanamycin; Tc, tetracycline; β -gal, β -galactosidase; 3Cba, 3-chlorobenzoate; -, no plasmid present

DNA ISOLATION

Plasmids were isolated using the alkaline lysis method (Birnboim and Doly 1979) for minipreps. For larger volumes, a modified, scaled up version was utilized that included the use of polyethylene glycol (PEG 6000) to further purify the plasmid (Krieg and Melton 1988). Single stranded M13mp18 and M13mp19 were isolated by the method of Dale et al. (1985).

DNA MANIPULATION FOR PLASMID MAPPING

Plasmids were digested using restriction endonucleases under conditions recommended by the manufacturer. Restriction enzymes used were: *AvaII*, *BamHI*, *BclI*, *BglII*, *EcoRI*, *HindIII*, *NheI*, *NotI*, *NruI*, *PstI*, *PvuII*, *SalI*, *SmaI* and *XbaI* (Bethesda Research Laboratories, BRL) and *SacI* (Boehringer Mannheim). The DNA fragment sizes were distinguished using agarose gel electrophoresis as described in Sambrook et al. (1989). Routinely 0.4-0.7% agarose (Bio Rad) gels submerged in Tris-acetate buffer (TAE) were employed. The fragment sizes based on electrophoretic mobility were calculated using the IBM PC program Bandsize (University of Toronto).

Plasmid pBRC40 DNA was cut with *EcoRI*, *NheI*, *NotI* or *SmaI* into fragments to be used as hybridization probes. These were electrophoretically separated on 0.7% low melting temperature agarose (Bio Rad), then each band was cut from

the gel and purified using glassmilk according to the protocol supplied by the manufacturer (Gene Clean, Bio 101 Inc.). The probes were labelled with a nonradioactive label, digoxigenin-11-dUTP, for 16 to 20 hours under reaction conditions recommended by the manufacturer (Boehringer Mannheim).

FILTER BLOTTING AND HYBRIDIZATION

Single and double restriction enzyme cleaved pBRC40 fragments were separated on agarose gels then transferred to nylon filters (Hybond N, Amersham) by the method of Southern (1975). The hybridization and detection of nonradioactive probes to the DNA on the filters was performed as recommended by the supplier (Boehringer Mannheim). Hybridizations were performed under stringent conditions at 68°C in 5x SSC (sodium chloride-sodium citrate) (Sambrook et al. 1989), 0.1% (W/V) N-laurylsarcosine (Na salt), 0.02% (W/V) sodium dodecyl sulphate (SDS) and 1% (W/V) blocking reagent (Boehringer Mannheim) for 16 hours. The colour reaction was left until the bands were easily detected (5 to 24 hours), then the results were documented by photocopying the wet nylon filter. The filters were then immediately stripped (Boehringer Mannheim) to be reprobed. Filters could be reprobed twice without losing their integrity.

PLASMID MAPPING STRATEGY

The plasmid pBRC40 was first restricted with a number of common restriction endonucleases available in the laboratory. The criterion used in choosing the enzymes for mapping the plasmid was number of restriction sites; endonucleases lacking restriction sites and enzymes that produced many indistinguishable fragments were excluded. The fragment sizes were determined as described earlier.

The relative positions of the fragment bands were determined using single and double endonuclease digestions according to standard procedures (Sambrook et al. 1989). Confirmation of the positions was accomplished by probing Southern blotted nylon filters of single and double restriction enzyme digested pBRC40 with digoxigenin-labeled fragments of *EcoRI*, *NheI* and *SmaI* cleaved pBRC40. The plasmid map of pBRC60 was produced by the inclusion of the transposon Tn5271 that was mapped using nucleotide sequence information and previously published restriction enzyme digests of pBRC60 (Wyndham et al. 1988).

DNA MANIPULATIONS FOR TRANSPOSON MAPPING AND CLONING

Plasmid pBRC60 was single or double digested with *HindIII*, *EcoRI*, *PstI*, *SalI*, *NarI*, *BamHI*, *NruI*, *BglII*, *SacI* and *SmaI* (Bethesda Res. Lab.) and resolved in 0.7% agarose in Tris borate EDTA buffer (Sambrook et al. 1989). The catabolic region of pBRC60 was mapped using the 10

restriction enzymes listed above by hybridizing [α - 32 P]-dCTP (Amersham)-labeled fragments to pBRC60 digests immobilized on nylon (Hybond, Amersham) (Wyndham et al. 1988).

Insertion sequence IS1071 on a 4.1 kb *Nru*I fragment of pBRC40 was cloned into the *Nru*I site of pDPL13 (Gendel et al. 1983) to give plasmid pBRN4029. The outside flanking ends of the two IS1071 sequences in Tn5271 were cloned as 1.2 and 1.9 kb *Eco*RI/*Nru*I fragments into *Eco*RI/*Sma*I cut pUC18 to give pBREN12 and pBREN19, respectively. The inside flanking ends were obtained from existing clones (pBRE11 and pBRH3, Wyndham et al. 1988). Transformations into *E. coli* JM109 or DH5 α were carried out either by the method of Hanahan (1984) or by electroporation with a Gene Pulser apparatus using the method suggested by the manufacturer (Biorad).

DNA SEQUENCING

*Hind*III fragment H4, cloned in pUC18 (pBRH4, Wyndham et al. 1988) was subcloned into the replicative forms of M13mp18 and M13mp19 (Sambrook et al. 1989), then sequenced in both directions using the commercial Sequenase kit (U.S. Biochem. Corp.), and [α - 35 S] dATP (Amersham). Purified plasmid pBRN4029 was then used for plasmid sequencing in both directions away from the H4 internal sequence. Plasmid sequencing was also used to characterize the termini of the two IS1071 elements cloned from pBRC60. Oligonucleotide

primers were prepared using a Cyclone DNA Synthesizer (BioSearch Inc.). Sequence analyses were performed using Microgenie version 6.0 (Beckman). The sequence of IS1071 was deposited with GENBANK under accession number M65135. Tn5271, IS1071, pBRC60 and pBRC40 are Plasmid Reference Center listings (E. Lederberg, Stanford).

RESULTS

MAPPING OF THE PLASMID pBRC60

Cleaving the plasmid with fifteen different restriction enzymes showed that *Xba*I does not cut the plasmid and *Ava*II, *Bcl*I and *Pst*I cut the plasmid into many similarly sized fragments. Thus the plasmid, pBRC40 was mapped using the restriction enzymes *Bam*HI, *Bgl*II, *Eco*RI, *Hind*III, *Nhe*I, *Nru*I, *Pvu*II, *Sac*I, *Sal*I and *Sma*I. The transposon Tn5271 map was then added to the pBRC40 map by insertion into the single IS1071 element on the plasmid. The location of the transposon and its spontaneous deletion *in vivo* to give pBRC40 were confirmed by comparing the pBRC40 and pBRC60 fragment sizes for each of the restriction enzymes, and by sequencing (see below). The map of the plasmid, pBRC60 is illustrated in Figure 2.1 and the size of each fragment is listed in Table 2.2. The fragments are not numbered in sequence, the system utilized includes the combination of fragments from both plasmids pBRC60 and its transposon deletion derivative pBRC40. Some of the fragments retained

the same designation number but changed size after the deletion of the transposon. A separate table of the pBRC40 fragment sizes is included (Table 2.3). The enzyme *EcoRI*, cut the plasmid in one region where other restriction sites were rare, therefore the order of the *EcoRI* fragments in this region could not be determined with the enzymes employed. The remainder of the *EcoRI* fragments were mapped because of their extensive use in studying the catabolic genes encoded on the transposon Tn5271. The enzymes *BamHI* and *SacI* cut the plasmid only once, and *HindIII* twice, providing restriction sites of potential use in developing a cloning vector from this plasmid.

MAPPING OF TRANSPOSON Tn5271

The catabolic region of the 88 kb plasmid pBRC60 is shown in Figure 2.2. Tn5271 (17 kb) is flanked by 3.2 kb direct repeat sequences designated IS1071. Within the internal 10.6 kb unique region are sequences required for the growth of *Alcaligenes* sp. BR60 on chlorobenzoates. Deletion of the catabolic region from pBRC60 occurs at a frequency of 1.6×10^{-3} cell⁻¹ generation⁻¹, in the absence of 3-chlorobenzoate, to give strain BR40 (Wyndham et al. 1988). The deletion plasmid pBRC40 (74 kb) retains one copy of IS1071 entirely within a 4.1 kb *NruI* fragment generated by the deletion event. The DNA sequence of this copy of IS1071 is shown in Figure 2.3. The sequence shown

Figure 2.1: Restriction Enzyme Digest Map of Plasmid pBRC60
Restriction map shows relative cleavage sites of restriction endonucleases determined by digestion of plasmid DNA with the enzymes in combination with probing with digoxigenin-labelled plasmid fragments.

Restriction endonucleases included in the map are:

B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; N, *Nhe*I; Nr, *Nru*I; S, *Sac*I; Sa, *Sal*I; Sm, *Sma*I; P, *Pvu*II; IS1071, an insertion element. Size of each restriction fragment is given in Table 2.2.

Table 2.2: Restriction Fragment Sizes of pBRC60

Restriction fragment sizes (kb) of plasmid pBRC60 and the corresponding identification number used on the restriction map (Figure 2.1).

<i>EcoRI</i>	<i>SalI</i>	<i>SmaI</i>	<i>NruI</i>	<i>PvuII</i>	<i>BglII</i>
E1, 16.0	Sa1, >30	Sm1, >30	Nr1, >30	P1, >30	Bg1, >30
E2, 12.5	Sa2, 19.3	Sm2, 15.1	Nr2, 23.8	P2, 28.0	Bg2, 24.5
E3, 10.1	Sa3, 7.2	Sm3, 12.2	Nr3, 8.3	P3, 11.5	Bg3, 11.6
E4, 7.9	Sa4, 4.4	Sm4, 10.0	Nr4, 8.2	P4, 7.0	Bg4, 8.4
E5, 7.1	Sa5, 4.3	Sm5, 9.3	Nr5, 6.5	P5, 4.5	Bg5, 6.1
E6, 6.4	Sa6, 4.0	Sm7, 7.2	Nr7, 3.8	P6, 4.1	
E7, 4.9	Sa7, 3.4	Sm8, 3.0	Nr8, 3.5	P7, 1.7	
E8, 4.7	Sa8, 3.4	Sm9, 0.9	Nr9, 3.3	P8, 1.6	
E9, 3.6	Sa9, 3.0	Sm10, 0.6	Nr10, 1.6	P9, 0.6	
E10, 3.5	Sa10, 2.5	Sm11, 0.3			
E11, 2.9	Sa11, 2.3				
E12, 1.8	Sa12, 1.9	<i>HindIII</i>	<i>BamHI</i>	<i>NheI</i>	<i>SacI</i>
E13, 1.6	Sa13, 1.7	H1, >30	B1, >30	N1, >30	S1, >30
E14, 1.5	Sa14, 0.5	H2, 8.4	B2, 8.9	N2, 27.4	S2, 20.0
E15, 1.1		H3, 4.4	B3, 3.3	N3, 10.8	S3, 4.5
E16, 0.8		H4, 1.1	B4, 0.8	N4, 7.1	S4, 2.7
E17, 0.6			B5, 0.7	N5, 4.9	S5, 1.4
E18, 0.6				N6, 3.1	S6, 0.3

TOTAL 88.4kb, according to *EcoRI* digested fragments

Table 2.3: Restriction Enzyme Fragment sizes of pBRC40
 Restriction fragment sizes (kb) of plasmid pBRC40. *Bam*HI and *Sac*I have unique restriction sites and are not included in this table.

<i>Eco</i> RI	<i>Sal</i> I	<i>Sma</i> I	<i>Nru</i> I	<i>Pvu</i> II	<i>Bgl</i> II	<i>Hind</i> III
E1,16.1	S1,>30	Sm1,>30	N1,27.8	P1,29.5	Bg1,>30	H1,>30
E2,12.9	S2,19.3	Sm2,15.1	N2,23.8	P2,27.7	Bg2,29.4	H4,1.08
E3,10.5	S3, 7.2	Sm5, 9.3	N3, 8.3	P3,11.5	Bg3, 6.1	
E4, 8.5	S5, 4.4	Sm6, 8.5	N6, 4.3	P4, 4.6		
E6, 6.4	S6, 4.0	Sm7, 7.2	N7, 3.8	P5,0.63		
E7, 5.0	S7, 3.4	Sm8, 3.0	N8, 3.5			
E8, 4.9	S9, 3.0	Sm9,0.88	N10,1.6			
E9, 3.8	S11,2.3					
E10,3.6	S12,0.5					
E12,1.9						
E15,1.1						
E16,0.72						
E17,0.63						

TOTAL 76.77kb, according to *Eco*RI digested fragments

Figure 2.2: Restriction Enzyme Digest Map of Tn5271
Restriction enzyme digest map of the catabolic transposon Tn5271 from *Alcaligenes* sp. BR60. Open reading frames within IS1071 copies (see Figure 2.3) are indicated by long arrows. The positions of inverted repeat sequences are indicated by vertical bars and short arrows. Only cloned restriction fragments or fragments used as probes are numbered.

represents the product of an intramolecular recombination at an unknown point between the two direct repeats of IS1071 present on pBRC60. This was confirmed by comparison of the sequence in Figure 2.3 with sequences at the four IS1071 termini of Figure 2.2:

IS1071- Left-Outside: 5'-GCCGGTCTCGCGGTC **GGGGTCTCCTCGTTT**-3'

IS1071- Left-Inside: 5'-CGTCGAGCAGGGCAT **GGGGTCTCCTCGTTT**-3'

IS1071- Right-Inside: 5'-CGGGCAGCTTGTCGC **GGGGTCTCCTCGTTT**-3'

IS1071- Right-Outside: 5'-GGTTTGTCTCGTAAA **GGGGTCTCCTCGTTT**-3'

The termini of IS1071 in Figure 2.3 correspond to the Left-Outside and Right-Outside sequences of Tn5271. There are no duplicated sequences flanking the IS1071 termini.

DNA nucleotide sequence similarity searches revealed that the 110 bp inverted repeat (IR) flanking IS1071 was related to the class II (Tn3-family) IRs (Figure 2.4). The regions of similarity in the outer 38 bp of the IR have been recognized as DNA-binding and cutting sites for the class II transposases (Ichikawa et al. 1990). There was no significant similarity of the IS1071 IR sequence extending from 39 to 110 bp with sequences flanking the 38 bp IRs of the Tn3-family transposons shown (Figure 2.4). However, an Integration Host Factor (IHF) consensus binding sequence occurs at nucleotides 71 to 83 within the left and right IR of IS1071 (Figure 2.4), at positions analogous to IHF binding sites near the Tny δ (also known as Tn1000) IR (Wiater and Grindley 1988).

Figure 2.3: Nucleotide sequence of IS1071.

Numbering is relative to the first G in the IR sequence (underlined). Short sequences on either side of the insertion element are shown. The translation product of the open reading frame beginning at position 146 is shown. There is a potential ribosome binding site (GAGG) at position 136, and a stem-loop structure underlined beginning at the A in the stop codon at position 3058.

Figure 2.4: Comparison of Inverted Repeat Sequences

Comparison of inverted repeat sequences defining the ends of class II transposable elements. L and R, left and right inverted repeats, defined as distal and proximal to the *tnpA* gene (Tsuda et al. 1989). Nucleotides which differ in the right inverted repeat compared to the left are indicated. The ends of the IRs are indicated by a shift to lower case letters. An * above the sequence indicates identity at this position to the left IR of IS1071. Dotted lines mark sequences in IS1071 and *Tnyδ* homologous to the IHF binding site consensus sequence. Sources: Tn21, Tn3 (Genbank), Tn4651 (Tsuda et al. 1989); *Tnyδ* (Wiater and Grindley 1988); Tn4430 (Mahillon and Lereclus 1988).

The large open reading frame within IS1071 (Figure 2.3) spans 2,910 bp, coding for a putative 970 amino acid polypeptide of molecular weight 108,413. This putative transposase exhibited significant similarity to the class II transposases of Tn3 and Tn21 (Figure 2.5). These alignments place greatest weight on the known alignment of the Tn3 and Tn21 TnpA sequences. There was 21% identity of amino acids in comparing either IS1071 to Tn3, or IS1071 to Tn21. Only 12% of amino acids were conserved for all three TnpA sequences. Including conservative substitutions raised similarities to 37% (with Tn3) or 39% (with Tn21).

DISCUSSION

Most of the degradative plasmids that have been isolated are members of the IncP incompatibility group (Sayler et al. 1990). In the same manner as IncP plasmids, the host range of pBRC60 appears to cover a broad spectrum of Proteobacteria (Fulthorpe and Wyndham 1991). The plasmid pBRC60 has been described as a possible member of the IncP group (Wyndham et al. 1988). This was predicted due to the loss of pBRC60 when the IncP plasmid R68.45 was introduced into *Alcaligenes* sp. strain BR60 by conjugation. This was then confirmed by Burlage et al. (1990) who probed pBRC60 with a number of genetic loci from IncP plasmids. They felt that pBRC60 was distantly related to the IncP plasmids used in their study due to the relatively weak

Figure 2.5: Alignment of TnpA Sequences

Alignment of TnpA sequences for Tn3, IS1071 and Tn21.

Identical amino acids are indicated by * ; similar amino acids by : . The numbers above each set of 3 sequence lines refer to the amino acid numbers for the IS1071 sequence. The Tn3 sequence is 1004 and the Tn21 sequence 988 amino acids long. Conserved blocks of amino acids are marked by horizontal lines. Data for the Tn3 and Tn21 transposase sequences were taken from Genbank.

25 50 75 100
  

signal found in their hybridizations. Their study also illustrated the importance of the availability of a good restriction enzyme map of pBRC60. Their published map of the IncP loci localized by hybridization to the pBRC60 plasmid placed the *oriV*-hybridizing fragments within the transposon Tn5271 (on fragment E5). However, a careful analysis of their hybridization results using our map, and a repetition of the probing using the plasmid RK2 *oriV* (R. C. Wyndham, personal communication), showed that the correct *oriV*-hybridizing fragments (E4 and E9/10) were actually outside Tn5271. The results are still intriguing because these *EcoRI* fragments are distant to each other on our map (Fig. 2.1). This raises the possibility that there are two *oriV* sites on pBRC60. While not an objective of this Thesis research, the nature of the plasmid replication site(s) is a potential avenue to be investigated in the future.

The chlorobenzoate catabolic genes on plasmid pBRC60 are subject to deletion at a high frequency due to recombination between IS1071 copies on the plasmid. Recombination of the catabolic genes with the chromosome of BR60 also occurs at a high frequency (Wyndham et al. 1988) and is a result of recombination between plasmid and chromosomal IS1071 copies (Nakatsu et al. 1991). This background of host-encoded recombination activity, combined with the fact that BR60 was recently isolated from an environment where natural genetic exchange is likely to be

important (Fulthorpe and Wyndham 1991), makes it difficult to attribute the structure of Tn5271 or the rearrangements of IS1071 in *Alcaligenes* sp. BR60 to transposition events. However, transposition of the cloned IS1071 element to generate cointegrates in a recombination deficient *E. coli* has been observed (Nakatsu et al. 1991; D. Wiens, personal communication). Plasmid pBRN4029 carrying a single copy of IS1071 formed cointegrates with the conjugative plasmid pRK2013 at a frequency of 5×10^{-4} per transconjugant in mating-out assays to a rifampicin resistant *E. coli* HB101 recipient. Restriction mapping of these cointegrates showed that they consist of the entire pBRN4029 replicon, flanked by direct repeats of IS1071, integrated into various sites on pRK2013. These cointegrates were quite stable in a *recA* host, although resolution was observed with subculturing in some cases. Therefore, it would seem likely that transposition has played a part in the mobilization of catabolic genes in Tn5271. Currently there is a research project in this laboratory characterizing transposition products of both Tn5271 and IS1071 in *Comamonas* species hosts in order to define target sites for transposition or recombination (A. Cashore, personal communication).

Class II transposable elements share structural features that clearly delineate them from other mobile genetic elements (Sherratt 1989). IS1071 has inverted repeat sequences and a transposase sequence that places it

among the class II elements, however it has no sequence homologous to class II resolvases. A search of the internal 10.6 kb sequence of the transposon (Appendix 1) has failed to reveal a candidate for a resolvase-like gene. Given the complementary nature of the activities of transposase and resolvase, and the evidence that resolvase catalyzes site-specific recombination between cis-oriented res sites (Krasnow and Cozzarelli 1983), it is unlikely that a resolvase gene linked to a specific transposon would exist outside of the inverted repeat defining the element. This indicates that IS1071 is a true insertion sequence carrying the coding information for replicative transposition only.

The transposase alignments of IS1071 with Tn3 and Tn21 revealed only 12 % conservation of amino acids, comparable to the identity observed when Tn3, Tn21, Tn501, Tn917, Tn2501, and Tn4430 sequences were all aligned (Mahillon and Lereclus 1988). These transposons are therefore representative of the known evolutionary limits of the class II elements. Schmitt et al. (1985) have proposed a scheme for the evolution of the class II transposons based on the observation that a single copy of an IR sequence and a transposase are the minimal requirements for transposition. Inverted repeats of the recognition sequence flanking the transposase may have been acquired by a process such as strand exchange during hairpin loop replication. Such an element would be structurally identical to IS1071 (Figure

2.6). One prediction of this scheme is that the ancestral inverted repeats, because of the non-specific events occurring at a hairpin loop replication fork, may have been longer than the minimum transposase recognition sequence of 38 bp. This is the observed case with IS1071 (110 bp IR), Tn2501 (48 bp IR) and the toluene transposon Tn4651 (46 bp IR). Acquisition of site-specific recombination systems (resolvase or integrase), antibiotic resistance determinants, or catabolic genes by an element like IS1071 would give rise to the Class II transposons depicted in Figure 2.6 (Krasnow and Cozzarelli 1983, Schmitt et al. 1985, Mercier et al. 1990).

The class II transposons have been very successful in promoting the fitness of a range of prokaryotes under very different environmental conditions. With the discovery of IS1071 there is now good evidence to suggest that the transposase gene alone confers a selective advantage on its host under some environmental conditions.

Figure 2.6: Evolution of class II transposable elements. An ancestral element is indicated by a transposase gene (A) and its recognition sequence (◀). IR sequences including the 38 bp recognition sequence indicated by shaded arrows (◀); The direction of transcription of *tnpA* (A), *tnpR* (R) and integrase (I) genes, where known, is indicated by arrows; •, *res* site; Ap, Su, Sm, Hg are ampicillin, sulfonamide, streptomycin and mercury resistance loci; Cba, Tol are chlorobenzoate and toluene catabolic loci; T and S are trans-acting resolution genes for the toluene transposon.

CHAPTER 3: BIOCHEMICAL PATHWAY FOR 3-CHLOROBENZOATE DEGRADATION

INTRODUCTION

Early investigations of the metabolism of benzoate and hydroxybenzoates by bacteria established that the initial steps of ring oxidation were critical in directing metabolites through different ring-fission pathways (Dagley et al. 1960). These pathways were found to be conserved within taxonomic groups and have long been recognized as valuable taxonomic characters for the identification of members of the Pseudomonadaceae and other bacteria (Dagley 1971, Palleroni 1984, Stanier and Ornston 1973, Stanier et al. 1966). Thus, for example, the fluorescent pseudomonads are characterized by convergent catabolic pathways for benzoate and para-hydroxybenzoate leading via ortho-ring-fission pathways to the common intermediate β -ketoadipate (Stanier and Ornston 1973). Other pseudomonads, for instance the *Acidovorans* complex, now classified as a separate family, Comamonadaceae (Willems et al. 1991a), express divergent pathways for these compounds, including the gentisate pathway and the protocatechuate meta-ring-fission pathway (Stanier et al. 1966).

Recent phylogenetic studies based on nucleic acid hybridization and sequence similarities have significantly altered traditional taxonomic groupings and in the process

have redefined relationships among bacterial genera and species expressing alternative ring-fission pathways (De Ley 1992). The challenge now is to understand the selective forces involved in the evolution of metabolic diversity, including aromatic ring-fission, within these new phylogenetic groups (Zavarzin et al. 1991, Harayama et al. 1991).

In contrast to the conserved pathways for aromatic ring-fission discussed above, genes specifying the biodegradation of aromatic hydrocarbons and chlorinated aromatic compounds are often carried on conjugative plasmids (Sayler et al. 1990), and in a few cases within transposable elements (Nakatsu et al. 1991; Tsuda and Iino 1987, 1988; van der Meer et al. 1991). Evidence is beginning to accumulate showing that these mobile elements transfer at reasonably high frequencies among bacterial populations in nature (Fulthorpe and Wyndham 1992). The implication is that metabolites which are generated following interspecific transfer of catabolic elements must be compatible with conserved metabolic pathways of the host. Gene transfer may create complex regulatory problems and metabolite inhibition that could limit the useful host-range of catabolic plasmids or transposons. This phenomenon has important consequences for both microbial evolution in natural environments, and microbial adaptation to toxic chemicals during waste treatment in the bioremediation industry. For example,

metabolic and genetic studies on chlorobenzene and chlorobenzoate metabolism in *Pseudomonas putida* or *P. cepacia* isolates have provided concise explanations of why chlorocatechol intermediates are metabolized through either ortho- or meta-ring-fission pathways (Higson and Focht 1992, Reineke and Knackmuss 1988, Schmidt and Knackmuss 1980, van der Meer et al. 1991).

Little is known concerning the host-ranges of mobile catabolic elements in nature. The vast majority of reports of the occurrence of plasmids specifying hydrocarbon or chlorinated aromatic chemical oxidation list one host (Sayler et al. 1990). Where an attempt has been made to determine the distribution of particular catabolic elements in natural isolates, the host-range has been found to be relatively narrow. For example, plasmids and sequences of DNA similar to the TOL or NAH catabolic plasmids, or to the chlorocatechol dioxygenase operon *clcABD*, have a worldwide distribution, but host isolates are limited to the fluorescent pseudomonads, *Alcaligenes eutrophus* or *A. paradoxus* (Chatfield and Williams 1986, Chatterjee and Chakrabarty 1983a, b, Don and Pemberton 1981, Sayler et al. 1990). This distribution is likely to reflect limitations on the expression of the catabolic genes and their coordinated regulation with conserved metabolic pathways.

Chapter 2 describes the structure of a plasmid, pBRC60, and its resident transposon, Tn5271, in *Alcaligenes* sp. BR60

which are involved in 3-chlorobenzoate metabolism. The pathway for chlorobenzoate metabolism associated with transposon Tn5271 has not been reported, although the absence of hybridization of the chlorocatechol ortho-ring-fission genes of *Pseudomonas putida* to pBRC60 has been reported (Wyndham et al 1988). In this Chapter, I demonstrate that Tn5271 encodes a previously unknown metabolic pathway for chlorobenzoates which, when expressed in the wild type strain BR60, involves oxidation at the distal carbon (3,4) positions of the aromatic ring. The substrate range of these enzymes includes 3,4-dichlorobenzoate. The orientation of dioxygenase attack on the ring directs metabolism through protocatechuate. I speculate that the nature of this pathway would influence the host-range of Tn5271, and in this Chapter data is assembled showing that natural selection for Tn5271 expression in contaminated environments has resulted in the distribution of Tn5271 in hosts from rRNA superfamilies II and III (the Γ and β Proteobacteria, respectively). The natural distribution of Tn5271 also appears to correlate with the phylogenetic distribution of the protocatechuate meta-ring-fission pathway. Factors which may contribute to this correlation and its importance in the bioremediation of chlorinated aromatics are discussed.

MATERIALS AND METHODS

BACTERIAL STRAINS AND GROWTH CONDITIONS

Alcaligenes sp. strains BR60 (pBRC60), BR40 (pBRC40), BR6020 (plasmid cured), BR6024 (tryptophan auxotroph, chloramphenicol resistant), and conditions for their growth have been described (Nakatsu et al. 1991, Chapter 2; Wyndham et al. 1988; Fulthorpe and Wyndham 1991). Descriptions of *Escherichia coli* strains HB101 and DH5 α , with plasmids pRK2013 or pGS65, and their growth conditions, have been described elsewhere (Nakatsu et al. 1991, Chapter 2; Selvaraj and Iyer 1985; Wyndham et al. 1988). Natural isolates capable of growth on 3-chlorobenzoate as sole source of carbon were recovered from samples of surface waters and sediments contaminated with chlorinated aromatics from two locations by other researchers in the laboratory. The first was a site of historical contamination from the Hyde Park chemical landfill in Niagara Falls, New York (Wyndham and Straus 1988). The second was experimentally contaminated lake water and sediments which had been inoculated with *Alcaligenes* sp. BR60 (Fulthorpe and Wyndham 1989, 1991, 1992). Isolates were characterized by these researchers using the Biolog rapid identification system (Klingler et al. 1992), with additional taxonomic criteria (Palleroni 1984; Smibert and Krieg 1981). A condensed rRNA cistron similarity dendrogram for rRNA superfamilies I, II and III, based on T_m(e), the melting temperature of DNA-rRNA

hybrids, was adapted from published dendrograms of De Ley (1992).

ANALYTICAL METHODS AND CHEMICALS

Oxygen uptake rates into whole cells were determined using an Orion polarographic oxygen electrode as described (Wyndham and Straus 1988), with 1 mM concentrations of substrates and a cell optical density of 0.5 at 600 nm. Whole cell protein content was determined by the method of Bradford (1976). Chemicals were supplied as follows: 3-chlorobenzoate, 3,4-dichlorobenzoate and 3-chloro-4-hydroxybenzoate (Aldrich Chem. Co.); hydroxy- and dihydroxybenzoates (Sigma Chem. Co.); 3-chlorocatechol and 4-chlorocatechol (Helix Chem. Co., Richmond, B.C.).

For extraction of metabolic intermediates, *Alcaligenes* sp. strain BR60 (pBRC60) was grown to late exponential phase on chlorinated substrates. Cells were removed by centrifugation and the medium was acidified to pH 2.5, then extracted with an equal volume of ethyl acetate, dried with anhydrous sodium sulphate and evaporated under dry N₂ using a Büchi 461 rotary evaporating unit (Brinkman Instruments, Westbury, N.Y.). The sample was treated with excess N,O-Bis-trimethylsilyl-acetamide (BSA) to generate trimethylsilyl (TMS) derivatives for gas chromatography-mass spectrometry (GC-MS) analysis. The gas chromatograph (Hewett Packard 5890 series II) with a 30m J&W DB5 column

was operated at 120°C for 2 min. then raised 10° per min. to 300°C, and linked to a mass spectrometer (Concept 2H, Kratos Analytical, Manchester, England) operated at 70eV energy accelerating at 8KV (C. Kazekoff operator).

Protocatechuate ring-fission pathways used by natural isolates recovered in this work were determined spectrophotometrically (Smibert and Krieg 1981). Additional data on the distribution of *ortho*- and *meta*-ring-fission pathways for protocatechuate throughout the taxa of rRNA superfamilies I, II and III were from Sections 4 and 5 of Bergey's Manual of Systematic Bacteriology (Palleroni 1984).

DNA MANIPULATIONS

The construct pBRCN5 was produced by ligating the 5.8 kb *Pst*I-*Hind*III fragment of Tn5271 with the broad-host-range, mobilizable vector pGS65 (Selvaraj and Iyer 1985) in which the kanamycin resistance gene was replaced by a synthetic multiple-cloning-site (Chapter 4). Following transformation into *E. coli* DH5 α as described (Nakatsu et al. 1991, Chapter 2), plasmids were mobilized into *Alcaligenes* sp. BR6024 via triparental matings with *E. coli* HB101 (pRK2013) and transconjugants capable of growth on 3-chlorobenzoate were selected.

The presence of Tn5271 in natural isolates from the Niagara region and in recipients of pBRC60 in lakewater was determined previously (Fulthorpe and Wyndham 1991, Wyndham

personal communications) by Southern-blot hybridization with the 7.5 kb *Hind*III restriction fragment (H2) spanning the left side of Tn5271 and part of IS1071 (Fig. 3.1a), to total genomic DNA or plasmid DNA of natural isolates digested with restriction enzyme *Eco*RI. Diagnostic fragments included E16 from within IS1071, and E13 and E14 from within the catabolic region (Nakatsu et al. 1991, Chapter 2). Selected isolates were mapped in more detail and the involvement of Tn5271 in 3-chlorobenzoate metabolism in these isolates was confirmed by spontaneous deletion of the catabolic region of the transposon (Fulthorpe and Wyndham 1991, Wyndham personal communication).

RESULTS AND DISCUSSION

Introduction of pBRCN5 (Figure 3.1a), carrying the 5.8 kilobase pair *Pst*I-*Hind*III restriction enzyme fragment of Tn5271, into derivatives of *Alcaligenes* sp. BR60 devoid of pBRC60 and Tn5271 (strains BR6020 and BR6023) resulted in the restoration of 3-chlorobenzoate catabolic activity. This established that the left side of the unique internal region of Tn5271 was required for chlorobenzoate metabolism in *Alcaligenes* sp. BR60 (discussed further in Chapter 4).

Oxygen uptake experiments with whole cell suspensions of *Alcaligenes* sp. BR60 (Table 3.1) showed equivalent respiration rates when resting cells initially grown on 3-chlorobenzoate were exposed to 3-chlorobenzoate or

Figure 3.1: Structural map and the metabolic pathway for 3-chlorobenzoate encoded by Tn5271.

3.1a. Restriction enzyme digest map of Tn5271 and the subclone pBRCN5. Abbreviations are: tnpA, the transposase gene of IS1071; N, *Nhe*I; H, *Hind*III; E, *Eco*RI; B, *Bam*HI; S, *Sal*I; P, *Pst*I; Xh, *Xho*I; Bg, *Bgl*II.

3.1b. Proposed pathway for 3-chlorobenzoate metabolism in *Alcaligenes* sp. BR60 (pBRC60:: Tn5271). The substrate was potentially oxygenated to 1-carboxy-3-chloro-3,4-dihydroxycyclohexa-1,5-diene (compound I) with subsequent chloride removal and re-aromatized to protocatechuate (compound II). Final degradation proceeded via the meta-ring fission pathway as indicated by the formation of 2-hydroxy-4-carboxy-muconic semialdehyde from protocatechuate (Smibert and Krieg 1981), following growth in 3-chloro- or 3,4-dichlorobenzoate.

3.1c. and d. Possible mass spectra of compounds I and II, respectively, extracted from an *Alcaligenes* sp. BR60 culture growing on 4 mM 3-chlorobenzoate. Mass spectra of compound II was confirmed by comparison to a known standard.

protocatechuate (3,4-dihydroxybenzoate). Similar activities were observed when BR60 was grown on 3,4-dichlorobenzoate. Oxygen was taken up by cell suspensions exposed to 3- or 4-hydroxybenzoate at less than half the rate observed when exposed to protocatechuate. There was no uptake of oxygen in the presence of either 3-chlorocatechol or 4-chlorocatechol. These results point to pathways involving either successive hydroxylations at the 3- and 4-positions, or direct dioxygenase attack at the 3- and 4- positions. Low oxygen uptake rates for 3-chloro-4-hydroxybenzoate suggested that 3-chlorobenzoate was not subject to hydroxylation in the *para*-position. The absence of oxygen uptake by chlorobenzoate-grown cells in the presence of gentisate (2,5-dihydroxybenzoate) suggested that an initial hydroxylation catalyzed by Tn5271-encoded enzymes did not occur to remove the chlorine at the *meta*-position (Johnston et al. 1972). In contrast, cells of benzoate grown strains BR60 and BR40 (a Tn5271 deletion derivative of BR60) respired oxygen in the presence of both gentisate and protocatechuate, suggesting that these pathways co-exist or are both induced in *Alcaligenes* sp. BR60, as they are in *Comomonas acidovorans* and *C. testosteroni* (Wheelis et al. 1967). Respiration measurements therefore suggest that the host expresses both gentisate and protocatechuate pathways, but that Tn5271 codes for the direct dioxygenation of chlorobenzoate to yield protocatechuate.

Table 3.1: Oxygen Uptake Rates by *Alcaligenes* sp. BR60

OXYGEN UPTAKE RATES ($\mu\text{mole O}_2 \text{ min.}^{-1}\text{mg. protein}^{-1}$) by
 ALCALIGENES SP. BR60 AND BR40 WITH VARIOUS CARBON SOURCES^a

STRAIN	BR40 ^b		BR60	
GROWTH SUBSTRATE	Ba	Ba	3Cba	3,4DCba
TEST SUBSTRATE				
BENZOATE	17	18	10	0
3-HYDROXYBENZOATE	7	9	14	8
4-HYDROXYBENZOATE	4	8	10	15
GENTISATE ^c	3	6	0	0
PROTocatechuate ^d	3	18	29	32
3-CHLOROBENZOATE	0	6	29	30
3,4-DICHLOROBENZOATE	0	4	22	16
3-CHLOROCATECHOL	0	0	0	0
4-CHLOROCATECHOL	0	0	0	0
3-CHLORO-4-HYDROXYBENZOATE	0	4	1	1

^aNo oxygen uptake occurred in the presence of 2-chlorobenzoate, 2,3-, 2,4-, 2,6- and 3,5- dihydroxybenzoate or 2,3-, 2,4-, 2,5-, 2,6- and 3,5- dichlorobenzoate.

^bBR40 is a BR60 derivative lacking Tn5271

^c2,5-dihydroxybenzoate

^d3,4-dihydroxybenzoate

3Cba: 3-chlorobenzoate; 3,4DCba: 3,4-dichlorobenzoate

The presence of intermediates of 3-chlorobenzoate metabolism in *Alcaligenes* sp. BR60 was determined using gas chromatography-mass spectrometry (GC-MS) (Figure 3.1b, c and d). Protocatechuate (compound II) was found in the culture medium of BR60 grown in 3-chlorobenzoate. Also present was an intermediate with mass and fragmentation pattern equivalent to 1-carboxy-3-chloro-3,4-dihydroxycyclohexa-1,5-diene (compound I). The latter metabolite is the predicted product of a 3-chlorobenzoate-3,4-dioxygenase mediated oxidation of the aromatic ring. The absence of 3- or 4-hydroxybenzoate in the culture medium provided additional evidence that these were not intermediate metabolites. The low oxygen uptake rates for these compounds would indicate that this step would be rate limiting and these metabolites would then accumulate in the medium.

Further metabolism of protocatechuate was determined to proceed through meta-ring-fission by a protocatechuate-4,5-dioxygenase to yield 2-hydroxy-4-carboxy-muconic semialdehyde. This activity was induced following growth of BR60 (pBRC60) and BR6024 (pBRCN5) on chlorobenzoates, and following growth of BR40 (pBRC40), BR6020 and BR6024 on para-hydroxybenzoate. The locus of the meta-ring-fission genes in these strains is unknown.

The reaction scheme presented in Figure 3.1b requires that the chlorine substituent at the initial site of oxidation be removed enzymatically or spontaneously with the

re-aromatization of the ring to protocatechuate. The reaction is analogous to the chlorobenzoate-1,2-dioxygenase pathway proposed for *Pseudomonas aeruginosa* strain that was able to eliminate a chlorine substituent ortho to the carboxyl group (Hickey and Focht 1990). The orientation of dioxygenase attack relative to the carboxyl group is the critical turning point in directing metabolism through either the catechol or protocatechuate pathways and it is likely to be a fundamental property of the dioxygenase enzyme. There may be closer mechanistic and evolutionary links between the Tn5271-encoded activity and dioxygenase-mediated attack at the 3- and 4-positions of para-sulphobenzoate or ortho-phthalate by *Comomonas testosteroni* strains (Locher et al. 1991; Nakazawa and Hayashi 1977; discussed further in Chapter 5). Both of these initial oxidations lead to a protocatechuate meta-ring-fission pathway and *C. testosteroni* is closely related to *Alcaligenes* sp. strain BR60.

Dioxygenase mediated attack that directs oxygen toward chlorine substituents in the 3- and 4- positions of benzoate is of interest for applications in the bioremediation of polychlorinated biphenyls, where 3- and 4-substituted chlorobiphenyls and chlorobenzoates are poorly metabolized or inhibit metabolism (Erickson and Mondello 1992, Sandossi et al. 1992). It may be possible to link Tn5271-encoded activity to chlorobiphenyl catabolic activity in *Comomonas*

testosteroni strains in order to engineer microbes to mineralize these recalcitrant chemicals.

Fulthorpe and Wyndham (1991, 1992) observed interspecific transfer of pBRC60 (Tn5271) in aquatic microcosms, but were unable to explain the determinants of fitness in alternative hosts isolated in these studies. The results of phenotypic identifications, including protocatechuate ring-fission pathways, on Tn5271-carrying isolates from the microcosms and from a contaminated creek in the Niagara watershed are presented in Figure 3.2. The data are presented in conjunction with the known phylogenetic distribution of families or genera among rRNA superfamilies I, II and III (De Ley 1992), and the known protocatechuate ring-fission pathways for type strains within these groups (Palleroni 1984). This historical information shows that the *ortho*- and *meta*-ring-fission pathways for protocatechuate are distributed throughout the three rRNA superfamilies shown, but that these pathways are generally deep-rooted. A notable exception is [*Alcaligenes*] *paradoxus* that utilizes the protocatechuate *ortho*-pathway although it is grouped within the Acidovorans Complex of rRNA superfamily III (the Comomonadaceae) that mainly uses the *meta*-pathway.

In 49 of 50 Tn5271 hosts from nature, protocatechuate was metabolized through the *meta*-ring-fission pathway. Most of these isolates belong to the Acidovorans Complex

Figure 3.2: The phylogenetic distribution of isolates expressing the Tn5271-encoded dioxygenase.

A condensed rRNA cistron similarity dendrogram based on $T_m(e)$, the melting temperature of DNA-rRNA hybrids, is adapted from published dendrograms of DeLey (1992). The $T_m(e)$ range encompassed by all taxa included in an rRNA branch is indicated by a box. When only one strain is included, a line is used. Column labeled R-F indicates the protocatechuate ring-fission pathway (either o, *ortho*-; or m, *meta*-) used by one or more taxa within an rRNA branch. Data were taken from published taxonomic data (Palleroni 1984) (indicated by letters alone), or determined spectrophotometrically (Smibert and Krieg 1981) (indicated by letters in parentheses). The *Pseudomonas fluorescens* *ortho*-pathway is used by type strains (Palleroni 1992), but the natural Tn5271 recipient PR24B uses an unknown pathway. The column labeled P shows a ◀ for the known host range of pBRC60, as determined using a Tn5-labeled derivative (Fulthorpe and Wyndham 1991). The column labeled Tn lists the numbers of independent isolates recovered on 3-chlorobenzoate media from aquatic microcosms (Fulthorpe and Wyndham 1991, 1992) or surface water runoff in the Niagara region (Wyndham and Straus 1988), carrying Tn5271. Tn5271-carrying isolates were characterized using the Biolog rapid identification system (Klingler et al. 1992), with additional taxonomic criteria (Smibert and Krieg 1981).

(Comamonadaceae) or to the authentic *Alcaligenes* species *faecalis* or *xylosoxidans* within superfamily III. The host range for successful expression of the Tn5271 chlorobenzoate-3,4-dioxygenase was narrower than the host-range of plasmid pBRC60, which replicated in type strains from rRNA superfamilies I, II and III (Fulthorpe and Wyndham 1991). Two isolates were identified to rRNA superfamily II, *Pseudomonas fluorescens* PR24B and *Xanthomonas maltophilia* PR42A. *Pseudomonas fluorescens* PR24B was apparently the single exception to the correlation of the meta-ring-fission of protocatechuate with the expression of Tn5271 in natural hosts. The type strain of *P. fluorescens*, as for other fluorescent pseudomonads, is known to metabolize catechol and protocatechuate through convergent ortho-ring-fission pathways (Palleroni 1984, Stanier and Ornston 1973). Research in our laboratory (R. C. Wyndham, personal communication) found that the natural isolate PR24B grows poorly on benzoate and 4-hydroxybenzoate and is therefore atypical among fluorescent pseudomonads. It also grows poorly on 3-chlorobenzoate under laboratory conditions (Fulthorpe and Wyndham 1991). Further characterization of the ring-fission-pathway induced in this isolate is required.

There are a number of factors that may limit the host range of these catabolic genes. One possibility is competition with indigenous enzymes such as toluene or

benzoate dioxygenases. In fluorescent pseudomonads, these enzymes are able to recognize 3-chlorobenzoate and produce chlorocatechol that is lethal to the cells in the absence of specialized chlorocatechol dioxygenases (Reineke and Knackmuss 1988). Alternatively, the necessary regulatory enzymes may not be present in these alternate hosts and may not be carried on the transposon Tn5271 (see discussion in Chapter 5).

Interspecific transfer and recombination processes promote the evolution of new phenotypes from existing metabolic pathways (Fulthorpe and Wyndham 1992, Reineke and Knackmuss 1988, van der Meer et al. 1991). The transposon Tn5271 encodes a 3-chlorobenzoate-3,4-dioxygenase activity that, when expressed in *Alcaligenes* sp. BR60, directs molecular oxygen into the chlorinated position of 3-chlorobenzoate with substrate specificity that includes 3,4-dichlorobenzoate. Selection for expression of this activity in contaminated environments has resulted in the distribution of Tn5271 in hosts from rRNA Superfamilies II and III. The distribution of Tn5271 correlates with the phylogenetic distribution of the protocatechuate meta-ring-fission pathway.

CHAPTER 4: CLONING AND EXPRESSION OF 3-CHLOROBENZOATE DEGRADATION GENES ENCODED ON Tn5271

INTRODUCTION:

There are a number of approaches that may be taken to understand the evolution of metabolic pathways in bacteria. One approach is to study the metabolism of chemicals that have been recently introduced into the environment and compare this with the metabolism of naturally occurring chemicals. For example, the biochemical pathways and organization of genes coding for enzymes involved in 3-chlorobenzoate (3Cba) degradation can be related to the elements involved in the metabolism of the parent compound benzoate. Comparisons like these have been invaluable in tracing the evolutionary roots of novel catabolic pathways. The catabolism of benzoate has been demonstrated to occur through a number of possible pathways. The most commonly studied route is via the β -ketoadipate pathway, where benzoate forms a *cis*-diol, 1-carboxy-2-hydroxy-1,2-dihydroxycyclohexa-3,5-diene, that is then converted to catechol (Neidle et al. 1987). This is the route commonly utilized by the fluorescent pseudomonads (Palleroni 1992). Alternatively, former pseudomonads of the Acidovorans complex (family Comamonadaceae, Willems et al. 1991a) first hydroxylate the benzoate ring in the *meta* or *para* positions then proceed to metabolize the compound through either

protocatechuate or gentisate (Wheelis et al. 1967).

When microbes encounter xenobiotic chemicals such as 3-chlorobenzoate some of the enzymes of the benzoate pathway may be recruited for metabolism because they have broad substrate specificities, recognizing the chlorinated compound. Other existing enzymes may require modification in order to recognize some of the novel metabolites produced during degradation. The enzymes involved in 3-chlorobenzoate degradation in *Pseudomonas putida* have been well characterized (Reineke and Knackmuss 1988) and their genetic organization has been studied (Frantz and Chakrabarty 1987). The pathway is illustrated in Figure 4.1a. In this case the specificity of the benzoate dioxygenase is relaxed enough to recognize 3-chlorobenzoate but not 4-chlorobenzoate, and then specialized enzymes are required to convert chlorocatechol to maleylacetate (Reineke and Knackmuss 1988). The comparison of the nucleotide sequence of the genes coding for chlorocatechol and catechol dioxygenases demonstrated that a common ancestry was probable (Neidle et al. 1988). There was also significant homology to chlorocatechol dioxygenase genes encoded on 2,4-dichlorophenoxyacetate (2,4-D) catabolic plasmid pJP4 of *Alcaligenes eutrophus* JMP134 (Ghosal and You 1988).

Until recently, the only report for an alternate route for 3-chlorobenzoate degradation was by Johnston et al. (1972) who found that a *Pseudomonas* sp. was able to directly

Figure 4.1: Degradation Pathways of 3-Chlorobenzoate

4.1a. Degradation of 3-chlorobenzoate (I) by pseudomonads of the fluorescent group proceeds through 2-hydroxy-5-chloro-1,2-dihydroxybenzoate (II), to 4-chlorocatechol (III) which is then *ortho* cleaved forming 3-chloro-*cis,cis*-muconate (IV). Alternatively the ring can be attacked on the opposite side to form 2-hydroxy-3-chloro-1,2-dihydroxybenzoate (V), then 3-chlorocatechol (VI) which is then *ortho* cleaved forming 2-chloro-*cis,cis*-muconate.

4.1b. Degradation of 3-chlorobenzoate (I) by *Alcaligenes* sp. strain BR60 also appears to occur by attack on both sides of the ring, potentially forming 1-carboxy-3-chloro-4,5-dihydroxycyclohexa-2,6-diene (VIII) or 1-carboxy-3-chloro-3,4-dihydroxycyclohexa-1,5-diene (XI), these would then re-aromatize to form 5-chloro-protocatechuate that may then be cleaved to form 1-hydroxy-3-carboxy-muconic chloro-semialdehyde (X) (Kersten et al. 1985) or the chloride substituent can be lost during aromatization to form protocatechuate (XII) which is then *meta* cleaved to form 2-hydroxy-4-carboxy-muconic semialdehyde (XIII).

Structures in [] have not been confirmed by comparison to standards.

dehalogenate 3-chlorobenzoate forming 3-hydroxybenzoate (3Hba). Degradation then proceeded through gentisate. This strain may belong to the Comamonadaceae, but there have been no further biochemical or genetic studies with this organism to determine the nature and evolution of the pathway.

In chapter 3 an alternate pathway for 3-chlorobenzoate degradation (Nakatsu et al. 1992, Chapter 3) utilized by the environmental isolate *Alcaligenes* sp. BR60 (Wyndham and Straus 1988) was described. In this organism degradation appears to proceed by a transposon encoded dioxygenase that attacks the aromatic ring in the *meta* and *para* carbon positions, possibly forming 1-carboxy-3-chloro-3,4-dihydroxycyclohexa-1,5-diene, with subsequent dehalogenation and formation of protocatechuate (Figure 4.1b). Protocatechuate is then cleaved by *meta*-ring-fission in the same manner as observed commonly in the Comamonadaceae. The extensive literature on metabolism in the Comamonadaceae, to which *Alcaligenes* sp. BR60 belongs, suggests that the normal pathway for benzoate metabolism involves an initial hydroxylation by a monooxygenase forming 3-hydroxybenzoate or 4-hydroxybenzoate (Wheelis et al. 1967). Therefore, the question arises as to the evolutionary origins of the dioxygenase activity in this organism. In order to study this directly, genes involved in the dioxygenation of 3-chlorobenzoate were isolated and their expression was studied.

Previous studies have demonstrated that dioxygenases that form *cis*-diols are multicomponent enzymes that are chromosomally or plasmid encoded (Fetzner et al. 1992, Neidle et al. 1987, Sinclair et al. 1986). The range of specificity of these dioxygenases is variable. For example, the benzoate dioxygenase of *Pseudomonas* sp. B13 is limited to benzoate and 3-chlorobenzoate, whereas the toluate dioxygenase specificity encoded by the TOL plasmid of *Pseudomonas putida* is broad, functioning on benzoate, 3-chlorobenzoate, 4-chlorobenzoate and 3,5-dichlorobenzoate (Lehrbach et al. 1984, Reineke and Knackmuss 1978a). A number of bacterial dioxygenases that function on aromatic compounds have been studied biochemically and a few of them genetically. The dioxygenases studied thus far have been composed of two or three different polypeptides comprising a short electron transport system with non-heme iron-sulphur terminal oxygenases (Fetzner et al. 1992, Gibson and Subramanian 1984). More information is required to determine the factors controlling the substrate specificity of these enzymes.

In this chapter I demonstrate that in *Alcaligenes* sp. strain BR60 there are genes involved in 3-chlorobenzoate degradation carried on the transposon Tn5271 (Nakatsu et al. 1991, Chapter 2; Wyndham et al. 1988). The genes were subcloned on a broad-host-range expression vector. When expressed in *E. coli* the dioxygenase activity appears to be

directed primarily to the 4,5 carbon positions and to a minor degree to the 3,4 positions, whereas in *Alcaligenes* sp. strain BR6024, dioxygenase activity appears to be is directed to both the 3,4 and 4,5 carbon positions.

MATERIALS AND METHODS

BACTERIAL STRAINS AND GROWTH CONDITIONS:

Alcaligenes sp. strains BR60 and BR6024 and their growth conditions have been described previously (Wyndham et al. 1988, Fulthorpe and Wyndham 1991). *Alcaligenes* sp. strain BR6024 (trp^- , Cm^r , 3Cba^- , Ba^-) was used as a recipient for mating experiments. *Escherichia coli* strains and their plasmids have been described: DH5 α (Hanahan 1983), CB1273 (pMMB66EH), CB1274 (pMMB66HE) (Fürste et al. 1986), HB101 (pGS65) (Selvarej and Iyer) and MM294 (pRK2013) (Figurski and Helinski 1979). *E. coli* strains were grown in TYE (1% tryptone, 1% yeast extract, 0.5% NaCl) with the addition of required antibiotics, ampicillin ($50 \mu\text{g}\cdot\text{ml}^{-1}$), chloramphenicol ($30 \mu\text{g}\cdot\text{ml}^{-1}$), kanamycin ($50 \mu\text{g}\cdot\text{ml}^{-1}$) and/or tetracycline ($15 \mu\text{g}\cdot\text{ml}^{-1}$), at 37°C with rotary shaking (New Brunswick Scientific) at 200 rpm.

CONSTRUCTION OF VECTOR pBW13

The vector pBW13 was constructed from the IncP, mobilizable plasmid pGS65 (Selvarej and Iyer 1985). The vector was modified by the removal of the *EcoRI* site and

Figure 4.2: Construction of pBW13

Construction of the Inc P, broad host range, mobilizable plasmid pBW13 from pGS65, by the deletion of the *EcoRI* site and the addition of a synthetic polylinker (indicated by

). The nucleotide sequence of the complementary oligonucleotide strands synthesized were:

5' AGCTTGGATCCGAGCTCGTTAACGAATTCTGCAGATC 3' and

3' ACCTAGGCTCGAGCAATTGCTTAAGACGTCTAGAGCT 5', coding for the restriction enzyme sites: *HindIII*, *BamHI*, *SacI*, *HpaI*, *EcoRI*, *PstI*, *BglII* and *XhoI*.

kanamycin resistance gene and the addition of a multiple cloning site (Figure 4.2). The *EcoRI* site present on pGS65 was first eliminated by restriction with *EcoRI* and treatment with mung bean nuclease (Boehringer Mannheim) in order to create blunt ends, then religated using T4 DNA ligase (Boehringer Mannheim) according to the methods outlined in Sambrook et al. (1989). Single stranded, synthetic oligonucleotides were made on a Cyclone DNA synthesizer (Biosearch Inc.). The two oligonucleotides were complementary strands that contained the nucleotide sequence for the restriction enzyme sites in the following order, *HindIII*, *BamHI*, *SacI*, *HpaI*, *EcoRI*, *PstI*, *BglII* and *XhoI*. The single stranded oligonucleotides were first annealed by incubating the strands at 65°C for 5 minutes in a heating block then the block was slowly cooled to room temperature. The double stranded oligonucleotide was then phosphorylated using bacteriophage T4 polynucleotide kinase (Bethesda Research Laboratories) then ligated to pGS65 that had been cleaved with *HindIII* and *XhoI*. The multiple cloning site was added to pGS65 between the *HindIII* and *XhoI* sites within the kanamycin resistance gene, therefore eliminating this marker. The existing *BglII* site on pGS65 could not be eliminated therefore cloning of *BglII* fragments was performed into the *BamHI* position of the multiple cloning site.

DNA MANIPULATIONS:

The methods used for plasmid isolation have been described previously (Nakatsu et al. 1991, Chapter 2). The non-repeated region of Tn5271 was cloned onto pBW13 in order to confirm that genes involved in 3-chlorobenzoate degradation were indeed encoded in this region and to determine if only a portion of this region was necessary. Subclones were made using the available restriction sites within Tn5271 for *Hind*III, *Not*I, *Eco*RI, *Hpa*I and *Pst*I. The DNA was obtained from clones of pBRC60 DNA constructed during the original genetic study of this isolate (Wyndham et al. 1988). The lack of a *Not*I cloning site in the vectors necessitated the conversion of the *Not*I site into a *Hind*III site using a *Hind*III synthetic linker (Boehringer Mannheim) according to the method outlined in Ausubel et al. (1989). Nested deletions of the constructed plasmid pBRCN5 containing the *Pst*I-*Hind*III fragment of Tn5271 were created by digestion from the *Hind*III site using exonuclease III (Bethesda Research Laboratories) following the procedure of Sambrook et al. (1989). The extent of the deletion was determined by restriction enzyme mapping of the clones.

The constructs were then mobilized into *Alcaligenes* sp. BR6024 using triparental mating of overnight cultures on filters supported on TYE agar plates. *E. coli* MM294 carrying the helper plasmid pRK2013 (Figurski and Helinski 1979) was mixed with equal volumes of the recipient and

donor cells. Transconjugants were selected by dilution plating onto minimal medium A (Wyndham 1986) with addition of 4 mM 3-chlorobenzoate or 10 mM succinate and 0.1 mM tryptophan, with the antibiotics tetracycline and chloramphenicol at concentrations indicated above. To confirm growth and degradation of 3-chlorobenzoate by transconjugants, the increase in optical density at 600 nm of cultures grown with 3-chlorobenzoate as sole carbon source was measured and substrate disappearance was determined by HPLC analysis following the methods previously described (Wyndham and Straus 1988).

GENE EXPRESSION:

Expression of genes contained in Tn5271 was performed using the *tac* promoter system of Fürste et al. (1986). The *Hind*III modified *Not*I-*Hind*III fragment of Tn5271 was cloned into pMMB66HE (Fürste et al. 1986) to determine polypeptide sizes and for *in vivo* experiments. In order to confirm that the right side *Eco*RI-*Not*I fragment did not code for a protein, it was cloned into pMMB66EH. The polypeptides were separated on 7.5% and 12% SDS-PAGE (sodium dodecyl sulphate-polyacrylamide) denaturing gels according to the method of Laemmli (1970). The gels were stained with 0.25% Coomassie Brilliant Blue-250 (U.S. Biochem. Corp.) dissolved in 40% methanol-10% glacial acetic acid. Sizes of the unique protein bands were determined by comparison to protein

standards of known molecular weights.

In vivo expression in *Alcaligenes* strain BR6024 was carried out in transconjugants containing the nested deletion constructs pBRCN8 to 12 (Figure 4.3c). Cultures were first grown in medium A plus succinate, tryptophan, tetracycline and chloramphenicol as indicated above, then inoculated into fresh medium with 3-chlorobenzoate alone or 3-chlorobenzoate plus succinate as carbon sources. Cultures grown in 3-chlorobenzoate alone were inoculated with 50 ml cultures that were concentrated by centrifugation. The supernatant was analysed by HPLC then extracted for analysis by gas chromatography-mass spectrometry (GC-MS) as described previously (Nakatsu et al. 1992, Chapter 3).

In vivo expression was also carried out using the pMMB66EH construct pBRCN13a that contained the *Hind*III modified *Not*I-*Hind*III fragment in the coding orientation in *E. coli* DH5 α . One millilitre of overnight culture was added to two flasks containing 100 ml of TYE-ampicillin medium, IPTG (isopropylthio- β -D-galactoside, 1 mM final concentration) was added to one flask in order to induce the genes. Cultures were grown for four hours, centrifuged and the supernatant was analysed by HPLC then extracted for analysis by GC-MS.

RESULTS:**ISOLATION OF *cba* GENES**

Figure 4.3 illustrates the transposon Tn5271 and the restriction enzyme sites utilized in subcloning. As indicated in the figure, only subclones containing the left side of Tn5271 were required to restore 3-chlorobenzoate degradation abilities to the 3-chlorobenzoate deficient mutant BR6024. The smallest fragment containing genes necessary for 3-chlorobenzoate degradation was the 5.2 kb *NotI-HindIII* fragment, pBRCN6. When the *ScaI* fragment was removed from the *PstI-HindIII* clone, it was no longer able to restore 3-chlorobenzoate degradation. Therefore, part or all of the DNA within the *ScaI* fragment was required. The absence of unique restriction sites on the right side of pBRCN6 made it necessary to employ nested deletions. Deletions of DNA to the right of the *EcoRI* site at the 8.1 kb position enabled growth. However, the colonies initially formed on selection plates after mating turned black. Subsequent culturing of these clones restored the normal off-white colony colour and growth using 3-chlorobenzoate. Growth experiments and HPLC analyses confirmed that the clones containing constructs pBRCN1, 2, 5, 8, 9 and 10 completely utilized 3-chlorobenzoate for growth. Clones of cutbacks into the *EcoRI* fragment between positions 7 and 8.1 kb (pBRCN-11 and -12; figure 4.3c) were not able to degrade 3-chlorobenzoate.

Figure 4.3: Cloning Map of Tn5271

4.3a. Physical map of Tn5271 with restriction enzyme sites marked that were used in subcloning. *Bst*EII (B), *Eco*RI (E), *Eco*RV (Ec), *Hind*III (H), *Hpa*I (Hp), *Not*I (N), *Pst*I (P), *Sca*I (S).

4.3b. Restricted Tn5271 fragments were cloned onto pBW13 and mated into *Alcaligenes* sp. strain BR6024 in order to determine restoration of 3-chlorobenzoate degradation (+), no growth (-), or nt, clone not tested for growth on 3-chlorobenzoate. ^a was modified for cloning by the conversion of the *Not*I site into a *Hind*III site by the addition of a synthetic polylinker. This fragment was also cloned in both orientations into the expression vector pMMB66HE creating pBRCN13a and 13b.

4.3c. Clones were created by manipulating pBRCN5 by the deletion of internal *Sca*I fragment pBRCN7 or by nested deletions from the *Hind*III site, pBRCN8-12. ^b *Eco*RI-*Not*I fragment was cloned onto pMMB66EH

PROTEIN EXPRESSION

The plasmids pBRCN13a and b were constructed by ligating the *Hind*III linker modified *Not*I-*Hind*III fragment (4.5 kb to 9.8 kb positions) into the expression vector pMMB66HE in opposite orientations. Two unique protein products approximately 51 kd and 42 kd were resolved in a 7.5% SDS-PAGE gel from the IPTG induced culture of pBRCN13a (Figure 4.4). This indicates that the fragment is translated from left to right in the orientation of the *Not*I to the *Hind*III site of Figure 4.3b. No unique proteins were observed in lysates of IPTG induced cultures carrying pBRCN13b with the *Not*I-*Hind*III fragment cloned in the opposite orientation. A clone of the fragment of DNA from the *Eco*RI site at 8.1 kb to the *Not*I site at 10.5 kb in pMMB66EH (outside the tentative coding region) failed to express unique proteins when induced with IPTG (Figure 4.4).

METABOLITES FORMED IN VIVO

The metabolites formed by *Alcaligenes* sp. strain BR6024 clones containing the nested deletion constructs revealed some interesting results when examined by GC-MS. Clones pBRCN8, pBRCN9 and pBRCN10 revealed similar results. When grown in the presence of 3-chlorobenzoate alone or succinate and 3-chlorobenzoate, cultures of these clones contained metabolites that had mass spectra corresponding to a single chloro-hydroxybenzoate, a chloro-dihydroxybenzoate and

Figure 4.4: 7.5% SDS-PAGE gel resolving unique polypeptides expressed by pBRCN13a.

Lane **A**. protein molecular weight standards (top to bottom, phosphorylase b, 97400 d, bovine serum albumin 68000 d and ovalbumin 43000 d). Lane **B**. pBRCN14 not induced. Lane **C**. pBRCN14 IPTG induced. Lane **D**. pBRCN13a not induced. Lane **E**. pBRCN13a IPTG induced. Lane **F**. pBRCN13b not induced. Lane **G**. pBRCN13b IPTG induced. Description of clones are given in Figure 4.3.

2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

A B C D E F G

97▶

68▶

43▶

protocatechuate. Only protocatechuate could be confirmed by comparison to an authentic sample. Relative amounts of each metabolite in extractions for GC-MS analysis differed according to the stage of growth of the clones, but usually the major metabolite was the chloro-hydroxybenzoate. When these clones were grown with 3-chlorobenzoate as the sole carbon source there was a greater proportion of protocatechuate in the medium.

The pBRCN11 clone did not grow when 3-chlorobenzoate was the sole carbon source but 3-chlorobenzoate was partially degraded. According to GC-MS analysis, the major product that accumulated in the medium was a single chloro-hydroxybenzoate. When succinate was added as an additional carbon source for this clone, the culture converted the majority of 3-chlorobenzoate into the chloro-hydroxybenzoate, but protocatechuate, a chloro-dihydroxybenzoate and a metabolite with mass spectrum indicative of a chloro-dihydrodiol of chlorobenzoate were also formed. There was no degradation of 3-chlorobenzoate by the pBRCN12 clone.

E. coli clones containing the construct pBRCN13a accumulated a purple-black metabolite when induced with IPTG in the presence of 3-chlorobenzoate. This colour did not form in the absence of IPTG, nor in clones containing the construct in the opposite orientation. GC-MS analysis revealed the presence of a major metabolite that had a mass

spectrum that corresponded to a chloro-dihydroxybenzoate, possibly 5-chloro-protocatechuate. Authentic samples of this chemical were not available to confirm the positions of the hydroxyl groups. Extracts of the growth medium of this clone also contained small amounts of metabolites with mass spectra corresponding to protocatechuate and a chloro-hydroxybenzoate.

DISCUSSION:

This chapter describes the isolation of genes involved in 3-chlorobenzoate degradation carried on transposon Tn5271. Previously we described the transposition functions of this element, a composite transposon flanked by a unique insertion element. The insertion element, IS1071, contains one long open reading frame encoding a putative transposase with amino acid sequence similarities to the Tn3 group transposases (*tnpA*) (Nakatsu et al. 1991, Chapter 2). In this chapter it is demonstrated that the entire non-repeated region of Tn5271 is not necessary for 3-chlorobenzoate degradation. The required area is confined to a 3.7 kb region flanked by *NotI* and *EcoRI* sites (Figure 4.3a). We do not know whether the remainder of the transposon has another function.

In chapter 3, the biochemical pathway utilized for 3-chlorobenzoate degradation in *Alcaligenes* sp. BR60 was

described (Nakatsu et al. 1992). Evidence indicated that degradation potentially occurred by means of a chlorobenzoate-3,4-dioxygenase activity to form 1-carboxy-3-chloro-3,4-dihydroxycyclohexa-1,5-diene, followed by removal of the chlorine substituent during rearomatization of the ring to protocatechuate. Both the chloro-dihydrodiol and protocatechuate were detected in BR60 culture supernatants during growth on 3-chlorobenzoate. Since the organism is able to grow on both benzoate and protocatechuate in the absence of Tn5271 we speculated that *cba* genes coding for chlorobenzoate-3,4-dioxygenase were carried on the transposon. In this report we note that the induction of the genes encoded on Tn5271 using the construct pBRCN13a in *E. coli* caused the accumulation of a major product with a mass spectrum indicating a chloro-dihydroxybenzoate and lesser amounts of protocatechuate and a chloro-hydroxybenzoate. Each of these metabolites could be the product of the breakdown of a dihydrodiol intermediate, either *in vivo* or as a result of dehydration during the extraction steps. This supports our earlier conclusion that the genes carried on pBRCN13a code for enzymes that dioxygenate 3-chlorobenzoate, however this observation differs from the observed metabolites formed in the wild type *Alcaligenes* sp. strain BR60 (Nakatsu et al. 1992, Chapter 3). In the wild type strain, only a possible 3-chloro-dihydrodiol intermediate and protocatechuate were

detected. When the chlorobenzoate dioxygenase genes were expressed in *E. coli*, or reintroduced into *Alcaligenes* sp. strain BR6024 on the pBRCN5 and 6 clones, both protocatechuate and the chloro-dihydroxybenzoate accumulated in the medium. In *E. coli*, relatively greater amounts of the chloro-dihydroxybenzoate metabolite were formed, while in *Alcaligenes* sp. strain BR6024 the same amounts of the two products were formed. This suggests that there is a difference in the activity of this enzyme when it is present in *E. coli* as compared to its expression in *Alcaligenes*. This also suggests that the dioxygenase can attack both the 3,4 and 4,5 positions of the benzoate ring as depicted in Figure 4.1b. The expression of *pca* genes coding for protocatechuate-3,4-dioxygenase also showed low specific activity when expressed in *E. coli* (Doten et al. 1987). We do not know if there are differences in the assembly of these dioxygenases in different hosts. Studies have found that there are differences in the expression of *Pseudomonas* genes when transferred to *E. coli*. Because of this, studies on the TOL plasmid encoded toluate-1,2-dioxygenase have been carried out using *Pseudomonas* as the host (Harayama et al. 1986). To our knowledge there have been no studies on the expression of the same dioxygenase enzymes in different hosts with respect to their substrate specificity and product distribution.

Laboratory adaptation may be an explanation of the

observed differences in the expression of chlorobenzoate dioxygenase genes in subclones of Tn5271 compared to the wild type strain. *Alcaligenes* sp. BR60 cultures used in biochemical studies were selected through many passages for their efficient use of 3-chlorobenzoate, while the DNA used to create the subclones pBRCN2-13 was cloned in earlier genetic studies of *Alcaligenes* sp. BR60 shortly after its recovery from the Hyde Park landfill site (Wyndham et al. 1988). There may be differences in dioxygenase expression, substrate specificity, or both that have developed with time in the laboratory. Analyses at the molecular level are required to determine whether changes have occurred in the nucleic acid sequences.

Another factor that may account for differences between the subclones and the wild type is that the recipient strain BR6024 used in our genetic studies was previously mutagenized in order to recover the tryptophan auxotroph (Fulthorpe and Wyndham 1991). Mutagenesis with N-methyl-N'-nitro-N-nitrosoguanidine may have caused unspecified mutations in the BR6024 genome. We have determined that BR6024 is impaired in its ability to metabolize benzoate and the phenotype is not restored by the presence of the Tn5271 genes. The location of the mutation is not known but may affect 3-chlorobenzoate degradation.

Lastly it is possible that the Tn5271 chlorobenzoate dioxygenase functions at both the 3,4 and 4,5 carbons of the

3-chlorobenzoate ring (Figure 4.1b) in the wildtype *Alcaligenes* sp. strain BR60. There may be more efficient utilization of the resulting 3-chloro-4,5-dihydroxybenzoate intermediate than protocatechuate in the wild type, so that the chlorinated dihydroxybenzoate does not accumulate enough to be detected by GC-MS or HPLC analyses. Kukor and Olsen (1992) found both 3- and 4- methylcatechol as products in the oxidation of *m*-cresol after purification of their phenol hydroxylase while crude extracts revealed only the presence of 3-methylcatechol in their original study (Kukor and Olsen 1990). Now that the genes encoding the Tn5271 chlorobenzoate dioxygenase are available on a broad-host-range expression vector, further biochemical studies require only the synthesis of the intermediates that are not commercially available.

The nested deletions produced some interesting results. After mating the constructs pBRCN8, 9 and 10 into *Alcaligenes* sp. strain BR6024, black-purple products accumulated around transconjugant colonies, but the cultures were able to recover the normal growth phenotype after subculturing. This suggests that one or more metabolites accumulated for a number of generations, after which the organism was able to adjust to efficiently metabolize the intermediate. GC-MS analysis showed that the metabolites in the medium were chloro-hydroxybenzoate, chloro-dihydroxybenzoate and protocatechuate. The auto-oxidation

of *cis*-dihydroxylated aromatic compounds such as protocatechuate and catechol results in purple-black discolouration in the medium (Parke 1992). The metabolite accumulating in the *E. coli* pBRCN13a clone that also turned the medium purple-black was identified as a chloro-dihydroxybenzoate. Kersten et al. (1985) showed that the protocatechuate-4,5-dioxygenase isolated from *Pseudomonas* (*Comamonas*) *testosteroni* is functional on substituted protocatechuates containing chlorine or methyl groups in the 5 carbon position. A protocatechuate-4,5-dioxygenase activity is expressed from a chromosomal gene in *Alcaligenes* sp. BR60, as indicated by growth on *p*-hydroxybenzoate and protocatechuate, and accumulation of the *meta*-ring-fission product of protocatechuate in strains lacking pBRC60 and Tn5271. The accumulation of a purple-black autooxidation product in Tn5271 subclones with deletions in the region of the *Hind*III site at 9.8 kb, but not in strains carrying larger subclones, may be due to the elimination of a transposon encoded protocatechuate dioxygenase activity that is disrupted by subcloning, then functionally replaced by genes on the chromosome. Alternatively, a gene involved in the regulation of protocatechuate dioxygenase activity may be lost. Further studies are in progress to investigate this region for possible function but the pBRCN14 clone containing the *Eco*RI-*Not*I fragment did not produce unique proteins when induced with IPTG.

The expression of the nested deletion clones pBRCN11 and 12 raised questions regarding the activities coded in the region from the 7 kb to 8.1 kb position of Tn5271. The major metabolite accumulating in the medium of the pBRCN11 clone was a chloro-hydroxybenzoate. This was most likely a spontaneous dehydration product formed from a chloro-dihydrodiol metabolite after acidification of the medium during ethyl acetate extraction (Gibson et al. 1970). Spontaneous dehydration of dihydrodiols to yield phenols has been observed by other investigators (Hickey and Focht 1990). The pBRCN11 clone also produced a metabolite with a mass spectrum equivalent to a chloro-dihydrodiol of cyclohexadiene-carboxylate, the expected product of a 3-chlorobenzoate-4,5-dioxygenase. This clone was not able to grow on 3-chlorobenzoate, despite the fact that protocatechuate and the chloro-dihydrodiol metabolite were found by GC-MS analysis. The absence of an accumulation of purple-black colour in transconjugants carrying pBRCN11, while vicinal diols and chloro-hydroxybenzoate were detected by GC-MS, suggests that the region around 8 kb encodes a chlorobenzoate dioxygenase component which metabolized 3-chlorobenzoate to the 3-chloro-4,5-dihydrodiol and 3-chloro-4-hydro-3,4-diol metabolites, but that these were not further metabolized. These diol structures could re-aromatize chemically during the extraction process either by the removal of H₂O or HCl, accounting for the presence of

minor amounts of hydroxylated compounds in the GC-MS analysis. *In vivo* cis-diol structures have been found to re-aromatize both enzymatically and spontaneously. Enzymatic re-aromatization of the cis-diol structures to catechol is conducted with dehydrogenases during benzoate degradation by *P. putida* and *A. calcoaceticus* (Neidle et al. 1992), whereas the conversion of 4-sulphobenzoate to protocatechuate occurs spontaneously in *C. testosteroni* (Locher et al. 1991). The accumulation of the cis-diol product in the medium in our studies and lack of growth by the transconjugants on 3-chlorobenzoate suggests that this region codes for a dehydrogenase and spontaneous re-aromatization is not sufficient in this organism for growth.

The clone carrying pBRCN12 did not grow on 3-chlorobenzoate and in this clone 3-chlorobenzoate remained unchanged. This suggests that information in the 7 kb region is essential for dioxygenase formation.

Two unique protein bands were distinguished in the pBRCN13a clone by SDS-PAGE corresponding to 51 and 42 kd (Figure 4.3). Previously described dioxygenases are multicomponent enzymes composed of two or three subunits. The two component oxygenase systems for 4-sulphobenzoate-3,4-dioxygenase (Locher et al. 1991) and 4-chlorophenylacetate-3,4-dioxygenase (Markus et al. 1986, Schweizer et al. 1987) appear to have some similarity in both polypeptide size and stereospecificity. Both are

composed of approximately 50 kd oxygenase subunits and 36 kd reductases, and both introduce oxygen at the distal 3,4-carbon positions of aromatic rings containing carboxyl groups linked to the 1- position.

In this report we demonstrate that Tn5271 carries genes for 3-chlorobenzoate dioxygenase within a 3.7 kb *NotI*-*EcoRI* fragment. This dioxygenase activity can be expressed in *E. coli* under the control of the *tac* promoter. The expression of this fragment produces two proteins approximately 51 and 42 kd in size. The orientation of the chlorobenzoate dioxygenase attack is distal to the carboxyl group and is therefore unlike the chlorobenzoate dioxygenases which initiate 3-chlorobenzoate metabolism via chlorocatechols in the fluorescent pseudomonads and *Alcaligenes eutrophus*. These polypeptides are likely involved in a two component dioxygenase system. This leaves us with an interesting question. The results suggest that spontaneous dehydrogenation does not occur but the region cloned appears to code for only two polypeptides. Is there a component missing in our protein analysis or is one component capable of both acting as a reductase component in the dioxygenase and also as a dehydrogenase? These questions are the topic of investigation in Chapter 5.

CHAPTER 5: NUCLEOTIDE SEQUENCE OF *cbaABC* GENES

INTRODUCTION

In the previous chapter the potential pathways that exist for 3-chlorobenzoate degradation by aerobic Proteobacteria were discussed. Evidence was presented that *Alcaligenes* sp. strain BR60 initiated the degradation of 3-chlorobenzoate by dioxygenation of the chlorinated benzoate ring. This information is carried on a 3.7 kb *NotI*-*EcoRI* fragment within the transposon Tn5271. The Acidovorans complex (Comomonadaceae, Willems et al. 1991a) has not been shown to degrade benzoate by an initial dioxygenase attack, therefore in this chapter we investigate the potential ancestry of these genes by comparing their deduced amino acid sequences to known sequences from other dioxygenases that attack aromatic rings to form *cis*-diols.

The nucleotide sequence and deduced amino acid sequence have been determined for two dioxygenases that attack benzoate, *benABC* from *Acinetobacter calcoaceticus* (Neidle et al. 1991) and *xylXYZ* from *Pseudomonas putida* (Harayama et al. 1991). The *benABC* genes code for an enzyme complex that is specific for benzoate, whereas the *xylXYZ* genes code for the *meta* pathway toluate-1,2-dioxygenase that has a broad specificity, enabling it to act on 3- and 4- substituted benzoates. Other dioxygenases involved in the conversion of aromatic hydrocarbons to *cis*-diols that have been studied at

the molecular level are biphenyl dioxygenases from *Pseudomonas* strain LB400 (Erickson and Mondello 1992) and *Pseudomonas pseudoalcaligenes* (Furukawa et al. 1989); and the benzene dioxygenase (Irie et al. 1987), toluene dioxygenase (Zylstra and Gibson 1989) and naphthalene dioxygenase (Kurkela et al. 1988) from various *Pseudomonas putida* strains. All these dioxygenases have been isolated from members of the fluorescent pseudomonads and are involved in the attack of the benzene ring in the 1,2- or 2,3- carbon positions. There is only limited amino acid sequence information available for a two component dioxygenase that attacks the benzoate ring in the 3,4- or 4,5- carbon positions. Locher et al. (1991) sequenced the N-terminal amino acids of the two components of 4-sulphobenzoate-3,4-dioxygenase.

There is more literature available on the biochemistry of dioxygenases that attack aromatic compounds. Dioxygenases that have been described are multicomponent systems that are usually composed of two or three proteins. It appears that in the two component systems such as 4-sulphobenzoate-3,4-dioxygenase (Locher et al. 1991), 2-halobenzoate-1,2-dioxygenase (Fetzner et al. 1992), phthalate oxygenase (Batie et al. 1987), 4-methoxybenzoate monooxygenase (Bernhardt et al. 1975), 4-chlorophenylacetate-3,4-dioxygenase (Markus et al. 1986, Schweizer et al. 1987) and benzoate dioxygenase (Yamaguchi and Fujisawa 1978, 1980,

1982; Neidle et al. 1987), one component is an iron-sulphur flavoprotein reductase and the other is an iron-sulphur terminal oxygenase. Three component systems such as toluene-2,3-dioxygenase (Subramanian et al. 1979, 1981, 1985), benzene dioxygenase (Axcell and Geary 1975, Crutcher and Geary 1979), naphthalene dioxygenase (Ensley and Gibson 1983; Haigler and Gibson 1990a, b) and toluene-4-monooxygenase (Whited and Gibson 1991) have, in addition to the terminal oxygenase, two components involved in electron transport, one element is a flavoprotein and the other is a small ferredoxin component.

In chapter 4 I demonstrated that the chlorobenzoate dioxygenase encoded on Tn5271 of *Alcaligenes* sp. BR60 appears to attack the benzoate ring in the 3,4- or 4,5-carbon positions. The expression of these genes produced two unique polypeptides, 41 and 52 kd in size, when resolved on an SDS-PAGE gel. In this chapter I find that three potential open reading frames (ORF's) are encoded in the 3.7 kb NotI-EcoRI fragment of Tn5271. I have designated these genes *cbaA*, *cbaB* and *cbaC*. The *cbaA* and *cbaB* genes code for amino acid residues that are consistent with conserved amino acid sequences found in other terminal oxygenase and reductase components of dioxygenases (Neidle et al. 1991). These conserved regions are the Rieske-type [2Fe-2S] centre in the terminal oxygenase component and chloroplast-like ferredoxin, FAD and NAD binding regions involved in electron

transport in the reductase component. The third ORF, *cbaC*, is possibly an oxidoreductase component also, because of similarities in its amino acid sequence to other oxidoreductases. This putative protein component may be a dehydrogenase involved in the re-aromatization of the ring after *cis*-diol formation. Protein expression in *E. coli* yielded two proteins the expected size of the *cbaA* and *cbaC* gene products (51 kd and 42 kd, respectively), but did not yield a 32 kd polypeptide fragment, the predicted product size of the *cbaB* gene. Mutation of the clone pBRCN5 within *cbaB* suggested that 3-chlorobenzoate degradation was affected in *Alcaligenes* sp. strain BR6024. This lead me to speculate that CbaB is not a stable protein and can not be easily maintained *in vitro*. Alternatively, the method employed for the resolution of unique proteins is not adequate to detect this protein.

MATERIALS AND METHODS

STRAINS AND GROWTH CONDITIONS

Alcaligenes sp. strain BR60, *Escherichia coli* DH5a and the strain harbouring the pUC18 construct pBRH2 have been described previously, along with their growth conditions (Wyndham and Straus 1988, Wyndham et al. 1988). The broad host range, mobilizable construct pBRCN5 containing the dioxygenase coding region of Tn5271, and pBRCN11 and 12, were described in Chapter 4.

Figure 5.1: Location of *cbaABC* on Tn5271

5.1a. Restriction enzyme map of Tn5271. B, *BstEII*; E, *EcoRI*; Ec, *EcoRV*; H, *HindIII*; N, *NotI*; P, *PstI*; S, *ScaI*.

5.1b. Focus on region of Tn5271 cloned into pBW13 forming pBRCN5 construct containing *cbaABC* genes, open reading frames designated by (→, upper strand), distance is marked in kilobases.

5.1c. Mutations of pBRCN5 used to determine function of genes, pBRCN7, pBRCN11 and 12 construction described in chapter 4, pBRCN15 and 16 mutations created by deletion of restriction enzyme sites *EcoRV* and *BstEII* respectively.

a.

b.

c.

DNA MANIPULATIONS

DNA was extracted for cloning and for single stranded sequencing templates using the methods described in Chapter 2. Potential ORF's were tested to determine if they were required for chlorobenzoate dioxygenase activity by creating site mutations. ORF3 was mutated by restricting pBRCN5 with *Bst*EII (Figure 5.1) and treatment with exonuclease III (Boehringer Mannheim) for one second, the site was then repaired and religated as described in Sambrook et al. (1989). The only available unique site within ORF2 was *Eco*RV but this site also existed within the vector, pBW13, therefore the *Hind*III-*Pst*I fragment (Chapter 4) was cloned into pUC18 then the *Eco*RV site was mutated by treatment with exonuclease III as described above. The modified fragment was then excised from pUC18 and recloned into pBW13 and conjugated into *Alcaligenes* sp. strain BR6024 as described previously (Chapter 4).

DNA SEQUENCING AND ANALYSIS

The *Hind*III fragment 2 of Tn5271, pBRH2 (Wyndham et al. 1988, Nakatsu et al. 1991) was restricted with *Eco*RI then cloned into the replicative forms of M13mp18 and M13mp19 (Yannisch-Perron et al. 1985). Subclones of the 2.9, 1.6 and 1.5 kb *Eco*RI fragments designated pBRE11, pBRE13 and pBRE14 respectively were used for sequencing (Figure 5.1). The method for sequencing strands in both orientations using

the commercial Sequenase kit (U.S. Biochem. Corp.) in conjunction with synthesized oligonucleotides has been described (Nakatsu et al. 1991, Chapter 2).

General sequence analysis was performed using Microgenie (Beckman). The National Center for Biotechnology (NCBI) BLAST (Basic Local Alignment Search Tool) network services were used to search for similarities to our nucleotide sequences and deduced amino acid sequences. The non-redundant components of their peptide sequence databases which includes SwissProt, PIR, GenPept (translations from GenBank) and GUpdate (cumulative daily updates to the major release), were searched using the Blastx and Blastp programs (Altschul et al. 1990).

RESULTS

NUCLEOTIDE SEQUENCE ANALYSIS

The nucleotide sequence from the NotI site within the EcoRI fragment 11 of pBRE11 and all of pBRE13 (EcoRI fragment 13) and pBRE14 (EcoRI fragment 14) are presented in Figure 5.2. Six potential open reading frames (ORFs) were deduced by computer analysis, three coding on the upper strand and three from the lower strand. Previous studies (Nakatsu et al. 1992, Chapter 4) demonstrated that unique protein products were produced only in the NotI to EcoRI orientation, the upper strand in this study, therefore only these deduced amino acid sequences are included in Figure

Figure 5.2: Nucleotide and deduced Amino Acid Sequence of *cbaABC*

Nucleotide sequence and deduced amino acid sequence of upper strand of *NotI-EcoRI* fragment of Tn5271 containing putative *cbaABC* genes. Potential ribosomal binding sites indicated by underlines. Tentative initiation codons are in bold letters.

cbaABC

4447 (NotI)
 GCG GCCGCTGCCC GCTGCATCAC CTCCGTCGTC GGCTCGGGTA AACACCTAGA
 TTTTCGATCT GAAAACCTCC GATAGTTACA AAACCAACTA CATGGATCGA ATCATCTCAG
 GTTTTCCCT CATAAGGGTT TGACACAAGT TGCTAAACCA ACAACATCCG CAAAATTTC

4631
AGGAGCAAGCT ATG GCA TTG ACC AAG GAC AAC GAA GAT TTG GTG CGA GTG GGC CGT
 Met Ala Leu Thr Lys Asp Asn Glu Asp Leu Val Arg Val Gly Arg

GGA ACC CCG ATG GGC GGC CTC ATG CGC GAG TAC TGG ATC CCC GCA TTG AAA TCG
 Gly Thr Pro Met Gly Gly Leu Met Arg Glu Tyr Trp Ile Pro Ala Leu Lys Ser

ACC GAG CTC GAG GCT GGC GGA AGT CCT GTG AGG CTG CTG CTC TTG GGC GAA AAG
 Thr Glu Leu Glu Ala Gly Gly Ser Pro Val Arg Leu Leu Leu Leu Gly Glu Lys

CTC GTC GCC TTC AGG GAG CCG TCG GGT GCG GTC GGC GTG ATG GAC AGC CGA TGC
 Leu Val Ala Phe Arg Glu Pro Ser Gly Ala Val Gly Val Met Asp Ser Arg Cys

CCC CAC CGT GGC GTT TCC CTG TTC ATG GGG CGC GTT GAA GAG GGC GGG CTG CGC
 Pro His Arg Gly Val Ser Leu Phe Met Gly Arg Val Glu Glu Gly Gly Leu Arg

TGC GTC TAC CAC GGT TGG AAA TTC AGC GCT GAG GGC AAA TGC GTT GAC ATG CCC
 Cys Val Tyr His Gly Trp Lys Phe Ser Ala Glu Gly Lys Cys Val Asp Met Pro

AGT GTC CGT CCT GAA GAC GAA TTC AAA AAC AGC GTG CGC GTG GCG CGA TAC CCG
 Ser Val Arg Pro Glu Asp Glu Phe Lys Asn Ser Val Arg Val Ala Arg Tyr Pro

GTG AAA GAA ATG GCG GGC GTT GTC TGG GTG TAC ATG GGT ACT CGA AAA GTT TTG
 Val Lys Glu Met Ala Gly Val Val Trp Val Tyr Met Gly Thr Arg Lys Val Leu

CCC GAG CTC CCG CCG CTG GAA GTG CTG AAT CTT CCG GAA AAC GAA GTG GAT GTC
 Pro Glu Leu Pro Arg Leu Glu Val Leu Asn Leu Pro Glu Asn Glu Val Asp Val

ATT TGC CTC CAG CGA AAG TCG AAC TGG TTG CAA AAC CTG GAA GGT GAA ATT GAT
 Ile Cys Leu Gln Arg Lys Ser Asn Trp Leu Gln Asn Leu Glu Gly Glu Ile Asp

ACC TCG CAT TTC AAC TTC CTT CAC GTG GGC GGA CTC CAC GCG GAC GAA GTG CCG
 Thr Ser His Phe Asn Phe Leu His Val Gly Gly Leu His Ala Asp Glu Val Pro

GAC GAT CAC CCA TTG AAA TAC ACC GCC CAA GTC GCT CCG CAG TAC CTG GTC AAG
 Asp Asp His Pro Leu Lys Tyr Thr Ala Gln Val Ala Pro Gln Tyr Leu Val Lys

GAA ACG GCA TTG GGC ACC TGC TAC GCG GCT CAA GTG CCG GCG GAA GAG GAT CAC
 Glu Thr Ala Leu Gly Thr Cys Tyr Ala Ala Gln Val Pro Ala Glu Glu Asp His

ACC TAC ACG CCG TTT GCG CAT TTC CTT TTC CCG TTC TGG GCG CTC ATT CCG CAA
 Thr Tyr Thr Arg Phe Ala His Phe Leu Phe Pro Phe Trp Ala Leu Ile Pro Gln

GCC GAT ATT GCG CAG AAC ATT TTG GCC AGA GCC TGG GTG CCC ATG GAC GAT GAG
 Ala Asp Ile Ala Gln Asn Ile Leu Ala Arg Ala Trp Val Pro Met Asp Asp Glu

CAC ACC ATG ATG TTC TTC TTC CGT TGG ACG GGA TCC AAG GCC AAA CGC CTC GAT
 His Thr Met Met Phe Phe Phe Arg Trp Thr Gly Ser Lys Ala Lys Arg Leu Asp

ACG CCA TTG AAG AGC GGT TCG CCC ATG CCC GGG GTC ACG CTG ACG GAC ATG AAG
 Thr Pro Leu Lys Ser Gly Ser Pro Met Pro Gly Val Thr Leu Thr Asp Met Lys

TAC AAG GAA AAC ACG ACG GAT TGG TAC GGC CGC TGG CAG CCG CTC GGC GAT GAG
 Tyr Lys Glu Asn Thr Thr Asp Trp Tyr Gly Arg Trp Gln Pro Leu Gly Asp Glu

AGC AAC GAT TGG CTC ATC GAT CGA GAT TTG CAA AAA GTA GGC CGG GTG TTC TCC
 Ser Asn Asp Trp Leu Ile Asp Arg Asp Leu Gln Lys Val Gly Arg Val Phe Ser

GGT ATT TAC GGA ATC CAC GCA CAG GAT CAG GCC ATG ACC GAC AGC ATG GGG CCG
 Gly Ile Tyr Gly Ile His Ala Gln Asp Gln Ala Met Thr Asp Ser Met Gly Pro

ATC ATT GAC CAC GGC CTG GAG CAA TTG GCA CCG ACC GAT CTG ATG ATC GTC AGA
 Ile Ile Asp His Gly Leu Glu Gln Leu Ala Pro Thr Asp Leu Met Ile Val Arg

ACC CGT CGC CGG ATT CTC AAG GCG CTC CGT GCG CAC GAA GCG AAT GGC ACC CTG
 Thr Arg Arg Arg Ile Leu Lys Ala Leu Arg Ala His Glu Ala Asn Gly Thr Leu

CCC CCG GGT GTT GAC GAG GCG GAT CAG TAC TTC ACG CCC AGG AGC GGC TAT TAC
 Pro Pro Gly Val Asp Glu Ala Asp Gln Tyr Phe Thr Pro Arg Ser Gly Tyr Tyr

CTC ACG CCG AAA AGC GTC GAC TGG GAG ACC GCT TAC GAA CAA AGG ATC GAG GGC
 Leu Thr Pro Lys Ser Val Asp Trp Glu Thr Ala Tyr Glu Gln Arg Ile Glu Gly

5929

CTC GTG CGG TGA A TCTGCAAGCA CTGCGCAATG ACGCGTCTT GGATGACCGC
 Leu Val Arg *

5983

GGAGGTGATG CC ATG GTC GCA ATC GAT CAA CAC GAT ACC TAC AGT GTT CCG GTG
 Met Val Ala Ile Asp Gln His Asp Thr Tyr Ser Val Arg Val

ATA TCC CGA TCG CAC CTG AGC AAG GAT ATT GTT CAG GTG GAA CTC GAG GAA TCG
 Ile Ser Arg Ser His Leu Ser Lys Asp Ile Val Gln Val Glu Leu Glu Glu Ser

AGC CAG CGA CCA TTG CCG GAC TAT GAG CCG GGA TCG CAC GTA GAC ATC TAT GTC
 Ser Gln Arg Pro Leu Pro Asp Tyr Glu Pro Gly Ser His Val Asp Ile Tyr Val

CAG GAT GAC CTT GTC CCG CAA TAC TCG CTG GTC AAG GCG TCC GAC GCG CAG GCC
 Gln Asp Asp Leu Val Arg Gln Tyr Ser Leu Val Lys Ala Ser Asp Ala Gln Ala

TCG TAC CAG ATC GCC TTC AAG GTC AAG CGC GAT CAG GGC AGT GCG ACG GAG CTG
 Ser Tyr Gln Ile Ala Phe Lys Val Lys Arg Asp Gln Gly Ser Ala Thr Glu Leu

ATG TGC GAG TTG CTC CGT GTC GGT GCA ACA ACC CGA ATC TCT GCG CCG AGG AAC
 Met Cys Glu Leu Leu Arg Val Gly Ala Thr Thr Arg Ile Ser Ala Pro Arg Asn

GCT TTT GCC CTG GAT CCG CAG GCG CGC GAG ACC GTC CTG ATT TGT GGC GGA ATA
 Ala Phe Ala Leu Asp Pro Gln Ala Arg Glu Thr Val Leu Ile Cys Gly Gly Ile

GGA ATA ACG CCC ATG GTT CAT ATG GCC ATG ACC CTG GTC AAA GCG AAG CCG CCA
 Gly Ile Thr Pro Met Val His Met Ala Met Thr Leu Val Lys Ala Lys Arg Pro

TGG AGC ATG CAC ATT GCC TCC CGG GAT GGC GAT GAG CTC GAT TTG CTG GGG CCG
 Trp Ser Met His Ile Ala Ser Arg Asp Gly Asp Glu Leu Asp Leu Leu Gly Pro

CTG GGA AGT TGC TCC GAG ATC AGC CGC TAC ATC TCG TCC CAG GGG GAC AGA TTG
 Leu Gly Ser Cys Ser Glu Ile Ser Arg Tyr Ile Ser Ser Gln Gly Asp Arg Leu

CCG ATT CCG GAC CTG GCC GAG CGC GCA CCT GCG AAT GCG CAT CTG TAC TTC TGC
 Pro Ile Arg Asp Leu Ala Glu Arg Ala Pro Ala Asn Ala His Leu Tyr Phe Cys

GGG CCT GAG GGC ATG TTG CAG GAA TTC ACG GTG AAT CTT GCC CCG AGT GGG GCG
 Gly Pro Glu Gly Met Leu Gln Glu Phe Thr Val Asn Leu Ala Arg Ser Gly Ala

CAG TAT GTC GTG CGC GAG GGC GAG ACG ATA TTG GAT GTC CTG CGC AAT GCC GGC
 Gln Tyr Val Val Arg Glu Gly Glu Thr Ile Leu Asp Val Leu Arg Asn Ala Gly

CAC CAT GTC ACG TCC TCC TGT CCG CAG GGG ATT TGT GGC ATG TGT GAA ACC ACT
 His His Val Thr Ser Ser Cys Arg Gln Gly Ile Cys Gly Met Cys Glu Thr Thr

TTG ATT TCC GGC GTG CCG GAC CAC AGG GAC CGC TTG CTG ACC GAT AGC GAA AAG
 Leu Ile Ser Gly Val Pro Asp His Arg Asp Arg Leu Leu Thr Asp Ser Glu Lys

GCC TCG GGT CGA ACC ATG TTG ATT TGC TGT TCA CGC GCG CTC AGC CCG GAA CTC
 Ala Ser Gly Arg Thr Met Leu Ile Cys Cys Ser Arg Ala Leu Ser Pro Glu Leu

6849 6852
 ACC CTG GAT CTG TGA A T ATG AGA ATA GGG GTC GTG GGC CTG GGG AAG GCA
 Thr Leu Asp Leu * Met Arg Ile Gly Val Val Gly Leu Gly Lys Ala

TTT ATG CTG ATG CTG CCG ACG TTT CTC ATG GAT CCG CCG GTT CAA CTG GTC GCT
 Phe Met Leu Met Leu Pro Thr Phe Leu Met Asp Arg Arg Val Gln Leu Val Ala

GCA TCG GAT ACC GAT CCG CTG AGC CTG CGC CAA TTC AAG GCG GAT TTC CCC GCG
 Ala Ser Asp Thr Asp Pro Leu Ser Leu Arg Gln Phe Lys Ala Asp Phe Pro Ala

GCG GCC GTG CAT GGC GAC ATC GAG TCG CTG TGC AAG AAC CCG GAT GTC GAG GTC
 Ala Ala Val His Gly Asp Ile Glu Ser Leu Cys Lys Asn Pro Asp Val Glu Val

GTC TAC ATC GGC ACC CCT CAC CAG TTT CAT GCT GTT CAC GCG GAA ATT GCG CTG
 Val Tyr Ile Gly Thr Pro His Gln Phe His Ala Val His Ala Glu Ile Ala Leu

AAC GCC GGA AAG CAT GTG CTC GTT GAG AAG CCC ATG GCG GTG ACC CTT GAG GAC
 Asn Ala Gly Lys His Val Leu Val Glu Lys Pro Met Ala Val Thr Leu Glu Asp

TGC TGC CGA ATG AAT GCC TGC GCG CAG CGC GCT GGG AAA TAC CTG ATC GTG GGC
 Cys Cys Arg Met Asn Ala Cys Ala Gln Arg Ala Gly Lys Tyr Leu Ile Val Gly

CAC AGC CAC AGC TTC GAT CAT CCG ATC TCC AGA GCC AAG GAG TTG ATC GAC TCC
 His Ser His Ser Phe Asp His Pro Ile Ser Arg Ala Lys Glu Leu Ile Asp Ser

GGA CGA TAT GGG CGC GTC AGG TTC ATT CAC TCG ATG AAC TAC ACG GAT TTC CTG
 Gly Arg Tyr Gly Arg Val Arg Phe Ile His Ser Met Asn Tyr Thr Asp Phe Leu

TAC CCG CCA AGA AGG GCG GAG GAG TTG AAT ACC GAC CTG GGG GGC GGC GTG GTC
 Tyr Arg Pro Arg Arg Ala Glu Glu Leu Asn Thr Asp Leu Gly Gly Gly Val Val

TTC AAC CAA GCG TCC CAT CAG CTC GAT GTG ATT CGC CTG CTG GCG CAA GGC CGC
 Phe Asn Gln Ala Ser His Gln Leu Asp Val Ile Arg Leu Leu Ala Gln Gly Arg

GTG GTG GAC GTT TCC AGC TAT CTG GGT CGC TGG GAC GCT GCC CCG CCG ACC GAA
 Val Val Asp Val Ser Ser Tyr Leu Gly Arg Trp Asp Ala Ala Arg Arg Thr Glu

GGG GCG TAC TCC GCA CTG ATC AAA TTC GAC AAT GGC GTG GCG GCC AAT GTC ACC
 Gly Ala Tyr Ser Ala Leu Ile Lys Phe Asp Asn Gly Val Ala Ala Asn Val Thr

TAC AGC GGC TAC GCC CAC TAC AAC TCT GAC GAG GTG ATG GGC TGG GTC AAT GAG
 Tyr Ser Gly Tyr Ala His Tyr Asn Ser Asp Glu Val Met Gly Trp Val Asn Glu

CTG GCG GAA AAG CAG CCG GAG CCG CCG CTT TTC CCG ACC CGA TTG CGA CTC GAT
 Leu Gly Glu Lys Gln Pro Glu Arg Arg Leu Phe Pro Thr Arg Leu Arg Leu Asp

GAG GTC TTG AAC AAG GGC AGT GAG GCG GCC TAC AAG AAT GAG ATG TCT TAT GGT
 Glu Val Leu Asn Lys Gly Ser Glu Ala Ala Tyr Lys Asn Glu Met Ser Tyr Gly

CGC GGG TGG TCG GCG CAT GGC TTG CCG AAA ATG GCC AGG AAC CAC CAG CAT TTC
 Arg Gly Trp Ser Ala His Gly Leu Pro Lys Met Ala Arg Asn His Gln His Phe

GGC CAT CTG GTG ATC AGC TGC GAA GGT GCC GAC ATC GTT CCC AAG GCG GAA AAA
 Gly His Leu Val Ile Ser Cys Glu Gly Ala Asp Ile Val Pro Lys Ala Glu Lys
 ATC GAA GTA TTT GCC AAT GGC GAA TAC CAC TGC GAG AAC TTC AGC ATG CCG GAT
 Ile Glu Val Phe Ala Asn Gly Glu Tyr His Cys Glu Asn Phe Ser Met Pro Asp
 TTT CCG CGC CAA GAA GTC ATG GAC GAA TTG TTT GAC GCT ATT TCC GGC ACG CGC
 Phe Pro Arg Gln Glu Val Met Asp Glu Leu Phe Asp Ala Ile Ser Gly Thr Arg
 CCC CCG GTT CAT ACC GGC GAG TGG GCC ATG GAA ACC CTC GAT CTT TGC ATT GCA
 Pro Pro Val His Thr Gly Glu Trp Ala Met Glu Thr Leu Asp Leu Cys Ile Ala
 CTG CTT GAA GGG GGG AAA CCT GAG TTG ATC GGG ATG CGC GCG ACG GCC GTC GCC
 Leu Leu Glu Gly Gly Lys Pro Glu Leu Ile Gly Met Arg Ala Thr Ala Val Ala

8045

TGT GCT GGT CGG TCC TCG GGG CGG TAG AAGTC GGGCTAATTC ACTTGTTGGC
 Cys Ala Gly Arg Ser Ser Gly Arg *

8104 (EcoRI)

GGTTTTGCTT TATTAATAGT TGGAGATGGC GATGAATTC

5.2. I designated these ORFs, *cbaA*, *cbaB* and *cbaC*.

The first ORF, *cbaA* has potential initiation codons at nucleotides 4631, 4685 and 4697 of Tn5271 where nucleotide 1 is the beginning of IS1071 (Chapter 2). The potential ribosome binding site preceeding nucleotide 4631 is AGGAG, which closely resembles the ribosomal binding site consensus sequence for *E. coli*, therefore I have chosen this as our tentative initiation codon. This ORF spans from nucleotide 4631 to 5929 and codes for 433 amino acids with a deduced molecular weight of 48,932 d. The second ORF, *cbaB* has an initiation codon at nucleotide 5983 with a potential ribosome binding site with the sequence GGAGG. The ORF spans to nucleotide 6849, coding for 288 amino acids with a deduced molecular weight of 31704. The third ORF, *cbaC* has initiation codons at nucleotides 6852 and 6888 with potential ribosome binding sites with the sequences GAA and AAGG respectively. The first possible initiation codon was used to analyse this ORF that would then span from nucleotide 6852 to 8045. This 398 amino acid polypeptide would have a deduced molecular weight of 44096.

The 50 base pairs upstream from *cbaA* was comprised of 54% A+T, much greater than the 42% found in the coding regions. This suggests that there is a promoter region upstream from *cbaA* but there was no apparent homology to promoter consensus sequences of *Pseudomonas* (Deretic et al. 1989, Ronald et al. 1992) or *E. coli* (Hawley and McClure

1983). In order to express these genes in *E. coli* it was necessary to use a *tac* promoter, therefore it appears that the promoter of this organism is not readily recognized by *E. coli*.

CODON USAGE AND G+C CONTENT

The G+C composition of *cbaA* is 58.7%, *cbaB* is 59.2% and *cbaC* is 59.4%. The number of deduced amino acids and the codons utilized are listed in Table 5.1. The most commonly used amino acids in: *cbaA* are leucine, glycine, valine, alanine and arginine; *cbaB* are leucine, serine, alanine and arginine; *cbaC* are alanine, leucine, glycine, valine and arginine. There is a preference for G or C in the wobble position (third nucleotide of a codon) for all the amino acids except: aspartic acid, glutamine, isoleucine and lysine in *cbaA*; aspartic acid, cysteine and isoleucine in *cbaB*; and histidine in *cbaC*.

SIMILARITIES TO OTHER OXYGENASES

The NCBI search of the peptide sequence databases listed the translated *vanA* component of the vanillate demethylase genes of a *Pseudomonas* species (Brunel and Davison 1988) as having the greatest similarity to *cbaA*, as illustrated in Figure 5.3a. The region of similarity has 35% identical and 62% similar matches over 59 amino acids.

There was also a strong match to the N-terminal

Table 5.1: Distribution of codons in the *cbaABC* genes.

Amino Acid	Codon	Number of Codons			Amino Acid	Codon	Number of Codons		
		<i>cbaA</i>	<i>cbaB</i>	<i>cbaC</i>			<i>cbaA</i>	<i>cbaB</i>	<i>cbaC</i>
Ala	GCT	5	1	6	Phe	TTT	1	1	6
	GCA	5	3	4		TTC	14	3	10
	GCC	7	9	14	Pro	CCT	2	2	2
	GCG	13	10	16		CCA	2	2	1
Arg	CGT	6	1	0		CCC	9	1	4
	CGA	6	4	4	Ser	CCG	15	8	10
	CGC	7	9	10		TCT	0	1	2
	CGG	5	6	9		TCA	0	1	0
Asn	AGA	2	1	3		AGT	2	4	1
	AGG	4	2	3	Thr	TCC	3	8	7
	AAT	2	3	7		TCG	5	7	5
	AAC	9	1	8		AGC	8	6	7
Asp	GAT	16	12	11		ACT	1	1	0
	GAC	12	8	11	Typ	ACA	0	1	0
Cys	TGT	0	5	1		ACC	13	8	9
	TGC	5	4	7		ACG	10	5	4
Gln	CAA	7	2	5		TGG	11	1	4
	CAG	7	11	5	Tyr	TAT	1	3	3
Glu	GAA	17	6	11		TAC	14	5	10
	GAG	14	11	16	Val	GTT	6	3	6
Gly	GGT	7	2	4		GTA	1	1	0
	GGA	5	4	2		GTC	11	11	12
	GGC	19	8	16	End	GTG	16	5	12
	GGG	4	5	12		TAA	0	0	0
His	CAT	2	3	9		TGA	1	1	0
	CAC	10	6	8		TAG	0	0	1
Ile	ATT	8	7	5					
	ATA	0	4	1					
	ATC	6	6	10					
Leu	TTA	0	0	0					
	CTT	3	2	4					
	CTA	0	0	1					
	TTG	12	9	7					
	CTC	15	5	5					
	CTG	12	14	18					
Lys	AAA	10	1	5					
	AAG	9	6	10					
Met	ATG	16	9	14					

Figure 5.3: Alignment of CbaA to VanA and Component A of Sulphobenzoate dioxygenase

5.3a. Alignment of deduced amino acid sequences of CbaA (residue 50 to 108) and VanA (residue 5 to 63) of vanillate demethylase (Brunel and Davison 1988). * Identical matches, | similar residues. There are 21/59 (35%), identical matches and 37/59 (62%) similar residues.

5.3b. Alignment of deduced amino acid sequence of *cbaA* (residue 1 to 35) and component A (residue 1 to 35) of 4-sulphobenzoate 3,4-dioxygenase (Locher et al. 1991). There are 17/35 (48%) identical matches and 22/35 (62%) similar residues.

a)

CbaA (50-108)

EKLVAFREPSGAVGVMDSRCPHRGVSLFMGRVEEGGLRCVYHGWKFS AEGKCVDMP SVR
 |||*|||*|||*****|*|*|||* * * * *||*||**|*
 ERMVIYRGAGQRVAALEDFCPHRGAPLSLGSIQDGKLVCGYHGLVMD CDGRTASMPAQR

VanA (5-63)

b)

CbaA	(1-35)	1	MALTKDNEDLVRVGRGTPMGGLMREYWIPALKSTE	35
			* * * * * * * * * * * * * * *	
ComponentA	(1-35)	1	MLTSENNQILTRVGPGTAMGXLFRHFXQPALLSEE	35

sequence of the terminal oxygenase component of the 4-sulphobenzoate-3,4-dioxygenase, which was sequenced from the purified protein (Figure 5.3b). There were 48% identical and 62% similar matches over the available 35 amino acid sequence.

The search also showed similarities of *cbaA* to the terminal oxygenase components of the biphenyl dioxygenase genes (Erickson and Mondello 1992, Furukawa et al. 1989). Alignments of these regions proved to be interesting due to their agreement with previously determined conserved regions for Rieske-type iron sulphur centres of the terminal oxygenase components of dioxygenases (Neidle et al. 1991), as illustrated in Figure 5.4.

The second ORF, *cbaB*, had the highest similarity to the ferredoxin-like protein component of vanillate demethylase (Brunel and Davidson 1988). There are three regions of homology in these components that have: 37% identical and 60% similar matches over 111 amino acids; 44% identical and 58% similar matches over 93 amino acids; and 50% identical and 77% similar matches over 18 amino acids. An alignment of the two sequences has 31.6% overall identical matches (Figure 5.5). The C-terminal region of this sequence has similarities to a number of plant type ferredoxins, 33% identical and 50% similar matches over 60 amino acids. This region has conserved cysteine residues, illustrated in Figure 5.6a, characteristic of the iron binding component

Figure 5.4: Alignment of oxygenase component of *cis*-diol forming dioxygenases.

Boxed cysteine and histidine residues are thought to be involved in the interaction of the Rieske type [2Fe-2S] centers. *, residues conserved in all the aligned sequences.

CbaA: chlorobenzoate dioxygenase, this study

VanA: vanillate demethylase (Brunel and Davison 1988)

BphA: biphenyl dioxygenase (Erickson and Mondello 1992)

TodC1: toluene dioxygenase (Zylstra and Gibson 1989)

NdoB: naphthalene dioxygenase (Kurkela et al. 1988)

BenA: benzoate dioxygenase (Neidle et al. 1991)

XylX: toluate dioxygenase (Harayama et al. 1991)

BnzA: benzene dioxygenase (Irie et al. 1987)

CbaA (59-93)	SGAVGVMSDR	C	P	H	RGVSLFMG-RVEEGGLR	C	VY	H	GW
VanA (14-48)	GQRVAALED	C	P	H	RGAPLSLG-SIQDGKLV	C	GY	H	GL
BphA (90-125)	DKSIKVFLNQ	C	R	H	RGMRICRSDAGNAKFT	C	SY	H	GW
TodC1 (86-121)	DASIAVFLNQ	C	R	H	RGMRICRADAGNAKFT	C	SY	H	GW
NdoB (71-106)	DGSIRAFLNV	C	R	H	RGKTLVSVEAGNAKGFV	C	SY	H	GW
BenA (85-120)	NGELNAMINA	C	S	H	RGAQLLGHKRGNKTTYT	C	PF	H	GW
XylX (80-115)	DGELNAFVNA	C	S	H	RGATLCRFRSGNKATHT	C	SP	H	GW
BnzA (86-121)	DASIAVFLNQ	C	S	H	RGMRICRADAGNAKFT	C	SY	H	GW

★ ★ ★★

★ ★ ★

Figure 5.5: Alignment of Amino Acid Sequence of CbaB and VanB

Alignment of deduced amino acid sequence of CbaB and VanB the ferredoxin-like protein of vanillate demethylase (Brunel and Davison 1988). The N-terminal region with potential FAD-isoalloxazine ring and NAD-ribose binding and the C-terminal region resembling plant-type ferredoxins show the highest homology. Amino acid residues presented in lower case type in the VanB sequence did not appear in alignments from databank search. Gaps were introduced into the CbaB sequence in order to achieve the best fit.

* indicates identical matches and ! indicates conservative matches of amino acids residues.

CbaB (1)

MVAIDQHDTYSVRVISRSHLSKDIVQVELEESSQRPLPDYEPGSHVDIYV

 *| | *| *| *| **| ***| *| *| *| *| *|
 mldl-----mirgl-rle-aprilGLELVATDGSPLPTFEAGAHLDLHL

VanB (1)

CbaB (51)

QDDLVRQYSLVKASDAQASYQIAFKVKRDQGSATELMCELLRVGATTRIS

 *** **| *| *| *| | | *| ****| *|
 PGGLVRPYSLCNAPGETHRYCLAVLLDPASRGGSRAVHEQLRVGQHLLTS

VanB (43)

CbaB (101)

APRNAFALDPQARETVLICGGIGITPMVHMAMTLVKAKRPWSMH--I-AS

 **** *| *| | | *| *****| **| *| | | *| | |
 APRNLFPLVAESSRSLLFAGGIGITPILAMAQVLaargdtfelhycvrer

VanB (93)

CbaB (148)

R--DGDE-LD-----L-L----GPLG-SCSEISR-----YISSQGD

 *| *| *| *| *| *| *| *| *| *| *| *|
 rlaafidwleastfaahvhlhaddgptpfdataallrdagdahlyvcgpgg

VanB (143)

CbaB (174)

RLPIRDLA-ERAPANA--HL---YFCG---PEGMLQEFTVNLARSGAQYV

 | | *| *| *| *| *| *| *| *| *| *| *| *|
 fme-hvlgcartagwdetrhreyfaapvqPAGDARAFEGRLARSGLTLO

VanB (193)

CbaB (215)

VREGETILDVLRNAGHHVTSSCRQGICGMCETTLLISGVPDHRDRLTLDSE

 *| | | **| **| | *| *****| *| *| *| *| *| *| *| *|
 VPAERSVAQVLDDAGVCIPLACEQGICGTCLTRVLDGEPEHRDSFLTDAE

VanB (242)

CbaB (265) KASGRTMLICCSRALSPELTDLU (288)

| *| *****| *| *| *|

VanB (292) RARNDQFTPCCSRARSACLVLDL- (314)

Figure 5.6: Proposed conserved regions in the reductase component of dioxygenases.

(*) indicates conserved amino acid in all the sequences compared. (-) indicates gap. (number) indicates the amino acid residue number.

CbaB: chlorobenzoate dioxygenase (this study)

VanB: vanillate demethylase (Brunel and Davison 1988)

XylA: xylene monooxygenase (Suzuki et al. 1991)

XylZ: toluate dioxygenase (Harayama et al. 1991)

BenC: benzoate dioxygenase (Neidle et al. 1991)

Fer1: ferredoxin, *Fischerella* sp. (Hase et al. 1978)

Fer2: ferredoxin, *Anabaena variabilis* (Chan et al. 1983)

5.6a. Conserved cysteines residues, enclosed in boxes, involved in binding [2Fe-2S] centre, characteristic of plant type ferredoxins used for electron transport. Amino acid residues used in alignments, CbaB (234-276); VanB (260-302); Fer1 (38-80); Fer2 (38-80); XylA (49-88); BenC (48-94); XylZ (37-82).

5.6b. Potential NAD-ribose binding region.

5.6c. Potential FAD-isoalloxazine ring-binding region

a)

PLANT TYPE FERREDOXIN

CbaB	TSS	C	RQGI	C	GM	C	ETTLISGVDPH----	RDRL-LTDSEKASGRTMLI	C	C
VanB	PLA	C	EQGI	C	GT	C	LTRVLDGEPEH----	RD-SFLTDAERARNDQFTP	C	C
Fer1	PYS	C	RAGA	C	ST	C	AGKLISGTVDQ----	SDQSFLDDDDQIEAG-YVLT	C	V
Fer2	PFS	C	RAGA	C	ST	C	AGKLVSGTVDQ----	SDQSFLDDDDQIEAG-YVLT	C	V
XylA	PHD	C	KVGS	C	GT	C	KYKLISGRVNE--LTSSAMGLSGDLYQSG-YRLG	C	Q	
BenC	PMD	C	REGE	C	GT	C	RAFCESGNYDMPEDNYIEKALTPEEAQQG-YVLA	C	Q	
XylZ	PLD	C	RDGA	C	GA	C	KCRAESGRYSL-GEEYIEDALSEAEAEQG-YVLT	C	Q	
		*	*	*	*		*	*	*	*

b)

POTENTIAL NAD-RIBOSE BINDING REGION

CbaB	(113-135)	ARETVLICGGIGITPMVHMAMTL
VanB	(106-126)	SSRSLLFAGGIGITPILAMAQVL
XylA	(216-238)	NATMVCVAGGTGLAPIKCVLQSM
BenC	(219-242)	KRPVLMLAGGTGIAPFLSMLQVL
XylZ	(206-229)	KRPLLLLAGGTGLAPFTAMLEKI

c)

FAD-ISOALLOXAZINE RING BINDING

CbaB	(56-62)	RQYSLVK
VanB	(48-54)	RPYSLCN
XylA	(158-164)	RSYSFSA
BenC	(166-172)	RSYSFSS
XylZ	(153-159)	RAYSFSS
		**

of plant type ferredoxins (Neidle et al. 1991). Examination of the two remaining conserved regions between *cbaB* and *vanB* to other ferredoxin-containing reductases revealed that these regions may be involved in FAD and NAD binding (Figure 5.6b, c).

The most significant homology of *cbaC* occurs with glucose-fructose oxidoreductase from *Zymomonas mobilis* (Kanagasundaram and Scopes 1992). There are 35% identical and 56% similarity matches over 92 amino acids (Figure 5.7a). There is a 42 amino acid sequence within this region that has 48% identical matches and this same region has 46% identical matches to D-galactose dehydrogenase from *Pseudomonas fluorescens* (Sperka et al. 1989) and 37% identical to biliverdin reductase from *Rattus norvegicus* (Fakhrai and Maines 1992) (Figure 5.7b).

MUTATIONS WITHIN *cbaB* and *cbaC*

The clone pBRCN16 was created by eliminating the *BstEII* restriction enzyme site within *cbaC* of pBRCN5. The introduction of this plasmid into *Alcaligenes* sp. strain BR6024 did not restore the 3-chlorobenzoate catalytic phenotype. The clone pBRCN15 was created by eliminating the *EcoRV* restriction enzyme site within *cbaB* of pBRCN5. The introduction of this plasmid into *Alcaligenes* sp. strain BR6024 did not initially produce 3-chlorobenzoate degrading transconjugants. After prolonged incubation (three weeks)

Figure 5.7: Alignment of CbaC to other reductases

Alignment of sequences from NCBI databanks, with potentially significant conserved residues. *, identical matches, | conservative matches.

5.7a: Alignment of CbaC and Gfo glucose-fructose oxidoreductase from *Zymomonas mobilis* (Kanagasundaram and Scopes 1992).

5.7b: Alignment of potentially significant conserved residues that may be associated with binding factors involved in reductase activity. The * and | designated above each sequence refers to identical and conservative matches of that sequence to CbaC. Gfo (Kanagasundaram and Scopes 1992); Gal-dh, D-galactose dehydrogenase from *Pseudomonas fluorescens* (Sparka et al. 1989); Rat, rat kidney biliverdin reductase (Fakhrai and Maines 1992).

a)

CbaC (52-	GDIESLCKNPDVEVVYIGTPHQFHAVHAEIALNAGKHVLVEKPMA
	 * * *** * ** * ***** *****
Gfo (139-	SNFDKIAKDPKIDAVYIILPNSLHAEFAIRSFKAGKHVMCEKPMA
CbaC -143)	VTLEDCCRMNACAQRAGKYLIVGHSHSFDHPISRAKELIDSGRYGRV
	 ** ** * * * * * * * ** *
Gfo -230)	TSVADCQRMIDAAKAANKKLMIGYRCHYDPMNRAAVKLIRENQLGKL

b)

CbaC (59-99)	KNPDVEVVYIGTPHQFHAVHAEIALNAGKHVLVEKPMAVTL
	* * *** * ** * ***** *****
Gfo (146-186)	KKPKIDAVYIILPNSLHAEFAIRSFKAGKHVMCEKPMATSV
	 * * ** * ** ***** **
Gal-dh (56-96)	NGPPVDAIAFCTPPQGRFALVQQALAAGKHVLVEKPPCATL
	 * ** * * ***** **
Rat (63-102)	RSQEIDVAYICSESSSHEDYIRQFLQAGKHVLVEYPMTLS
consensus	***** * * AGKHV--E-P

of these plates, colonies did appear on the 3-chlorobenzoate selection plates. When these putative transconjugants were patched to fresh 3-chlorobenzoate selection plates, growth of 10% (5/50) of the patched colonies occurred. Patching of pBRCN15 transconjugants selected for growth on succinate in the presence of the antibiotic markers only, onto 3-chlorobenzoate, produced no growth even after prolonged incubation.

DISCUSSION

The G+C content of *cbaABC* is in the expected range for the Comamonadaceae, which has been determined to be between 57 to 70 % (Willems et al. 1991a). In this family the G+C content of the genus *Comomonas* is 59.7-68.7, and the lower range is found in the species *C. terrigena* (Willems et al. 1991b). With further defining of the family Comomonadaceae and of the genus *Comomonas*, *Alcaligenes* sp. strain BR60 may eventually prove to belong to this genus.

There is no available information at this time on codon usage by species within the Comomonadaceae, therefore I was not able to determine whether these genes evolved from the Comomonadaceae or whether there was a greater resemblance in codon usage to other groups. The fact that these genes are borne on a transposon suggests that they could have originated in an alternative host. Harayama et al. (1991) found that there was an exceptional use of the codons AGA

and AGG for arginine in the *xylX* genes of *Pseudomonas putida*. These codons are rarely used by *Pseudomonas* (West and Iglewski 1988) which led them to speculate that these codons were acquired or maintained during divergence from the ancestral operon. These codons were also used in the *Alcaligenes* sp. BR60 genes *cbaABC* in addition to other codons rarely used in *Pseudomonas* (West and Iglewski 1988). The codons considered rare in *P. aeruginosa* (West and Iglewski 1988), but used at a greater than 10 in 1000 frequency in *cbaABC*, are arginine-CGA, alanine-GCA, isoleucine-ATT, leucine-TTG, glycine-GGG and valine-GTT. In *cbaA* and *B*, glycine-GGA is a rare codon that is used. In *cbaA* and *C*, alanine-GCT is used. In *cbaB* and *C*, asparagine-AAT is used. In *cbaA* arginine-CGT and AGG are used. In *cbaB*, rare codons that are used are: cysteine-TGT, isoleucine-ATA and serine-AGT. And finally, in *cbaC* leucine-CTT and phenylalanine-TTT are rare codons that are used. This use of codons that are rarely used by pseudomonads suggest that these genes do differ from *Pseudomonas* genes.

A search for amino acid sequences similar to the putative *cbaA* gene product revealed the highest scoring alignment with the *vanA* gene of vanillate demethylase (Brunel and Davison 1988) in the N-terminal region. There was also a strong conservation of amino acid sequence of the N-terminal amino acid sequence of the *cbaA* product to

component A of 4-sulphobenzoate-3,4-dioxygenase (Locher et al. 1991). Only the first 35 amino acids have been sequenced thus far for this polypeptide, but there were 48% identical matches to the first 35 amino acids deduced for *cbaA*. Alignments of the N-terminal have been performed on other terminal oxygenase components of *cis*-diol forming dioxygenases (Neidle et al. 1991). These alignments show limited homology in this region, suggesting that the strong alignment of the N-terminus of *cbaA* and component A of the 4-sulphobenzoate-3,4-dioxygenase will likely be extended into the remainder of the protein once the sequence of the latter becomes available.

The terminal oxygenase component is involved in the binding and activation of oxygen, which usually involves a metal cofactor such as iron. Neidle et al. (1991) proposed that the speculated Rieske-type iron sulphur centre in the terminal oxygenase component of benzoate and toluate dioxygenases is coordinated by cysteine and histidine residues. The same cysteine and histidine pattern is also found in the *cbaA* sequence (Figure 5.4). Although the entire protein has not been sequenced, the UV-visible absorption spectrum of component A of the 4-sulphobenzoate-3,4-dioxygenase has been examined and it reveals the presence of a Rieske-type iron-sulphur centre (Locher et al. 1991). I speculated earlier that there may be an evolutionary relationship between this dioxygenase and the

BR60 3-chlorobenzoate dioxygenase because both host organisms fall into the Comamonadaceae, and because the relative regiospecificity of the reactions was similar (Chapter 3 and 4). If this were true then further biochemical studies should show that the conserved cysteine and histidine residues deduced in the *cbaA* product are involved in the Rieske iron-sulphur centre.

The estimated size of CbaA deduced from the amino acid sequence is 49 kd, in the same size range as other dioxygenase components that have been studied (Fetzner et al. 1992). The expression of the DNA *NotI-HindIII* fragment in the clone pBRCN13a, that contains *cbaA*, revealed a polypeptide approximately 51 kd in size on a 7.5% SDS-PAGE gel (Figure 4.4). This is within the estimated size of CbaA and is most likely the expressed product of *cbaA*. The elimination of the *ScaI* fragment of pBRCN5 produced the construct pBRCN7 (Chapter 4, Figure 4.3). This construct eliminated most of *cbaA* (Figure 5.1) and was not able to restore 3-chlorobenzoate catabolic activities to *Alcaligenes* sp. strain BR6024 thus demonstrating that this protein is likely expressed and is essential in chlorobenzoate dioxygenase activity.

The second ORF designated *cbaB* also revealed homology to the vanillate demethylase genes (Brunel and Davison 1988). In this case the homology exists with *vanB*, the ferredoxin-like protein component. The strongest alignments

occur at the N-terminal and C-terminal regions of the genes. Due to the difference in the coding lengths, 288 residues in CbaB and 314 residues in VanB, there are a number of gaps in the sequence alignments in the centre (Figure 5.5). This would suggest that the central region is either not included among the functional domains of the protein, or that the divergence in the nature of the substrate and the type of reaction (dioxygenase vs monooxygenase) is reflected in this region. The ferredoxin-like region is involved in transferring electrons to the terminal oxygenase. This is similar to the benzoate and toluate dioxygenases (Neidle et al. 1991) because the reductase components of these enzymes apparently also use a chloroplast-type ferredoxin. In three-component dioxygenases that have a separate ferredoxin component, for example biphenyl dioxygenase, the [2Fe-2S] centres resemble the Rieske-type centres (Erickson and Mondello 1992), similar to those found in the terminal oxygenase components described above. Plant-like ferredoxins are characterized by four cysteines that coordinate the [2Fe-2S] centre (Neidle et al. 1991). These [2Fe-2S] centres are found in the N-terminal end of ferredoxins from plants, BenC, XylZ and XylA, whereas in VanB and CbaB they are found in the C-terminal region. This may or may not be of stereospecific or evolutionary significance.

The targets of vanillate demethylase (Dagley 1986) and

chlorobenzoate dioxygenase (Chapter 3) are the *meta*-, or *meta*-and *para*-positions, respectively, whereas the targets of benzoate and toluate dioxygenase (Neidle et al. 1991) and xylene monooxygenase (Suzuki et al. 1991) are the first carbon and/or *ortho* positions. There might be an evolutionary relationship between the vanillate demethylase genes and the chlorobenzoate dioxygenase genes based on the evidence of conservation of sequence between CbaA and VanA, and CbaB and VanB, found in the database searches. This finding opens up a very interesting line of experimental work concerning the mechanistic and evolutionary relationships between dioxygenase and monooxygenase reactions, but this work is beyond the scope of this thesis.

In addition to the ferredoxin sequences, Neidle et al. (1991) proposed conservation of amino acid residues involved in NAD-ribose binding and FAD-isoalloxazine ring binding regions encoded on the reductase component of benzoate dioxygenase, toluate 1,2-dioxygenase, xylene monooxygenase and vanillate demethylase. These cofactors are required in the electron transport chain of these oxidoreductases. The proposed conserved residues indicative of these binding sites are limited but they are also observed in the sequence of the translated *cbaB* gene (Figure 5.6b and c). These regions are observed towards the N-terminus of CbaB. Additional sequences that may be involved in the binding of the FAD-PP_i and NADP-PP_i moieties have been proposed, but

the conserved residues in these regions were not observed in the CbaB sequence.

The size of CbaB based on the deduced amino acid sequence is approximately 32 kd. The expression of the construct pBRCN13a in *E. coli* did not reveal a unique polypeptide of this size when separated on an SDS-PAGE gel (Figure 4.4). This polypeptide may have been translated but the unique band may have been overlooked due to either poor expression or masking by other proteins of similar size. The reductases in general are not stable when exposed to oxygen (Locher et al. 1991) and the CbaB protein may have become denatured and degraded following cell lysis. The deletion of the *EcoRV* site of pBRCN5 mutated the N-terminal region of *cbaB*. The initial results, lack of growth of the *Alcaligenes* sp. strain BR6024 transconjugants on 3-chlorobenzoate, suggested that the gene was necessary for catabolism. The growth of colonies on selection plates after prolonged incubation could have occurred due to a number of reasons. I speculate that the gene is required for 3-chlorobenzoate catabolism due to the prolonged incubation required for transconjugant formation; this is not observed in clones lacking this mutation. Transconjugants generally appear on 3-chlorobenzoate selection plates in one to three days. Transconjugants selected solely for the plasmid using the antibiotic marker and an alternate carbon source were also unable to grow on

3-chlorobenzoate. The transconjugants that did appear after prolonged incubation may have repaired the deletion around the *EcoRV* site. Alternatively, the mutation may not have disrupted the translation of the entire protein, despite the missing residues. I have not been able to express this protein in functional clones, therefore I presently can not confirm the presence or absence of the translated product of this disrupted gene.

The translation of the third ORF, *cbaC*, showed similarity to the glucose-fructose oxidoreductase from *Zymomonas mobilis* (Kanagasundaram and Scopes 1992). This enzyme carries out an oxidation - reduction reaction without an external cofactor. Instead it contains a non-dissociable NADP. We have no biochemical information available on *CbaC* regarding the cofactors involved in reduction, to confirm its function. There was similarity of *CbaC* to other reductases present in the databanks as well, both prokaryotic and eukaryotic. The alignment of these sequences all occurred in the same region, suggesting that the conserved residues may be involved in the binding of NAD(P) or another essential function of the oxidoreductases. The alignments of these sequences are illustrated in Figure 5.7b.

The similarity to other oxidoreductases in the databank search leads me to believe that the *cbaC* product is a reductase component. The deduced amino acid sequence of

CbaC codes for a polypeptide of approximately 44 kd. The expression of the construct pBRCN13a produces a unique polypeptide that is approximately 42 kd, close to the predicted size of the product of *cbaC*.

The dioxygenases studied thus far have either one or two electron transport components. If two components are present, one is a low molecular weight (ca 10 to 20 kd) ferredoxin and the other is a flavoprotein. If one component is used it generally contains both a ferredoxin domain and a flavin domain, resulting in a larger protein (ca 30 to 45 kd). After the *cis*-diol is formed a dehydrogenase is used in some microbes to re-aromatize the ring forming a dihydroxylated aromatic structure. Examination of the results of the expression of the nested deletions pBRCN11 and pBRCN12 suggested that *cbaC* may be involved in dehydrogenase activity. The extent of the deletion forming the construct pBRCN12 could not be confirmed without nucleotide sequencing. Restriction digestion of the clone revealed that the 6.6 kb (Tn5271 coordinate) *EcoRI* site was still present but the 7.2 kb *BstEII* site was eliminated. This deletion possibly includes the 3'-end of *cbaB*. If this is the case then it would account for the lack of dioxygenase activity in the *Alcaligenes* sp. strain BR6024 pBRCN12 transconjugant (Chapter 4). This would also be additional evidence that *cbaB* is required for dioxygenase activity.

If the dioxygenase activity is coded in *cbaAB*, and if *cbaC* encodes a dehydrogenase, then there should have been an accumulation of chloro-dihydroxy-cyclohexadiene or its dehydrated form, chloro-hydroxybenzoate, in the medium of clones carrying mutations in *cbaC*. This product did accumulate in the medium when the construct pBRCN11 was present (Chapter 4). Restriction enzyme mapping and sizing indicated this construct has a deletion at the 3'-end of *cbaC*, between the 7.8 kb *PvuII* and 8.1 kb *EcoRI* sites. This suggests that *cbaC* is involved in dehydrogenase activity, which would account for the sequence similarities to other oxidoreductases.

If the deletions in pBRCN11 and pBRCN12 are both within *cbaC*, then an interesting scenario arises from these data. It would suggest that *cbaC* is involved in both the reductase activity of the dioxygenase and the dehydrogenase activity required to rearomatize the ring. In essence the two electrons donated to the substrates (3-chlorobenzoate and oxygen) would be recovered from the dihydrodiol intermediate in the next enzyme reaction, without diffusing away from the reaction centre. This would be a highly efficient combination from the biochemical perspective. Further sequencing and expression studies on mutants will be required to test this hypothesis.

The formation of *cis*-diols in the degradation of aromatic compounds is catalyzed by multicomponent

dioxygenases. The terminal oxygenase components studied to date are composed of identical subunits or two non-identical α - and β - subunits. The terminal oxygenase component of the chlorobenzoate dioxygenase described in this investigation appears to be composed of a single subunit encoded by *cbaA*. More extensive biochemical studies are required in order to determine if this is indeed the organization of subunits and the number involved. Electron transport to the terminal oxygenase can be either a one or two component system. In the chlorobenzoate dioxygenase there appears to be one electron transport component, with conserved amino acid residues for NAD and FAD binding, and a ferredoxin-like arrangement of amino acid residues, as found in other aromatic ring dioxygenase systems. The CbaC protein may also be involved in dioxygenase-reductase activity, but the *in vivo* expression of pBRCN11 and 12 suggests that this protein is a cis-diol dehydrogenase.

CHAPTER 6: A Summary

The major goals of this project were to provide an understanding of the structure and function of a catabolic transposon involved in the growth of a microbe in an environment contaminated with xenobiotic chemicals. The organism *Alcaligenes* sp. strain BR60 was chosen as the microbe because it was isolated from Bloody Run Creek, a habitat that has been exposed to a mixture of halogenated and non-halogenated chemicals for about two decades. 3-Chlorobenzoate was chosen as the chemical of interest for several reasons. It is an abundant contaminant at the site, it has a relatively simple structure with a single chlorinated benzoate ring, and it is known to be produced during the metabolism of more complex environmental contaminants such as polychlorinated biphenyls. The choice of this aromatic substrate also allowed me to investigate the relationship of the evolved degradation pathway used by this organism to catabolic pathways that have been described for other microbes.

This Chapter is a brief summary of the results of this investigation in answer to the stated objectives of Chapter one. Included are some potential avenues of future research that have arisen due to the findings of this project.

Objective 1: To determine the physical structure of the plasmid, pBRC60 and the transposon, Tn5271.

In Chapter 2, background into the structure of elements carrying the genes involved in 3-chlorobenzoate degradation was given. The physical structure of the plasmid, pBRC60, was given in the form of a restriction enzyme map. The chlorobenzoate catabolic transposon Tn5271 in *Alcaligenes* sp. BR60 was characterized by establishing the nucleotide sequence of the flanking direct repeats in this element, and observing its relationship to both class I and class II transposable elements.

Objective 2: To determine the biochemical pathway utilised by *Alcaligenes* sp. BR60 to degrade 3-chlorobenzoate.

Chapter 3 demonstrates that the transposon Tn5271 encodes a 3-chlorobenzoate-3,4-dioxygenase activity that, when expressed in the wildtype strain, *Alcaligenes* sp. BR60, directs molecular oxygen into the chlorinated position of 3-chlorobenzoate with substrate specificity that includes 3,4-dichlorobenzoate. In Chapter 4 it was shown that this enzyme is not strictly regiospecific, allowing it to also direct oxygen into the non-chlorinated 4,5 positions of the benzoate ring. This is the first description of this route for 3-chlorobenzoate degradation described for a bacterium. Numerous projects could be developed in order to understand more fully the biochemical aspects of this pathway.

The orientation of dioxygenase attack on the ring directed metabolism through protocatechuate or chloro-protocatechuate. It appeared that the nature of this pathway influenced the host-range of Tn5271, and in Chapter 3 data were assembled in order to show that natural selection for Tn5271 expression in contaminated environments has resulted in the distribution of Tn5271 in hosts from rRNA superfamilies II and III (the Γ and β Proteobacteria, respectively). This natural distribution of Tn5271 correlates with the phylogenetic distribution of the protocatechuate meta-ring-fission pathway.

Objective 3: To determine the genes present on Tn5271 which are involved in 3-chlorobenzoate degradation.

Chapter 4 demonstrated that Tn5271 carries genes for 3-chlorobenzoate dioxygenase within a 3.7 kb *NotI*-*EcoRI* fragment. This dioxygenase activity was expressed in *E. coli* by cloning this fragment on an expression vector in order to place the genes under the control of the *tac* promoter. When expressed under the control of this promoter in *E. coli*, the dioxygenase activity appeared to be directed primarily to the 4,5 carbon positions and to a minor degree to the 3,4 positions, whereas in *Alcaligenes* sp. strain BR6024, dioxygenase activity was directed to both the 3,4 and 4,5 carbon positions.

The expression of this fragment produced two proteins

approximately 51 and 42 kd in size when resolved on a SDS-polyacrylamide gel. Nucleotide sequencing studies in chapter 5 revealed that three open reading frames (ORF's) are encoded in the 3.2 kb NotI-EcoRI fragment of Tn5271. These were designated *cbaA*, *cbaB* and *cbaC*. Although two polypeptides were expressed from this NotI-EcoRI fragment in *E. coli*, there were three genes that appeared to be essential for the conversion of 3-chlorobenzoate into protocatechuate or chloro-dihydroxybenzoate. This left interesting questions that have not been resolved. Is there a component missing in our protein analysis or is *cbaB* not required for dioxygenase activity in *E. coli*? Is one gene, *cbaC*, capable of producing an enzyme that can act both as a reductase component in the dioxygenase and also as a dehydrogenase?

Objective 4: To determine the relationships of these genes to genes for other aromatic catabolic pathways that have been characterized.

The *cbaA* and *cbaB* genes code for amino acid sequences that are consistent with conserved amino acid sequences found in other hydroxylase and reductase components of dioxygenases that act on aromatic compounds to form *cis*-diols. These conserved regions are believed to be a Rieske-type [2Fe-2S] centre in the terminal oxygenase component and chloroplast-like ferredoxin, FAD and NAD binding regions

involved in electron transport through the reductase component. The third ORF, *cbaC*, also appears to code for a reductase-like component due to similarities in the translated amino acid sequence to other reductases. This component may be involved in the dehydrogenation of the *cis*-diol structure.

Comparisons of the deduced amino acid sequences of *cbaA* and *cbaB* to sequences in the NCBI databank showed that the closest similarities are with the vanillate demethylase and 4-sulphobenzoate dioxygenase genes. This evolutionary relationship may be based on the fact that all of these enzymes function on the distal region of the benzene ring. The availability of more sequence and biochemical information on these enzymes will enable investigators to examine these evolutionary relationships.

FUTURE RESEARCH PROJECTS

The observation that this catabolic pathway, involving a chlorobenzoate dioxygenase, has a host range for functional expression in the environment that includes primarily the β -Proteobacteria is of great importance in the bioremediation of chlorinated aromatics. This will be especially important when attempts are made in the future to expand the catabolic substrate range of these microbes. Precise knowledge of existing aromatic catabolic pathways in microbes is required in order to introduce genes that will

complement them. This may be a factor that is presently limiting researchers in their attempts to genetically engineer microbes capable of degrading more complex chemicals such as polychlorinated biphenyls (PCBs). The isolation of the genes responsible for chlorobenzoate-3,4-dioxygenase activity, described in this Thesis, will enable researchers to test the use of this route in order to expand the substrate range of bacteria that are members of the Γ and β proteobacteria. This was not previously possible using the *clcABD* operon, which is apparently only functional in the fluorescent pseudomonads and *Alcaligenes eutrophus*.

There are many questions that arise from the expression of the *cbaABC* pathway in the β proteobacteria that are all potential areas for future research. Is the failure of expression of these genes in the fluorescent pseudomonads due to the lack of suitable regulatory factors? Is it possible that these genes are expressed in alternate groups such as the fluorescent pseudomonads, but that dead end metabolite such as chloro-dihydroxybenzoate (as observed in chapter 4) is formed, preventing the final mineralization of the chemical? Are the enzymes functional but there are existing enzymes that are producing lethal metabolites killing the host? Can regulatory genes and/or additional genes encoding subsequent parts of the *cbaABC* pathway be introduced in order to produce an alternative 3-chlorobenzoate catabolic pathway in a wider range of

Proteobacteria, including *E. coli*?

The availability of the restriction map of pBRC60 has already proved to be of benefit to this laboratory. It has allowed the investigation of the instability and rearrangement of Tn5271 within the host *Alcaligenes* sp. strain BR60 and related strains of the family Comamonadaceae. The transposition properties of Tn5271, and the mobility, host range and incompatibility functions of pBRC60 are presently being investigated by other researchers in our group.

There are also some intriguing questions arising from the research of others in the laboratory which relate to the pathway discovered in this research. Fulthorpe (1991) found positive hybridizations of fragments within Tn5271 to fragments mapped outside of the transposon on the plasmid (in particular to the Bgl5 fragment at about 4 o'clock on the pBRC60 map, Figure 2.1). This could have occurred because there are partial copies of sequences within Tn5271 in other parts of the plasmid. Although Tn5271 has been described as a novel transposon with a class I composite form but a class II type transposase, there was no known resolvase found coded within the non-repeated region of Tn5271 by sequence analysis. This gene is necessary in resolving the cointegrates formed following transposition of class II elements. This gene is expected to be found within the transposing element allowing it to maintain its

transposition abilities once it enters a new replicon. Perhaps the resolvase gene(s) is outside the Tn5271 element as depicted in Chapter 1, if there is a greater complexity to the actual structure of this transposon than described here. Tn5271 may be a part of another transposon, as found in the nested toluene catabolic transposons Tn4651 and Tn4653 (Tsuda et al. 1989).

There are number of questions that can be asked from a community ecology perspective. Do the Γ and β proteobacteria dominate in any particular environments? Are equivalent numbers of bacteria found in the natural habitat using both known 3-chlorobenzoate pathways? Under what environmental and laboratory conditions do these pathways dominate? Presently there are research projects in our laboratory attempting to answer some of these questions.

The simplicity of bacterial structure limits the use of traditional approaches used in the study of ecological distributions and abundances of more structurally complex organisms such as eukaryotes. This simplicity presents a challenge, but also allows us, through an understanding of molecular detail, to study the ecology of an organism in a much more exciting venue than can be matched in more complex organisms. This study provides basic information regarding one member of the vast microbial community present in freshwater. It provides insight into how this organism copes with just one chemical amongst thousands of man-made

chemicals that are distributed widely in the environment. Finally, it provides the detailed molecular information required to apply DNA-based ecological methods to these prokaryotes in their natural habitats. I fully believe it will form part of the base of the pyramid of knowledge that is essential if we are to understand how microorganisms grow and evolve in natural ecosystems exposed to chemical contaminants.

APPENDICES

APPENDIX 1: Nucleotide Sequence of Tn5271

Nucleotide sequence starts at the first nucleotide of the 5' end of IS1071 and continues to the 3' end of the non-repeated region. The nucleotide sequence of the second copy of IS1071 is not included since the sequences are identical.

Tn5271

10	20	30	40	50	60
GGGGTCTCCT	CGTTTTCACT	GCAATAAGTG	ACGGTACGCA	AAGCTAGCAC	TGGCGCGGGG
70	80	90	100	110	120
GTGGTCTGGG	TAGACCGTTG	ATTTCATTGA	CTTTCCTGTT	CGCTTTGTAA	ACGGGTATGG
130	140	150	160	170	180
TGGCCTCCCA	CTTTTGAGGT	TCACGATGCA	GGGTGGGCAC	ACAACGTTTT	TGGGGATGCG
190	200	210	220	230	240
TGGGCTCCCC	CGCGATATCA	GCGACTTCGA	GATGAAGGCA	TTTTTCACCT	TCGATGGTGC
250	260	270	280	290	300
CGAGCGCGAC	GCAATCAATG	CACGCCGAGG	TGATTCCCAC	AAGCTTGGTC	TGGCGCTCCA
310	320	330	340	350	360
TATTGGTTTC	CTGCGCATGA	GTGGGCGTTT	GCTCGGTGCC	TTTCGGGTAA	TTCCAGTAGC
370	380	390	400	410	420
CTTGTGGCGC	CACCTTGGCA	ACGAGCTTGG	CATTGCAGCA	CCAGAAGTCG	CCTCGCTGAG
430	440	450	460	470	480
AGCCATGTAT	GAACGCGGGC	GCACGCTATT	CGATCACCAA	CAAGTAGCCT	GCACGGTCCT
490	500	510	520	530	540
TGGATTCCAG	TGGATGAGCG	AGCACCAGCG	CCGCTCACTG	GTACGTGAAC	TGCGCGACGA
550	560	570	580	590	600
AGTGGCCGGC	TGCGACCGCG	ATCAGCTACT	CGTGCGGGCG	CGTCAATGGC	TGTACAAGAA
610	620	630	640	650	660
CAAGCTGGTG	ATCGTGCACG	AGCGGGCAAT	TCGGACACTG	ATTGCGGGCG	CACTTGCCCA

670	680	690	700	710	720
GCTTGAAGTT	GAAACAGGCA	CGCCCATCGC	CGCCAGCGTT	GATCCAGCAA	CACCTTGATCG
730	740	750	760	770	780
CTGGCGAGCC	TCAGTTTCAG	AGCTGCGCCC	AGATGGACAA	ACCCAGCAGA	GTTGGCTATG
790	800	810	820	830	840
GGCTGCACCG	GCGAAACACT	CAACCCGCCA	AATCAGCGAG	GTA CTGGAGC	GCATCGACCT
850	860	870	880	890	900
GCTTTACACG	CTGGACGTTT	ATAAGCACCT	GGCAGACATC	CCCGATCTCA	TCTTGCGCCG
910	920	930	940	950	960
CTACGCGCGC	CGACTTGTCT	CCAGGCCGCC	CTCAGCCGGA	GCCAAGATCA	AAGAGCCAGC
970	980	990	1,000	1,010	1,020
GCGCACCGTG	GAGGTCGCAT	GCTTTCTTCG	GTATTGCCTG	TTCACCACCA	CAGACCAGTT
1,030	1,040	1,050	1,060	1,070	1,080
GATCCTTATG	GTGCAGCGCC	GGATCGCCGA	TCTGTGGCGT	CAGGCTGCCG	CCGATGTCCC
1,090	1,100	1,110	1,120	1,130	1,140
CGCTACCGTC	AATTGGGCGG	CAATGTACAA	AACGCTGCTC	GGCGAACTTG	TTGCCTTGAG
1,150	1,160	1,170	1,180	1,190	1,200
CGCGCAAGGT	GCGGTGCCAG	ATGCTGAGTT	GCGTGCCCGT	CTTGAAGCCT	TGATCACCAG
1,210	1,220	1,230	1,240	1,250	1,260
AACCCAGAAA	CGCAAACCAC	CGAGCAGGGC	CTCCCTGGTC	CGCGAGGGAT	TGATTGATGG
1,270	1,280	1,290	1,300	1,310	1,320
AATTCGCCCC	GTGCGGTGCT	TGCTCGTCGC	CATTGCAAAG	CTGCCCTGGC	AGGCCACCGG
1,330	1,340	1,350	1,360	1,370	1,380
CGAGCATCCT	GCCATCGAGT	ACCTTGCCAA	GCTGCAAGCT	TTATATCTCA	AAGGATCCAG
1,390	1,400	1,410	1,420	1,430	1,440
AAAGCTGCCA	GTGGAAGTGG	TGGCACCAAG	TCTGGGAATG	ATCTGGCAGG	TTTCGATCTC
1,450	1,460	1,470	1,480	1,490	1,500
CAGCCCAGAC	CGGGAACGGG	CGTTTCAGGC	GTTGGAGGTG	GCCACCCTGT	TTGCCCTGCG
1,510	1,520	1,530	1,540	1,550	1,560
CCGCGCGGTG	CGCAATGGCT	CGGTCTGGAT	TGAGCACAGC	CTGAGCTTTC	GGGGTCGTGC

1,570	1,580	1,590	1,600	1,610	1,620
GCGCTTGTTT	TTCACGGACG	AGCGTTGGCA	GGCAGAGTCC	AAGAAACACT	ATGCCCCTCT
1,630	1,640	1,650	1,660	1,670	1,680
ATCGTTACCC	AGCAAGGCTG	CCACTTTCTT	GAAGCCTTTG	CTGGCCAGAG	TAACTGCCGG
1,690	1,700	1,710	1,720	1,730	1,740
TGTCGATGCG	GTGGCCGCTG	CAGCCCGCAG	TGGCGTACTG	CGCGTGGATG	ATGAACTCCA
1,750	1,760	1,770	1,780	1,790	1,800
TTTGTGCGCA	TTGCCCCGAG	AGGACGAAGA	CCCAGAAAGT	ACCAAGCTGC	GCGCGGCTTT
1,810	1,820	1,830	1,840	1,850	1,860
GGATCACCGC	ATCGGTGAGG	TTCAATTGCC	GGAAGTGATT	CTGGCCGTTG	ACGCCCAGGT
1,870	1,880	1,890	1,900	1,910	1,920
GCGCTTTAGC	TGGATCATGC	TCGGACGTGA	GCCGCGCTCT	ACCGACGAGC	TGCTGATGGT
1,930	1,940	1,950	1,960	1,970	1,980
CTATGCCGGC	ATCATGGCCC	ACGGCACCAG	TCTGACTGCG	GTCGAATGCG	CGCGCATGAT
1,990	2,000	2,010	2,020	2,030	2,040
TCCGCAATTG	TCTGCCACCA	GCATTCGCCA	GGCCATGCGC	TGGGCGCGGG	ACGAACGGCG
2,050	2,060	2,070	2,080	2,090	2,100
TCTGAGCCAG	GCCTGCCAGG	CTGTGCTGGA	ATTGATGAGC	CGACACCGCA	TTGCCGCCAC
2,110	2,120	2,130	2,140	2,150	2,160
CTGGGGGCGG	TCCGATTTGG	CATCTTCTGA	CATGATGAGC	ATGGAGACCA	CCAAACGGGT
2,170	2,180	2,190	2,200	2,210	2,220
GTGGCAAGCC	CGGCTTGATC	CTCGGCGCAA	CACACCTTCC	ATTGGAATCT	ACTCCCATGT
2,230	2,240	2,250	2,260	2,270	2,280
AAAAGACCGG	TGGGGCATCT	TCCATGCGCA	GCCCTTTGTG	CTCAATGAGC	GCCAGGCGGG
2,290	2,300	2,310	2,320	2,330	2,340
CGTGGCCATT	GAAGGTGTCA	TCCGCCAAGA	AAAGCTGGAG	ACCAGCCAGC	TTGCTGTGGA
2,350	2,360	2,370	2,380	2,390	2,400
TACCCATGGC	TACACCGACT	TTGCCATGTC	ACATGCCCCG	TTGCTTGTTT	TTGATCTTTG
2,410	2,420	2,430	2,440	2,450	2,460
CCCGCGGTTG	AAGGAACTCA	AACAGCGCCA	CCTCTTTGTG	CCACGCGGCA	CCAAAGTGCC

2,470	2,480	2,490	2,500	2,510	2,520
CGCAGAAATC	GCTGCGGTGT	GCGAAGCCAA	TGTCGACGTC	GCTTTGATCG	AAAAGCATTG
2,530	2,540	2,550	2,560	2,570	2,580
GGATAGTCTG	GTGCACCTGG	CAGCCTCGGT	CATGAGCGGA	CATGCCAGTG	CGGTGGCAGC
2,590	2,600	2,610	2,620	2,630	2,640
TCTTGCGCGG	TTCGGTTCTG	CCGCCCAGGG	CGATCCAATC	TATGAGGCTG	GCGTGCAATT
2,650	2,660	2,670	2,680	2,690	2,700
GGGGCGGTTG	CTGCGTACGG	CGTTTTTGGC	TGACTACTTT	GTCAAGGACG	CTTTCAGGAA
2,710	2,720	2,730	2,740	2,750	2,760
CGAGTTGCGC	CGGGTGCTCA	ATCGGGGCGA	GGCTGTTAAC	GCCCTCAAGC	GCCCATTTA
2,770	2,780	2,790	2,800	2,810	2,820
TACCGGCCGG	ATCAGCCCGG	CGCAGGCCAA	ACGTGTCGAT	GAAATGCAGG	CTGTGGCCGA
2,830	2,840	2,850	2,860	2,870	2,880
TGCGTTGAGC	CTGATGGCCA	ACATCGTGAT	GGCGTGGAAT	ACCTCACAGA	TGCAGGCCGT
2,890	2,900	2,910	2,920	2,930	2,940
CCTGGATCGC	TGGTGAACC	GCCGCCAGGT	CATTCCACCG	GAACTGATCG	GGAAGATTGC
2,950	2,960	2,970	2,980	2,990	3,000
GCCCACCAGG	CTGGAGAGCA	TCAACTTGCG	GGGTGTGTTT	CGCTTCCCGG	TTGACCGCTA
3,010	3,020	3,030	3,040	3,050	3,060
TGCTGACCAA	ATCCTGCCTT	CGCGGCCAAA	TGCATCGATA	ACTGGCACCA	ATGGATGAAA
3,070	3,080	3,090	3,100	3,110	3,120
CCGACCACGG	TTTGACGCCA	CGAATCGCAG	ATTGAAAGT	GAACAGGAAA	GTCAATGAAA
3,130	3,140	3,150	3,160	3,170	3,180
TCAACGATCT	ACCAACACCA	CCTCCGCGCC	AGTGCTAGCT	TTTCGTACCG	TCACTTATTG
3,190	3,200	3,210	3,220	3,230	3,240
CACTGAAAAC	GAGGAGACCC	CATGCCCTGC	TCGACGCCTT	GCTGGGCTGC	GTGCTGCTGC
3,250	3,260	3,270	3,280	3,290	3,300
GTGAGGCCGG	TGTACCTCGA	TGTCCCGAG	CCCAACCGCG	AGGCCATCGC	GATGGCCGAG
3,310	3,320	3,330	3,340	3,350	3,360
CGGCCAGGGC	ATGGTGACGG	GCTTTCCGAC	CGCGCGCATG	TACCGCGGGC	CGGCGCCGCG

3,370	3,380	3,390	3,400	3,410	3,420
GCTGGACCTG	GGGCGTTGTT	ATGCCTTGAC	GGCCTTGGAG	GTCGGCTGAG	CCTGGTGGCT
3,430	3,440	3,450	3,460	3,470	3,480
TGATCAGGGC	AGGCGTTGGC	GCTACTCCTG	AGGAGATTGG	CGATGCCAGT	GCGCCTTGAC
3,490	3,500	3,510	3,520	3,530	3,540
GTAGAGCGAA	GAATCGATAT	CCACCCTGTT	GATGTTCTCC	TTCAGGCGCT	GGAGAAGGCG
3,550	3,560	3,570	3,580	3,590	3,600
AACCAGGGTC	TCTCGCTCCG	AGGCGGTCAG	CACAGACAGG	AGCGAATTTT	CGCGATGGAG
3,610	3,620	3,630	3,640	3,650	3,660
CGCGTACAGA	ATGATCTTGT	CGTGCATTGC	CTGGCCTTCT	TCGGTCAGCT	GCGTGAGCCT
3,670	3,680	3,690	3,700	3,710	3,720
GACCGGCTTG	TCGCGGTCCG	CCGTTGCCAG	AACGATCAGG	CCCTTGGCCA	ACATTGCTT
3,730	3,740	3,750	3,760	3,770	3,780
CATCACCTTG	CTGACGGAGC	CCTTGTCCAT	GCCAGTGAAC	TGGCAAATGC	TCTGTGCCGT
3,790	3,800	3,810	3,820	3,830	3,840
GGTTTCGCCT	GCCACTGCTA	TGAGCGCCAG	GACCGGCCAG	GTCTCGATGC	CGATGCCGAA
3,850	3,860	3,870	3,880	3,890	3,900
ATTCTTTAAA	TAGTCTTTTG	AGGCACCCTG	CGCGAGCTTG	TTGGCGATGG	CGGTAAAAAA
3,910	3,920	3,930	3,940	3,950	3,960
TGCCGGGGTG	TAGGTATCGA	GATTGAGGGT	CACCGGCTT	TTTCGAGGAT	CTCTTGCAAG
3,970	3,980	3,990	4,000	4,010	4,020
CATTCAGTAT	CCAGCGTTGA	TGCGTATAGG	CCCCAGACTC	GGTGGGGGCG	GTAAAAGCCA
4,030	4,040	4,050	4,060	4,070	4,080
TAAACCCTTT	CCGGATACCG	CTGCAACCAG	CTATCGACCC	CGTTGACATG	GCGCTTGCGC
4,090	4,100	4,110	4,120	4,130	4,140
CAGGCAGCCG	TGTTGAAAGA	GATCTGCAGG	GATCACAACA	ATGCTGCGCG	GTCCTCTGTC
4,150	4,160	4,170	4,180	4,190	4,200
TCCTCCGGGC	CATTGCAGCA	GGCTCATGAT	TCGATGCCGA	GCTTAGCACC	TTGGTTCCTGA
4,210	4,220	4,230	4,240	4,250	4,260
GCAAATGGCC	AGGGTCCTTT	GGTTCGTATT	GGCCACCAAC	AGCATTGCGC	GTGACGCCCC

4,270	4,280	4,290	4,300	4,310	4,320
ATTGAGCGCC	GCGCTCCTCC	GGAGACCGGC	AATTCCTCG	GCCGGTGCAC	CGGCTGGCGC
4,330	4,340	4,350	4,360	4,370	4,380
TGCCGCATTC	GCGCTTCTC	CTCCCGCCCG	GCCTTGCCGC	TGCCCCGACG	CGATCTGGCG
4,390	4,400	4,410	4,420	4,430	4,440
GCGTCGGCAT	GAGCCGCACT	GAGCGGCAGC	GCGCGCCGCG	CCACCGCCGC	TTCACGCGCC
4,450	4,460	4,470	4,480	4,490	4,500
CGCACAGCGG	CCGCTGCCCCG	CTGCAGCACC	TGCCGTGCTC	GGCTCGGGTA	AACACCTAGA
4,510	4,520	4,530	4,540	4,550	4,560
TTTTCGATCT	GAAAACCTCC	GATAGTTACA	AAACCAACTA	CATGGATCGA	ATCATCTCAG
4,570	4,580	4,590	4,600	4,610	4,620
GTTTTCCCT	CATAAGGGTT	TGACACAAGT	TGCTAAACCA	ACAACATCCG	CAAAATTTCA
4,630	4,640	4,650	4,660	4,670	4,680
GGAGCAAGCT	ATGGCATTGA	CCAAGGACAA	CGAAGATTG	GTGCGAGTGG	GCCGTGGAAC
4,690	4,700	4,710	4,720	4,730	4,740
CCCATGGGC	GGCCTCATGC	GCGAGTACTG	GATCCCCGCA	TTGAAATCGA	CCGAGCTCGA
4,750	4,760	4,770	4,780	4,790	4,800
GGCTGGCGGA	AGTCCTGTGA	GGCTGCTGCT	CTTGGGCGAA	AAGCTCGTCG	CCTTCAGGGA
4,810	4,820	4,830	4,840	4,850	4,860
GCCGTGCGGT	GCGGTGCGCG	TGATGGACAG	CCGATGCCCC	CACCGTGGCG	TTCCCTGTT
4,870	4,880	4,890	4,900	4,910	4,920
CATGGGGCGC	GTTGAAGAGG	GCGGGCTGCG	CTGCGTCTAC	CACGGTTGGA	AATTCAGCGC
4,930	4,940	4,950	4,960	4,970	4,980
TGAGGGCAAA	TGCGTTGACA	TGCCCAGTGT	CCGTCCTGAA	GACGAATTCA	AAAACAGCGT
4,990	5,000	5,010	5,020	5,030	5,040
GCGCGTGGCG	CGATACCCCG	TGAAAGAAAT	GGCGGGCGTT	GTCTGGGTGT	ACATGGGTAC
5,050	5,060	5,070	5,080	5,090	5,100
TCGAAAAGTT	TTGCCCCGAGC	TCCCGCGGCT	GGAAGTGCTG	AATCTTCCGG	AAAACGAAGT
5,110	5,120	5,130	5,140	5,150	5,160
GGATGTCATT	TGCCTCCAGC	GAAAGTCGAA	CTGGTTGCAA	AACCTGGAAG	GTGAAATTGA

5,170	5,180	5,190	5,200	5,210	5,220
TACCTCGCAT	TTCAACTTCC	TTCACGTGGG	COGACTCCAC	GCGGACGAAG	TGCCGGACGA
5,230	5,240	5,250	5,260	5,270	5,280
TCACCCATTG	AAATACACCG	CCCAAGTCGC	TCCGCAGTAC	CTGGTCAAGG	AAACGGCATT
5,290	5,300	5,310	5,320	5,330	5,340
GGGCACCTGC	TACGCGGCTC	AAGTGCCGGC	GGAAGAGGAT	CACACCTACA	CGCGGTTTGC
5,350	5,360	5,370	5,380	5,390	5,400
GCATTTCTT	TTCCCGTTCT	GGGCGCTCAT	TCCGCAAGCC	GATATTGCGC	AGAACATTTT
5,410	5,420	5,430	5,440	5,450	5,460
GGCCAGAGCC	TGGGTGCCCA	TGGACGATGA	GCACACCATG	ATGTTCTTCT	TCCGTTGGAC
5,470	5,480	5,490	5,500	5,510	5,520
GGGATCCAAG	GCCAAACGCC	TCGATACGCC	ATTGAAGAGC	GGTTCGCCCA	TGCCCCGGGT
5,530	5,540	5,550	5,560	5,570	5,580
CACGCTGACG	GACATGAAGT	ACAAGGAAAA	CACGACGGAT	TGGTACGGCC	GCTGGCAGCC
5,590	5,600	5,610	5,620	5,630	5,640
GCTCGGCGAT	GAGAGCAACG	ATTGGCTCAT	CGATCGAGAT	TGCAAAAAG	TAGGCCGGGT
5,650	5,660	5,670	5,680	5,690	5,700
GTTCTCCGGT	ATTTACGGAA	TCCACGCACA	GGATCAGGCC	ATGACCGACA	GCATGGGGCC
5,710	5,720	5,730	5,740	5,750	5,760
GATCATTGAC	CACGGCCTGG	AGCAATTGGC	ACCGACCGAT	CTGATGATCG	TCAGAACCCG
5,770	5,780	5,790	5,800	5,810	5,820
TCGCCGGATT	CTCAAGGCGC	TCCGTGCGCA	CGAAGCGAAT	GGCACCCCTGC	CCCCGGGTGT
5,830	5,840	5,850	5,860	5,870	5,880
TGACGAGGCG	GATCAGTACT	TCACGCCCAG	GAGCGGCTAT	TACCTCACGC	CGAAAAGCGT
5,890	5,900	5,910	5,920	5,930	5,940
CGACTGGGAG	ACCGCTTACG	AACAAAGGAT	CGAGGGCCTC	GTGCGGTGAA	TCTGCAAGCA
5,950	5,960	5,970	5,980	5,990	6,000
CTGCGCAATG	ACGCCGTCTT	GGATGACCGC	GGAGGTGATG	CCATGGTTCG	AATCGATCAA
6,010	6,020	6,030	6,040	6,050	6,060
CACGATACCT	ACAGTGTTTCG	GGTGATATCC	CGATCGCACC	TGAGCAAGGA	TATTGTTTTCAG

6,070	6,080	6,090	6,100	6,110	6,120
GTGGAAC TCG	AGGAATCGAG	CCAGCGACCA	TTGCCGGACT	ATGAGCCGGG	ATCGCACGTA
6,130	6,140	6,150	6,160	6,170	6,180
GACATCTATG	TCCAGGATGA	CCTTGTCGGG	CAATACTCGC	TGGTCAAGGC	GTCCGACGCG
6,190	6,200	6,210	6,220	6,230	6,240
CAGGCCTCGT	ACCAGATCGC	CTTCAAGGTC	AAGCGCGATC	AGGGCAGTGC	GACGGAGCTG
6,250	6,260	6,270	6,280	6,290	6,300
ATGTGCGAGT	TGCTCCGTGT	CGGTGCAACA	ACCCGAATCT	CTGCGCCGAG	GAACGCTTTT
6,310	6,320	6,330	6,340	6,350	6,360
GCCCTGGATC	CGCAGGCGCG	CGAGACCGTC	CTGATTTGTG	GCGGAATAGG	AATAACGCCC
6,370	6,380	6,390	6,400	6,410	6,420
ATGGTTCATA	TGGCCATGAC	CCTGGTCAAA	GCGAAGCGGC	CATGGAGCAT	GCACATTGCC
6,430	6,440	6,450	6,460	6,470	6,480
TCCCGGGATG	GCGATGAGCT	CGATTGCTG	GGGCCGCTGG	GAAGTTGCTC	CGAGATCAGC
6,490	6,500	6,510	6,520	6,530	6,540
CGCTACATCT	CGTCCCAGGG	GGACAGATTG	CCGATTCGGG	ACCTGGCCGA	GCGCGCACCT
6,550	6,560	6,570	6,580	6,590	6,600
GCGAATGCGC	ATCTGTACTT	CTGCGGGCCT	GAGGGCATGT	TGCAGGAATT	CACGGTGAAT
6,610	6,620	6,630	6,640	6,650	6,660
CTTGCCCGCA	GTGGGGCGCA	GTATGTCGTG	CGCGAGGGCG	AGACGATATT	GGATGTCCTG
6,670	6,680	6,690	6,700	6,710	6,720
CGCAATGCCG	GCCACCATGT	CACGTCCTCC	TGTCGGCAGG	GGATTGTGTT	CATGTGTGAA
6,730	6,740	6,750	6,760	6,770	6,780
ACCACTTTGA	TTTCCGGCGT	GCCGGACCAC	AGGGACCGCT	TGCTGACCGA	TAGCGAAAAJ
6,790	6,800	6,810	6,820	6,830	6,840
GCCTCGGGTC	GAACCATGTT	GATTGCTGT	TCACGCGCGC	TCAGCCCGGA	ACTCACCTTG
6,850	6,860	6,870	6,880	6,890	6,900
GATCTGTGAA	TATGAGAATA	GGGGTCGTGG	GCCTGGGGAA	GGCATTATG	CTGATGCTGC
6,910	6,920	6,930	6,940	6,950	6,960
CGACGTTTCT	CATGGATCGG	CGGGTTCAAC	TGGTCGCTGC	ATCGGATACC	GATCCGCTGA

6,970	6,980	6,990	7,000	7,010	7,020
GCCTGCGCCA	ATTCAAGGCG	GATTTCCCCG	CGGCGGCCGT	GCATGGCGAC	ATCGAGTCGC
7,030	7,040	7,050	7,060	7,070	7,080
TGTGCAAGAA	CCCGGATGTC	GAGGTCGTCT	ACATCGGCAC	CCCTCACCAG	TTTCATGCTG
7,090	7,100	7,110	7,120	7,130	7,140
TTCACGCGGA	AATTGCGCTG	AACGCCGGAA	AGCATGTGCT	CGTTGAGAAG	CCCATGGCGG
7,150	7,160	7,170	7,180	7,190	7,200
TGACCCTTGA	GGACTGCTGC	CGAATGAATG	CCTGCGCGCA	GCGCGCTGGG	AAATACCTGA
7,210	7,220	7,230	7,240	7,250	7,260
TCGTGGGCCA	CAGCCACAGC	TTCGATCATC	CGATCTCCAG	AGCCAAGGAG	TTGATCGACT
7,270	7,280	7,290	7,300	7,310	7,320
CCGGACGATA	TGGGCGCGTC	AGGTTCAATC	ACTCGATGAA	CTACACGGAT	TTCTGTACCC
7,330	7,340	7,350	7,360	7,370	7,380
GGCCAAGAAG	GGCGGAGGAG	TTGAATACCG	ACCTGGGGGG	CGGCGTGGTC	TTCAACCAAG
7,390	7,400	7,410	7,420	7,430	7,440
CGTCCCATCP	GCTCGATGTG	ATTGCGCTGC	TGGCGCAAGG	CCGCGTGGTG	GACGTTTCCA
7,450	7,460	7,470	7,480	7,490	7,500
GCTATCTGGG	TCGCTGGGAC	GCTGCCCCGC	GGACCGAAGG	GGCGTACTCC	GCACTGATCA
7,510	7,520	7,530	7,540	7,550	7,560
AATTCGACAA	TGGCGTGGCG	GCCAATGTCA	CCTACAGCGG	CTACGCCAC	TACAACTCTG
7,570	7,580	7,590	7,600	7,610	7,620
ACGAGGTGAT	GGGCTGGGTC	AATGAGCTGG	GGGAAAAGCA	GCCGGAGCGG	CGGCTTTTCC
7,630	7,640	7,650	7,660	7,670	7,680
CGACCCGATT	GCGACTCGAT	GAGGTCTTGA	ACAAGGGCAG	TGAGGCGGCC	TACAAGAATG
7,690	7,700	7,710	7,720	7,730	7,740
AGATGTCTTA	TGGTCGCGGG	TGGTCGGCGC	ATGGCTTGCC	GAAAATGGCC	AGGAACCACC
7,750	7,760	7,770	7,780	7,790	7,800
AGCATTTCGG	CCATCTGGTG	ATCAGCTGCG	AAGGTGCCGA	CATCGTTCCC	AAGGCGGAAA
7,810	7,820	7,830	7,840	7,850	7,860
AAATCGAAGT	ATTGCCAAT	GGCGAATACC	ACTGCGAGAA	CTTCAGCATG	CCGGATTTTC

7,870	7,880	7,890	7,900	7,910	7,920
CGCGCCAAGA	AGTCATGGAC	GAATTGTTTG	ACGCTATTTT	CGGCACGCGC	CCCCCGGTTT
7,930	7,940	7,950	7,960	7,970	7,980
ATACCGGCGA	GTGGGCCATG	GAAACCCTCG	ATCTTTGCAT	TGCACTGCTT	GAAGGGGGGA
7,990	8,000	8,010	8,020	8,030	8,040
AACCTGAGTT	GATCGGGATG	CGCGCGACGG	CCGTGCGCTG	TGCTGGTCGG	TCCTCGGGGC
8,050	8,060	8,070	8,080	8,090	8,100
GGTAGAAGTC	GGGCTAATTC	ACTTGTTGGC	GGTTTTGCTT	TATTAATAGT	TGGAGATGGC
8,110	8,120	8,130	8,140	8,150	8,160
GATGAATTCA	AGACGTATGG	TGCTTTCTCA	TGGGGCGGCG	CTTGTGGGCG	CGGCTTCCAC
8,170	8,180	8,190	8,200	8,210	8,220
TGGCTTGCTG	TTGCCGCAAT	GGGCGCGTGC	CAATCGGCCA	AGGTCAAGGT	CGGCCTGATG
8,230	8,240	8,250	8,260	8,270	8,280
CTGCCGTTCA	CCGGCACCTT	TGCTCAACCC	GGTATTGCGA	GCGAAAACGG	GTTTCGCCTG
8,290	8,300	8,310	8,320	8,330	8,340
GCCCTTGAGC	AGAACGGGCT	CAAGCTGGGT	GGGCGCGAGA	TCGAGTTCTT	CAAGGTCTGA
8,350	8,360	8,370	8,380	8,390	8,400
GACGAATCGA	ACCCGGCCAA	GGGCATCGAG	AACGCCACCC	GCCTGGTTCA	GCGCGACAA
8,410	8,420	8,430	8,440	8,450	8,460
GTCGACGTGC	TGATCGGCAC	CGTGCACTCG	GGCGTGCAGA	TGGGCATCCA	CAAGGTGGCG
8,470	8,480	8,490	8,500	8,510	8,520
CGCGAGTCGG	GCGTGCTCAA	CCTCATCCCC	AACGCTGGCG	TGCACATGGC	CACGCGCGCG
8,530	8,540	8,550	8,560	8,570	8,580
CAATGTGCCG	CCAACGTGTT	TCGGGCGTCC	TTCACCAATG	CCCAGCCGAC	GCTGGCGCTG
8,590	8,600	8,610	8,620	8,630	8,640
GGCAAGGTCA	TGGTCGATCG	CGGCCACAAG	AAAGCCGTGT	GGATCACCTG	GAAGTACGCG
8,650	8,660	8,670	8,680	8,690	8,700
GCCGGCGACG	AGGCCTTCGA	GGGCTTCGAC	AAGAGTTACA	CCGCCGCCGG	CACCATCGTC
8,710	8,720	8,730	8,740	8,750	8,760
AAGAAGCTCT	CCCTGCCGTT	CCCCAACGTG	GAGTTTCAGG	CCCTGCTGAC	CGAGATCGCC

8,770	8,780	8,790	8,800	8,810	8,820
TCGCTCAAGC	CCGACGCCGT	GGCCTGCTTC	TTCGCCGGCG	GCGGTGCGGC	CAAGTTCCTG
8,830	8,840	8,850	8,860	8,870	8,880
CGTGACTACG	CGGACTCGGG	CCTGGGCAAA	TCTATTCCCC	TGTATGGCTC	GGGCTTCCTG
8,890	8,900	8,910	8,920	8,930	8,940
ACCGAGGGCG	TGCTGGATGC	CGCCGGCGAT	GCCGCCAACG	GTGTGCTCAC	GACGCTGCAC
8,950	8,960	8,970	8,980	8,990	9,000
TATGCCGATT	CGCTGGACAC	CGAGCGCAAC	AAGCAGTTCA	GGACGCAATA	CGCCAAGCGA
9,010	9,020	9,030	9,040	9,050	9,060
TTCAAGCTGC	AACCCGATGT	GTACGCCATG	CAGGGCTACG	ACACCGGCCT	GCTGCTGATT
9,070	9,080	9,090	9,100	9,110	9,120
CGCGGGGCGA	ATGCGTCAAC	GGCGACTTTG	GCAACAAGCG	TGCCTGTTTG	CGCGGCTGGA
9,130	9,140	9,150	9,160	9,170	9,180
GAGCAGATCG	ACAGCCCGCG	TGGCAAGTGG	AAGATGAGCC	CGCGCACAAC	CGATCCAGGA
9,190	9,200	9,210	9,220	9,230	9,240
CATGTACCTG	CGCCAGGTGG	TCAACAAGGA	CAACAAGGTG	ATCGGCATTG	CGCAAAAAGC
9,250	9,260	9,270	9,280	9,290	9,300
CGTGGCCGAC	CCTGGCACCG	GTTGCCGCAT	AGGGTGATTG	GGCCGCGCGC	CGGCTGCGTC
9,310	9,320	9,330	9,340	9,350	9,360
TGTGCGGCGG	AGTTCGTCCC	ATGGGGGGGC	CGCCGCATTT	TTTTCTGCCT	TTGATTTCCC
9,370	9,380	9,390	9,400	9,410	9,420
TGTATGGATT	CCGCCAATCT	TCTGATCCAG	CTGCTCAACA	GCGTGCAGTA	CGGCCTGCTG
9,430	9,440	9,450	9,460	9,470	9,480
CTGTTCATGC	TGGCCGCCGG	CCTGACGCTG	ATCTTCGGCA	TCATGGGCGT	GGTCAACCTC
9,490	9,500	9,510	9,520	9,530	9,540
GCGCACGGCA	GCTTCTACAT	GATCGGTGCC	TACCTTGCGT	GGTACTTCAC	CTCCATGCTG
9,550	9,560	9,570	9,580	9,590	9,600
GGCAGCCTCA	CGCTGGCCAT	CGTGCTGGGC	ACGGTGCTCG	CCGTGCTGCT	CGGCTGGGCG
9,610	9,620	9,630	9,640	9,650	9,660
CTGGAATGGC	TGCAGTTTCG	GCACTTCTAC	CACCGCGACC	ACCTCGACCA	GGGTGCTGCT

9,670	9,680	9,690	9,700	9,710	9,720
CACCTTCGGC	CTGATCTACA	TCTTCGAA	GA	GATGCGCTCC	ATCTTCTGGG
					GCGACGAAGT
9,730	9,740	9,750	9,760	9,770	9,780
GCATGCGGTG	CAGGTGCCCC	AGGTGCTGAA	CTGGGGCATC	CGGCTCACC	G
					ACAAGCTTGA
9,790	9,800	9,810	9,820	9,830	9,840
TTACCCGGTC	TACCGTCTCG	TCATGTCGGG	CGTGTGCATC	GCGCTGGCCA	T
					TCGGCCTGTA
9,850	9,860	9,870	9,880	9,890	9,900
CCTGCTCATC	AACAAGACGC	GCCTGGGCAT	GAAGATCCGC	GCCGGCGCCT	T
					TCAACCGCGA
9,910	9,920	9,930	9,940	9,950	9,960
GATGGCCGAA	TCCCTGGGCG	TGAACATCCG	CCTCATCCAC	GGCGTGGTGT	T
					TCGCGCTCGG
9,970	9,980	9,990	10,000	10,010	10,020
CGTGGGCCTG	GCCGCCATCG	CCGGCATGAT	CGCCTCGCCC	ATCGCCAGCG	T
					TGTACCCAGG
10,030	10,040	10,050	10,060	10,070	10,080
CATGGGCAGC	TCGGTGCTCA	TCATGTGCCT	CGTGGTGGTG	GTCATCGGCG	G
					GCATCGGCTC
10,090	10,100	10,110	10,120	10,130	10,140
GGTGCGCGGC	GCGCTCGTGG	CGGCGCTGCT	GGTGGGCCTG	GTCGACACCT	T
					TCGGCAAGGT
10,150	10,160	10,170	10,180	10,190	10,200
GCTGGTGCCG	CAGATCGCCG	GCATGGCGGT	CTACATGCTC	ATGGCCCTGG	T
					TGCTGCTCTA
10,210	10,220	10,230	10,240	10,250	10,260
CAAACCCGAA	GGGCTGTTCA	AACAATGAGC	ACTGCCAAAT	CCCAACTCGA	A
					GTGCTGCCG
10,270	10,280	10,290	10,300	10,310	10,320
CGCAGCATCC	AGTTCCTGCT	CGCGCTCGGC	GCGATCGCGC	TCGCGGCCTT	T
					TCCCCTGCTG
10,330	10,340	10,350	10,360	10,370	10,380
CCCATCGCGG	GCGCCGACTT	CTACCTCGGC	ATGCTGTGCG	AGATGATGAT	C
					CTGGCCATC
10,390	10,400	10,410	10,420	10,430	10,440
TTGCGGATGA	GCCTGGACCT	GCTGCAGGGC	GTCACCGGCC	TGGTCTCGCT	G
					GGGCCATGCG
10,450	10,460	10,470	10,480	10,490	10,500
GCCTTCTTCG	GCCTGGCCGG	CTACATCCTG	GCCTTCGTCA	CGCCGGCCGA	C
					CGCGCCGATC
10,510	10,520	10,530	10,540	10,550	10,560
GCCCTGTGGT	GGTCGCTGCC	GCTGGCGGCC	GCCGGTGCCG	GCCTGGCCGC	G
					CTGGTTCATC

10,570	10,580	10,590	10,600	10,610	10,620
GGCTTCTTCG	TGGTCCGCAC	CCACGGGCATC	TACTTCATCA	TGGTCACCAT	GGCCTTCGCG
10,630	10,640	10,650	10,660	10,670	10,680
CAGTGTATTT	CCTGTTCTTC	GACAACAAGG	CGCTCGGCGG	TTCCGACGGC	ATCTACGTCA
10,690	10,700	10,710	10,720	10,730	10,740
ACTTCCGCCC	CGACGCCACG	GCGCTGGGCC	TGGACCTGGA	CAACAAGACC	GTCTTCTACT
10,750	10,760	10,770	10,780	10,790	10,800
ACTTCACGCT	CGCCGTGCTG	CTGCTGCTGT	ACTTTGCGCT	GCGCCGCCTG	CTGTTCTCGC
10,810	10,820	10,830	10,840	10,850	10,860
CCTTCGGCCG	CGCGCTCGCC	GGCATCCGCG	TCAACGAGCA	CCGCATGCGC	GCCGTCGGCT
10,870	10,880	10,890	10,900	10,910	10,920
ACGGCAGCTT	CGGCTACAAG	CTCACGGCCT	TCACGCTGGC	CGGTGCGCTG	GCCGGCGTCG
10,930	10,940	10,950	10,960	10,970	10,980
CGGGCTACCT	GTGGGCGCGC	CAGACGGGCT	TCGTGAACCC	CGAGCTCATG	GGCTTTCACA
10,990	11,000	11,010	11,020	11,030	11,040
TGAGCGCGCA	CGCCATCATG	ATGGTGATCC	TGGGTGGCAT	GGGCAACTTC	GCGGGCGCCA
11,050	11,060	11,070	11,080	11,090	11,100
TCGTGCGCGC	CTTCGCCTTC	GAGTACGTCA	TGCACGTGTT	CAAGGACATG	ACCAAGCACT
11,110	11,120	11,130	11,140	11,150	11,160
GGCAGCTGCT	CATGGGCGGC	TTCATCGTGC	TGGTGGTGCT	GGTCGCGCCG	CGCGGCCTGC
11,170	11,180	11,190	11,200	11,210	11,220
TGGGCCTGAT	CGATCGCTTC	ACCCTACGCG	GGAGCAAGCC	ATGAACGAGT	TGCTGCTCAG
11,230	11,240	11,250	11,260	11,270	11,280
CGCGAAGAAG	CTCACCAAAC	GCTTCGGCGG	CCTGGCCGCC	GTCAACGACG	TCTCGGTGGA
11,290	11,300	11,310	11,320	11,330	11,340
CCTGCACCGC	GGCCACATCC	ACGCCATCAT	CGGGCCCCAAC	GGCGCGGGCA	AGTCCACGCT
11,350	11,360	11,370	11,380	11,390	11,400
GACCAACCTG	CTCTCAGGCG	ATCTGTCGGT	CAGCAGTGGG	CAGATCATGC	TCAACGGCCA
11,410	11,420	11,430	11,440	11,450	11,460
GACCATCACG	GGCCGCTCGC	CCGAGCGCAT	CTCGCGCCAG	GGCCTGGGCC	GCAGCTACCA

11,470	11,480	11,490	11,500	11,510	11,520
GAAGACCAAC	ATCTTCCTGC	CCTTCACGGT	CTGGGACAAC	GTGCGCCTGG	CCGCGCAATC
11,530	11,540	11,550	11,560	11,570	11,580
GCGCGAACGC	CAGGCCCCGT	TCAACCCGCG	CCACTGGTTC	GGCAAGGCCG	ACAAGCTCAG
11,590	11,600	11,610	11,620	11,630	11,640
CGAGGTGAAC	CGGCGCTGCG	AACGCGCGCT	GGAGCTTGCG	GGCCTGACGG	CGCGCGCGCA
11,650	11,660	11,670	11,680	11,690	11,700
CCGTGGCCGC	GGCCACCAGC	CACGGCGAGC	AGCGCCAGCT	CGAAATCGCC	ATGACCCTGG
11,710	11,720	11,730	11,740	11,750	11,760
CCACCGAACC	CAGCGTGCTG	CTGCTCGACG	AGCCTCTGGC	CGGCATGGGC	GCGGCCGAGG
11,770	11,780	11,790	11,800	11,810	11,820
CCGAGTCCAT	GGTGGCCCTG	CTGCAGCGAC	TGAAGGCCGA	ACACGCCATG	CTGCTCGTGG
11,830	11,840	11,850	11,860	11,870	11,880
AGCACGACAT	GGACGCCGTG	TTCGCGCTGG	CCGACGTGCT	GACCGTGATG	GTCAACGGGC
11,890	11,900	11,910	11,920	11,930	11,940
AGGTGATCGC	CAGCGGAACC	CCCGCGCAGA	TCCGCGCGGA	CGCCGGCGTC	CAAGCCGCCT
11,950	11,960	11,970	11,980	11,990	12,000
ACCTCGGCCA	GGAGCATTGA	TGAGCGCACT	GATCGAAGCC	AAGGGCCTGC	ACACGCACTA
12,010	12,020	12,030	12,040	12,050	12,060
CGGCCAGAGC	CACATCCTGC	GCGGCGTGGA	TTTTCGCGTC	GACGTGCGCC	AGACCATCGG
12,070	12,080	12,090	12,100	12,110	12,120
CCTGATGGGC	CGCAACGGCA	TGGGCAAGAC	CACGCTGCTG	AAGTCGCTCA	TGGGCATCGT
12,130	12,140	12,150	12,160	12,170	12,180
CAAGCCCAGC	GGCGGCACGG	TGCAGGTCAA	GGGCCGGGAC	ATGACCGGCG	CCACGCCCTA
12,190	12,200	12,210	12,220	12,230	12,240
CGCGGTGGCG	CGCGAAGGCA	TCGCCTACGT	GCCCGAAGGC	CGGGGCATCT	TCGGCAACCT
12,250	12,260	12,270	12,280	12,290	12,300
GAGCGTGAAG	GAAAACCTGG	TGATGGCGGC	GCGAGCCGGC	ACGCGGGGCC	AGCGCGAATG
12,310	12,320	12,330	12,340	12,350	12,360
GACCTACGAG	CGCGTGCTGG	AGACCTTCCC	CCGCCTGGCG	GAGCGGTTGA	ACCACGGCGG

12,370	12,380	12,390	12,400	12,410	12,420
GCAGCAGTTG	AGCGGCGGCG	AACAGCAGAT	GCTCACCATC	GGCCGTGCAC	TGATGACCAA
12,430	12,440	12,450	12,460	12,470	12,480
CCCGGACCTG	CTGATCCTCG	ACGAGCCACC	GAGGGCCTCG	CCCCGCTGAT	CGCGCGCGAG
12,490	12,500	12,510	12,520	12,530	12,540
ATCTGGCGCA	TCTGCGGACT	GATCGCCGAG	AGCGGCATCA	GCAGCGTGAT	CGTGGACAAG
12,550	12,560	12,570	12,580	12,590	12,600
AACTGGAAGC	ACGTGACGCA	GATCACCGAC	CGCAACCTGA	TCCTGGTCAA	GGGCGAGCTG
12,610	12,620	12,630	12,640	12,650	12,660
GTGTTGAGG	GCACGACGCC	CGAGCTGCTG	GCACAGCCTG	GGGTCTTGGA	GCAGCACCTG
12,670	12,680	12,690	12,700	12,710	12,720
GGGGTGTGAG	GCCATCCTGC	CGCGGGCCGC	ACCACCGCGC	GGGCCCGCGG	CAGATCGACT
12,730	12,740	12,750	12,760	12,770	12,780
GCCCCCTCAGG	CGCTCAGCCG	CCGGCCTGGT	CGCGCTCGCG	CAGGCGGAAG	CGCTGGATCT
12,790	12,800	12,810	12,820	12,830	12,840
TGCCGGTGGC	GGTCTTGGGC	AGCTCGTCCA	GGAACTCGAT	GAAGCGCGGG	TACTTGTAAG
12,850	12,860	12,870	12,880	12,890	12,900
GCGCCAGGCG	CTCCTTCACG	AAGGCCTGCA	GGTCCGCGGC	ACCGGCCTGC	TGGCCGGGCT
12,910	12,920	12,930	12,940	12,950	12,960
TGAGCACCAC	GTAGGCCTTG	CTCTTGGTCA	GGCCTTCGCC	GTCCGTGTGG	CCGATCACCG
12,970	12,980	12,990	13,000	13,010	13,020
CGGCTTCCAG	CACGGCCGGG	TGCTGGACCA	GGGTGGCCTC	GACCTCGAAG	GGGCTGACGT
13,030	13,040	13,050	13,060	13,070	13,080
AGATGCCGCT	GACCTTGAGC	ATGTCGTCGT	TGCGCCCGGC	GTAGGTGTAG	GTGCCGTTGG
13,090	13,100	13,110	13,120	13,130	13,140
GGTTGCGCTG	GTAATTGTGC	CCGGCCTTGG	TCCAGCCGCC	CTGGAAGGTG	TCGCGGCTCT
13,150	13,160	13,170	13,180	13,190	13,200
TGTCGCGGTT	GTTCCAGTAC	ATCAACGCCG	CGCTGGGGCC	CTTGATGTAG	AGGTCGCCGA
13,210	13,220	13,230	13,240	13,250	13,260
TGTCGATCGA	ACCATCGGCG	TTGGGCGCGA	GCGCCTGGCC	GTCTTCGCCA	CGCAACTCCA

13,270 13,280 13,290 13,300 13,310 13,320
 CCTCGTAACC CGGCACGGGG TGGCCCGTGC TGCCGTAGTG CACCTCGCCC GGCACGTTGG

 13,330 13,340 13,350 13,360 13,370 13,380
 ACAGGAAGAT GTGCAGCATC TCGGTGGAGC CGATGCCGTC GATGATTTCC ACCCCGAAGT

 13,390 13,400 13,410 13,420 13,430 13,440
 GCGCGCTCCA GCGCTCGCCG ATGTCTTGCG GCAAGGCTTC GCCGGCCGAG GAGCACAGGC

 13,450 13,460 13,470 13,480 13,490 13,500
 GCAGCGCCAC CGCTTCCTTG GCCGGCAGGT CGGCACTGGC GAGCATGCCG CCGTAGCCGG

 13,510 13,520 13,530 13,540 13,550 13,560
 TGGGCGCTCC GCAGAAGATG GTGGGACGGT GCTCCGTCAT GCGCTTGAAG CAGGCCTGCG

 13,570 13,580 13,590 13,600 13,610 13,620
 GCGTCGGCCG CTCGGCCATC AGCACCACGG TGGCGCCAC GCTGAGCGGG AAGCTCAGCG

 13,630 13,640 13,650 13,660 13,670 13,680
 CATTGCCAG GCCGTAGGCG AAGAAGAGCT TGGCGGCCGA GAGCACGACG TCGTCTTCGC

 13,690 13,700 13,710 13,720 13,730 13,740
 GCATGCCAG CACCGCCTTG CCATAGAGCT CGCAGGTCCA GTAAAGGTTG GCCTGGGAGT

 13,750 13,760 13,770 13,780 13,790 13,800
 GCACCGTGCC CTTGGGGGCA CCGGTCGAGC CGCTGGAGTA GAGCCAGAAG GCCGGCTCGT

 13,810 13,820 13,830 13,840 13,850 13,860
 CGCCGTGCGT GTCGATGGGG TCGGCCGGGC GCGCGCGGGC GAGCAGGTGG TCGAAGGCGT

 13,870 13,880 13,884
 GGGCGCCCGC GGGCAGCTTG TCGC repeat sequence IS1071

APPENDIX 2: GC-MS ANALYSES

APPENDIX 2A: GC-MS Analysis of Medium of 3-chlorobenzoate grown *Alcaligenes* sp. strain BR60. Mass spectra only of peaks from gas chromatogram relevant to 3-chlorobenzoate degradation are included.

711194 Scan 511 RT=11:36

7t1194 Scan 511 RT 11:36 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
10	59	75679	0.2	N	136	185	120324	0.3	N
14	63	300400	0.6	N	138	187	29583	0.1	N
16	65	47576	0.1	N	144	193	70588	0.1	N
20	69	153315	0.3	N	148	197	50114	0.1	N
23	72	81033	0.2	N	149	198	15679	0.0	N
24	73	2922890	6.1	N	150	199	140861	0.3	N
25	74	280235	0.6	N	151	200	50864	0.1	N
26	75	439231	0.9	N	152	201	29885	0.1	N
28	77	201023	0.4	N	158	207	6412	0.0	N
30	79	26595	0.1	N	162	211	157519	0.3	N
34	83	15460	0.0	N	164	213	99867	0.2	N
40	89	33389	0.1	N	166	215	27313	0.1	N
42	91	589420	1.2	N	170	219	239233	0.5	N
43	92	40432	0.1	N	171	220	36517	0.1	N
44	93	421286	0.9	N	172	221	141502	0.3	N
45	94	8523	0.0	N	176	225	62996	0.1	N
46	95	142284	0.3	N	178	227	1371589	2.9	N
50	99	18246	0.0	N	179	228	309990	0.6	N
54	103	72098	0.2	N	180	229	549252	1.2	N
56	105	99065	0.2	N	181	230	126029	0.3	N
61	110	41142	0.1	N	182	231	63225	0.1	N
66	115	27321	0.1	N	192	241	20962	0.0	N
68	117	67519	0.1	N	194	243	66353	0.1	N
70	119	89283	0.2	N	204	253	15340	0.0	N
72	121	41842	0.1	N	208	257	1292898	2.7	N
76	125	61741	0.1	N	209	258	286693	0.6	N
77	126	62533	0.1	N	210	259	538540	1.1	N
78	127	93614	0.2	N	211	260	87994	0.2	N
79	128	20272	0.0	N	232	281	49020	0.1	N
80	129	28040	0.1	N	236	285	57799	0.1	N
82	131	91735	0.2	N	238	287	311565	0.7	N
84	133	147777	0.3	N	239	288	22543	0.0	N
86	135	59823	0.1	N	240	289	48407	0.1	N
88	137	30321	0.1	N	250	299	38273	0.1	N
92	141	20181	0.0	N	251	300	26686	0.1	N
94	143	741491	1.6	N	252	301	7348279	15.4	N
95	144	260995	0.5	N	253	302	1666996	3.5	N
98	147	460027	1.0	N	254	303	2967948	6.2	N
99	148	115568	0.2	N	255	304	699057	1.5	N
100	149	116506	0.2	N	256	305	222918	0.5	N
104	153	98538	0.2	N	257	306	39817	0.1	N
105	154	31223	0.1	N	260	309	20595	0.0	N
106	155	111922	0.2	N	266	315	40062	0.1	N
107	156	85883	0.2	N	267	316	3332613	7.0	N
108	157	63289	0.1	N	268	317	853861	1.8	N
109	158	18604	0.0	N	269	318	1290006	2.7	N
114	163	43675	0.1	N	270	319	292223	0.6	N
115	164	64008	0.1	N	271	320	104157	0.2	N
116	165	33520	0.1	N					
118	167	119100	0.2	N					
120	169	289871	0.6	N					
121	170	33782	0.1	N					
122	171	72343	0.2	N					
130	179	43028	0.1	N					
132	181	58490	0.1	N					
134	183	194054	0.4	N					
135	184	56827	0.1	N					

7/11/9 Scan 596 RT=13:01

7t1194 Scan 601 RT 13:06 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
9	58	8612	0.0	N	176	225	108346	0.2	N
10	59	122567	0.3	N	180	229	13119	0.0	N
20	69	142175	0.3	N	188	237	292771	0.6	N
23	72	152162	0.3	N	189	238	58471	0.1	N
24	73	8549340	17.9	N	190	239	57864	0.1	N
25	74	76239	1.6	N	202	251	40056	0.1	N
26	75	648653	1.4	N	204	253	19978	0.0	N
28	77	255656	0.5	N	216	265	161511	0.3	N
34	83	65322	0.1	N	217	266	34401	0.1	N
40	89	22427	0.0	N	218	267	1107310	2.3	N
42	91	89707	0.2	N	219	268	207622	0.4	N
44	93	38753	0.1	N	220	269	98838	0.2	N
54	103	50258	0.1	N	223	272	294882	0.6	N
56	105	196098	0.4	N	224	273	118512	0.2	N
60	109	39548	0.1	N	230	279	143750	0.3	N
68	117	71827	0.2	N	231	280	33103	0.1	N
70	119	171282	0.4	N	232	281	1463431	3.1	N
72	121	10868	0.0	N	233	282	473899	1.0	N
82	131	186669	0.4	N	234	283	284149	0.6	N
84	133	390080	0.8	N	235	284	68726	0.1	N
85	134	66234	0.1	N	246	295	33279	0.1	N
86	135	127010	0.3	N	248	297	86965	0.2	N
88	137	448509	0.9	N	249	298	122780	0.3	N
89	138	63281	0.1	N	258	307	29911	0.1	N
96	145	26928	0.1	N	260	309	110196	0.2	N
98	147	643336	1.3	N	262	311	1723007	3.6	N
99	148	110264	0.2	N	263	312	515269	1.1	N
100	149	179124	0.4	N	264	313	234047	0.5	N
101	150	22631	0.0	N	265	314	50259	0.1	N
102	151	40052	0.1	N	290	339	59231	0.1	N
114	163	182922	0.4	N	292	341	1118260	2.3	N
115	164	52810	0.1	N	294	343	68080	0.1	N
116	165	709530	1.5	N	304	353	107351	0.2	N
117	166	114229	0.2	N	305	354	129034	0.3	N
118	167	36783	0.1	N	306	355	5283203	11.1	N
120	169	55443	0.1	N	307	356	1565239	3.3	N
128	177	28468	0.1	N	308	357	769677	1.6	N
130	179	276691	0.6	N	309	358	200584	0.4	N
131	180	43516	0.1	N	310	359	38895	0.1	N
132	181	72832	0.2	N	320	369	211465	0.4	N
142	191	121572	0.3	N	321	370	12336964	25.8	N
143	192	84727	0.2	N	322	371	4029413	8.4	N
144	193	10694012	22.4	N	323	372	1872622	3.9	N
145	194	1575683	3.3	N	324	373	412472	0.9	N
146	195	658581	1.4	N	325	374	101391	0.2	N
147	196	52381	0.1	N					
156	205	61859	0.1	N					
158	207	372941	0.8	N					
159	208	55347	0.1	N					
160	209	115626	0.2	N					
161	210	23051	0.0	N					
162	211	65203	0.1	N					
170	219	28841	0.1	N					
172	221	75632	0.2	N					
173	222	50591	0.1	N					
174	223	1175442	2.5	N					
175	224	277969	0.6	N					

71195 Scan 626 RT=13:32

7t1194 Scan 629 RT 13:35 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags
20	69	134566	0.3	N
21	70	17522	0.0	N
24	73	1491825	3.2	N
25	74	142607	0.3	N
26	75	638071	1.4	N
27	76	22761	0.0	N
28	77	136449	0.3	N
44	93	127987	0.3	N
46	95	38564	0.1	N
70	119	94889	0.2	N
78	127	69823	0.1	N
82	131	35318	0.1	N
84	133	15369	0.0	N
94	143	4634	0.1	N
94	147	240734	0.5	N
99	148	42517	0.1	N
100	149	93653	0.2	N
105	154	23231	0.0	N
120	169	66975	0.1	N
132	181	15579	0.0	N
141	190	17935	0.0	N
142	191	83270	0.2	N
144	193	46927	0.1	N
158	207	18819	0.0	N
160	209	13076	0.0	N
178	227	189651	0.4	N
179	228	27274	0.1	N
180	229	150807	0.3	N
181	230	23888	0.1	N
182	231	26174	0.1	N
208	257	156218	0.3	N
209	258	33288	0.1	N
210	259	57088	0.1	N
232	281	43650	0.1	N
240	289	137554	0.3	N
241	290	29877	0.1	N
242	291	60598	0.1	N
252	301	522908	1.1	N
253	302	91927	0.2	N
254	303	213034	0.5	N
255	304	23699	0.1	N
260	309	20658	0.0	N
267	316	172940	0.1	N
268	317	47814	0.1	N
269	318	58071	0.1	N
278	327	21795	0.0	N
321	370	20243	0.0	N
322	371	302512	0.6	N
323	372	90097	0.2	N
324	373	38129	0.1	N
342	391	74776	0.2	N
344	393	33700	0.1	N
357	406	321338	0.7	N
358	407	113675	0.2	N
359	408	128247	0.3	N

APPENDIX 2B: GC-MS Analysis of Medium of 3-chlorobenzoate-grown *Alcaligenes* sp. strain BR6024 containing pBRCN8. Mass spectra only of peaks from gas chromatogram relevant to 3-chlorobenzoate degradation are included.

711550 Scan 253 RT=7.37

7t1550 Scan 253 RT 7:36 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags
3	40	1445721	2.3	N
6	43	114112	0.2	N
7	44	544768	0.9	N
8	45	191601	0.3	N
10	47	29512	0.0	N
13	50	58329	0.1	N
14	51	48146	0.1	N
22	59	8676	0.0	N
35	72	26897	0.0	N
36	73	224410	0.4	N
37	74	43001	0.1	N
38	75	648666	1.1	N
39	76	206941	0.3	N
40	77	70470	0.1	N
54	91	20933	0.0	N
67	104	64615	0.1	N
69	106	186773	0.3	N
70	107	63364	0.1	N
74	111	1020555	1.7	N
75	112	69994	0.1	N
76	113	310917	0.5	N
102	139	1535817	2.5	N
103	140	95463	0.2	N
104	141	516430	0.8	N
105	142	27532	0.0	N
132	169	524425	0.8	N
133	170	63073	0.1	N
134	171	172648	0.3	N
141	178	49725	0.1	N
176	213	3453099	5.6	N
177	214	521689	0.8	N
178	215	1188012	1.9	N
179	216	168391	0.3	N
180	217	23315	0.0	N
191	228	372462	0.6	N
192	229	54751	0.1	N
193	230	128623	0.2	N

7:1550 Scan 467 RT=11:31

7t1550 Scan 467 RT 11:31 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags
1	38	69596	0.1	N
3	40	1304349	2.1	N
6	43	28843	0.0	N
7	44	416599	0.7	N
8	45	123960	0.2	N
26	63	43516	0.1	N
36	73	1038050	1.7	N
37	74	47622	0.1	N
38	75	422261	0.7	N
40	77	91068	0.1	N
54	91	178094	0.3	N
56	93	110828	0.2	N
58	95	38376	0.1	N
90	127	10993	0.0	N
96	133	26500	0.0	N
106	143	455076	0.7	N
107	144	139139	0.2	N
110	147	191987	0.3	N
111	148	23145	0.0	N
112	149	8664	0.0	N
132	169	80152	0.1	N
146	183	79694	0.1	N
148	185	13439	0.0	N
174	211	60014	0.1	N
176	213	40066	0.1	N
188	225	18112	0.0	N
190	227	565361	0.9	N
191	228	98694	0.2	N
192	229	221748	0.4	N
220	257	216492	0.3	N
221	258	41983	0.1	N
222	259	89819	0.1	N
249	286	152249	0.2	N
264	301	4256809	6.8	N
265	302	970193	1.6	N
266	303	1692988	2.7	N
267	304	348696	0.6	N
268	305	95828	0.2	N
278	315	13175	0.0	N
279	316	2282702	3.7	N
280	317	610062	1.0	N
281	318	1032657	1.7	N
282	319	204663	0.3	N
283	320	58484	0.1	N

7(1550 Scan 542 RT=12:53

7t1550 Scan 542 RT 12:53 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags
3	40	1392979	2.2	N
6	43	71377	0.1	N
7	44	536222	0.9	N
8	45	618817	1.0	N
9	46	37528	0.1	N
10	47	26108	0.0	N
21	58	13230	0.0	N
22	59	108219	0.2	N
35	72	116790	0.2	N
36	73	7082900	11.4	N
37	74	620799	1.0	N
38	75	605371	1.0	N
39	76	6425	0.0	N
40	77	75603	0.1	N
46	83	37386	0.1	N
68	105	37730	0.1	N
82	119	67031	0.1	N
94	131	98267	0.2	N
96	133	213795	0.3	N
98	135	58347	0.1	N
100	137	183594	0.3	N
110	147	421940	0.7	N
111	148	79117	0.1	N
112	149	82172	0.1	N
126	163	79622	0.1	N
128	165	203344	0.3	N
142	179	85038	0.1	N
154	191	32197	0.1	N
155	192	35306	0.1	N
156	193	2919706	4.7	N
157	194	445974	0.7	N
158	195	141218	0.2	N
170	207	175190	0.3	N
171	208	10927	0.0	N
186	223	349363	0.6	N
187	224	70237	0.1	N
200	237	137514	0.2	N
201	238	17902	0.0	N
228	265	85044	0.1	N
230	267	282057	0.5	N
231	268	65892	0.1	N
235	272	26396	0.0	N
242	279	58055	0.1	N
244	281	806977	1.3	N
245	282	236743	0.4	N
246	283	141413	0.2	N
261	298	42755	0.1	N
274	311	317993	0.5	N
275	312	117996	0.2	N
276	313	34561	0.1	N
304	341	733602	1.2	N
305	342	17234	0.0	N
316	353	59336	0.1	N
318	355	2633092	4.2	N
319	356	848265	1.4	N
320	357	435582	0.7	N
321	358	88035	0.1	N
332	369	113011	0.2	N
333	370	8687410	13.9	N
334	371	2736608	4.4	N
335	372	1412237	2.3	N
336	373	308508	0.5	N
337	374	52792	0.1	N

711550 Scan 620 RT=14:19

7t1550 Scan 620 RT 14:18 +EI LRP

Normalised to largest peak in current list

Peak	Mass	Intensity	%	Flags
3	40	1363430	100.0	N
7	44	413901	30.4	N
8	45	90772	6.7	N
36	73	852734	62.5	N
37	74	70383	5.2	N
38	75	193460	14.2	N
40	77	30799	2.3	N
110	147	13038	1.0	N
190	227	208457	15.3	N
191	228	28502	2.1	N
192	229	57534	4.2	N
352	389	166015	12.2	N
353	390	32947	2.4	N
354	391	65181	4.8	N
367	404	457278	33.5	N
368	405	131322	9.6	N
369	406	203261	14.9	N
370	407	49714	3.6	N

APPENDIX 2C: GC-MS Analysis of Medium of 3-chlorobenzoate-grown *Alcaligenes* sp. strain BR6024 containing pBRCN11. Mass spectra only of peaks from gas chromatogram relevant to 3-chlorobenzoate degradation are included.

Runname : 711756 Acquired Nominal data (Zero baseline).
Unsmoothed data only:

711756 Scan 468 RT=11:32

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags
1	38	31493	0.1	N
2	39	43270	0.1	N
3	40	837166	1.6	N
4	41	13220	0.0	N
6	43	345776	0.7	N
7	44	378909	0.7	N
8	45	642109	1.3	N
17	47	104439	0.2	N
14	51	10927	0.0	N
20	57	34555	0.1	N
21	58	49717	0.1	N
22	59	123925	0.2	N
24	61	40259	0.1	N
26	63	437694	0.9	N
7	64	8430	0.0	N
28	65	36582	0.1	N
29	66	15243	0.0	N
32	69	67840	0.1	N
35	72	85846	0.2	N
36	73	3403048	6.7	N
37	74	288787	0.6	N
38	75	698348	1.4	N
39	76	76854	0.2	N
40	77	163161	0.3	N
41	78	6378	0.0	N
42	79	54262	0.1	N
46	83	37627	0.1	N
47	84	11066	0.0	N
52	89	81062	0.2	N
53	90	64096	0.1	N
54	91	722843	1.4	N
55	92	145549	0.3	N
56	93	409028	0.8	N
58	95	141463	0.3	N
60	97	26486	0.1	N
61	98	40476	0.1	N
62	99	33263	0.1	N
66	103	133985	0.3	N
68	105	134391	0.3	N
69	106	81943	0.2	N
70	107	42765	0.1	N
73	110	97463	0.2	N
74	111	61661	0.1	N
75	112	33796	0.1	N
76	113	30312	0.1	N
78	115	54894	0.1	N
80	117	75528	0.1	N
81	118	10864	0.0	N
82	119	253894	0.5	N
83	120	124365	0.2	N
84	121	72511	0.1	N
88	125	96145	0.2	N
89	126	134273	0.3	N
90	127	148149	0.3	N
91	128	87524	0.2	N
92	129	104139	0.2	N
94	131	87781	0.2	N
96	133	183860	0.4	N
97	134	101423	0.2	N
98	135	89145	0.2	N
99	136	32835	0.1	N
100	137	62482	0.1	N
101	138	25468	0.1	N
102	139	29427	0.1	N
104	141	13400	0.0	N
105	142	28263	0.1	N
106	143	1319596	2.6	N
107	144	465806	0.9	N
108	145	15699	0.0	N
109	146	24647	0.0	N
110	147	909407	1.8	N

3 of/de 3

PM-1 3 1/2" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

7t1756 Scan 468 RT 11:32 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
111	148	133012	0.3	N					
112	149	162316	0.3	N					
113	150	36537	0.1	N					
116	153	128008	0.3	N					
117	154	27574	0.1	N					
118	155	140898	0.3	N	190	227	2140354	4.2	N
119	156	124663	0.2	N	191	228	467977	0.9	N
120	157	78602	0.2	N	192	229	950468	1.9	N
121	158	28443	0.1	N	193	230	171673	0.3	N
125	162	13465	0.0	N	194	231	81616	0.2	N
126	163	54755	0.1	N	199	236	45810	0.1	N
127	164	55705	0.1	N	200	237	51365	0.1	N
128	165	28427	0.1	N	204	241	48480	0.1	N
130	167	167095	0.3	N	206	243	101848	0.2	N
132	169	376481	0.7	N	207	244	27114	0.1	N
133	170	55146	0.1	N	208	245	34690	0.1	N
134	171	88323	0.2	N	214	251	221429	0.4	N
142	179	62116	0.1	N	215	252	24414	0.0	N
144	181	27974	0.1	N	216	253	51111	0.1	N
146	183	232016	0.5	N	218	255	24525	0.0	N
147	184	83505	0.2	N	220	257	1621468	3.2	N
148	185	212326	0.4	N	221	258	352924	0.7	N
149	186	38583	0.1	N	222	259	698825	1.4	N
150	187	33950	0.1	N	223	260	143133	0.3	N
156	193	85732	0.2	N	224	261	39705	0.1	N
157	194	280526	0.6	N	244	281	71325	0.1	N
158	195	52495	0.1	N	248	285	60057	0.1	N
160	197	90274	0.2	N	250	287	790836	1.6	N
161	198	15750	0.0	N	251	288	279125	0.5	N
162	199	120725	0.2	N	256	293	88213	0.2	N
163	200	49383	0.1	N	262	299	82673	0.2	N
164	201	21037	0.0	N	263	300	124617	0.2	N
170	207	27178	0.1	N	264	301	14948785	29.5	N
172	209	234535	0.5	N	265	302	3290906	6.5	N
173	210	27513	0.1	N	266	303	6068117	12.0	N
174	211	261447	0.5	N	267	304	1267198	2.5	N
175	212	42422	0.1	N	268	305	450205	0.9	N
176	213	145694	0.3	N	269	306	68335	0.1	N
178	215	27624	0.1	N	278	315	119352	0.2	N
182	219	446927	0.9	N	279	316	7352667	14.5	N
184	221	123572	0.2	N	280	317	1661953	3.3	N
188	225	87388	0.2	N	281	318	2860852	5.6	N
189	226	26786	0.1	N	282	319	581875	1.1	N
					283	320	204920	0.4	N

7i1756 Scan 545 RT=12:57

Normalised to larger non-reference peak in scan

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
3	40	809431	1.6	N	120	157	20615	0.0	N
4	41	11071	0.0	N	126	163	199214	0.4	N
6	43	304687	0.6	N	127	164	86866	0.2	N
7	44	507854	1.0	N	128	165	704509	1.4	N
8	45	1722717	3.5	N	129	166	125356	0.3	N
9	46	125850	0.3	N	130	167	20959	0.0	N
10	47	115619	0.2	N	140	177	38179	0.1	N
18	55	26699	0.1	N	141	178	39431	0.1	N
20	57	63619	0.1	N	142	179	209449	0.4	N
21	58	81330	0.2	N	143	180	33431	0.1	N
22	59	253002	0.5	N	144	181	51950	0.1	N
24	61	33419	0.1	N	154	191	98863	0.2	N
26	63	40677	0.1	N	155	192	87373	0.2	N
32	69	70696	0.1	N	156	193	7859268	15.7	N
34	71	33682	0.1	N	157	194	1179300	2.4	N
35	72	172009	0.3	N	158	195	431288	0.9	N
36	73	10785890	21.6	N	159	196	60771	0.1	N
37	74	941940	1.9	N	168	205	8579	0.0	N
38	75	826339	1.7	N	170	207	280904	0.6	N
39	76	48881	0.1	N	171	208	67803	0.1	N
40	77	138661	0.3	N	172	209	72427	0.1	N
42	79	51896	0.1	N	173	210	21163	0.0	N
46	83	224200	0.4	N	174	211	39234	0.1	N
47	84	8625	0.0	N	184	221	31517	0.1	N
48	85	23946	0.0	N	185	222	22106	0.0	N
52	89	70355	0.1	N	186	223	981629	2.0	N
53	90	37929	0.1	N	187	224	197742	0.4	N
54	91	85257	0.2	N	188	225	93068	0.2	N
55	92	8525	0.0	N	198	235	27023	0.1	N
56	93	48183	0.1	N	200	237	230447	0.5	N
58	95	33863	0.1	N	201	238	36266	0.1	N
66	103	66844	0.1	N	202	239	34972	0.1	N
68	105	166119	0.3	N	214	251	30649	0.1	N
70	107	68461	0.1	N	220	257	59470	0.1	N
72	109	81445	0.2	N	228	265	129029	0.3	N
78	115	34958	0.1	N	229	266	34753	0.1	N
80	117	63110	0.1	N	230	267	868602	1.7	N
81	118	21077	0.0	N	231	268	178518	0.4	N
82	119	151229	0.3	N	232	269	67473	0.1	N
84	121	38467	0.1	N	235	272	43635	0.1	N
86	123	49800	0.1	N	236	273	137953	0.3	N
89	126	59024	0.1	N	242	279	70790	0.1	N
94	131	167022	0.3	N	243	280	24828	0.0	N
96	133	394347	0.8	N	244	281	1191808	2.4	N
97	134	34341	0.1	N	245	282	341789	0.7	N
98	135	137632	0.3	N	246	283	253821	0.5	N
99	136	18533	0.0	N	247	284	39330	0.1	N
100	137	656253	1.3	N	260	297	43713	0.1	N
101	138	75214	0.2	N	261	298	110320	0.2	N
102	139	6338	0.0	N	262	299	13085	0.0	N
106	143	36545	0.1	N	274	311	937037	1.9	N
108	145	16195	0.0	N	275	312	293136	0.6	N
110	147	637840	1.3	N	276	313	141543	0.3	N
111	148	99744	0.2	N	302	339	25231	0.1	N
112	149	183105	0.4	N	304	341	954013	1.9	N
113	150	49779	0.1	N	306	343	18011	0.0	N
114	151	66968	0.1	N	316	353	77660	0.2	N
					317	354	52200	0.1	N
					318	355	3986136	8.0	N
					319	356	1309017	2.6	N
					320	357	659069	1.3	N
					321	358	111584	0.2	N
					322	359	15750	0.0	N
					332	369	172968	0.3	N
					333	370	11099967	22.2	N
					334	371	3512611	7.0	N
					335	372	1749695	3.5	N
					336	373	387925	0.8	N
					337	374	51273	0.1	N

711756 Scan 574 RT=13:28

7t1756 Scan 574 RT 13:28 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
7	44	319202	0.7	N	137	174	35141	0.1	N
8	45	664865	1.4	N	139	176	24427	0.1	N
10	47	93251	0.2	N	141	178	88155	0.2	N
13	50	10847	0.0	N	142	179	167585	0.4	N
14	51	33895	0.1	N	143	180	27183	0.1	N
20	57	8569	0.0	N	144	181	20692	0.0	N
21	58	50685	0.1	N	151	188	24298	0.1	N
22	59	65189	0.1	N	152	189	18592	0.0	N
24	61	18210	0.0	N	153	190	203278	0.4	N
28	65	30505	0.1	N	154	191	66558	0.1	N
32	69	59742	0.1	N	156	193	63949	0.1	N
35	72	86579	0.2	N	165	202	23102	0.0	N
36	73	5684847	12.0	N	167	204	104400	0.2	N
37	74	501368	1.1	N	179	216	52008	0.1	N
38	75	967705	2.0	N	181	218	57640	0.1	N
39	76	250919	0.5	N	183	220	56968	0.1	N
40	77	257755	0.5	N	190	227	70403	0.1	N
41	78	34500	0.1	N	192	229	33669	0.1	N
53	90	38051	0.1	N	193	230	116546	0.2	N
54	91	103717	0.2	N	194	231	60745	0.1	N
55	92	34074	0.1	N	195	232	75787	0.2	N
56	93	80960	0.2	N	197	234	24665	0.1	N
65	102	311798	0.7	N	207	244	41685	0.1	N
66	103	152261	0.3	N	208	245	125682	0.3	N
67	104	98452	0.2	N	209	246	275013	0.6	N
68	105	165365	0.4	N	210	247	97354	0.2	N
78	115	46713	0.1	N	211	248	35136	0.1	N
79	116	13522	0.0	N	223	260	95224	0.2	N
80	117	63438	0.1	N	224	261	273174	0.6	N
81	118	37025	0.1	N	225	262	8044406	17.0	N
82	119	176312	0.4	N	226	263	1678908	3.6	N
83	120	61322	0.1	N	227	264	479550	1.0	N
84	121	23945	0.1	N	228	265	74077	0.2	N
90	127	56199	0.1	N	252	289	77679	0.2	N
91	128	39182	0.1	N	254	291	11170	0.0	N
92	129	57073	0.1	N	255	292	15762	0.0	N
93	130	626410	1.3	N	264	301	61398	0.1	N
94	131	134823	0.3	N	266	303	18664	0.0	N
95	132	35096	0.1	N	280	317	31161	0.1	N
96	133	106750	0.2	N	299	336	144632	0.3	N
97	134	58657	0.1	N	300	337	23938	0.1	N
98	135	61205	0.1	N	327	364	317930	0.7	N
105	142	31849	0.1	N	328	365	95892	0.2	N
106	143	77833	0.2	N	329	366	18809	0.0	N
107	144	328692	0.7	N	334	371	200297	0.4	N
108	145	73945	0.2	N	335	372	58371	0.1	N
109	146	69429	0.1	N	342	379	707416	1.5	N
110	147	619094	1.3	N	343	380	190515	0.4	N
111	148	121032	0.3	N	344	381	71360	0.2	N
112	149	86015	0.2	N	354	391	28853	0.1	N
120	157	18407	0.0	N	369	406	200902	0.4	N
121	158	21043	0.0	N	370	407	46330	0.1	N
123	160	52632	0.1	N	371	408	66647	0.1	N
126	163	421564	0.9	N					
127	164	50328	0.1	N					
135	172	548638	1.2	N					
136	173	86839	0.2	N					

7i1756 Scan 619 RT=14:18

Chloro-dihydroxybenzoate

7t1756 Scan 619 RT 14:17 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags
3	40	702515	1.5	N
6	43	108404	0.2	N
7	44	346044	0.7	N
8	45	645156	1.4	N
9	46	15304	0.0	N
10	47	23219	0.0	N
22	59	52457	0.1	N
32	69	22788	0.0	N
35	72	66392	0.1	N
36	73	3570671	7.6	N
37	74	328145	0.7	N
38	75	339414	0.7	N
39	76	23435	0.0	N
40	77	61569	0.1	N
54	91	21387	0.0	N
56	93	73337	0.2	N
58	95	23672	0.1	N
68	105	15315	0.0	N
82	119	33711	0.1	N
94	131	27764	0.1	N
96	133	16005	0.0	N
110	147	119908	0.3	N
126	163	23169	0.0	N
134	171	59457	0.1	N
142	179	15385	0.0	N
149	186	29227	0.1	N
162	199	35186	0.1	N
170	207	19138	0.0	N
190	227	979801	2.1	N
191	228	129571	0.3	N
192	229	385390	0.8	N
193	230	56779	0.1	N
220	257	69197	0.1	N
222	259	63614	0.1	N
238	275	47393	0.1	N
239	276	490020	1.0	N
240	277	125190	0.3	N
241	278	21680	0.0	N
264	301	112543	0.2	N
265	302	15700	0.0	N
266	303	24510	0.1	N
278	315	54749	0.1	N
308	345	40612	0.1	N
352	389	392256	0.8	N
353	390	123446	0.3	N
354	391	166017	0.4	N
355	392	34837	0.1	N
356	393	57437	0.1	N
367	404	792110	1.7	N
368	405	324412	0.7	N
369	406	395252	0.8	N
370	407	116523	0.2	N
371	408	46628	0.1	N

APPENDIX 2D: GC-MS Analysis of Medium of 3-chlorobenzoate-grown *Alcaligenes* sp. strain BR6024 containing pBRCN12.

Mass spectra only of peaks from gas chromatogram relevant to 3-chlorobenzoate degradation are included.

Runname : 711755 Acquired Nominal data (Zero baseline).
Unsmoothed data only:

711755 Scan 259 RT=7.43

7c1755 Scan 259 RT 7:42 +EI LRP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
1	39	27630	0.1	N	141	178	289679	0.6	N
2	39	8656	0.0	N	142	179	39827	0.1	N
3	40	805213	1.7	N	156	193	46283	0.1	N
6	43	261018	0.6	N	160	197	32526	0.1	N
7	44	259799	0.5	N	162	199	55976	0.1	N
8	45	526877	1.1	N	175	212	159204	0.3	N
10	47	147887	0.3	N	176	213	20695531	43.8	NS
13	50	282232	0.6	N	177	214	2853005	6.0	N
14	51	97698	0.2	N	178	215	7865533	16.7	N
21	58	23986	0.1	N	179	216	1060186	2.2	N
22	59	146444	0.3	N	180	217	266096	0.6	N
23	60	10883	0.0	N	190	227	92643	0.2	N
24	61	63414	0.1	N	191	228	2183202	4.6	N
26	63	71018	0.2	N	192	229	129658	0.7	N
32	69	43390	0.1	N	193	230	866890	1.8	N
35	72	52705	0.1	N	194	231	126815	0.3	N
36	73	485356	1.0	N					
37	74	288649	0.6	N					
38	75	1525872	3.2	N					
39	76	539462	1.1	N					
40	77	170747	0.4	N					
47	84	8731	0.0	N					
48	85	113205	0.2	N					
50	87	36954	0.1	N					
54	91	121748	0.3	N					
56	93	60735	0.1	N					
58	95	25016	0.1	N					
66	103	29956	0.1	N					
67	104	225313	0.5	N					
68	105	40462	0.1	N					
69	106	452956	1.0	N					
70	107	237841	0.5	N					
73	110	100053	0.2	N					
74	111	3664781	7.8	N					
75	112	265150	0.6	N					
76	113	1190442	2.5	N					
77	114	63029	0.1	N					
82	119	69799	0.1	N					
96	133	16014	0.0	N					
97	134	236562	0.5	N					
98	135	86212	0.2	N					
99	136	22433	0.0	N					
101	138	44652	0.1	N					
102	139	6542303	13.8	N					
103	140	461751	1.0	N					
104	141	2173057	4.6	N					
105	142	153252	0.3	N					
116	153	45225	0.1	N					
118	155	108886	0.2	N					
119	156	66524	0.1	N					
120	157	15613	0.0	N					
126	163	18509	0.0	N					
132	169	3965212	8.4	N					
133	170	512126	1.1	N					
134	171	1301853	2.8	N					
135	172	163213	0.3	N					
136	173	37258	0.1	N					

APPENDIX 2E: GC-MS Analysis of Medium of 3-chlorobenzoate-grown, IPTG induced *E. coli* strain containing pBRCN13a.

Mass spectra only of peaks from gas chromatogram relevant to 3-chlorobenzoate degradation are included.

Runname : 711735 Acquired Nominal data (Zero baseline).
Unsmoothed data only:

711735 Scan 258 RT=7.42

751735 Scan 253 RT 7.42 +EI LRP

Normalised to largest peak in current list

Peak	Mass	Intensity	%	Flags
3	40	459763	8.6	N
6	43	94925	1.7	N
7	44	164419	2.9	N
8	45	193169	3.4	N
10	47	78836	1.4	N
13	50	79339	1.4	N
14	51	29777	0.5	N
22	59	47413	0.8	N
32	69	1239768	21.8	N
36	73	184812	3.3	N
37	74	83355	1.5	N
38	75	491179	8.6	N
39	76	149409	2.6	N
40	77	86899	1.5	N
48	85	45923	0.9	N
54	91	23651	0.4	N
56	93	25737	0.5	N
63	100	80117	1.4	N
67	104	51433	0.9	N
69	106	97425	1.7	N
74	111	1080387	19.0	N
75	112	82381	1.5	N
76	113	360079	6.3	N
82	119	412097	7.3	N
94	131	382848	6.7	N
97	134	13030	0.2	N
98	135	11024	0.2	N
102	139	1751137	30.8	N
103	140	114678	2.0	N
104	141	600467	10.6	N
105	142	40131	0.7	N
118	155	20937	0.4	N
132	169	1472282	25.9	N
133	170	132144	2.3	N
134	171	417444	7.3	N
135	172	55576	1.0	N
141	178	66433	1.2	N
144	181	219938	3.9	N
156	193	24076	0.4	N
175	212	18767	0.3	N
176	213	5680735	100.0	N
177	214	777213	13.7	N
178	215	1969828	34.7	N
179	216	284650	5.0	N
180	217	69181	1.2	N
182	219	255134	4.5	N
191	228	606490	10.7	N
192	229	100428	1.8	N
193	230	234590	4.1	N
194	231	197590	3.3	N
206	243	13315	0.2	N

701735 Scan 62.5 RT=14:24

7c1735 Scan 625 RT 14:24 -EI LFP

Normalised to largest peak in current list

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
3	40	497223	2.9	N	136	173	186188	1.1	N
6	43	157716	0.9	N	140	177	86671	0.5	N
7	44	322898	1.9	N	141	178	18142	0.1	N
8	45	1029058	6.0	N	142	179	166208	1.0	N
9	46	47041	0.3	N	144	181	276975	1.6	N
10	47	80872	0.5	N	146	183	47122	0.3	N
20	57	34161	0.2	N	148	185	40385	0.2	N
21	58	61104	0.4	N	154	191	95539	0.6	N
22	59	105907	0.6	N	155	192	24021	0.1	N
32	69	1259593	7.3	N	156	193	220178	1.3	N
34	71	13479	0.1	N	158	195	38420	0.2	N
35	72	184355	1.1	N	160	197	126203	0.7	N
36	73	9353955	54.3	N	161	199	62827	0.4	N
37	74	748393	4.3	N	162	199	334363	1.9	N
38	75	540282	3.1	N	163	200	55754	0.3	N
39	76	15223	0.1	N	164	201	113957	0.7	N
40	77	222812	1.3	N	168	205	24028	0.1	N
42	79	42255	0.2	N	170	207	125243	0.7	N
44	81	25671	0.1	N	171	208	18555	0.1	N
46	83	47636	0.3	N	172	209	62724	0.4	N
52	89	23765	0.1	N	174	211	42229	0.2	N
54	91	70906	0.4	N	175	212	41994	0.2	N
56	93	291416	1.7	N	176	213	115723	0.7	N
58	95	106173	0.6	N	178	215	70098	0.4	N
63	100	81891	0.5	N	182	219	262734	1.5	N
66	103	67035	0.4	N	184	221	61224	0.4	N
68	105	135004	0.8	N	185	222	18296	0.1	N
70	107	24000	0.1	N	186	223	29356	0.2	N
78	115	10819	0.1	N	188	225	69228	0.4	N
80	117	148614	0.9	N	189	226	99850	0.6	N
82	119	469334	2.7	N	190	227	11259945	65.3	N
94	131	502005	2.9	N	191	228	1543391	9.0	N
95	132	54649	0.3	N	192	229	4101533	23.8	N
96	133	206825	1.2	N	193	230	562339	3.3	N
97	134	23255	0.1	N	194	231	353829	2.1	N
98	135	102268	0.6	N	195	232	13519	0.1	N
100	137	56147	0.3	N	198	235	48635	0.3	N
102	139	34291	0.2	N	200	237	230421	1.3	N
104	141	33568	0.2	N	201	238	41056	0.2	N
106	143	105108	0.6	N	202	239	24036	0.1	N
108	145	8416	0.0	N	204	241	346282	2.0	N
110	147	558189	3.2	N	205	242	75247	0.4	N
111	148	72323	0.4	N	206	243	161771	0.9	N
112	149	89056	0.5	N	207	244	6462	0.0	N
114	151	51041	0.3	N	208	245	49241	0.3	N
116	153	123768	0.7	N	218	255	35596	0.2	N
118	155	73892	0.4	N	219	256	17764	0.1	N
120	157	20487	0.1	N	220	257	930761	5.4	N
124	161	64587	0.4	N	221	258	214297	1.2	N
125	162	24086	0.1	N	222	259	359426	2.1	N
126	163	79464	0.5	N	223	260	62451	0.4	N
127	164	40459	0.2	N	228	265	82998	0.5	N
128	165	23956	0.1	N	230	267	29523	0.2	N
130	167	100223	0.6	N	232	269	168905	1.0	N
132	169	451302	2.6	N	234	271	171439	1.0	N
134	171	458149	2.9	N	235	272	35109	0.2	N
135	172	50768	0.3	N	236	273	81414	0.5	N

7:1735 Scan 625 RT 14:24 +EI LRP

Normalised to largest peak in current list

Peak	Mass	Intensity	Δ	Flags
242	279	107613	0.6	N
243	280	50749	0.3	N
244	281	130709	0.8	N
248	285	39535	0.2	N
250	287	34295	0.2	N
262	299	137161	0.8	N
263	300	27493	0.2	N
264	301	1417307	8.2	N
265	302	337442	2.0	N
266	303	594545	3.5	N
267	304	99737	0.6	N
268	305	83583	0.5	N
269	306	189350	1.1	N
270	307	113525	0.7	N
272	309	80813	0.5	N
276	313	66857	0.4	N
278	315	898091	5.2	N
279	316	317509	1.8	N
280	317	381772	2.2	N
281	318	121392	0.7	N
282	319	132560	0.8	N
294	331	160059	0.9	N
295	332	13390	0.1	N
296	333	23612	0.1	N
308	345	755284	4.4	N
309	346	223794	1.3	N
310	347	339234	2.0	N
311	348	87300	0.5	N
312	349	15643	0.1	N
316	353	47753	0.3	N
332	369	60667	0.4	N
336	373	87468	0.5	N
338	375	2118083	12.3	N
340	377	1216315	7.1	N
342	379	35982	0.2	N
344	381	97244	0.6	N
346	383	24388	0.1	N
349	386	12874	0.1	N
350	387	152339	0.9	N
351	388	153937	0.9	N
352	389	7742176	44.9	N
353	390	2371227	13.8	N
354	391	3601418	20.9	N
355	392	1023008	5.9	N
356	393	442370	2.6	N
357	394	119142	0.7	N
358	395	25942	0.2	N
366	403	305195	1.8	N
367	404	17232984	100.0	N
368	405	5507826	32.0	N
369	406	8039815	46.7	N
370	407	2245534	13.0	N
371	408	954355	5.5	N
372	409	184740	1.1	N
373	410	24528	0.1	N
394	431	26682	0.2	N

711735 Scan 465 RT=11:29

7t1735 Scan 465 RT 11:28 +EI LPP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags	Peak	Mass	Intensity	%	Flags
2	39	11041	0.0	N	117	154	20732	0.0	N
3	40	463714	1.0	N	118	155	68143	0.1	N
6	43	190908	0.4	N	119	156	83223	0.2	N
7	44	224583	0.5	N	120	157	67242	0.1	N
8	45	554301	1.2	N	126	163	20731	0.0	N
9	46	26130	0.1	N	127	164	29886	0.1	N
10	47	105551	0.2	N	130	167	98592	0.2	N
20	57	26719	0.1	N	132	169	594929	1.2	N
21	58	39568	0.1	N	133	170	24110	0.1	N
22	59	114930	0.2	N	134	171	65909	0.1	N
26	63	373903	0.8	N	142	179	28570	0.1	N
32	69	1269803	2.7	N	144	181	254039	0.5	N
35	72	56541	0.1	N	145	182	18904	0.0	N
36	73	2947829	6.2	N	146	183	136426	0.3	N
37	74	277678	0.6	N	147	184	37123	0.1	N
38	75	455787	1.0	N	148	185	85678	0.2	N
39	76	53334	0.1	N	156	193	34151	0.1	N
40	77	163032	0.3	N	160	197	24692	0.1	N
42	79	10706	0.0	N	162	199	47030	0.1	N
46	83	40998	0.1	N	174	211	142557	0.3	N
49	86	35872	0.1	N	175	212	35173	0.1	N
52	89	57685	0.1	N	176	213	68381	0.1	N
54	91	594383	1.2	N	182	219	503783	1.1	N
55	92	30440	0.1	N	184	221	34148	0.1	N
56	93	273771	0.6	N	188	225	44082	0.1	N
58	95	80211	0.2	N	190	227	1052487	2.2	N
60	97	17587	0.0	N	191	228	236337	0.5	N
61	98	21181	0.0	N	192	229	411555	0.9	N
62	99	28909	0.1	N	193	230	72511	0.2	N
63	100	111883	0.2	N	194	231	160883	0.3	N
66	103	80516	0.2	N	206	243	63422	0.1	N
68	105	84048	0.2	N	218	255	16020	0.0	N
69	106	38783	0.1	N	220	257	701371	1.5	N
73	110	69626	0.1	N	221	258	146818	0.3	N
74	111	47269	0.1	N	222	259	258865	0.5	N
75	112	6444	0.0	N	223	260	48211	0.1	N
76	113	39096	0.1	N	232	269	170511	0.4	N
78	115	22984	0.0	N	244	281	141862	0.3	N
80	117	33855	0.1	N	248	285	24098	0.1	N
82	119	451923	0.9	N	249	286	242642	0.5	N
84	121	32761	0.1	N	250	287	17066	0.0	N
88	125	63295	0.1	N	251	288	32117	0.1	N
89	126	109058	0.2	N	262	299	32594	0.1	N
90	127	99133	0.2	N	264	301	5868459	12.3	N
91	128	46482	0.1	N	265	302	1359668	2.8	N
92	129	49140	0.1	N	266	303	2418579	5.1	N
94	131	427366	0.9	N	267	304	535935	1.1	N
96	133	104997	0.2	N	268	305	183311	0.4	N
98	135	50646	0.1	N	278	315	6440	0.0	N
100	137	30544	0.1	N	279	316	2765360	5.8	N
106	143	776139	1.6	N	280	317	681425	1.4	N
107	144	314105	0.7	N	281	318	1118163	2.3	N
108	145	17773	0.0	N	282	319	363100	0.8	N
110	147	385206	0.8	N	283	320	98378	0.2	N
111	148	67982	0.1	N	294	331	59422	0.1	N
112	149	83865	0.2	N					
116	153	59838	0.1	N					

7h1735 Scan 542 RT=12:53

7:1735 Scan 542 RT 12:53 +E1 LPP

Normalised to largest non-reference peak in scan

Peak	Mass	Intensity	%	Flags
3	40	447015	0.9	N
6	43	44107	0.1	N
7	44	163073	0.3	N
8	45	254113	0.5	N
22	59	18367	0.0	N
32	69	1240197	2.6	N
36	73	1487134	3.1	N
37	74	112657	0.2	N
38	75	119777	0.3	N
40	77	65979	0.1	N
63	100	83621	0.2	N
82	119	413975	0.9	N
94	131	380718	0.8	N
96	133	34166	0.1	N
100	137	62997	0.1	N
110	147	64502	0.1	N
128	165	65785	0.1	N
132	169	390600	0.8	N
144	181	191536	0.4	N
156	193	953238	2.0	N
157	194	123065	0.3	N
158	195	51948	0.1	N
182	219	237534	0.5	N
186	223	71615	0.2	N
194	231	131278	0.3	N
206	243	18300	0.0	N
230	267	98258	0.2	N
232	269	139154	0.3	N
244	281	180327	0.4	N
245	282	37488	0.1	N
274	311	77815	0.2	N
282	319	99363	0.2	N
294	331	56437	0.1	N
318	355	341780	0.7	N
319	356	95852	0.2	N
320	357	42360	0.1	N
332	369	45781	0.1	N
333	370	955639	2.0	N
334	371	333691	0.7	N
335	372	140560	0.3	N
336	373	15819	0.0	N
344	381	39610	0.1	N

REFERENCES

- Aldrich, T. A., B. Frantz, J. F. Gill, J. J. Kilbane and A. M. Chakrabarty. 1987. Cloning and complete nucleotide sequence determination of the *catB* gene encoding *cis,cis*-muconate lactonizing enzyme. *Gene* 52: 185-195.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith and K. Struhl (eds.). 1989. **Short protocols in molecular biology**. John Wiley & Sons. New York.
- Axcell, B. A. and P. J. Geary. 1975. Purification and some properties of a soluble benzene-oxidizing system from a strain of *Pseudomonas*. *Biochem. J.* 146: 173-183.
- Batie, C. J., E. La Haie and D. P. Ballou. 1987. Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from *Pseudomonas cepacia*. *J. Biol. Chem.* 262: 1510-1518.
- Bayley, S. A., C. J. Duggleby, M. J. Worsey, P. A. Williams, K. G. Hardy, and P. Broda. 1977. Two modes of loss of the TOL function from *Pseudomonas putida* mt-2. *Mol. Gen. Genet.* 154: 203-204.
- Bernhardt, F.-H., E. Heymann and P. S. Taylor. 1978. Chemical and spectral properties of putidamonooxin, the iron-containing and acid-labile sulfur-containing monooxygenase of a 4-methoxybenzoate O-demethylase from *Pseudomonas putida*. *Eur. J. Biochem.* 92: 209-223.
- Birnboim, H. C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7: 1513-1523.
- Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* 72: 248-250.
- Brunel, F. and J. Davison. 1988. Cloning and sequencing of *Pseudomonas* genes encoding vanillate demethylase. *J. Bacteriol.* 170: 4924-4930.
- Burlage, R. S., L. A. Bemis, A. C. Layton, G. S. Sayler and F. Larimer. 1990. Comparative genetic organization of

- incompatibility group P degradative plasmids. *J. Bacteriol.* 172: 6818-6825.
- Busse, H. J., T. Elbanna, H. Oyaizu, G. Auling. 1992. Identification of xenobiotic-degrading isolates from the beta subclass of the proteobacteria by a polyphasic approach including 16S rRNA partial sequencing. *Int. J. Syst. Bacteriol.* 42: 19-26.
- Chakrabarty, A. M., D. A. Friello, and L. H. Bopp. 1978. Transposition of plasmid DNA segments specifying hydrocarbon degradation and their expression in various microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 75: 3109-3112.
- Chakrabarty, A. M. and I. C. Gunsalus. 1971. Degradative pathways specified by extrachromosomal gene clusters in *Pseudomonas*. *Genetics* 68: S10.
- Chan, T.-M., M. A. Hermodson, E. L. Ulrich and J. L. Markley. 1983. Nuclear magnetic resonance studies of two-iron-two-sulfur ferredoxins. 2. Determination of the sequence of *Anabaena variabilis* ferredoxin II, assignment of aromatic resonances in proton spectra, and effects of chemical modifications. *Biochem.* 22: 5988-5995.
- Chapman, P. J. 1987. Constructing microbial strains for degradation of halogenated aromatic hydrocarbons. in: *Environmental Biotechnology*. G. S. Omenn (ed.). Plenum Press. New York. pp. 81-95.
- Chatfield, L. K. and Williams, P. A. 1986. Naturally occurring TOL plasmids in *Pseudomonas* strains carry either two homologous or two non-homologous catechol 2,3-dioxygenase genes. *J. Bacteriol.* 168: 878-885.
- Chatterjee, D. K., S. T. Kellogg, S. Hamada and A. M. Chakrabarty. 1981. A plasmid specifying total degradation of 3-chlorobenzoate by a modified ortho pathway. *J. Bacteriol.* 146: 639-646.
- Chatterjee, D. K. and A. M. Chakrabarty. 1982. Genetic rearrangement in plasmids specifying total degradation of chlorinated benzoic acids. *Mol. Gen. Genet.* 188: 279-285.
- Chatterjee, D. K. and A. M. Chakrabarty. 1983a. Genetic homology between independently isolated chlorobenzoate-degradative plasmids. *J. Bacteriol.* 153: 532-534.

- Chatterjee, D. K. and A. M. Chakrabarty. 1983b. Restriction mapping of chlorobenzoate degradative plasmid and molecular cloning of degradative genes. *Gene* 27: 173-181.
- Chatterjee, D. K., S. T. Kellogg, S. Hamada and A. M. Chakrabarty. 1981. A plasmid specifying total degradation of 3-chlorobenzoate by a modified *ortho* pathway. *J. Bacteriol.* 146: 639-646.
- Crutcher, S. E. and P. J. Geary. 1979. Properties of the iron-sulphur proteins of the benzene dioxygenase system from *Pseudomonas putida*. *Biochem. J.* 177: 393-400.
- Dagley, S. 1971. Catabolism of aromatic compounds by microorganisms. *Adv. Microbial Physiol.* 6: 1-46.
- Dagley, S. 1986. Biochemistry of aromatic hydrocarbon degradation by pseudomonads. in: *the Bacteria* vol. X. J. R. Sokatch (ed.). Academic Press. New York. pp. 527-555.
- Dagley, S., W. C. Evans and D. W. Ribbons. 1960. New pathways in the oxidative metabolism of aromatic compounds by micro-organisms. *Nature (London)* 188: 560-566.
- Dale, R. M. K., B. A. McClure and J. P. Houchins. 1985. A rapid single stranded cloning strategy for producing sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. *Plasmid* 13: 31-46.
- De Ley, J. 1992. The proteobacteria: ribosomal RNA cistron similarities and bacterial taxonomy. in: A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer. *The Prokaryotes*. Springer-Verlag. New York. pp. 2111-2140.
- Deretic, V., W. M. Konyecsni, C. D. Mohr, D. W. Martin and N. S. Hibler. 1989. Common denominators of promoter control in *Pseudomonas* and other bacteria. *Biotech.* 7: 1249-1254.
- Don, R. H. and J. M. Pemberton. 1981. Properties of six pesticide degradation plasmids isolated from *Alcaligenes eutrophus* and *Alcaligenes paradoxus*. *J. Bacteriol.* 145: 681-686.
- Don, R. H. and J. M. Pemberton. 1985. Genetic and physical map of the 2,4-dichlorophenoxyacetic acid-degradative plasmid pJP4. *J. Bacteriol.* 161: 466-468.

- Don, R. H., A. J. Weightman, H.-J. Knackmuss and K. N. Timmis. 1985. Transposon mutagenesis and cloning analysis of the pathways for degradation of 2,4-dichlorophenoxyacetic acid and 3-chlorobenzoate in *Alcaligenes eutrophus* JMP134 (pJP4). *J. Bacteriol.* 161: 85-90.
- Dorn, E., M. Hellwig, W. Reineke and H.-J. Knackmuss. 1974. Isolation and characterization of a 3-chlorobenzoate degrading pseudomonad. *Arch. Microbiol.* 99: 61-70.
- Dorn, E. and H.-J. Knackmuss. 1978a. Chemical structure and biodegradability of halogenated aromatic compounds. Two catechol 1,2-dioxygenases from a 3-chlorobenzoate-grown pseudomonad. *Biochem. J.* 174: 73-84.
- Dorn, E. and H.-J. Knackmuss. 1978b. Chemical structure and biodegradability of halogenated aromatic compounds. *Biochem. J.* 174: 85-94.
- Doten, R. C., D. J. Mitchell, K.-L. Ngai and L. N. Ornston. 1987. Cloning and genetic organization of the *pca* gene cluster from *Acinetobacter calcoaceticus*. *J. Bacteriol.* 169: 3168-3174.
- Engesser, K. H. and P. Schulte. 1989. Degradation of 2-bromo-, 2-chloro- and 2-fluorobenzoate by *Pseudomonas putida* CLB 250. *FEMS Microbiol. Lett.* 60: 143-148.
- Ensley, B. D. and D. T. Gibson. 1983. Naphthalene dioxygenase: purification and properties of a terminal oxygenase component. *J. Bacteriol.* 155: 501-511.
- Erickson, B.D. and Mondello, F.J. 1992. Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl dioxygenase, a multicomponent polychlorinated-biphenyl-degrading enzyme in *Pseudomonas* strain LB400. *J. Bacteriol.* 174: 2903-2912.
- Evans, W. C. and G. Fuchs. 1988. Anaerobic degradation of aromatic compounds. *Annu. Rev. Microbiol.* 42: 289-317.
- Fakhrai, H. and M. D. Maines. 1992. Expression and characterization of cDNA from rat kidney biliverdin reductase. *J. Biol. Chem.* 267: 4023-4029.
- Fetzner, S., R. Müller and F. Lingens. 1992. Purification and some properties of 2-halobenzoate 1,2-dioxygenase, a two component enzyme system from *Pseudomonas cepacia* 2CBS. *J. Bacteriol.* 174: 279-290.

- Figurski, K. H. and D. W. Helinski. 1979. Replication of an origin containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci. U.S.A.* 76: 1648-1652.
- Focht, D. D. and D. Shelton. 1987. Growth kinetics of *Pseudomonas alcaligenes* C-0 relative to inoculation and 3-chlorobenzoate metabolism in soil. *Appl. Environ. Microbiol.* 53: 1846-1849.
- Frantz, B. and A. M. Chakrabarty. 1986. Degradative plasmids in *Pseudomonas*. in: *The Bacteria*. vol. 10. J. R. Sokatch (ed.). Academic Press. New York. pp. 295-323.
- Frantz, B. and A. M. Chakrabarty. 1987. Organization and nucleotide sequence determination of a gene cluster involved in 3-chlorocatechol degradation. *Proc. Natl. Acad. Sci. U.S.A.* 84: 4460-4464.
- Frantz, B., K.-L. Ngai, D. K. Chatterjee, L. N. Ornston and A. M. Chakrabarty. 1987. Nucleotide sequence and expression of *clcD*, a plasmid-borne diene lactone hydrolase gene from *Pseudomonas* sp. strain B13. *J. Bacteriol.* 169: 704-709.
- Fulthorpe, R. R. and R. C. Wyndham. 1989. Survival and activity of a 3-chlorobenzoate-catabolic genotype in a natural system. *Appl. Environ. Microbiol.* 55: 1584-1590.
- Fulthorpe, R. R. and R. C. Wyndham. 1991. Transfer and expression of the catabolic plasmid pBRC60 in wild bacterial recipients in a freshwater ecosystem. *Appl. Environ. Microbiol.* 57: 1546-1553.
- Fulthorpe, R. R. and Wyndham, R. C. 1992. Involvement of a chlorobenzoate catabolic transposon, Tn5271, in community adaptation to chlorobiphenyl, chloroaniline and 2,4-dichlorophenoxyacetic acid in a freshwater ecosystem. *Appl. Environ. Microbiol.* 58: 314-325.
- Fürste, J. P., W. Pansegrau, R. Frank, H. Blöcker, P. Scholz, M. Bagdasarian and E. Lanka. 1986. Molecular cloning of the plasmid RP4 primase region in a multi-host range tac P expression vector. *Gene* 48: 119-131.
- Furukawa, K., N. Hayase, T. Kazunari and N. Tomizuka. 1989. Molecular relationship of chromosomal genes encoding biphenyl/polychlorinated biphenyl catabolism: some soil bacteria possess a highly conserved *bph* operon. *J. Bacteriol.* 171: 5467-5472.

- Gendel, S., N. Straus, D. Pulleyblank and J. Williams. 1983. Shuttle cloning vector for the cyanobacterium *Anacystis nidulans*. *J. Bacteriol.* 156: 148-154.
- Ghosal, D. and I.-S. You. 1988. Nucleotide homology and organization of chlorocatechol oxidation genes of plasmids pJP4 and pAC27. *Mol. Gen. Genet.* 211: 113-120.
- Ghosal, D. and I.-S. You. 1989. Operon structure and nucleotide homology of the chlorocatechol oxidation genes of plasmids pJP4 and pAC27. *Gene* 83: 225-232.
- Ghosal, D., I.-S. You, D. K. Chatterjee and A. M. Chakrabarty. 1985. Genes specifying degradation of 3-chlorobenzoic acid in plasmids pAC27 and pJP4. *Proc. Natl. Acad. Sci. USA.* 82: 1638-1642.
- Gibson, D. T., M. Hensley, H. Yoshioka and T. J. Mabry. 1970. Formation of (+)-cis-2,3-dihydroxy-1-methylcyclohexa-4,6-diene from toluene by *Pseudomonas putida*. *Biochem.* 9: 1626-1630.
- Gibson, D. T. and V. Subramanian. 1984. Microbial degradation of aromatic hydrocarbons. in: *Microbial degradation of organic compounds*. D. T. Gibson (ed.). Marcel Dekker Inc. New York. pp. 181-252.
- Haigler, B. E. and D. T. Gibson. 1990a. Purification and properties of NADH-ferredoxin_{NAP} reductase, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. *J. Bacteriol.* 172: 457-464.
- Haigler, B. B. and D. T. Gibson. 1990b. Purification and properties of NADH-ferredoxin_{NAP}, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. *J. Bacteriol.* 172: 465-468.
- Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166: 557.
- Hanahan, D. 1984. Techniques for transformation of *Escherichia coli*. in: *DNA Cloning: A Practical Approach*, Vol. 1. D. M. Glover (ed.). IRL Press. Washington D.C. p. 115.
- Harayama, S., M. Rekik and K. N. Timmis. 1986. Genetic analysis of a relaxed substrate specificity aromatic ring dioxygenase, toluate 1,2-dioxygenase, encoded by TOL plasmid pWWO of *Pseudomonas putida*. *Mol. Gen. Genet.* 202: 226-234.

- Harayama, S., M. Rekik, A. Bairoch, E. L. Neidle, E. L. and L. N. Ornston. 1991. Potential DNA slippage structures acquired during evolutionary divergence of *Acinetobacter calcoaceticus* chromosomal *benABC* and *Pseudomonas putida* TOL *pWW0* plasmid *xylXYZ* genes encoding benzoate dioxygenases. *J. Bacteriol.* 173: 7540-7548.
- Hase, T., S. Wakabayashi, H. Matsubara, K. K. Rao, D. O. Hall, H. Widmer, J. Gysi and H. Zuber. 1978. submitted to the Atlas, accession number A00252. NCBI databanks.
- Haugland, R. A., U. M. X. Sangodkar and A. M. Chakrabarty. 1990. Repeated sequences including RS1100 from *Pseudomonas cepacia* AC1100 function as IS elements. *Mol. Gen. Genet.* 220: 222-228.
- Hawley, D. K. and W. R. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* 8: 2237-2255.
- Hedges, R. W. and A. F. Jacob. 1974. Transposition of ampicillin resistance from RP4 to other replicons. *Mol. Gen. Genet.* 132: 31-40.
- Hernandez, B. S., F. K. Higson, R. Kondrat and D. D. Focht. 1991. Metabolism of and inhibition by chlorobenzoates in *Pseudomonas putida* P111. *Appl. Environ. Microbiol.* 57: 3361-3366.
- Hickey, W. J. and D. D. Focht. 1990. Degradation of mono-, di- and tri-halogenated benzoic acids by *Pseudomonas aeruginosa* JB2. *Appl. Environ. Microbiol.* 56: 3842-3850.
- Higson, F. K. and D. D. Focht. 1992. Degradation of 2-methylbenzoic acid by *Pseudomonas cepacia* MB2. *Appl. Environ. Microbiol.* 58: 2501-2504.
- Ichikawa, H., K. Ikeda, J. Amemura and E. Ohtsubo. 1990. Two domains in the terminal inverted-repeat sequence of transposon Tn3. *Gene* 86: 11-17.
- Irie, S., S. Doi, T. Yorifugi, M. Takagi and K. Yano. 1987. Nucleotide sequencing and characterization of the genes encoding benzene oxidation enzymes of *Pseudomonas putida*. *J. Bacteriol.* 169: 5174-5179.
- Jacoby, G. A., J. E. Rogers, A. F. Jacob and R. W. Hedges. 1978. Transposition of *Pseudomonas* toluene-degrading genes and expression in *Escherichia coli*. *Nature (London)*. 274: 179-180.

- Johnston, H. W., G. G. Briggs and M. Alexander. 1972. Metabolism of 3-chlorobenzoic acid by a pseudomonad. *Soil Biol. Biochem.* 4: 187-190.
- Kanagasundaram, V. and R. K. Scopes. 1992. Cloning and sequence analysis and expression of the structural gene encoding glucose-fructose oxidoreductase from *Zymomonas mobilis*. *J. Bacteriol.* 174: 1439-1447.
- Kersten, P. J., P. J. Chapman and S. Dagley. 1985. Enzymatic release of halogens or methanol from some substituted protocatechuic acids. *J. Bacteriol.* 162: 693-697.
- Kleckner, N. 1981. Transposable elements in prokaryotes. *Annu. Rev. Genet.* 15: 341-404.
- Klingler, J. M., R. P. Stow, D. C. Obenhuber, T. O. Groves, S. K. Mishra and D. L. Pierson. 1992. Evaluation of the Biolog automated microbial identification system. *Appl. Environ. Microbiol.* 58: 2089-2092.
- Krasnow, M. A. and N. R. Cozzarelli. 1983. Site specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented sites. *Cell* 32: 1313-1324.
- Kreig, N. R. and J. G. Holt (eds.). 1984. *Bergey's Manual of Systematic Bacteriology*. Williams and Wilkins. Baltimore, London. pp.361-373.
- Kreig, P. and D. Melton. 1988. Quick large scale plasmid preparation. *Promega Biotec*, Technical bulletin no. 009.
- Kukor, J. J. and R. H. Olsen. 1990. Molecular cloning, characterization, and regulation of a *Pseudomonas pickettii* PK01 gene encoding phenol hydroxylase and expression of the gene in *Pseudomonas aeruginosa* PA01c. *J. Bacteriol.* 172: 4624-4630.
- Kukor, J. J. and R. H. Olsen. 1992. Complete nucleotide sequence of *tbuD*, the gene encoding phenol/cresol hydroxylase from *Pseudomonas pickettii* PK01, and functional analysis of the encoded enzyme. *J. Bacteriol.* 174: 6518-6526.
- Kurkela, S., H. Lehvaeslaiho, E. T. Palva and T. H. Teeri. 1988. Cloning, nucleotide sequence and characterization of genes encoding naphthalene dioxygenase of *Pseudomonas putida* strain NCIB9816. *Gene* 73: 355-362.

- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* (London) 227: 680.
- Lehrbach, P., J. Zeyer, W. Reineke, H.-J. Knackmuss and K. N. Timmis. 1984. Enzyme recruitment *in vitro*: use of cloned genes to extend the range of haloaromatics degraded by *Pseudomonas* sp. strain B13. *J. Bacteriol.* 158: 1025-1032.
- Locher, H. H., T. Leisinger and A. M. Cook. 1991. 4-Sulphobenzoate 3,4-dioxygenase: purification and properties of a desulphonative two-component enzyme system from *Comamonas testosteroni* T-2. *Biochem. J.* 274: 833-842.
- Mahillon, J. and D. Lereclus. 1988. Structural and functional analysis of Tn4430: identification of an integrase like protein involved in the co-integrate resolution process. *EMBO J.* 7: 1515-1526.
- Markus, A., D. Krekel and F. Lingens. 1986. Purification and some properties of component A of the 4-chlorophenylacetate 3,4-dioxygenase from *Pseudomonas* species strain CBS. *J. Biol. Chem.* 261: 12883-12888.
- McClure, N. C., A. J. Weightman and J. C. Fry. 1989. Survival of *Pseudomonas putida* UWC1 containing cloned catabolic genes in a model activated-sludge unit. *Appl. Environ. Microbiol.* 55: 2627-2634.
- Mercier, J., J. Lachapelle, F. Couture, M. Lafond, G. Vezina, M. Boissinot, and R. C. Levesque. 1990. Structural and functional characterization of *tnpI*, a recombinase locus in Tn21 and related β -lactamase transposons. *J. Bacteriol.* 172: 3745-3757.
- Michiels, T., G. Cornelis, K. Ellis and J. Grinstead. 1987. Tn2501, a component of the lactose transposon Tn951, is an example of a new category of class II transposable elements. *J. Bacteriol.* 169: 624-631.
- Nakatsu, C., J. Ng, R. Singh, N. Straus and C. Wyndham. 1991. Chlorobenzoate catabolic transposon Tn5271 is a composite class I element with flanking class II insertion sequences. *Proc. Natl. Acad. Sci. USA.* 88: 8312-8316.
- Nakatsu, C. H., N. A. Straus and R. C. Wyndham. 1992. The phylogenetic distribution of a transposable dioxygenase (submitted to *Proc. Natl. Acad. Sci. U.S.A.*).

- Nakazawa, T. and E. Hayashi. 1977. Phthalate metabolism in *Pseudomonas testosteroni*: accumulation of 4,5-dihydroxyphthalate by a mutant. *J. Bacteriol.* 131: 42-48.
- Nakazawa, T., E. Hayashi, T. Yokota, Y. Ebina and A. Nakazawa. 1978. Isolation of TOL and RP4 recombinants by integrative suppression. *J. Bacteriol.* 134: 270-277.
- Neidle, E. L., M. K. Shapiro and L. N. Ornston. 1987. Cloning and expression in *Escherichia coli* of *Acinetobacter calcoaceticus* genes for benzoate degradation. *J. Bacteriol.* 169: 5496-5503.
- Neidle, E. L., C. Hartnett, S. Bonitz and L. N. Ornston. 1988. DNA sequence of the *Acinetobacter calcoaceticus* catechol 1,2-dioxygenase I structural gene *catA*: evidence for evolutionary divergence of intradiol dioxygenases by acquisition of DNA sequence repetitions. *J. Bacteriol.* 170: 4874-4880.
- Neidle, E. L., C. Hartnett, L. N. Ornston, A. Bairoch, M. Rekik and S. Harayama. 1991. Nucleotide sequences of the *Acinetobacter calcoaceticus* *benABC* genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. *J. Bacteriol.* 173: 5385-5395.
- Neidle, E. L., C. Harnett, L. N. Ornston, A. Bairoch, M. Rekik and S. Harayama. 1992. *Cis*-diol dehydrogenases encoded by the TOL pWWO plasmid *xyiL* gene and the *Acinetobacter calcoaceticus* chromosomal *benD* gene are members of the short-chain alcohol dehydrogenase superfamily. *Eur. J. Biochem.* 204: 113-120.
- Palleroni, N. J. 1984. Family 1. Pseudomonadaceae. in: *Bergey's Manual of Systematic Bacteriology*, Vol. 1. Krieg, N. R. (ed.) Williams & Wilkins. Baltimore. pp. 141-219.
- Palleroni, N. J. 1992. Introduction to the family Pseudomonadaceae in: *The Prokaryotes*. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K.-H. Schleifer (eds.). Springer Verlag. New York. pp. 3071-3085.
- Parke, D. 1992. Application of *p*-toluidine in chromogenic detection of catechol and protocatechuate, diphenolic intermediates in catabolism of aromatic compounds. *Appl. Environ. Microbiol.* 58: 2694-2697.
- Pemberton, J. M. and P. R. Fisher. 1877. 2,4-D plasmids and persistence. *Nature (London)* 268: 732-733.

- Perkins, E. J., M. P. Gordon, O. Caceres and P. I. Lurquin. 1990. Organization and sequence analysis of the 2,4-dichlorocatechol oxidative operons of plasmid pJP4. *J. Bacteriol.* 172: 2351-2359.
- Reineke, W. and H.-J. Knackmuss. 1978a. Chemical structure and biodegradability of halogenated aromatic compounds. Substituent effects on 1,2-dioxygenation of benzoic acid. *Biochim. Biophys. Acta* 542: 412-423.
- Reineke, W. and H.-J. Knackmuss. 1978b. Chemical structure and biodegradability of halogenated aromatic compounds. Substituent effects on dehydrogenation of 3,5-cyclohexadiene-1,2-diol-1-carboxylic acid. *Biochim. Biophys. Acta* 542: 424-429.
- Reineke, W. and H.-J. Knackmuss. 1988. Microbial degradation of haloaromatics. *Annu. Rev. Microbiol.* 42: 263-287.
- Ronald, S., M. A. Farinha, B. J. Allan and A. M. Kropinski. 1992. Cloning and physical mapping of transcriptional regulatory (sigma) factors from *Pseudomonas aeruginosa*. in: *Pseudomonas, molecular biology and biotechnology*. E. Galli, S. Silver and B. Wiholt (eds.). ASM publications. Washington, D.C. pp. 249-257.
- Sambrook, J., E. F. Fritsch and T. Maniatis. 1989. *Molecular cloning, a laboratory manual*, 2nd edition. 3 vols. Cold Spring Harbor Laboratory Press. Cold Spring Harbour, New York.
- Sandossi, M., M. Sylvestre and D. Ahmad. 1992. Effects of chlorobenzoate transformation on the *Pseudomonas testosteroni* biphenyl and chlorobiphenyl degradation pathway. *Appl. Environ. Microbiol.* 58: 485-495.
- Savard, P., L. Peloquin and M. Sylvestre. 1986. Cloning of *Pseudomonas* sp. strain CBS3 genes specifying dehalogenation of 4-chlorobenzoate. *J. Bacteriol.* 168: 81-85.
- Sayler, G. S., S. W. Hooper, A. C. Layton and J. M. Henry King. 1990. Catabolic plasmids of environmental and ecological significance. *Microbial Ecol.* 19: 1-20.
- Schmidt, E. and H.-J. Knackmuss. 1980. Chemical structure and biodegradability of halogenated aromatic compounds. *Biochem. J.* 192: 339-347.
- Schmidt, E., G. Remberg and H.-J. Knackmuss. 1980. Chemical structure and biodegradability of halogenated aromatic

- compounds. Halogenated muconic acids as intermediates. *Biochem. J.* 192: 331-337.
- Schmitt, R., P. Rogowsky, S. E. Halford and J. Grinsted. 1985. Transposable elements and evolution. in: *Evolution of Prokaryotes*. K. H. Schleifer and E. Stackebrandt (eds.). Academic Press. London. pp. 91-114.
- Scholten, J. D., K. H. Chang, P. C. Babbitt, H. Charest, M. Sylvestre and D. Dunawaymariano. 1991. Novel enzymic hydrolytic dehalogenation of a chlorinated aromatic. *Science* 253: 182-185.
- Schweizer, D. A. Markus, M. Seez, H. H. Ruf and F. Lingens. 1987. Purification and some properties of component B of the 4-chlorophenylacetate 3,4-dioxygenase from *Pseudomonas* species strain CBS3. *J. Biol. Chem.* 262: 9340-9346.
- Selvaraj, G. and V. N. Iyer. 1985. A small mobilizable Inc P group plasmid vector packable into bacteriophage capsids in vitro. *Plasmid* 13: 70-74.
- Sherratt, D. 1989. Tn3 and related transposable elements: site specific recombination and transposition. in: *Mobile DNA*. D. E. Berg and M. M. Howe (eds.). Amer. Soc. Microbiol. Washington D.C. pp. 163-184.
- Sinclair, M. I., P. C. Maxwell, B. R. Lyon and B. W. Holloway. 1986. Chromosomal location of TOL plasmid DNA in *Pseudomonas putida*. *J. Bacteriol.* 168: 1302-1308.
- Smibert, R. M. and N. R. Krieg. 1981. General characterization. in: *Manual of Methods for General Bacteriology*. P. Gerhardt, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg and G. B. Phillips. (eds.). Amer. Soc. Microbiol. Washington D.C. pp. 409-443
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.
- Sperka, S., E. Zehelein, S. Fiedler, S. Fischer, R. Sommer, P. Buckel. 1989. Complete nucleotide sequence of *Pseudomonas fluorescens* D-galactose dehydrogenase gene. *Nucleic Acids Res.* 17: 5402.
- Stanier, R. Y. and L. N. Ornston. 1973. The β -ketoadipate pathway. *Adv. Microbial Physiol.* 9: 89-151.

- Stanier, R. Y., N. J. Palleroni and M. Duodoroff. 1966. The aerobic pseudomonads: a taxonomic study. *J. Gen. Microbiol.* 43: 159-271.
- Subramanian, V., T.-N. Liu, W.-K. Yeh and D. T. Gibson. 1979. Toluene dioxygenase: purification of an iron-sulfur protein by affinity chromatography. *Biochem. Biophys. Res. Commun.* 91: 1131-1139.
- Subramanian, V., T.-N. Liu, W.-K. Yeh, M. Narro and D. T. Gibson. 1981. Purification and properties of NADH-ferredoxin-TOL reductase, a component of toluene dioxygenase from *Pseudomonas putida*. *J. Biol. Chem.* 256: 2723-2730.
- Subramanian, V., T.-N. Liu, W.-K. Yeh, C. Serdar, L. P. Wackett and D. T. Gibson. 1985. Purification and properties of ferredoxin_{TOL}: a component of toluene dioxygenase from *Pseudomonas putida* F1. *J. Biol. Chem.* 260: 2355-2363.
- Suzuki, M., T. Hayakawa, J. P. Shaw, M. Rekik and S. Harayama. 1991. Primary structure of xylene monooxygenase: similarities to and differences from the alkane hydroxylation system. *J. Bacteriol.* 173: 1690-1695.
- Syvanen, M. 1988. Bacterial insertion sequences. in: *Genetic Recombination*. R. Kucherlapati and G. R. Smith (eds.). Amer. Soc. Microbiol., Washington D.C. pp. 331-356.
- Taege, K., H.-J. Knackmuss and E. Schmidt. 1988. Biodegradability of mixtures of chloro- and methyl substituted aromatics: simultaneous degradation of 3-chlorobenzoate and 3-methylbenzoate. *Appl. Microbiol. Biotechnol.* 28: 603-608.
- Tomasek, P. H., B. Frantz, U. M. X. Sangodkar, R. A. Haugland and A. M. Chakrabarty. 1987. Characterization and nucleotide sequence determination of a repeat element isolated from a 2,4,5-T degrading strain of *Pseudomonas cepacia*. *Gene* 76: 227-238.
- Tsuda, M. and T. Iino. 1987. Genetic analysis of a transposon carrying toluene degrading genes on a TOL plasmid pWW0. *Mol. Gen. Genet.* 210: 270-276.
- Tsuda, M. and T. Iino. 1988. Identification and characterization of Tn4653, a transposon covering the toluene transposon Tn4651 on TOL plasmid pWW0. *Mol. Gen. Genet.* 213: 72-77.

- Tsuda, M. and T. Iino. 1990. Naphthalene degrading genes on plasmid NAH7 are on a defective transposon. *Mol. Gen. Genet.* 223: 33-39.
- Tsuda, M., K.-I. Minegishi and T. Iino. 1989. Toluene transposons Tn4651 and Tn4653 are class II transposons. *J. Bacteriol.* 171: 1386-1393.
- van der Meer, J. R., A. J. B. Zehnder and W. M. de Vos. 1991. Identification of a novel composite transposable element, Tn5280, carrying chlorobenzene dioxygenase genes of *Pseudomonas* sp. strain P51. *J. Bacteriol.* 173: 7077-7083.
- Weisshaar, M. P., F. C. H. Franklin and W. Reineke. 1987. Molecular cloning and expression of the 3-chlorobenzoate-degrading genes from *Pseudomonas* sp. strain B13. *J. Bacteriol.* 169: 394-402.
- West, S. E. H. and B. H. Iglewski. 1988. Codon usage in *Pseudomonas aeruginosa*. *Nucleic Acids Res.* 16: 9323-9335.
- Wheelis, M. L., N. J. Palleroni and R. Y. Stanier. 1967. The metabolism of aromatic acids by *Pseudomonas testosteroni* and *P. acidovorans*. *Arch. Microbiol.* 59: 302-314.
- Whited, G. M. and D. T. Gibson. 1991. Toluene-4-monooxygenase, a three component enzyme system that catalyses the oxidation of toluene to *p*-cresol in *Pseudomonas mendocina* KR1. *J. Bacteriol.* 173: 3010-3016.
- Wiater, L. A. and N. D. F. Grindley. 1988. Gamma-delta transposase and integration host factor bind cooperatively at both ends of gamma-delta. *EMBO J.* 7: 1907-1911.
- Willems, A., J. De Ley, M. Gillis and K. Kersters. 1991a. *Comamonadaceae*, a new family encompassing the *Acidovorans* rRNA complex, including *Variovorax paradoxus* gen. nov., comb. nov., for *Alcaligenes paradoxus* (Davis 1969). *Int. J. Syst. Bacteriol.* 41: 445-450.
- Willems, A., B. Pot, E. Falsen, P. Vandamme, M. Gillis, K. Kersters and J. De Ley. 1991b. Polyphasic taxonomic study of the emended genus *Comamonas*: relationship to *Aquaspirillum aquaticum*, E. Falsen group 10, and other clinical isolates. *Int. J. Syst. Bacteriol.* 41: 427-444.

- Williams, P. A. and K. Murray. 1974. Metabolism of benzoate and methylbenzoates by *Pseudomonas putida* (arvilla) mt-2: evidence for the existence of a TOL plasmid. J. Bacteriol. 120: 416-423.
- Wyndham, R. C. 1986. Evolved aniline catabolism in *Acinetobacter calcoaceticus* during continuous culture of river water. Appl. Environ. Microbiol 51: 781-789.
- Wyndham, R. C., R. K. Singh and N. A. Straus. 1988. Catabolic instability, plasmid gene deletion and recombination in *Alcaligenes* sp. BR60. Arch. Microbiol. 150: 237-243.
- Wyndham, R. C. and N. A. Straus. 1988. Chlorobenzoate catabolism and interaction between *Alcaligenes* and *Pseudomonas* species from Bloody Run Creek. Arch. Microbiol. 150: 230-236.
- Yamaguchi, M. and H. Fujisawa. 1978. Characterization of NADH-cytochrome C reductase, a component of benzoate 1,2-dioxygenase system from *Pseudomonas arvilla* C-1. J. Biol. Chem. 253: 8848-8853.
- Yamaguchi, M. and H. Fujisawa. 1980. Purification and characterization of an oxygenase component of benzoate 1,2-dioxygenase system from *Pseudomonas arvilla* C-1. J. Biol. Chem. 255: 5058-5063.
- Yamaguchi, M. and H. Fujisawa. 1982. Subunit structure of an oxygenase component of benzoate 1,2-dioxygenase system from *Pseudomonas arvilla* C-1. J. Biol. Chem. 257: 12497-12502.
- Yannisch-Perron, C., J. Vieira and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.
- Zaitsev, G. M. and B. P. Baskunov. 1985. Utilization of 3-chlorobenzoic acid by *Acinetobacter*. Mikrobiologiya 54: 203-208. (english translation 158-164).
- Zavarzin, G. A., E. Stackebrandt and R. G. E. Murray. 1991. A correlation of phylogenetic diversity in the proteobacteria with the influences of ecological forces. Can. J. Microbiol. 37: 1-6.
- Zylstra, G. J. and D. T. Gibson. 1989. Toluene degradation by *Pseudomonas putida* F1. Nucleotide sequence of the todC1C2BADE genes and their expression in *Escherichia coli*. J. Biol. Chem. 264: 14940-14946.

END

1 2 - 1 2 - 9 3

FIN