Analysis of cloned aromatic catabolic genes from *Klebsiella pneumoniae*

SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF LEICESTER

NEIL DAVID ROBSON

JUNE 1992

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I would first like to thank Dr. R.A Cooper for his helpful advice and encouragement throughout the course of this work. I also gratefully acknowledge the members of Lab 108 for all of the assistance, helpful advice and friendship offered. Finally I must particularly mention the members of my family who I could never even begin to thank enough for their support.

Declaration

This thesis, submitted for the degree of Doctor of Philosophy entitled: Analysis of cloned aromatic catabolic pathways from *Klebsiella pneumoniae*, is based upon work conducted by the author in the Department of Biochemistry between October 1st 1988 and September 30th 1991.

All of the work recorded in this thesis is original unless otherwise acknowledged in the text or by references.

None of this work has been submitted for another degree in this or any other University.

Signed : Neil Ridley

Date : 26-6-92

Abbreviations

A_{680}	Absorbance at 680 nm
Amp	Ampicillin
bp	Base pairs
CHM	5-Carboxymethyl-2-hydroxymuconate
CHMS	5-Carboxymethyl-2-hydroxymuconate semialdehyde
COHED	5-Carboxymethyl-2-oxo-hex-3-ene-1,6-dioate
DHB	Dihydroxybenzoate
EDTA	Ethylenediaminetetraacetic acid
ExoIII	Exonuclease III
FPLC	Fast Protein Liquid Chromatography
HB	Hydroxybenzoate
HHDD	2-hydroxyhepta-2,4-diene-1,7-dioate
HHED	2,4-dihydroxy-hepta-2-ene-1,7-dioate
HPC	Homoprotocatechuate (3,4-dihydroxyphenylacetate)
IPTG	Isopropyl-β-D-thiogalactoside
Kbp	Kilo base pairs
λ	Wavelength
MOPS	3-[N-morpolino] propanesulphonic acid
NAD(P)	Nicotinamide adenine dinucleotide (phosphate)
OHED	2-oxo-hepta-3-ene-1,7-dioate
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
SDS	Sodium dodecylsulphate
SSA	Succinic semialdehyde
TEMED	N,N,N',N'-tetramethylethyldiamine
Tris	Trishydroxymethylaminomethane

ABSTRACT

The organisation of two aromatic catabolic pathways operational in *Klebsiella pneumoniae* has been investigated through the analysis of plasmid-based genomic clones.

A single genomic clone encoding all of the activities required for the metabolism of 3-hydroxybenzoate was isolated on an 8.0Kbp *Sph*I fragment via a cloning strategy determined by the analysis of an existing partial clone. The presence of a plasmid containing this fragment allowed strains of *Escherichia coli* to grow on 3-hydroxybenzoate as sole carbon and energy source. Sub-cloning experiments revealed that the genes involved in this pathway;*mhbM* encoding 3-hydroxybenzoate mono-oxygenase, *mhbD* encoding 2,5-dihydroxybenzoate dioxygenase, *mhbI* (maleylpyruvate isomerase), *mhbH* (fumarylpyruvate hydrolase) and the regulator gene *mhbR*, were arranged in the order *mhbRDHMI*. Preliminary analysis of the regulatory system has suggested that the expression of the *mhb* genes is under positive control.

High level expression of the 2,5-dihydroxybenzoate dioxygenase encoded by *mhbD* was detected from several sub-clones. This allowed the development of a three-step purification procedure which resulted in a preparation suitable for N-terminal amino acid sequence determination. Full nucleotide sequence was obtained for the *mhbD* gene with assignment of the reading frame made on the basis of the protein's N-terminal sequence. Comparison of this sequence with existing databases failed to detect strong homologies with other dioxygenases.

Two discrete genomic clones encoding genes involved in the catabolism of homoprotocatechuate (HPC); *hpcB* encoding HPC dioxygenase, *hpcC* (CHMS dehydrogenase), *hpcD* (CHM isomerase), *hpcE* (COHED decarboxylase), *hpcF* (HHDD isomerase), *hpcG* (OHED hydratase), *hpcH* (HHED aldolase) and *hpcR* (regulator gene), were subjected to restriction analysis and the gene organisation investigated by the sub-cloning of specific fragments. The genes were found to be arranged in the order *hpcR(EF)CBDGH* where the precise relationship of the *hpcE* and *hpcF* genes was not defined. This gene order is identical to that determined for the corresponding system in *Escherichia coli* C.

TABLE OF CONTENTS

Chapter 1

CHAPTER 2

Index of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Index of Tables

CHAPTER 1

Introduction

INTRODUCTION

The relevance of studying aromatic catabolism

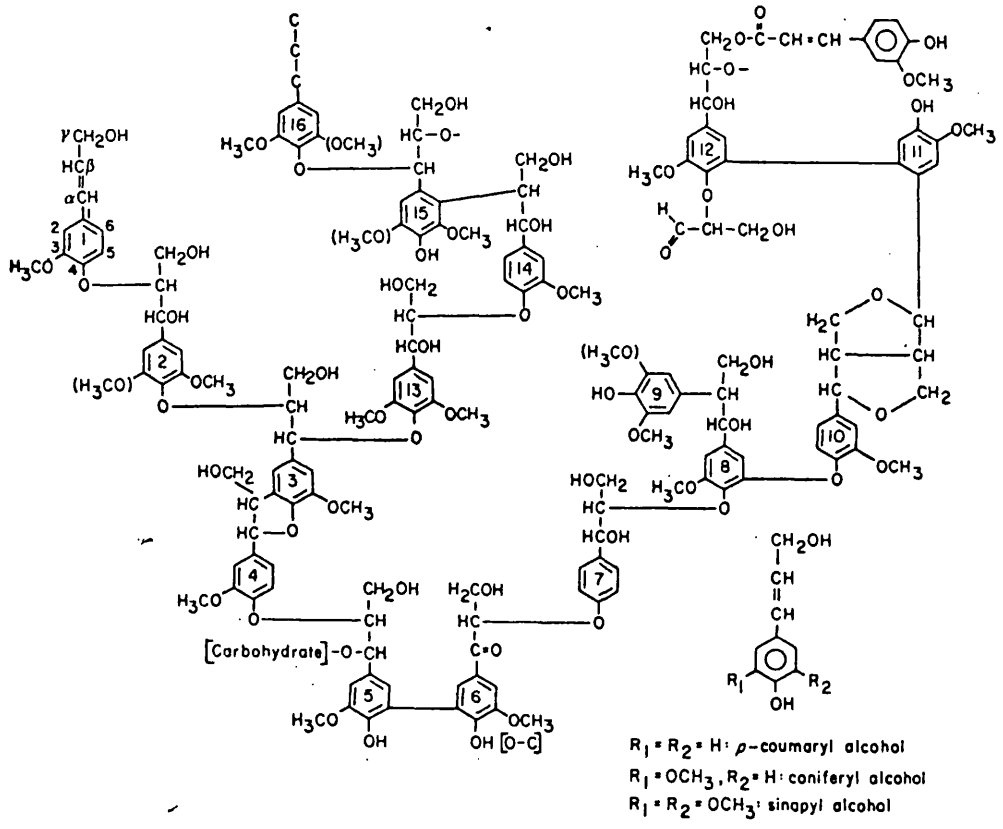
The importance of aromatic catabolism with respect to the functioning of the earth's carbon cycle becomes apparent if it is considered that, next to glucosyl residues, the benzene ring is the most widely distributed unit of chemical structure in the biosphere. A great deal of what might be termed aromatic carbon is contained within the lignins, complex biopolymers of higher plants more abundant than protein. At this point it should be noted that proteins also contain aromatic groups by virtue of the presence of the amino acids tyrosine and phenylalanine.

Lignins are produced through the coupling of three precursor aromatic alcohols: *p*-coumaryl, coniferyl and sinapyl and a representation of a lignin structure is shown in Figure 1.1. Disruption of the lignin structure is of further importance in the context of the carbon cycle because lignin physically protects most of the world's cellulose and hemicelluloses from enzymic hydrolysis. The breakdown of lignin is primarily initiated by the action of various fungi (Reviewed by Kirk and Farrell, 1987) and as a consequence a wide range of low molecular weight aromatic molecules enter the environment. The breakdown of plant tissues and the action of processes involved in the formation of oil, coal and tar have produced further benzenoid compounds. In view of the quantity of carbon represented by these low molecular weight aromatic compounds it is perhaps not surprising that microorganisms have developed systems which allow their utilization as foodstuffs.

The importance of aromatic catabolism in the modern world is not just based upon the recycling of carbon. Many aromatic compounds are generally regarded as noxious substances which can have detrimental effects if they are introduced into the environment either in large quantities or at inappropriate locations. Here one thinks of the types of compounds which are found for example in oil and coal and which can pose a pollution threat. Metabolism of such compounds therefore has a detoxifying effect which from an anthropocentric

Figure 1.1

Schematic structural formula for lignin illustrating the major interunit linkages (Kirk and Farrell, 1987). The three precursor alcohols are shown at the lower right side.



perspective is of particular value in the context of the global carbon cycle.

In recent years a whole range of novel man-made aromatic compounds have entered the environment as a consequence of their usage as herbicides and pesticides . Such compounds may, by virtue of their toxicity, pose a serious health risk. In a number of cases it has been possible to identify bacterial strains which are able to metabolise these compounds. In several instances the capacity of bacteria in stressed environments to accommodate novel substrates is remarkable. A prime example of this is the discovery of organisms which are able to degrade polychlorinated biphenyls (PCBs). These compounds have potential toxic, mutagenic and teratogenic effects on human populations and are known to bioaccumulate. There were no known sources of PCB prior to their commercial synthesis in 1929 yet natural systems are able to mediate their breakdown. However, the ability to degrade man-made aromatic compounds is still limited and a number of compounds remain recalcitrant . If the understanding of the processes involved in aromatic catabolism, including its regulation, is increased then a number of options for pollution management may become available. For example it may become possible to "design" specialised organisms which are able to remove previously non-degradable compounds or which remove more efficiently compounds which currently persist for long enough to pose a pollution hazard.

Studies of aromatic catabolism therefore offer insights into both fundamental and applied aspects of the global carbon cycle.

The processes of aromatic catabolism

Aromatic compounds are degraded by both aerobic and anaerobic processes. This section will focus solely on the processes of aerobic aromatic catabolism in which cleavage of the ring involves the insertion of both atoms of molecular oxygen by ring fission dioxygenases.

The benzenoid nucleus of aromatic compounds is a particularly stable structure. This stability results from the resonance energy associated with the bond organisation found in aromatic rings. The ring fission event is generally preceded by substitution reactions which destabilise the ring and render it susceptible to cleavage. In the vast

majority of cases compounds suitable for ring fission have two hydroxyl groups substituted onto the ring. Hydroxyl groups may be introduced either singly by the action of a monooxygenase (hydroxylase) or two may be introduced simultaneously by the action of a dioxygenase (dihydroxylase). Exceptions to the dihydroxy-substitution rule include the fission of 5-chlorosalicylate (5-chloro-2-hydroxybenzoate) by a specific dioxygenase from *Bacillus brevis* (Crawford *et al*, 1979) and the cleavage of 5-aminosalicylate at moderate efficiency by 2,5-dihydroxybenzoate (gentisate) 1,2-dioxygenases from *Pseudomonas testosteroni* or *P.acidovorans* (Harpel and Lipscomb, 1990a).

The dihydroxy aromatic nucleus may be cleaved in a number of ways depending on the pattern of substitution. In cases where the ring fission substrate contains adjacent (*ortho*) hydroxyls cleavage may occur between the hydroxyls (*ortho*-fission/intradiol-cleavage) or to one side of one of the hydroxyls (*meta*-fission/extradiol-cleavage). This is shown in Figure 1.2a where catechol (1,2-dihydroxybenzene) is the ring fission substrate. If the hydroxyls on the ring are *para*-substituted then ring fission occurs between one of the hydroxyls and a side chain. This has been referred to as distal extradiol-cleavage. This is shown in Figure 1.2b with gentisate (2,5-dihydroxybenzoate) as the substrate.

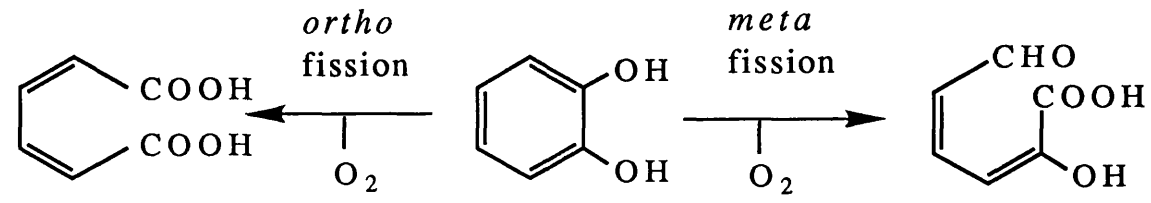
Central pathways of aromatic catabolism

The catabolism of a wide range of aromatic compounds has now been investigated in a number of different bacterial species. It has emerged that aromatic catabolism is no different from many other branches of metabolism in that a wide range of substrates are channelled through a limited number of central pathways. This type of organisation is beneficial for the cell because of the energy savings resulting from the reduced genetic load and through the simplification of the regulatory mechanisms required to allow growth on a range of different compounds. In the case of aromatic catabolism it has been observed that many benzenoid compounds converge on one of a small range of ring-fission substrates such as catechol, gentisate and protocatechuate or their derivatives (Ribbons and Eaton, 1982). Well defined pathways exist for the further metabolism of each these

Figure 1.2

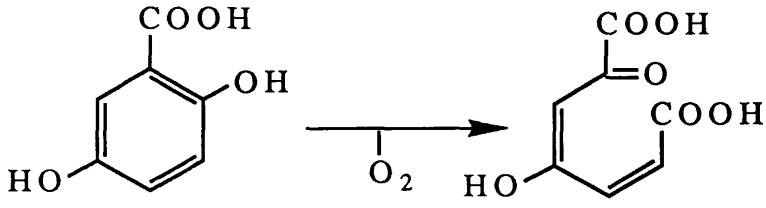
The possible routes for cleavage of different ring-fission substrates

a) *Ortho*-(intradiol) and *meta*-(extradiol) fission of catechol



Cis, cis-muconate | Catechol | 2-Hydroxymuconate semialdehyde

b) Distal *meta*-(extradiol) fission of gentisate



Gentisate | Maleylpyruvate

compounds. The concept of substrate channelling is embodied by the β-ketoadipate pathway which has been found in a number of different bacteria. This pathway is involved in the degradation of compounds which are converted to protocatechuate, catechol or halocatechol. Cleavage of the ring by *ortho*-fission initiates the β-ketoadipate pathway which is depicted in Figure 1.3. The very high degree of pathway convergence suggested by Figure 1.3 is slightly misleading. In some species, for example *Acinetobacter calcoaceticus* (Doten *et al*, 1987) the terminal reactions common to all three arms of the pathway are catalysed by two sets of isofunctional enzymes which are induced independently by metabolites specific to the catechol and protocatechuate arms.

Distribution of bacterial aromatic catabolic pathways

As might be expected the metabolic capacities of bacteria with respect to aromatic compounds is tailored to those compounds that each normally encounters. Hence it is not surprising that soil micro-organisms such as the pseudomonads have the greatest catabolic repertoire since they are likely to be exposed to the widest spectrum of compounds. It should be noted that pathways which are involved in the degradation of the more unusual benzenoid compounds are frequently plasmid-encoded. The benefit of this is that perhaps the majority of the bacterial cells in the population are spared the energetic expense involved in the maintenance of unusual catabolic genes. Instead the metabolic plasmids are retained at low frequency in the population and can, as a result of their self-transmissibility, be mobilised into the wider population when the particular pathway is required.

The ability of enteric organisms such as *Escherichia coli* to catabolise aromatic compounds went unrecognised until utilisation of 3- and 4-hydroxyphenylacetate (Cooper and Skinner, 1980) was reported. Further degradative capabilities have been reported for 3-phenylpropionate (Burlingame and Chapman, 1983), phenylacetate (Cooper *et al*, 1985), and phenylethylamine (Parrott *et al*, 1987) but it is apparent that the range of metabolisable aromatics is narrow. The compounds which are degraded by strains of *E.coli* appear to be produced by the partial metabolism of the aromatic amino acids (Spoelstra, 1978) and may be frequently encountered in the

Figure 1.3

The β-ketoadipate pathway showing convergence of the discrete arms.

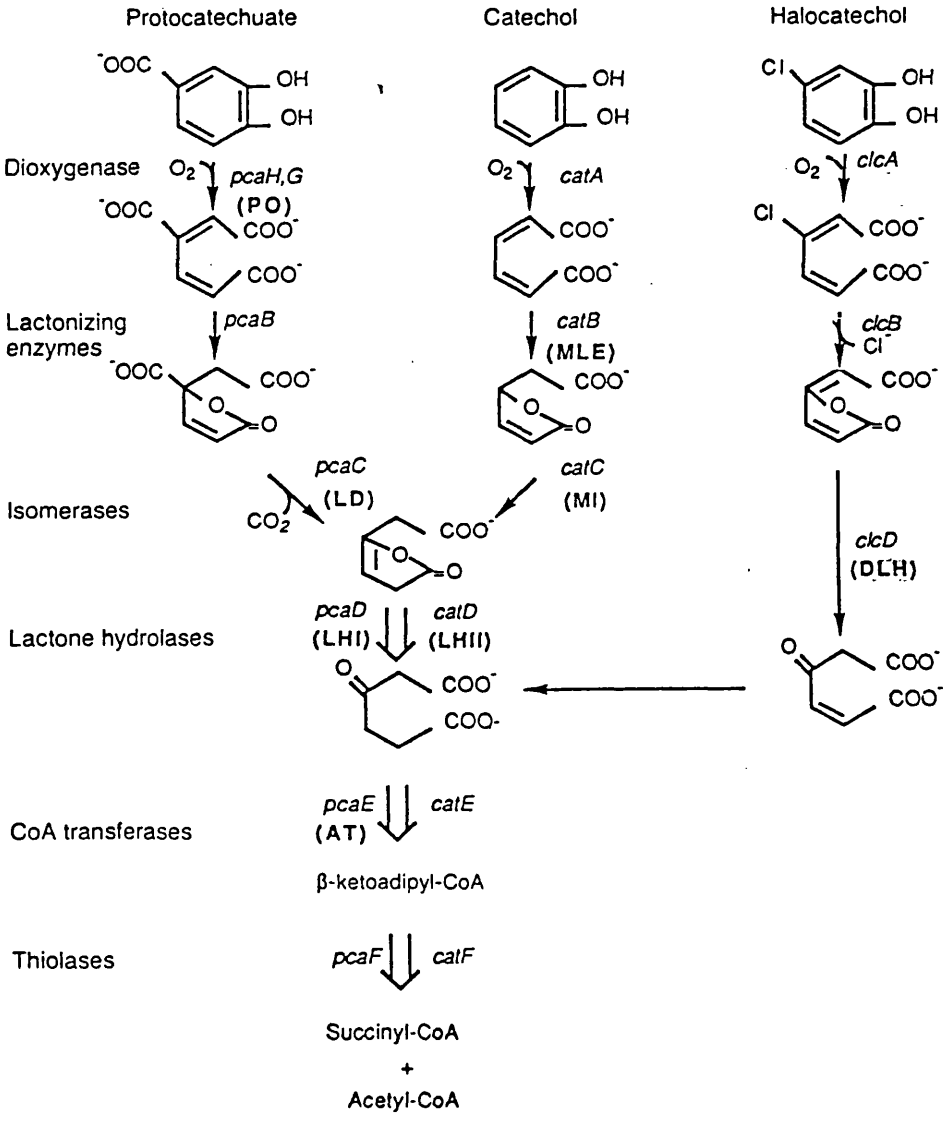
The genes and enzymes relating to each step are indicated (genes in italics; enzymes in bold).

Key to enzymes:

PO= protocatechuate 3,4-dioxygenase
MLE= muconate lactonizing enzyme
LD= γ-carboxymuconolactone decarboxylase
MI= muconolactone isomerase
DLH= dienelactone hydrolase
LHI= lactone hydrolaseI
LHII= lactone hydrolaseII
AT= β-ketoadipate-succinyl-CoA transferaseI

Figure 1.3

The β-ketoadipate pathway showing convergence of the discrete arms (Allewell, 1989).



intestinal-faecal environment. The pathways involved are chromosomally encoded.
Other members of the Enterobacteriaceae can metabolise a more diverse range of compounds. Particularly notable in the context of this thesis are klebsiellae. In 1969 it was reported that *K. aerogenes* was able to dissimilate benzoate and catechol via the β-ketoadipate pathway (Grant and Patel, 1969; Patel and Grant, 1969) and more recently the metabolism of protocatechuate by the β-ketoadipate pathway in this organism has been investigated (Doten and Ornston, 1987). The additional catabolic diversity of klebsiellae is perhaps not surprising in view of the fact that they are frequently isolated from samples of living, decaying or composted wood and bark (Deschamps *et al.*, 1983) where aromatics are abundant. There has also been a report that an isolate of *K.pneumoniae* ,obtained from a polluted stretch of the Hudson River, contained a plasmid designated pAC21 which allowed partial metabolism of a chlorinated biphenyl (Kamp and Chakrabarty, 1979). Therefore, by choosing to study aromatic catabolism in an organism such as *K.pneumoniae* it is possible to combine a fairly accessible genetic system with a fairly wide catabolic capacity.

Ring fission dioxygenases

Ring fission dioxygenases have attracted considerable attention as they catalyse the critical step in aromatic catabolism i.e the opening of the ring. The dioxygenases have tended to be divided into three groups on the basis of the nature of the ring fission they catalyse. Hence there are the *ortho*-fission dioxygenases, the *meta*-fission dioxygenases and the dioxygenases which do not precisely fit into either category. Most notable in this latter grouping are the gentisate 1,2-dioxygenases. In addition to studies of individual dioxygenases attention has also focused on the possible relationships between different dioxygenase enzymes. Examination of a range of different dioxygenases has suggested that there are distinct differences between the *ortho*- and *meta*-fission enzymes. The intradiol dioxygenases which mediate *ortho*-fission may be composed of either one or two sub-units and contain varying amounts of ferric iron at

their catalytic sites. The presence of the ferric iron frequently gives these enzymes a characteristic red-brown colouration. Extradiol enzymes catalysing *meta*-fission reactions are usually composed of a single sub-unit type and contain ferrous iron as a prosthetic group. More detailed comparisons of various dioxygenases have been made as amino acid and nucleotide sequence has become available. Such analyses culminated in the suggestion by Harayama *et al* (1988) that bacterial aromatic ring-cleavage enzymes could be classified into two different gene families. One representing the *ortho*- and the other the *meta*-cleavage enzymes. The suggestion of an *ortho*-fission gene family was made on the basis of the comparison of sequences from three separate intradiol dioxygenases (Neidle *et al*, 1988). The overall sequence homologies of catechol 1,2-dioxygenase I (CatA), catechol 1,2-dioxygenase II (ClcA) and protocatechuate 3,4-dioxygenase (PcaA) were such that it appeared that all three proteins were derived from a common ancestor (Neidle *et al*, 1988). Evidence for similar relationships between *meta*-fission enzymes came from consideration of the sequences of two catechol 2,3-dioxygenases and two 2,3-dihydroxybiphenyl dioxygenases from different *Pseudomonas* species (Harayama *et al*, 1987; Taira *et al*, 1988). The sequence of the *nahC* gene encoding 1,2-dihydroxynaphthalene dioxygenase was subsequently shown to share considerable homology with that of a 2,3-dihydroxybiphenyl dioxygenase (Harayama *et al*, 1988). A more detailed statistical analysis suggested a lesser but still significant degree of homology between the *nahC* product and the catechol 2,3-dioxygenases encoded by *xylE* and *nahH*. This implied a common evolutionary origin for this group of dioxygenases. It was not, however, possible to detect any significant global or localised similarities between the two dioxygenase groups. As the sequences of more dioxygenases have become available the concept of there being two gene superfamilies has weakened somewhat. No significant homologies were detected between the nucleotide sequence of the protocatechuate 4,5-dioxygenase from *P.paucimobilis* and other *meta*- (or *ortho*-) dioxygenases (Noda *et al*, 1990). It should perhaps be pointed out that this dioxygenase is somewhat unusual as it is a *meta*-cleavage enzyme composed of two dissimilar sub-units and so the extent of sequence homology might be expected to be lower. The homoprotocatechuate dioxygenase of *E.coli* C is more typical in the

sense that it is made up of a single sub-unit type yet its sequence shows no significant homology with those of other dioxygenases (Roper and Cooper, 1990b).

Examination of ring fission dioxygenases has not however been confined to the analysis of nucleotide and amino acid sequence. A number of studies have been performed which have investigated the ways in which ring fission is brought about (for example Arciero and Lipscomb, 1986; Harpel and Lipscomb, 1990b). The latter study addressed the mechanism of gentisate 1,2-dioxygenase by a combination of Electron Paramagnetic Resonance Spectroscopy (EPR), which investigates interactions involving the active site Fe^{2+}, and the use of substrate analogues. It emerged that the mechanism proposed shared common features with the mechanisms proposed for extra and intradiol catecholic dioxygenases. One aspect shared by all of the dioxygenases was that the iron appears to be able to participate in the reaction mechanisms via direct binding of the organic substrate. This allows the iron to affect the reactions by controlling both the distribution of electron density within the organic substrate and its orientation relative to the O_2 co-substrate. Where the Fe^{2+} and Fe^{3+} dioxygenases appear to differ is in the precise manner in which the iron is utilised. In the Fe^{2+} dioxygenases it appears that the O_2 is activated by binding directly to the Fe^{2+} which is simultaneously engaged in an electron rich chelate with the organic substrate. In the Fe^{3+} dioxygenases the substrate is activated for direct attack by O_2 by virtue of the Lewis acidity of the Fe^{3+} . Mechanistic examination of gentisate 1,2-dioxygenase has started to probe the precise nature of the general requirement for dihydroxylation of the aromatic ring prior to ring fission. This is a difficult question to address in cases where the hydroxyl groups are adjacent because it becomes difficult to choose between possible mechanistic roles for the hydroxyls because of their close proximity to the iron. Whilst their major role might be to form an iron chelate which would define a specific substrate orientation they might also participate directly in the mechanism via electronic effects such as electron donation to the ring or ketonization. The *para* configuration of the hydroxyls in gentisate allows the effects to be distinguished for its dioxygenase. For this enzyme the iron binding and orientational effects are carried out by the carbon 1 carboxylate of gentisate rather than the second hydroxyl group for the

catecholic dioxygenases. Electronic effects unique to the carbon 5 hydroxyl of gentisate cannot be effectively replaced by any other substituent. The fact that only an amine substitution at this position allows even partial replacement of the hydroxyl suggests that ketonization of the C5 substituent is mechanistically important. A proposal was then made that ketonization of one of the hydroxyl groups may be essential for the oxygen insertion reactions of all catecholic and gentisate dioxygenases (Harpel and Lipscomb, 1990b).

Evolution of pathways of aromatic catabolism

This is a subject which has received greatly increased attention as nucleotide and amino acid sequence information has become available. Molecular biophysics is also playing a role as crystallographic data becomes available for more and more enzymes, particularly those involved in the β-ketoadipate pathway (Allewell, 1989). One subset of the comparisons which have been made to suggest that pathways may be evolutionarily related has already been described, namely that of the ring fission dioxygenases. A second subset which have probed the relatedness of *meta*-fission pathways is described in the next section. A third area which has been well investigated has been the β-ketoadipate pathway.

The β-ketoadipate pathway affords many opportunities for sequence comparisons being distributed among a very wide range of bacteria. It is possible to make comparisons between isofunctional enzymes in the different branches of the pathway and also between the enzymes catalysing the reactions within an individual arm. Scrutiny of the available sequences has revealed several surprising features. It appears that what would seem to be distantly related genes may have substantial sequence similarity. For example, there is a strong degree of similarity between the genes for muconolactone isomerase and β-ketoadipate enol-lactone hydrolase II (LHII) (Yeh and Ornston, 1980) of *Acinetobacter*. Similarity also exists between the γ-carboxymuconolactone decarboxylase (LD) and β-ketoadipate-succinyl-CoA transferase I (AT) (Yeh and Ornston, 1982). Conversely functionally related genes such as lactone hydrolases I and II (LHI, LHII) may have remarkably little in common (Yeh *et al*, 1980). This is

an interesting point because in the view of Dagley (1975) one of the major points against the evolution of catabolic systems via gene duplication and subsequent mutation is that the mechanisms involved in two consecutive reactions may be very different. Yet here is a case where the greatest similarity is between very different enzymes although it must be stated that the balance of evidence is that the gene duplication models are not generally applicable.

Analysis of the regulatory genes in different bacteria has also been informative. Whereas the evolutionary homologue of a gene (*catR*) codes for a transcriptional activator in *P.putida* it appears to encode a repressor (*catM*) in *A.calcoaceticus* (Neidle *et al.*, 1989). The differences in regulatory mechanism among species have been associated with differences in the organisation of genes in operons. It has been suggested that these differences are evidence for selection first at the level of catalysis and subsequently at the regulatory level (Stanier and Ornston, 1973).

In organisms such as *A.calcoaceticus* the genes involved in both the protocatechuate and catechol arms of the β-ketoadipate pathway are maintained in the chromosome. The genes are organised such that there are separate operons of isofunctional genes *catDEF* and *pcaDFE* (Doten *et al*, 1987). Amino acid sequence comparisons have indicated that the isofunctional enzymes are homologous and indeed hybridization has been demonstrated between the DNA fragments containing the *cat* and *pca* clusters. Stable maintenance of these clusters requires that they should be divergent to a degree sufficient to prevent frequent recombination between them. Therefore selective demand for sequence divergence to avoid the loss of genes through recombination may have contributed to sequence rearrangements that have been observed within and among genes.

The assembly of aromatic catabolic pathways

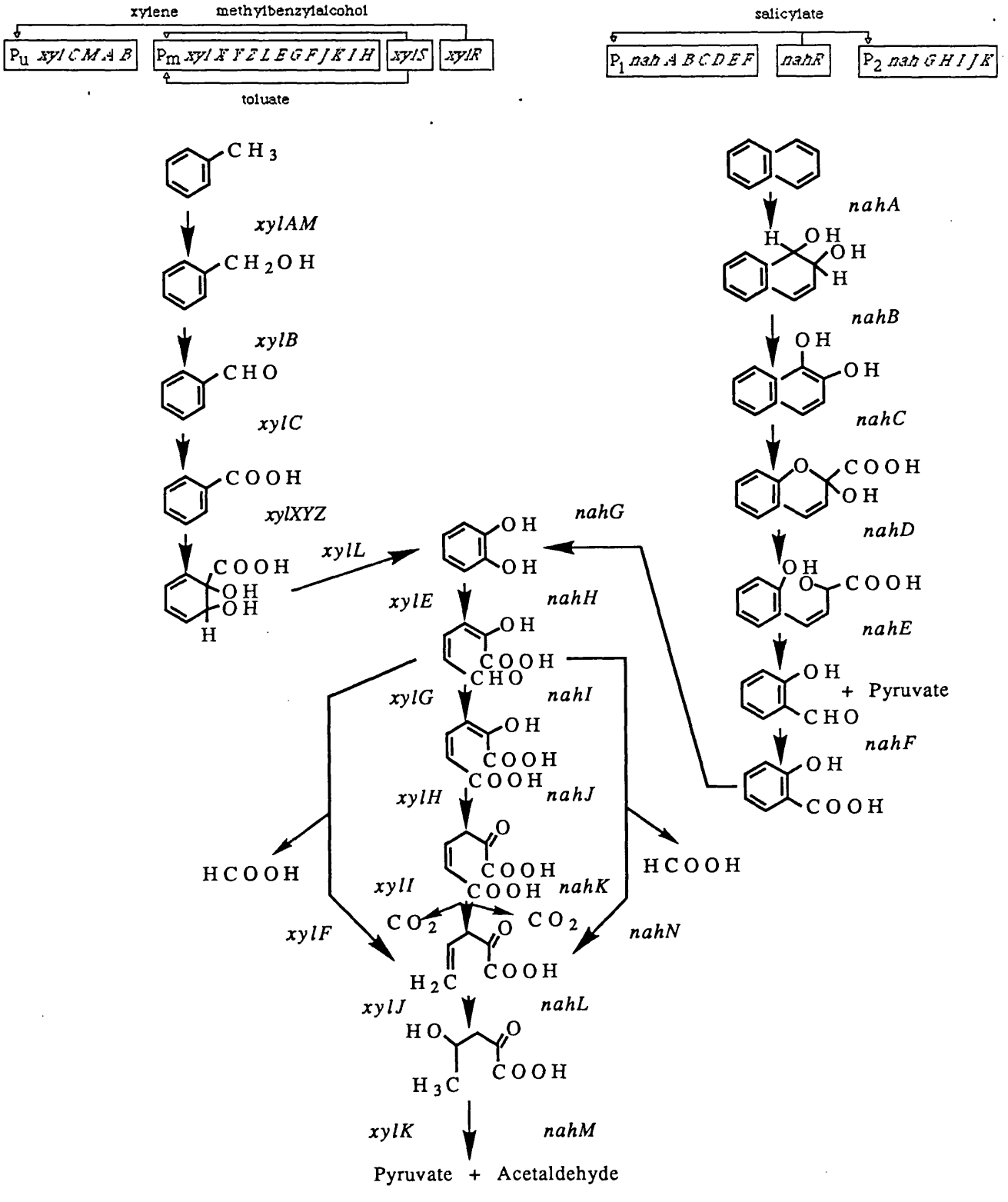
Many studies of aromatic catabolic pathways have focused on the relatedness of the proteins/enzymes/regulatory systems involved in benzenoid metabolism. In many cases the similarities may extend well beyond the individual pathway component and this has led to proposals of the existence of a number of so-called catabolic modules. It has been proposed that the catabolic functions were acquired as a

result of the horizontal spread of ancestral DNA segments. A typical module, identified on the basis of the similarity between the DNA sequences of the *xylXYZL* genes from the TOL plasmid pWWO and the isofunctional genes present in the chromosome of *Acinetobacter calcoaceticus*, is involved in the transformation of benzoate to catechol (Neidle *et al*, 1987).

The most intensively studied catabolic module encodes the reactions involved in the assimilation of catechol via *meta* cleavage. This pathway element is included in systems which degrade toluene (TOL) and related compounds, naphthalene (NAH) and phenol . The schemes for the degradation of toluene and naphthalene are depicted in Figure 1.4. It is interesting to note the broad similarities in the operon structures of the two pathways. In the TOL system there are two operons; one which allows the oxidation of toluene to benzoate and a second metabolising benzoate through to central metabolites via *meta*-fission. In the case of the NAH one operon encodes enzymes which specify the conversion of naphthalene to salicylate whilst the second metabolises salicylate to the same central metabolites via the same *meta*-pathway. On the basis of detailed analyses of the genetic organisation of the two pathways on the plasmids pWWO and NAH7 Harayama *et al* (1987) suggested a model for the evolutionary history of the *meta*-cleavage pathway. This scheme involves the movement of the gene cluster encoding the catechol oxidative system into a region of DNA downstream of the salicylate oxidative gene (ancestral *nahG*) and/or the toluate oxidative genes. This led to the formation of an operon structure which would allow the coordinated expression of all pathway enzymes for the oxidation of toluate and/or salicylate providing a selective advantage for cells using such compounds as substrate. The reversal of the orders of the *nahJ-nahK* or *xylI-xylH* is attributed to a simple DNA rearrangement. As part of a comparison of the *meta*-pathway operons on the NAH plasmid pWW60-22 and the TOL plasmid pWW53-4 a similar model was proposed for the assembly of the NAH pathway (Assinder and Williams, 1988). In this case it was proposed that the NAH sequence resulted from the chance combination in a single replicon of three pre-evolved metabolic modules. The first of these modules allowed naphthalene to be metabolised to salicylate; the second included the enzyme salicylate hydroxylase under the control of NahR whilst the third included the

Figure 1.4

Catabolic pathways and genes of TOL plasmid pWWO and NAH plasmid NAH7 (Harayama *et al*, 1987). The upper part of the figure shows the gene order of the TOL and NAH operons and their regulation whilst the lower portion depicts the reactions involved..



meta- pathway enzymes. The *meta*-pathway in this plasmid was believed to share a common origin with that present in the TOL pathway with the equivalents of the *xylD* and *xylL* genes being lost from NAH through redundancy. The fusion of the latter two modules occurring as the result of the deletion of DNA between the *nahG* and *nahH* genes would produce a single operon under the control of NahR. The precise mechanisms by which the genetic modules might be transferred between organisms have not been defined. It is possible that the required genetic transfer could be achieved by a variety of mechanisms including plasmid-mediated conjugation, phage-mediated transduction, conjugative transposition of mobile genetic elements and transformation by unassociated DNA fragments. Interestingly on the TOL plasmid pWWO the TOL genes are associated with a transposon like structure. Replicons such as TOL which are conjugative and have a wide-host range have a high potential for picking up new genes from different hosts and, under appropriate selective pressure, for evolving new functions which may subsequently be transmitted among a wide variety of bacteria.

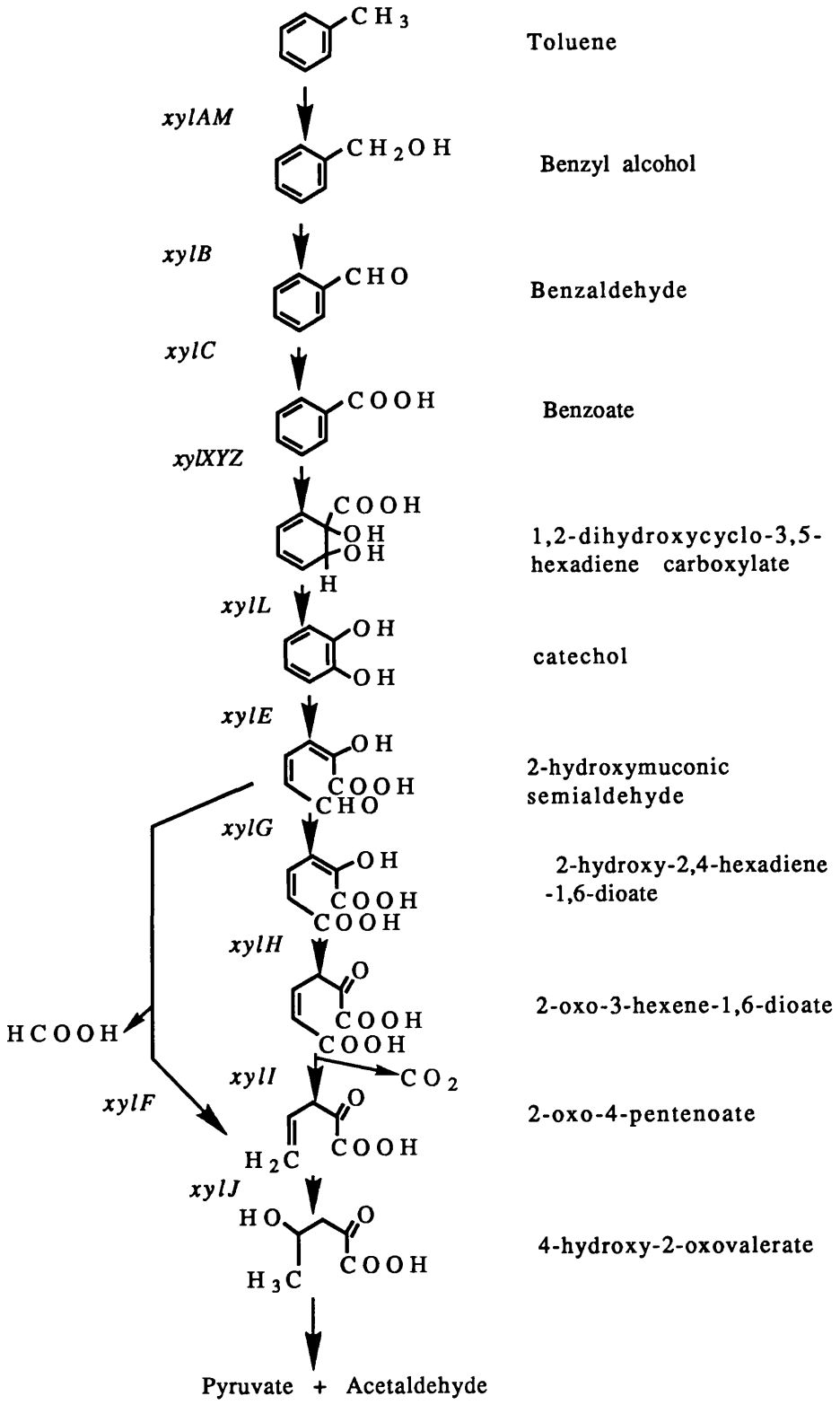
The TOL *meta*-fission pathway

The TOL *meta*-fission pathway is considered in some detail because it is perhaps the most extensively characterised aromatic pathway. The data generated during such studies have been of key significance in the design of modified metabolic routes.

Analysis of the ways in which compounds such as toluene, *m*- and *p*-toluate and *m*- and *p*-xylene were degraded led to the identification of the classical TOL *meta*-fission pathway. The reactions involved in this catabolic route are depicted in Figure 1.5 where it is apparent that metabolism occurs via a bifurcating pathway which results in the production of acetaldehyde and pyruvate. The reactions have historically been divided into two discrete segments. This division is based on the observation that the genes for the TOL pathway enzymes are clustered in two operons (Nakazawa, *et al*, 1980; Inouye *et al*, 1981; Franklin *et al*, 1981). The first stages of metabolism are catalysed by the enzymes of the so-called upper pathway. In the upper pathway hydrocarbons such as toluene and the xylenes are

Figure 1.5

Pathway for the degradation of toluene encoded by *Pseudomonas* plasmid pWWO.



sequentially oxidised to benzoate and toluates respectively. The products of the upper pathway then enter the second stage of metabolism which is mediated by the enzymes of the lower pathway. This results in the production of Krebs' cycle intermediates via *meta* cleavage of the ring fission substrate catechol. The pathway diverges into hydrolytic and dehydrogenative routes at the ring fission product and reconverges later at 2-hydroxypent-2,4-dienoate. It was shown by Harayama *et al* (1987) that different starting compounds were metabolised by different branches of the pathway. The pathway used being determined by the affinities of the enzymes for the pathway intermediates. Benzoate and *p*-toluate being degraded via the *xylGHI* branch and *m*-toluate being degraded by the *xylF* pathway.

The best characterised of all of the TOL plasmids is pWWO which was first described in 1974 (Williams and Murray, 1974) in the strain *P.putida* mt-2. This is a large catabolic plasmid with a size of around 117 Kbp about 40Kbp of which is needed for the catabolic pathway and the regulatory genes. Although this plasmid has been regarded as the archetypal TOL plasmid it is slightly unusual in that it expresses a single *xylE* gene whilst other TOL plasmids have two of these genes (Chatfield and Williams, 1986). In pWWO the upper and lower pathway operons were shown by transposon mutagenesis to be separated by some 14 Kbp (Franklin *et al*, 1981). The order of the genes of the upper pathway was determined to be *xylCAB* (Harayama *et al*, 1986). However, subsequent analysis of the upper operon by gene cloning and maxicell analysis has suggested that there may in fact be five upper pathway genes (Harayama *et al*, 1989). The five genes are believed to occur in the order *xylCMABN* where *xylM* and *xylA* are believed to encode the sub-units of xylene monooxygenase (XO). The role of the XylN product has not, so far, been defined. The architecture of the lower pathway operon of pWWO has also been described. It has emerged that there may, in fact be thirteen genes in the *meta*-cleavage operon (Harayama and Rekik, 1990) rather than the nine which are represented in the catabolic scheme depicted in Figure 1.5. The "additional" genes are *xylX*, *Y* and *Z*, which are believed to represent the sub-units of the toluate 1,2-dioxygenase which was formerly designated XylD, and *xylT* and *Q* whose roles have not been determined.

Relatively few of the enzymes of this pathway have been purified so the enzymological picture is far from complete. Proteins which have been purified include the catechol 2,3-dioxygenase (XylE) (Nakai *et al*, 1983) and the 2-hydroxymuconic semialdehyde hydrolase (XylF) (Duggleby and Williams, 1986). Overall, however, the best characterised segment of the pathway is that involved in the dehydrogenative route (Harayama *et al*, 1989). This study included the purification of three enzymes 4-oxalocrotonate isomerase (XylH), 4-oxalocrotonate decarboxylase (XylI) and 2-oxopent-4-enoate hydratase (XylJ) which act consecutively. The most interesting findings of these investigations were that the sub-unit of XylH is exceptionally small with a molecular mass of 3,500 and that the XylI and XylJ proteins appear to form a complex *in vivo*. A possible explanation for the latter observation emerged from the characterisation of the intermediates of this part of the pathway. This showed that the substrate of the hydration reaction 2-hydroxypent-2,4-dienoate is unstable and it was suggested that the complex between the decarboxylase, which produces this compound, and the hydratase assures its efficient transformation *in vivo*.

The investigations of the ways in which aromatic compounds are catabolised by pWWO have also probed the way in which the pathway is regulated. It has emerged that the TOL genes are part of a tightly regulated system with control at the level of transcription. Two regulatory genes designated *xylR* and *xylS* have been identified which are transcribed from physically close but functionally divergent promoters. Transcription from the upper operon is positively regulated by XylR which is activated in the presence of toluene or its alcohol derivatives. Expression of the genes of the lower pathway is also positively regulated but in this case by XylS which is activated in the presence of effectors such as benzoates and toluates. Upper pathway substrates are , however, able to activate the lower pathway if both *xylR* and *xylS* are present. An important finding relating to operon control was that expression of the upper pathway genes in an *E.coli* host was increased if the the host *ntrA* gene was functional(Dixon,1986). The NtrA protein is a sigma factor which allows certain promoter sequences, for example those involved in nitrogen regulation, to be recognised by RNA polymerase. Analysis of the expression of nitrogen metabolism genes has suggested a

requirement for the involvement of an activator protein at such promoters. Typical activators are the NifA and NtrC from *K.pneumoniae*. The observation that either of these products could replace XylR for expression of the upper pathway genes suggested that XylR acted as a transcriptional activator. Promoters showing the appropriate sequence characteristics are found in the *xylS* and upper pathway operons and the expectation is that XylR activates transcription from these sites. Ramos *et al* (1987) proposed a model for the regulation of the TOL genes which is depicted in Figure 1.6. It was proposed that when a substrate such as toluene enters the cell it combines with, and activates XylR which is produced constitutively. The activated XylR binds to the *xylCAB* operon promoting binding of RNA polymerase containing the *ntrA* σ factor. This leads to the production of upper pathway products which induce the expression of the lower pathway genes by activation of XylS. However, XylR can also activate the lower pathway by promoting transcription at the *xylS* operon in concert with the *ntrA* σ factor. Evidence has been obtained which suggests that simple overproduction of XylS is sufficient to induce lower pathway expression even in the absence of the usual inducers. The fact that XylS can activate in the presence and absence of inducers suggested that XylS might actually exist in a dynamic equilibrium between active and inactive forms. Binding of inducer generates the active form from the inactive form whilst overproduction simply increases the total amount of active XylS.

Expansion of the range of compounds which can be degraded by an aromatic catabolic pathway

One possible way in which the processes of aromatic catabolism may be exploited is through the expansion of the range of compounds which may be degraded. For example it might well be possible to construct an organism to degrade otherwise recalcitrant compounds. In this section the attempts which have been made to expand the substrate range of the TOL pathway will be described. Two general strategies can be envisaged for the experimental evolution of new catabolic activities (Ramos *et al*, 1987). These involve either the assembly of a new route through the combination of appropriate

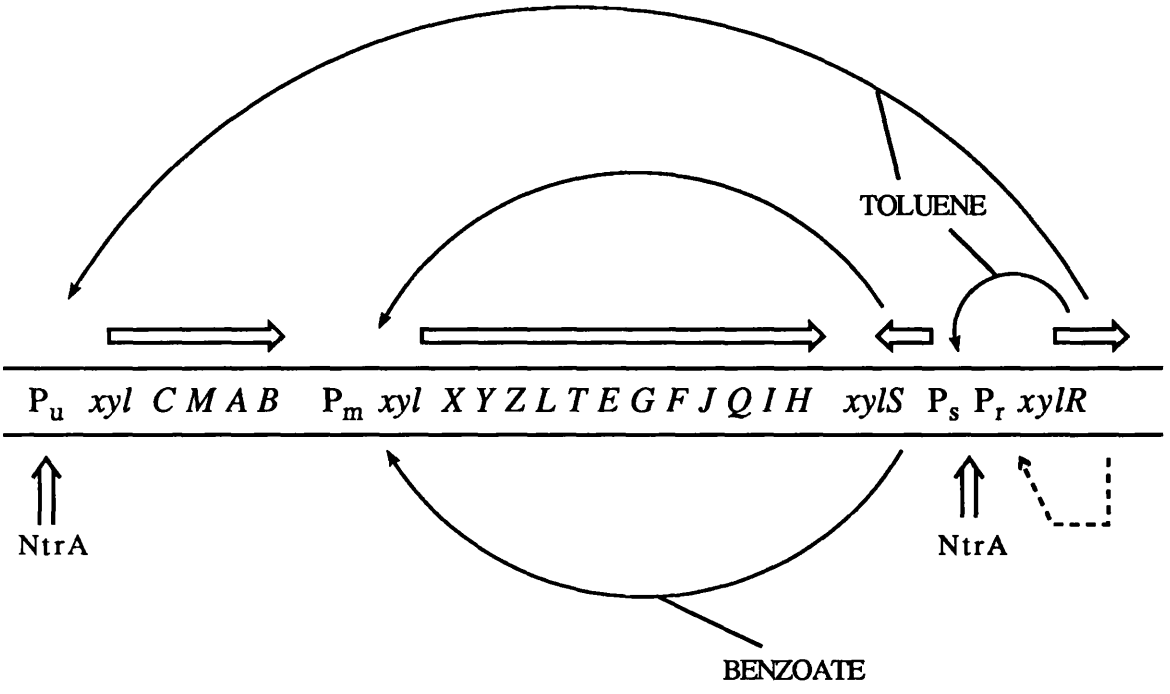


Figure 1.6

Proposed model for gene regulation in pWWO. Open arrows above gene blocks indicate direction of transcription. Dotted lines indicate negative regulatory circuits whilst unbroken lines indicate positive circuits. Promoters requiring the presence of the NtrA protein are marked. Toluene combines with the XylR protein to affect two promoters, P_u and P_s. Increased production of XylS can then activate P_m. Alternatively, benzoate in the presence of the constitutive amount of XylS is able to activate P_m. The *xylR* gene has been shown to be subject to autoregulation.

Adapted from Burlage *et al* (1989)

sections of different pathways or the restructuring of an existing pathway.

The former approach is exemplified by the construction of a *Pseudomonas* sp. strain B13 derivative which was able to grow on 4-chlorobenzoate (4-CB) (Reineke and Knackmuss, 1979). The parent strain was able to metabolise 3-chlorobenzoate (3-CB) but not 4-chlorobenzoate. This was shown to be the result of the narrow substrate specificity of the toluate 1,2-dioxygenase (TO) which is the first enzyme of the pathway (Reineke and Knackmuss, 1978). In contrast the TO encoded by the TOL plasmid pWWO had a relaxed substrate specificity . The introduction of the TOL plasmid into strain B13 would appear to provide the necessary function which would allow metabolism of 4-CB. There was however, a further complication which necessitated mutant selection on 4-CB following the introduction of pWWO before an organism could be isolated which was able to grow on 4-CB. This was that whilst the 4-chlorocatechol generated from the 4-CB by the TO can be metabolised completely by the chromosomal *ortho*-pathway metabolism via the plasmid encoded *meta*-pathway leads to the production of lethal intermediates. Once inactivation of the *meta*-pathway had resulted then the new combination of functions could allow growth on 4-CB .

The archetypal TOL plasmid pWWO has also been manipulated to allow development of systems which are able to metabolise certain alkylbenzoates and alkyltoluenes. In this instance the pathway modification involves the acquisition of new enzymic and regulatory specificities through the mutational alteration of existing proteins. Initially attempts focused on the isolation of organisms able to utilize 4-ethylbenzoate (4-EB) (Ramos *et al*, 1987). Analysis of the critical steps in the TOL pathway that prevent 4-EB metabolism revealed two main factors. The first of these was that 4-EB fails to induce synthesis of the catabolic enzymes and the second was that one of the metabolic intermediates irreversibly inactivates the catechol 2,3-dioxygenase (C2,3O) which mediates ring fission. Therefore 4-EB catabolism would require broadening of the effector range of the regulator protein XylS and the generation of a C2,3O resistant to inactivation. Regulator mutants were selected by the use of a vector in which the P_m promoter of the TOL *meta*-cleavage operon was fused upstream of a promoterless tertracycline resistance gene. Mutant *xylS* alleles were

identified from cells which showed 4-EB induced tetracycline resistance when they contained this plasmid and one encoding the *xylS* gene. The mutant C2,3O was then isolated by plating of *P.putida* cells harbouring pWWO and the plasmid bearing the *xylS* 4EBmutant onto 4-EB plates. A plasmid designated pWWO-EB was isolated from the cells able to grow on 4-EB which contained the mutant *xylE6* allele.

This work was extended by further manipulation of the TOL pathway to allow utilisation of 2-ethyl and 3- and 2-chlorotoluene by the construction of a plasmid designated pWWO-EB62 (Abril *et al*, 1989). In this instance toluene oxidase was again identified at the bottleneck which prevented degradation of these compounds. The mutant TO is able to oxidise 2-ethyltoluene to 2-ethylbenzyl alcohol which can be further transformed to 2-ethylbenzoate by TOL plasmid encoded dehydrogenases. The 2-ethylbenzoate can of course be metabolised by the modified pathway constructed by Ramos *et al*, (1987). The three mutations which allow 2-ethyltoluene degradation each occurred at a frequency of 10^{-8}-10^{-9} per cell per generation making it unlikely that a plasmid similar to pWWO-EB62 could be isolated from nature.

The potential of combining the approaches of the modification of existing pathways with the introduction of new pathway elements into cells was also demonstrated.Introduction of pWWO-EB62 into *Pseudomonas* sp. strain B13 resulted in the eventual isolation of cells able to grow on 3-chlorotoluene.

It is apparent from the studies described above that directed evolution of aromatic catabolism requires a detailed knowledge of existing pathways. Only through a thorough understanding of both the biochemistry and genetics of pathway elements does the approach become feasible. In the case of the TOL pathway the prospects are enhanced by the flexibility of pathway induction and the very broad substrate specificity of several pathway enzymes. It is to be hoped that as more information becomes available about other pathways that they may be exploited ultimately for the removal of otherwise recalcitrant compounds.

The degradation of 3-hydroxybenzoate (3-HB)

Two main pathways have been described for the aerobic bacterial catabolism of 3-HB (Yano and Arima, 1958). One involves ring hydroxylation *para* to the existing hydroxyl group to yield 2,5-dihydroxybenzoate (gentisate) which then enters the gentisate pathway which is described in more detail in the following section. Purification of the 3-HB-6-monooxygenase has been reported on two occasions (Groseclose and Ribbons, 1973; Wang *et al*, 1987). In each case it appears that hydroxylation may be achieved using either NADH or NADPH as cofactor. In the second pathway hydroxylation occurs at the C4 position of the aromatic ring resulting in the production of the ring-fission substrate 3,4-dihydroxybenzoate (3,4-DHB; protocatechuate). Protocatechuate may then be degraded by either *ortho*- or *meta*-fission routes In organisms such as *Pseudomonas testosteroni* ring *meta*-fission is mediated by protocatechuate 4,5-dioxygenase and subsequent metabolism produces pyruvate and formate. The analysis of 3-HB metabolism in two *Pseudomonas* species, both members of the *acidovorans* group suggested that the precise route of catabolism had taxonomic significance (Wheelis et *al*, 1967). Whilst *P.testosteroni* metabolised 3-HB via 3,4-DHB, *P.acidovorans* employed the gentisate pathway. These metabolic schemes are depicted in Figure 1.7. Interestingly, although *P.testosteroni* apparently synthesises no gentisate from 3-HB , low levels of gentisate 1,2-dioxygenase and subsequent pathway enzymes are expressed (Harpel and Lipscomb, 1990a). The precise role for the gentisate pathway in *P.testosteroni* has not so far been defined although preliminary experiments have suggested that it may be involved in the metabolism of methoxy-substituted aromatic compounds.

The gentisate (2,5-dihydroxybenzoate; 2,5-DHB) pathway

A wide range of aromatic compounds have been shown to be degraded by catabolic routes in which 2,5-DHB serves as the ring fission substrate. Catalogues of such compounds (e.g Harpel and Lipscomb, 1990a) include both single and multiple ring species. A selection of these compounds are included in Table 1.1. Examination of

Figure 1.7

Scheme depicting routes for the metabolism of 3-hydroxybenzoate by *P.testosteroni* and *P.acidovorans* (Harpel and Lipscomb, 1990a)

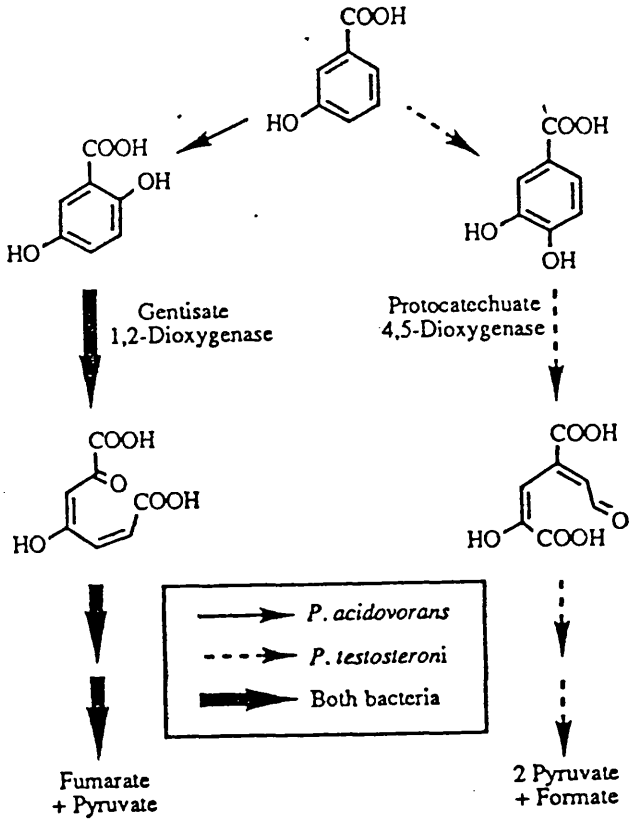


Table 1.1

Compounds degraded via 2,5-dihydroxybenzoate (gentisate) from Harpel and Lipscomb (1990a).

Single ring compounds	
Hydroxybenzoates	2-hydroxybenzoate (salicylate) 3-hydroxybenzoate 4-hydroxybenzoate
Aminobenzoates	2-aminobenzoate
Cresols	3-hydroxytoluene (*m*-cresol)
Multiple ring compounds	
Flavonones	
Naphthalene derivatives-	β-naphthol naphthalene disulphonates

the catabolism of compounds including 2,5- and 3,5-xylenol has suggested that these compounds are degraded respectively via the substituted gentisates 4-methylgentisate and 3-methylgentisate.
Disruption of the aromatic ring of 2,5-DHB is achieved under aerobic conditions by the action of 2,5-DHB 1,2-dioxygenase. The compound generated by this ring fission event is maleylpyruvate. This compound is hydrolysed directly by some species to produce pyruvate and maleate (Hopper *et al*, 1968; Hopper *et al*, 1971). In other organisms the maleylpyruvate undergoes isomerisation to fumarylpyruvate which is subsequently hydrolysed to give fumarate and pyruvate (Lack, 1959). The alternative pathways for maleylpyruvate degradation are shown in Figure 1.8. It has been observed that organisms which oxidise aromatic substrates to substituted gentisates tend to utilise direct hydrolytic fission of maleylpyruvate rather than isomerisation. In Gram-negative organisms the isomerisation of maleylpyruvate requires the presence of reduced glutathione (GSH) as a cofactor. In a number of Gram-positive organisms including *Bacillus megaterium* GSH is not required (Hagedorn, 1980). Purification of the various proteins involved in the gentisate pathway has been reported. The best characterised protein is undoubtedly the gentisate 1,2-dioxygenase which has been purified both from a *Bacillus* sp. (Crawford *et al*, 1975) and from both *P.testosteroni* and *P.acidovorans* (Harpel and Lipscomb, 1990a). The findings of a more detailed analysis of the enzymes from *Pseudomonas* have been referred to in the section on ring-fission dioxygenases.
Investigations of the metabolism of 3-HB by *Klebsiella pneumoniae* has revealed that the degradation of this compound occurs via the gentisate pathway with isomerisation of maleylpyruvate preceding hydrolysis (Jones, 1985;Jones and Cooper, 1990). This pathway was deduced through the analysis of the patterns of substrate oxidation, enzyme assays of cell-free extracts of both wild-type and mutant cells and through the isolation of pathway intermediates. The reaction scheme is shown in Figure 1.9. It has been suggested that the isomerisation of maleylpyruvate (compound **III**) generates predominantly the enol-form of fumarylpyruvate (**IV**) rather than the keto-form (**V**). This is because compound **IV** is stabilised through its arrangement of conjugated bonds. However it is likely that the substrate for the fumarylpyruvate hydrolase reaction is in fact the

Figure 1.8
Possible routes for the metabolism of 2,5-dihydroxybenzoate (gentisate) following ring fission to maleylpyruvate

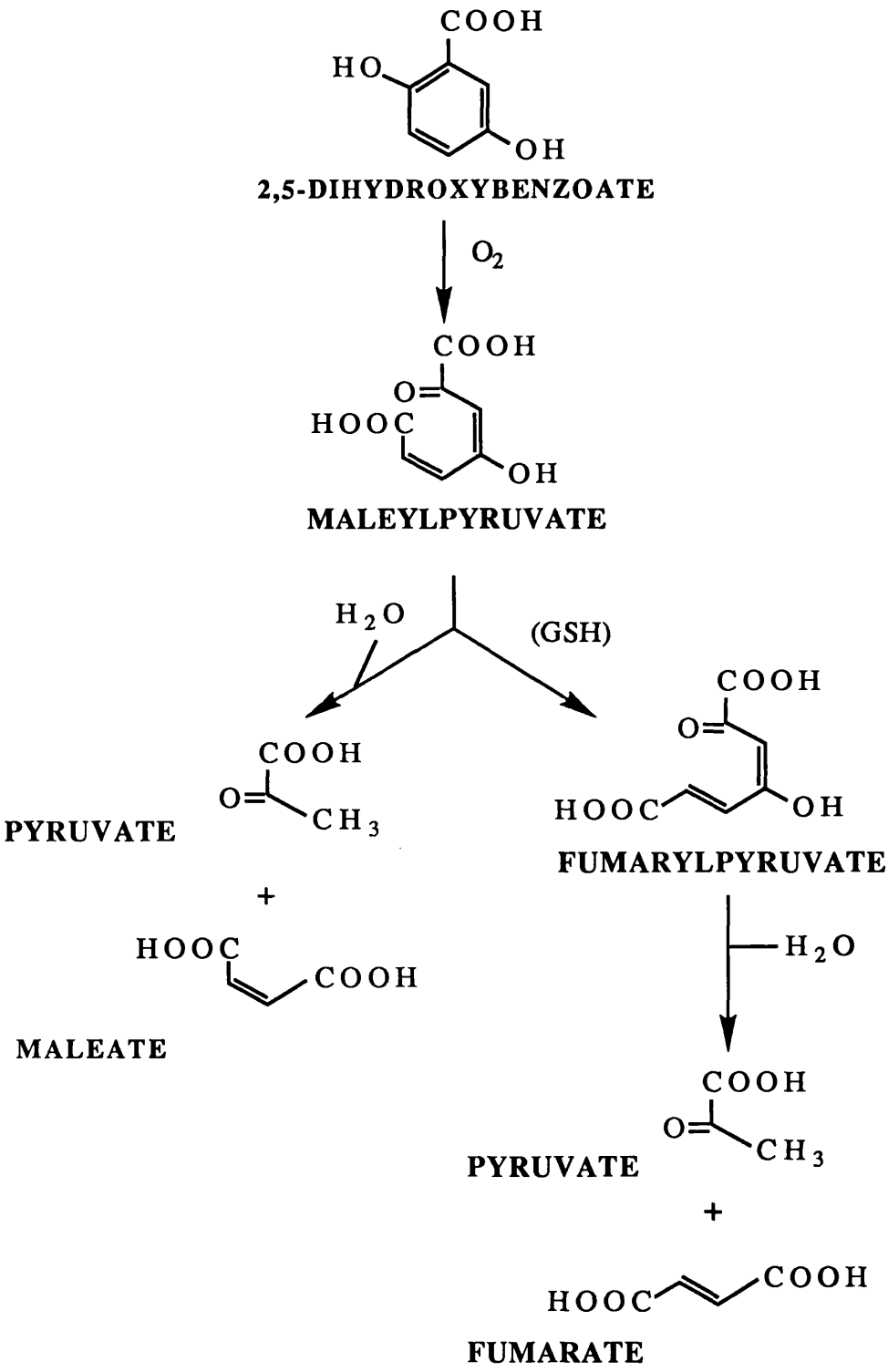
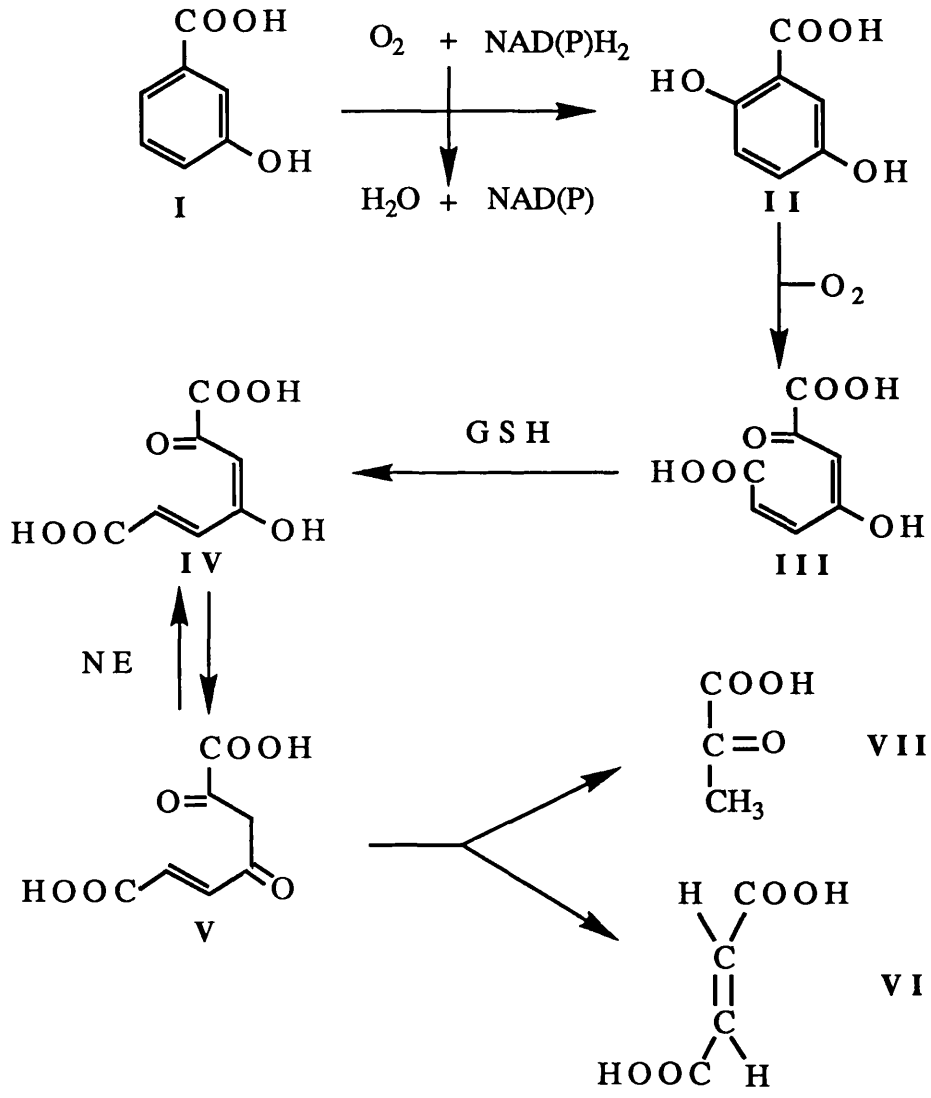


Figure 1.9

The gentisate pathway for 3-hydroxybenzoate catabolism. 3-Hydroxybenzoate (**I**) is converted to pyruvate (**VII**) and fumarate (**VI**) via 2,5-dihydroxybenzoate [gentisate] (**II**), maleylpyruvate [enol] (**III**), fumarylpyruvate [enol] (**IV**) and fumarylpyruvate (**V**) by the sequential action of 3-hydroxybenzoate 6-monooxygenase, 2,5-dihydroxybenzoate dioxygenase, maleylpyruvate isomerase, a non-enzymic (**NE**) keto-enol tautomerisation and fumarylpyruvate hydrolase.



keto-form. It has thus been postulated that the enol-form of fumarylpyruvate undergoes a non-enzymic (NE) tautomerisation prior to hydrolysis.

Induction of the 3-HB catabolic enzymes was also examined. Analysis of wild-type cells of *K.pneumoniae* revealed that either 3-HB or 2,5-DHB could induce expression. However, the analysis of a mutant defective for the *mhbM* gene encoding 3-HB monooxygenase indicated that expression of the remaining enzymes was induced by 2,5-DHB but not by 3-HB. This suggested that 3-HB was not able to function as an inducer for all of the *mhb* genes if further metabolism was prevented. In contrast 2,5-DHB was able to induce all of the *mhb* genes in the systems examined. This suggested that the induction of the *mhb* genes required either 2,5-DHB or a subsequent intermediate. The fact that induction occurred in mutant strains defective for the *mhbI* gene encoding maleylpyruvate isomerase suggested that fumarylpyruvate was not the inducing compound.

In contrast with many other aromatic catabolic systems there has been very little analysis of the 3-HB degradative system at the genetic level. Jones and Cooper (1990) were able to demonstrate close linkage of the *mhbM* and *mhbI* genes in *K.pneumoniae* through transductional analysis. This investigation revealed that transduction of a mutant defective in both of the genes with selection for growth on 2,5-DHB resulted, in all cases tested, in the simultaneous acquisition of the ability to grow on 3-HB. The close linkage suggested that it might well be possible to obtain clones containing multiple *mhb* genes on small restriction fragments. The methodology employed to obtain clones including *mhb* genes from *K.pneumoniae* involved purification of the maleylpyruvate isomerase and the subsequent screening of genomic libraries using an oligonucleotide derived from N-terminal amino acid sequence. This resulted in the isolation of a clone which expressed two of the catabolic functions namely 3-hydroxybenzoate-6-monooxygenase (MhbM) and maleylpyruvate isomerase (MhbI). This plasmid, designated pSP01, represents the starting point for the analysis of the *mhb* genes from *K.pneumoniae* described in subsequent chapters.

The homoprotocatechuate (HPC) pathway

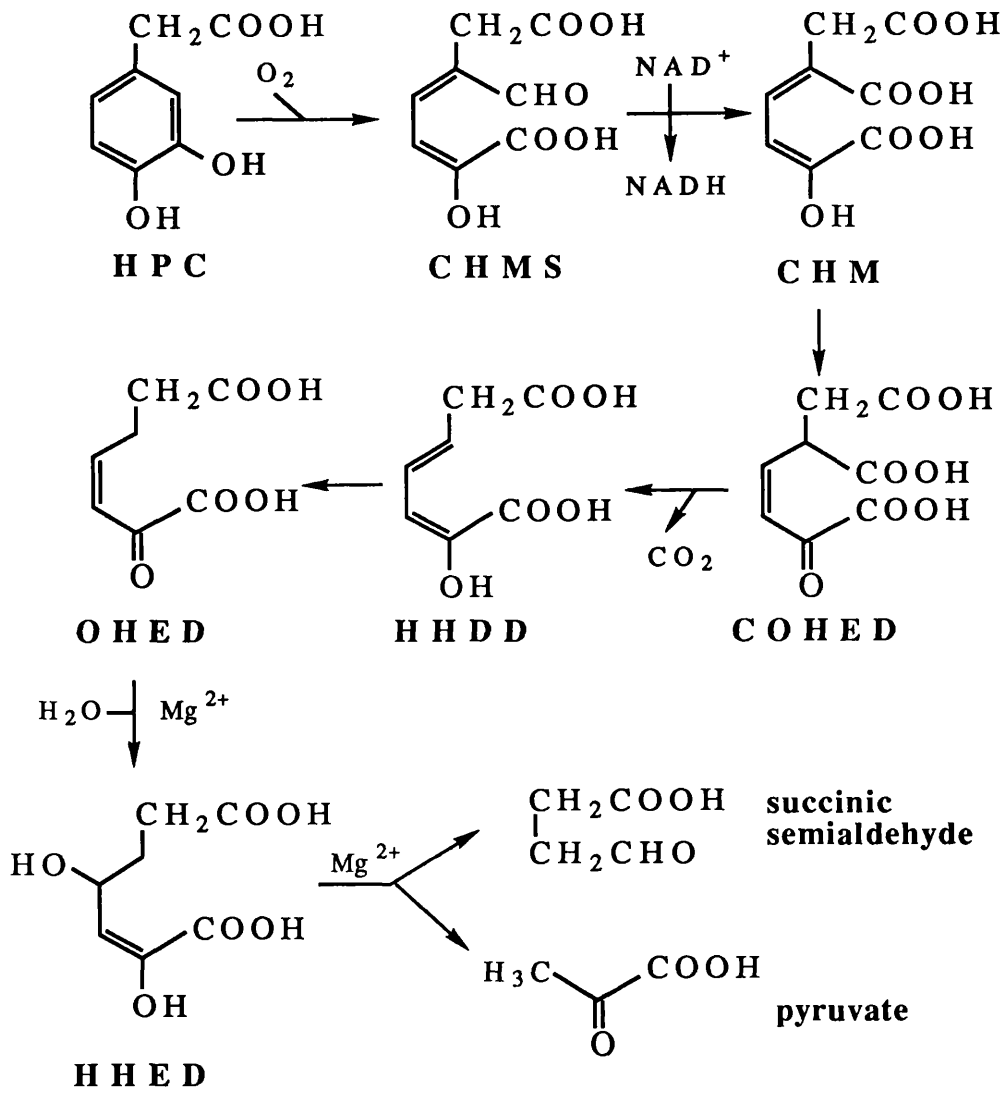
Homoprotocatechuate (HPC; 3,4-dihydroxyphenylacetate) was shown to be an important intermediate in aromatic catabolism during the investigation of 4-hydroxyphenylacetate (4-HPA) metabolism by *Pseudomonas ovalis* (Adachi *et al*, 1964). HPC is, of course, a classical substrate for ring fission because it is dihydroxylated. The HPC dioxygenase was crystallised from *P.ovalis* by Kita *et al* (1965) and it was suggested that HPC is cleaved by a *meta*-fission reaction to give 5-carboxymethyl-2-hydroxymuconate semialdehyde (CHMS). This was confirmed by Sparnins *et al* (1974) during the investigation of 4-HPA catabolism by two species of *Pseudomonas* and one species of *Acinetobacter*. It was demonstrated that in these organisms the 4-HPA is ultimately metabolised to pyruvate and succinic semialdehyde. Sparnins *et al* also identified 5-carboxymethyl-2-hydroxymuconate (CHM), 2-hydroxyhepta-2,4-diene-1,7-dioate (HHDD) and 4-hydroxy-2-ketopimelic acid (HKP) as intermediates of the pathway. Dagley (1975) proposed a pathway for the dissimilation of HPC which included an NAD^+ dependent dehydrogenation followed by three keto-enol isomerisations. This scheme remained in place until experiments involving the analysis of the cloned genes for HPC degradation from *E.coli* C were performed (Jenkins and Cooper, 1988). It emerged from the study of sub-clones which could be used to convert HPC to selected pathway intermediates that there were only two isomerisation steps prior to the aldol fission which produced pyruvate and succinic semialdehyde. Nuclear magnetic resonance studies suggested that the substrate for the aldol cleavage reaction was in fact the enol-compound 2,4-dihydroxy-hepta-2-ene-1,7-dioate (HHED) rather than keto-compound HKP. The scheme for the catabolism of HPC by *E.coli* C proposed by Jenkins and Cooper is depicted in Figure 1.10.

Analysis of the HPA degradative pathway of enteric bacteria

Several members of the Enterobacteriaceae have been found to be able to utilise 3- and 4-HPA. These organisms include *E.coli* strains B, C and W and *Klebsiella pneumoniae*. These hydroxyphenylacetates are

Figure 1.10

Pathway for the degradation of homoprotocatchuate (HPC) in *Escherichia coli* C (Jenkins and Cooper, 1988)



Key to compounds:
HPC = homoprotocatechuate (3,4-dihydroxyphenylacetate)
CHMS = 5-carboxymethyl-2-hydroxymuconate semialdehyde
CHM = 5-carboxymethyl-2-hydroxymuconate
COHED =5-carboxymethyl-2-oxo-hex-3-ene-1,6-dioate
HHDD = 2-hydroxyhepta-2,4-diene-1,7-dioate
OHED = 2-oxo-hepta-3-ene-1,7-dioate
HHED = 2,4-dihydroxy-hepta-2-ene-1,7-dioate

further hydroxylated in these organisms to give HPC which is then metabolised by the enzymes of the HPC pathway. Aldol fission of HHED, the terminal intermediate of the HPC pathway, generates pyruvate and succinic semialdehyde (SSA). The oxidation of SSA to produce succinate requires the action of the NAD-linked SSA dehydrogenase which is the product of the *sad* gene (Donnelly and Cooper, 1981). Analysis of the induction of the components involved in the degradation of HPA has suggested that there are three distinct regulatory groups which are defined as follows:

Group 1: Genes involved in the uptake and hydroxylation of 3- and 4-HPA *hpaA* (permease), *hpaB* (hydroxylase) and *hpaR* (regulator)

Group 2: Genes involved in the catabolism of HPC to pyruvate and SSA. Expression of the genes *hpcB* to *hpcH* is induced by HPA and HPC and is controlled by a specific regulator encoded by *hpcR*

Group 3: SSA dehydrogenase encoded by the *sad* gene. Expression of this gene is induced by SSA (Donnelly and Cooper, 1981).

Analysis of the *hpc* genes and associated gene products

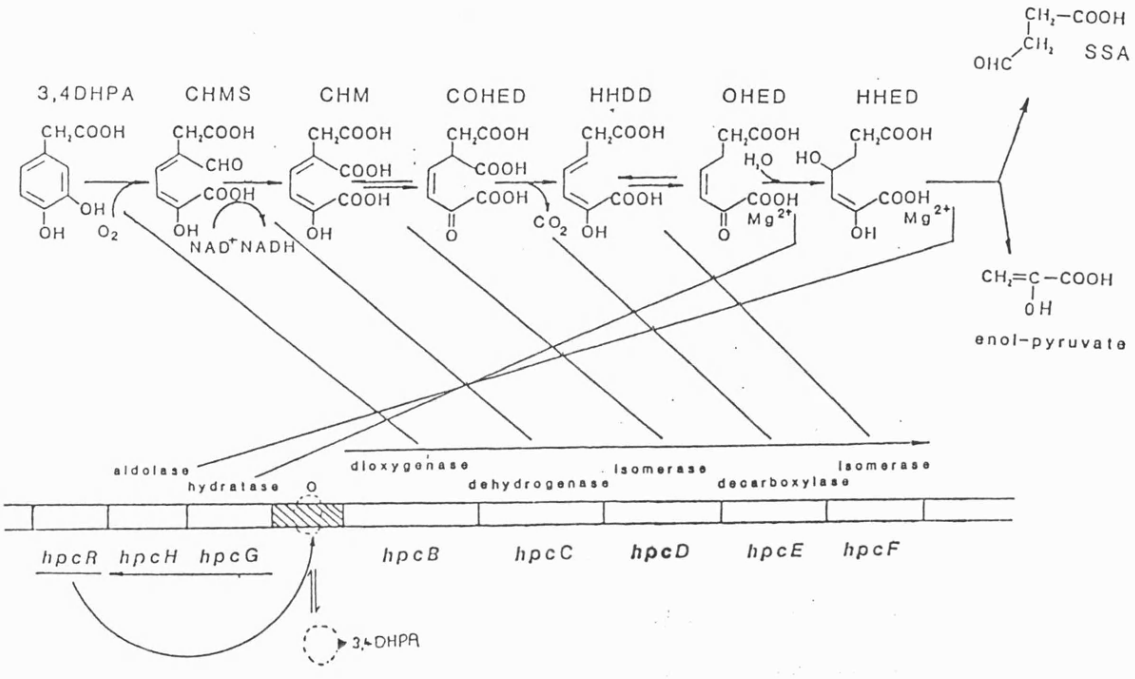
Genomic clones have been isolated which include *hpc* genes from *E.coli* C (Jenkins, 1987; Jenkins and Cooper, 1988) and from *K.pneumoniae* M5a1 (Fawcett, 1989; Martin *et al*, 1991). Studies aimed at the determination of the organisation of the genes from *K.pneumoniae* are described in Chapter 6 of this thesis. These represent the first stages of the analysis of the *hpc* genes at the genetic level in this organism. In contrast the *hpc* genes from *E.coli* C have now been particularly well studied. Three genomic clones which express some or all of the *hpc* genes were isolated from *Sau*3A libraries of *E.coli* C DNA in the vectors pBR322 and pBR328. Two of these clones designated pJJ200 and pJJ210 were found to include all of the *hpc* catabolic genes including the regulatory *hpcR* gene. However *E.coli* K-12 strains which do not themselves express any of the *hpc* genes were unable to grow on 3- or 4-HPA or on HPC when harbouring pJJ200 or pJJ210. This implies the absence of some important function from these clones. It was suggested that uptake of the aromatic compounds by the cells

might represent a limiting factor (Fawcett, 1989). The demonstration of the presence of the *hpcR* gene through the requirement for the presence of 4-HPA to induce expression implies that some import of aromatic compounds occurs. The third genomic clone designated pJJ801 was shown to express only the genes involved in the first five reactions of the HPC pathway allowing conversion of HPC to OHED. It was also demonstrated that the *hpc* gene expression in *E.coli* K-12 strains harbouring this plasmid was constitutive whereas inducible expression of the *hpcC* gene was observed when pJJ801 was introduced into the *E.coli* C strain 221 (defective in *hpcC*). This demonstrated that the plasmid was able to respond to regulator protein supplied in *trans* by the host and suggested that the clone lacked the *hpcR* gene. Restriction mapping of the clones allowed more precise localisation of the *hpc* gene cluster. The restriction mapping of the genomic clones provided the basis for the sub-cloning experiments which allowed a model to be proposed for the organisation of the *hpc* genes. This analysis suggested that the HPC catabolic genes were arranged in two blocks whose expression was controlled by a pair of centrally arranged divergent promoters. Transcription of the gene blocks thus occurred in opposite directions. The first block contained five genes arranged in the order *hpcBCDEF* whilst the second comprised the two remaining catabolic genes arranged *hpcGH*. The regulatory gene, *hpcR*, was placed downstream of *hpcGH*. The model for the genetic organisation is depicted in Figure 1.11.

The model proposed by Jenkins has now been revised in the light of further experiments performed by Fawcett (1989) and by Roper (1990). The first notable difference highlighted by these studies was that it was possible to demonstrate that all of the *hpc* catabolic genes were transcribed in the same direction. This rendered the divergent promoter model untenable. This conclusion was based on a combination of sub-cloning experiments in conjunction with Southern blot and nucleotide sequence analysis which localised the 5' terminus of each of the *hpcC* and *hpcG* genes. The significance of the latter aspect was that it defined the orientation of these genes thus providing reference points for further analysis. The model proposed by Fawcett retains the idea of there being two blocks of the catabolic genes. The order of the genes in this model was *hpc(EF)CD* in the first block followed by *hpcBGH* with overall transcription proceeding from

Figure 1.11

Organisation of the *hpc* genes from *Escherichia coli* C (Jenkins and Cooper,1988). Arrows indicate the predicted directions of transcription of the gene blocks which is divergent from a centrally located operator/promoter region.



KEY TO GENES

hpcB= HPC dioxygenase

hpcC= CHMS dehydrogenase

hpcD= CHM isomerase

hpcE= COHED decarboxylase

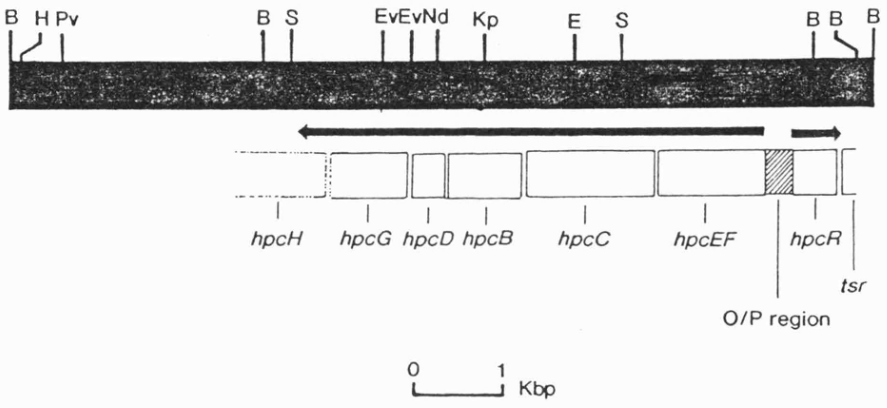
hpcF= HHDD isomerase

hpcG= OHED hydratase

hpcH= HHED aldolase

Figure 1.12

Revised model for the organisation of the *hpc* genes from *E.coli* C (Roper, 1990). Arrows indicate the direction of transcription from the operator/promoter region (O/P).



KEY TO GENES

hpcB= HPC dioxygenase

hpcC= CHMS dehydrogenase

hpcD= CHM isomerase

hpcE= COHED decarboxylase

hpcF= HHDD isomerase

hpcG= OHED hydratase

hpcH=HHED aldolase

tsr= serine chemoreceptor

Key to restriction enzyme abbreviations.

B= *Bam*HI; E=*Eco*RI; EV=*Eco*RV; Kp=*Kpn*I; N=*Nde*I; Pv=*Pvu*II; S=*Sal*I

hpc(EF) through to *hpcH*. In the studies conducted by Fawcett there was no data which allowed the relative positions of the *hpcE* and *hpcF* genes to be determined. This model was in broad accord with Northern blot analysis performed by Fawcett which suggested that there were transcripts with sizes of 4.5, 2.7 and 1.6Kbp that were produced from the insert present in pJJ200. The first two were suggested to belong to the catabolic gene units with the third attributed to the *hpcR* gene.

The further refinement of the model for the organisation of the *hpc* genes of *E.coli*C continued with the work of Roper (1990). This provided the most precise evidence since it was based upon nucleotide sequencing of several of the *hpc* genes. Interpretation of the nucleotide sequencing information was made possible because of the availability of N-terminal sequence for many of the pathway enzymes. The purifications of the HpcB,C,D and G proteins have been described (respectively Roper and Cooper,1990b; Fawcett *et al*, 1989; Roper and Cooper, 1990a and Ferrer and Cooper, 1988). The inability of Fawcett to determine the relative positions of the *hpcE* and *F* genes has been explained by further analysis by Roper (1990) which concluded that both functions are encoded in a single gene. This was based upon three main lines of evidence.The first of these was that it was not possible to generate sub-clones using unidirectional deletion methods which expressed one or other but not both of the activities attributed to the HpcE and F products. This is somewhat indirect evidence but the finding was confirmed when it was shown that it was possible to purify a single protein with a relative molecular mass of 44,000 which expressed both activities. Nucleotide sequencing around the expected location of *hpc(EF)* identified a single open reading frame which would gave rise to a protein similar in size to that which was purified. The nucleotide sequence analysis also suggested that the relative positions of the *hpcB* and *hpcD* genes had been incorrectly determined previously. In the model proposed by Roper which is depicted in Figure 1.12 the overall gene order is proposed to be *hpcR, hpc(E/F)CBDGH*. The status of the transcriptional units proposed by Fawcett is uncertain in the light of the revised gene order. Transcript mapping identified a promoter sequence occurring upstream of *hpc(EF)* but no additional promoter regions have been so far identified. Comparison of the models shown in Figures 1.11 and

in Figures 1.11 and 1.12 indicates that the location of the regulator gene *hpcR* has been re-evaluated as the result of subsequent sub-cloning experiments. The HpcR protein was originally proposed by Jenkins (1987) to be a repressor and subsequent investigations have confirmed the hypothesis that the *hpc* genes are negatively regulated. The investigation of the HPC degradative system of *K.pneumoniae* began with the purification of the HpcC protein (CHMS dehydrogenase)(Fawcett *et al*, 1989). Comparison of the basic characteristics of the protein from *K.pneumoniae* M5a1 and from *E.coli* C indicated very similar affinities for substrates and sub-unit molecular masses. The enzymes also displayed similar pH optima. The most striking comparison however, was that the two proteins displayed 90% identity over the first twenty amino-terminal residues. A similar degree of identity was observed when the amino terminal sequences of the HpcG (OHED hydratase) proteins were compared (R.A Cooper pers.comm). These results suggested that the *hpc* genes in the two organisms might be closely related. However, the fact that restriction fragments from the clones from *E.coli* C including *hpc* genes had not hybridised to *K.pneumoniae* chromosomal DNA suggested that overall the degree of conservation across the whole gene cluster might not be very high. Probing of various chromosomal DNAs digested with *Bam*HI with an oligonucleotide corresponding to a portion of the amino terminal sequence of HpcC showed that the hybridising fragment from *K.pneumoniae* was appreciably smaller than that from *E.coli* C. This indictated that the restriction maps of the *hpc* gene clusters in each of the two organisms were not identical.

In order to investigate the organisation of the *hpc* genes from *K.pneumoniae* cloning of the cluster would be required. Once this had been accomplished then more detailed comparisons could be made between the gene clusters of the two organisms and between various aromatic pathways of *K.pneumoniae*. Two separate instances in which *hpc* genes from *K.pneumoniae* were cloned have been reported (Fawcett, 1989; Martin *et al*, 1991). In both instances very similar cloning strategies were employed. In each case oligonucleotides representing portions of the amino-terminus of the *hpc*C gene were used to screen genomic libraries of *K.pneumoniae* which had been constructed in pBR-based vectors. Fawcett had constructed a library which incorporated *Hin*dIII fragments as this enzyme had been shown

to give rise to a large hybridising fragment. The benefit of this being the increased probability of cloning several or all of the *hpc* genes on a single fragment. The clones isolated by Fawcett (1989) which were designated pTF100 and pTF102 both included a 14 Kbp insert and both appeared to express all of the *hpc* genes including *hpcR*. However, as with the plasmids pJJ200 and pJJ210 derived from *E.coli* C, cells of the *E.coli* K-12 strain 5K harbouring pTF100 or 102 were unable to grow on 3- or 4-HPA or HPC as sole carbon and energy source. Fawcett proceeded to show that the restriction patterns of pTF100/102 were very similar but not identical and that they were quite different from those of the clones from *E.coli* C. Fawcett did not carry out any further analysis of the organisation of the *hpc* genes carried by pTF100/102. An analysis of the organisation of the genes present on pTF100 is described in Chapter 6. In the case of Martin *et al* the genomic library was constructed using the endonuclease *Bam*HI. The clone isolated by this group was reported to contain a 7 Kbp *Bam*HI and to express the *hpcB*, *C*, *D*, *E* and *G* genes constitutively.The size of the cloned fragment was not in accord with Southern blot analysis using an almost identical probe which had suggested that the hybridising *Bam*HI fragment had a size of approximately 4.5 Kbp. The organisation of the *hpc* genes on this particular clone has not so far been reported .

Aims of the project

The basic aim of the project was to initiate a detailed molecular analysis of aromatic degradation by *Klebsiella pneumoniae* through the analysis of the systems involved in the degradation of 3-hydroxybenzoate and 3,4-dihydroxyphenylacetate.

The analysis of the 3-HB catabolic pathway would require the analysis of the existing partial clone pSP1 and the subsequent design and execution of a cloning strategy which would allow the complete pathway to be obtained. Basic characterisation of the new clone(s) would follow and would include the determination of the order of the genes. Interest would also focus on the nature of the regulation of the genes of the pathway. The dioxygenase involved in the gentisate pathway, gentisate 1,2-dioxygenase, has been purified from several sources but no nucleotide sequence has so far determined. As this dioxygenase is somewhat different from the standard *ortho*- and *meta*-fission enzymes nucleotide sequencing of the dioxygenase gene was proposed.

Analysis of the existing clones of the HPC pathway would concentrate primarily on the determination of the gene order so that comparisons could be made with that deduced for *Escherichia coli* C.

CHAPTER 2

Materials and methods

Materials and methods

Bacterial strains and vectors and plasmids used in this study are given in Tables 2.1 and 2.2 respectively.

Table 2.1 Bacterial strains used

STRAIN	GENOTYPE	SOURCE
Escherichia coli C	prototroph	Lab.stock
E.coli C221	*hpcC*, *recA*	Jenkins, 1987
E.coli 5K	F$^-$, *supE*44, *tonA*21, *hsdR*, *rpsL*, *thr*-1, *leu-B*6, *thi*-1, λ$^-$	Lab. stock Hubacek and Glover (1970)
E.coli NM522	*supE*, *thi*, *hsd* 5, (*lac-proAB*), [F' *proAB lacI*^{q} ΔZM15]	Lab.stock Gough and Murray (1983)
E.coli JM101	*thi*, (*lac-pro*AB), [F', *traD*36, *proAB*, *lac* I^q ΔZM15]	Yanisch-Perron *et al* (1985)
Klebsiella pneumoniae M5al	prototroph	Lab. stock
K.pneumoniae MI214	*mhbI*	Jones and Cooper (1990)

Table 2.2 Vectors and plasmids used in this study

VECTORS

Plasmids

Vector	Relevant features	Reference
pBR322	Ap^R, Tc^R	Soberon *et al*, 1980
pBR328	Ap^R, Tc^R, Cm^R	Soberon *et al*, 1980
pUC18/pUC19	Ap^R, *lacZ'*	Yanisch-Perron *et al* (1985)

Bacteriophage M13 based

	Reference
M13 mp18/mp19	Yanisch-Perron *et al*, 1985

Clones/Sub-clones used

Clones of *hpc* genes from *K.pneumoniae* M5a1

Clone	Vector	Genes Encoded	Reference
pTF100/102	pBR322	*hpcB,C,D,E,F,G,H,R*	Fawcett, 1989
pJF1	pBR328	*hpcB,C,D,G*	Fernandez/Cooper (pers.comm)

Clones of *hpc* genes from *E.coli*C

Clone	Vector	Genes Encoded	Reference
pJJ002	pBR328	*hpcB*	Jenkins, 1987
pJJ003	pBR328	*hpcBC*	"
pJRJ003	pBR328	*hpcBCDEF*	"

pJJ801	pBR328	*hpcBCDEFG*	"
pDR9304	pUC19	*hpcB*	Roper, 1990

Clones of *mhb* genes from *K.pneumoniae* M5a1

Clone	Vector	Genes Encoded	Reference
pSP1	pBR328	*mhbMI*	S.Parrott and R.A.Cooper (pers.comm)
pNDR20	pBR328	*mhbMDIHR*	THIS STUDY

Growth media and conditions

Bacterial cells were grown in either complex or minimal media. Complex media (Luria broth (LB)) was as described by Miller (1972) and minimal media was as described by Hareland *et al* (1975). For the preparation of bacteriophage DNA bacterial cells were grown on the complex medium 2YT(Maniatis et al, 1982) and plated out using R-top agar. Cells prepared for measurement of enzyme activities were grown at 30°C whereas for DNA preparations they were grown at 37°C. Liquid cultures were incubated in an orbital shaker at 200r.p.m. Where appropriate liquid media was solidified by the addition of 1% (w/v) Bacto-agar (Difco). Carbon sources and particular growth requirements were sterilised separately and added aseptically to the media to final concentrations of :

Aromatic carbon sources - 5mM, Glycerol - 20mM,

Thiamine - 10μg/ml , Amino acids- 80μg/ml

Liquid minimal cultures were also supplemented with LB to a final concentration of 0.025% (w/v).

Minimal medium cultures were routinely grown by inoculating flasks with 0.02 volumes of an overnight culture grown on LB. Luria broth cultures were inoculated from individual colonies/patches/plaques using a sterile loop or sterile toothpicks. Antibiotics were added as required to LB media to final concentrations of: ampicillin (100μg/ml) ,except for the selection of transformants of *Klebsiella pneumoniae* when double this concentration was used, chloramphenicol (50μg/ml) or tetracycline (12.5μg/ml). Antibiotics were used in minimal cultures at one quarter of these concentrations.

When blue/white colony or plaque selection was used to detect the presence of inserts in the multiple cloning sites of pUC or M13mp vectors by the inactivation of the α– fragment of the β-galactosidase gene isopropyl-β-D-thiogalacto-pyranoside (IPTG; 0.3mM) was incorporated into the media. The agar plates were also supplemented with 50μl of 3% (w/v) 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal;) in dimethylformamide, spread onto the surface of the plate immediately before the addition of the bacterial culture. For the screening of plaques IPTG and X-gal were added directly to the Top-agar.

Preparation of cell-free extracts

Cell-free extracts were prepared from cells in the mid-to-late exponential phase of growth. Bacteria from liquid culture were harvested by centrifugation at 10,000g for 10 minutes at 4°C. The cell pellets were routinely resuspended in 0.04 volumes of 0.1M sodium phosphate buffer pH 7.5. The cell suspension maintained at 0°C was then disrupted by ultrasonication in an MSE 100W ultrasonic disintegrator at a peak amplitude of 7μm. Sonication was generally carried out for 45 seconds. Unbroken cells and cell wall material was removed by centrifugation at 20,000g for 15 minutes at 4°C. Where the removal of membranous material was necessary extracts were ultracentrifuged at 120,000g for 90 minutes at 4°C.

Estimation of protein concentrations

The protein concentration of cell extracts was measured using the biuret method (Gornall *et al*, 1949). Protein concentrations of extracts prepared from cells grown in minimal media was usually in the range

1-5mg/ml whilst for cells grown on Luria broth it was 4-8mg/ml. A standard curve of protein concentration was constructed using bovine serum albumin as reference. The concentration of protein in column fractions was estimated through measurement of the ultra-violet absorbance at 260/280nm (Layne, 1959)

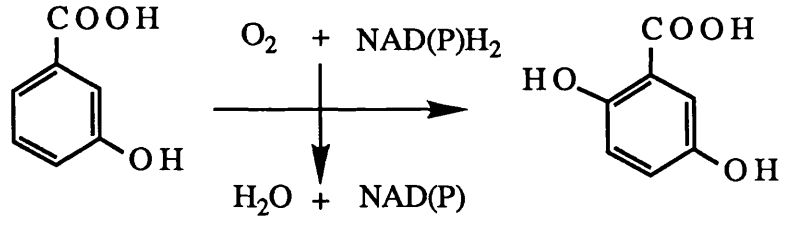
Enzyme assays

Enzyme assays were conducted at 30°C in disposable plastic cuvettes for those reactions where the wavelength monitored was greater than 320nm . Where the wavelength to be used was less than 320nm matched quartz cuvettes were used. In all cases the path length of the cuvettes was 1cm and unless otherwise stated the total volume of the reaction mixture was 1ml. For assays of HHDD isomerase activity reactions were carried out in a total volume of 0.8ml in order to reduce the amount of D_2O required. All assays were carried out using a Pye-Unicam SP1800 recording spectrophotometer.

Preparation of maleylpyruvate and fumarylpyruvate

Maleylpyruvate was generated from 2,5-DHB by the action of a partially purified preparation of 2,5-DHB dioxygenase. Fumarylpyruvate was prepared according to the method of Lack(1959) by the acid catalysed isomerisation of maleylpyruvate. Once the dioxygenase reaction had proceeded to completion the mixture was acidified by the addition of sulphuric acid to a final concentration of 50mM. The acid catalysed isomerisation was allowed to proceed at room temperature for 7 days.

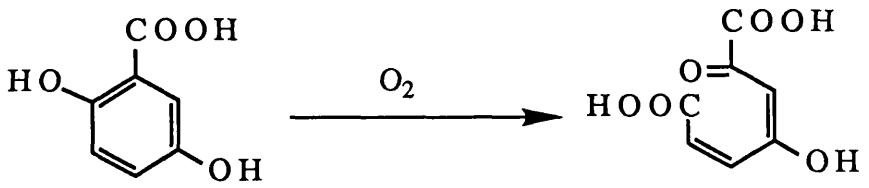
Assay for 3-hydroxybenzoate-monooxygenase



3-HB **2,5-DHB**

Activity was measured spectrophotometrically at 340nm by following the 3-hydroxybenzoate-dependent oxidation of NADPH . The 1ml reaction mixture comprised of 0.15µmol NADPH and 0.2µmol 3-hydroxybenzoate in 0.1M sodium phosphate buffer pH 7.5. The reaction was started by the addition of 10-50 µl of cell extract. A molar extinction coefficient of 6220 $M^{-1}.cm^{-1}$ was assumed for NADPH.

Assay for 2,5-dihydroxybenzoate dioxygenase

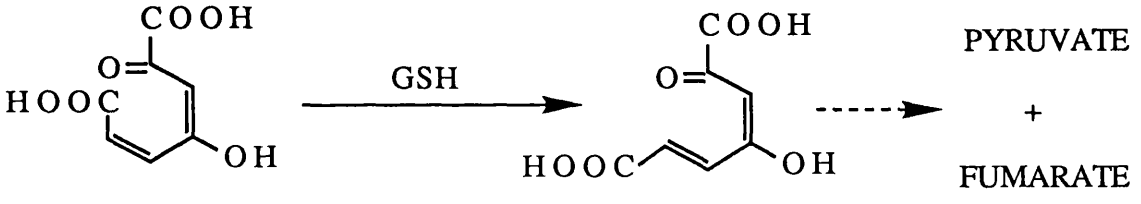


2,5-DHB **maleylpyruvate**

The activity of this enzyme was measured spectrophotometrically by monitoring the increase in absorbance at 330nm associated with the formation of maleylpyruvate. The reaction mixture contained 0.15μmol of 2,5-DHB in 0.1M sodium phosphate buffer pH 7.5 and the reaction was started by the addition of 3-50μl of iron-treated cell extract. An extinction coefficient of 13,000 $M^{-1}.cm^{-1}$ was assumed for maleylpyruvate.

Assay for maleylpyruvate isomerase

(Lack, 1961; Jones and Cooper,1990)



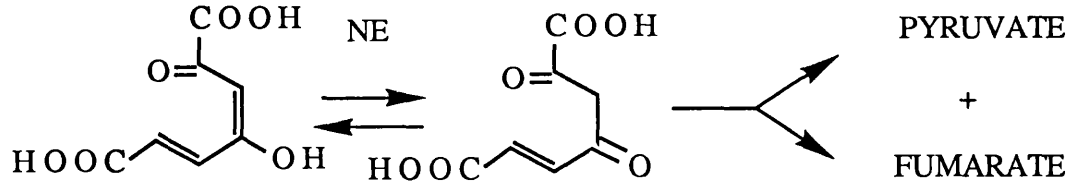
maleylpyruvate **fumarylpyruvate**

Maleylpyruvate isomerase activity was measured by following the reduced-glutathione (GSH) dependent fall in absorbance at 330nm accompanying the isomerisation of maleylpyruvate to fumarylpyruvate whose accumulation was prevented by the presence of an extract with excess fumarylpyruvate hydrolase activity. The exact method of assay depended upon the manner by which maleylpyruvate was generated. If maleylpyruvate was produced from 2,5-DHB with a partially purified 2,5-DHB dioxygenase preparation then fumarylpyruvate hydrolase activity was provided by a crude extract of *K.pneumoniae* MI214 which is defective for the *mhbI* gene encoding maleylpyruvate isomerase. When the plasmid pNDR21, expressing 2,5-DHB dioxygenase and fumarylpyruvate hydrolase only, became available an extract of *E.coli* 5K (pNDR21) was used to provide the requisite activities. The final reaction mixture comprised 0.1μmol

maleylpyruvate with 0.1μmol glutathione in 1ml of 0.1M sodium phosphate buffer pH 7.5, with the fumarylpyruvate hydrolase containing extract present. The reaction was started by the addition of 5-50μl of cell extract. An extinction coefficient of 13,000 $M^{-1}.cm^{-1}$ was assumed for maleylpyruvate.

Assay for fumarylpyruvate hydrolase

(Lack, 1961)



fumarylpyruvate

Fumarylpyruvate hydrolase activity was measured spectrophotometrically through the fall in absorbance at 335nm as fumarylpyruvate was hydrolysed. The assay mixture contained 0.1μmol of neutralised fumarylpyruvate in 0.1 M sodium phosphate buffer pH 7.5 and the reaction was started by the addition of cell extract. An extinction coefficient of 9,000 $M^{-1}.cm^{-1}$ at 335 nm was assumed for fumarylpyruvate.

Preparation of intermediates of the HPC pathway

E.coli 5K strains containing sub-clones encoding appropriate sections of the HPC pathway were utilised for the preparation of particular pathway intermediates (Jenkins,1987).

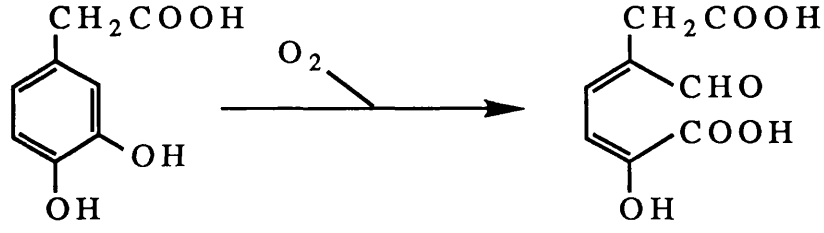
The general method employed was as follows : cell-free extracts were prepared from 100ml overnight cultures of *E.coli* 5K containing the appropriate plasmid grown on Luria broth-ampicillin. The minimum amount of extract required for complete conversion of HPC to CHMS was determined and double this quantity was used in the subsequent reactions to ensure quantitative conversion of HPC. The reaction mixture was gently shaken at 30°C in a 250ml flask to ensure good aeration until the reaction was complete as determined by the absence of HPC. The crude extract was used in a reaction mixture that consisted of 0.1M sodium phosphate buffer pH 7.5 and the following i) HPC (50μmol) and extract from *E.coli* 5K (pJJ002 or the higher

expressing pDR9304) for the production of CHMS; ii) HPC (50µmol), NAD (67µmol) and extract from *E.coli* (pJRJ005) for the production of CHM; iii) as for ii) but with extract from *E.coli* 5K (pJRJ002) instead of *E.coli* 5K (pJRJ005) for the production of HHDD; iv) as for ii) but with $MgCl_2$ (125µmol) and extract from *E.coli* 5K (pJJ212) instead of *E.coli* 5K (pJRJ005) for production of HHED.

Concentrated HCl (1ml) was added to the reaction and the precipitated protein removed by centrifugation. The resulting supernatant was extracted three times with an equal volume of diethyl ether and the pooled organic phases were pooled and dried over anhydrous sodium sulphate. The dried organic phase was then evaporated to dryness at 30°C in a rotary evaporator. The residue was then dissolved in 1ml of water and stored at -20°C or was freeze-dried to remove the water and kept as a solid. When HHDD was prepared D_2O was used instead of water so as to reduce the spontaneous isomerisation of HHDD to OHED which occurs in aqueous solution.

Assay for HPC 2,3-Dioxygenase

(Cooper and Skinner, 1980)

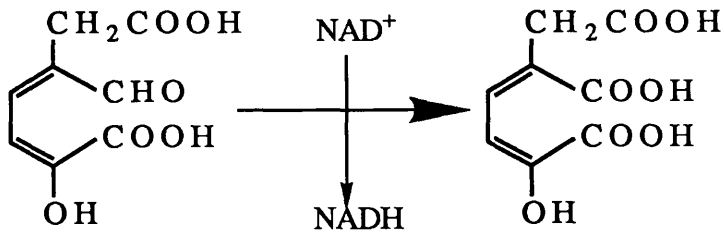


HPC **CHMS**

The *meta* ring cleavage of HPC by HPC dioxygenase was monitored via the increase in absorbance at 380nm resulting from the formation of CHMS. The reaction mixture contained 0.2µmol HPC in 0.1M sodium phosphate buffer pH 7.5 and was started by the addition of 5-50µl of extract. A molar extinction coefficient at pH 7.5 of 31,800 $M^{-1}.cm^{-1}$ was assumed for CHMS.

Assay for CHMS dehydrogenase

(Cooper and Skinner, 1980)

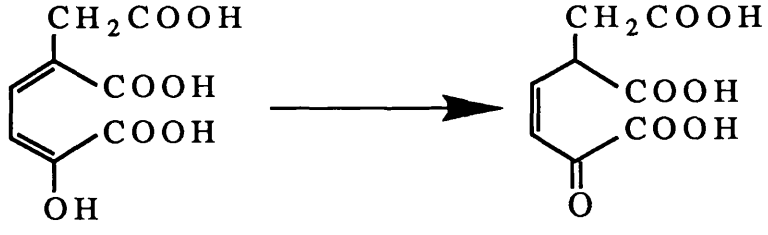


CHMS **CHM**

CHMS dehydrogenase activity was assayed by measuring the decrease in absorbance at 380nm resulting from the NAD^+-dependent dehydrogenation of CHMS to CHM. The reaction was conducted in 0.1 M sodium phosphate buffer pH 7.5 and the reaction mixture contained 0.035µmol of CHMS and 5-50µl of extract. The reaction was started by the addition of 0.2µmol of NAD^+. An extinction coefficient of 31,800 $M^{-1}.cm^{-1}$ was assumed for CHMS at pH 7.5.

Assay for CHM isomerase

(Garrido-Pertierra and Cooper, 1981)

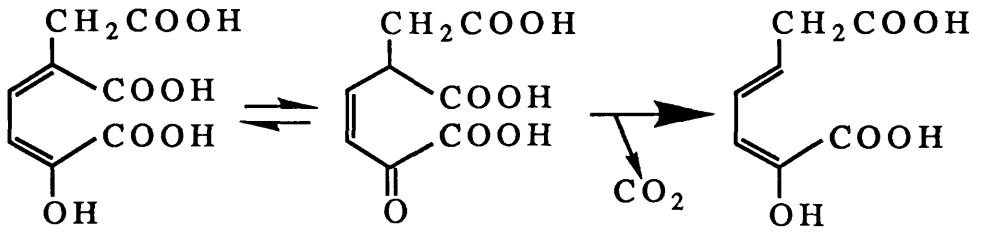


CHM **COHED**

CHM isomerase activity was monitored through the decrease in absorbance at 300nm resulting from the loss of CHM. The assay mixture comprised of 0.05µmol of CHM in 0.1 M sodium phosphate buffer pH 7.5. The spontaneous rate of isomerisation was monitored for approximately 1 minute before the addition of 5-50µl of extract. A molar extinction coefficient at pH 7.5 of 20,000 $M^{-1}.cm^{-1}$ was assumed for CHM.

Assay for COHED decarboxylase

(Garrido-Pertierra and Cooper, 1981; Roper, 1990)

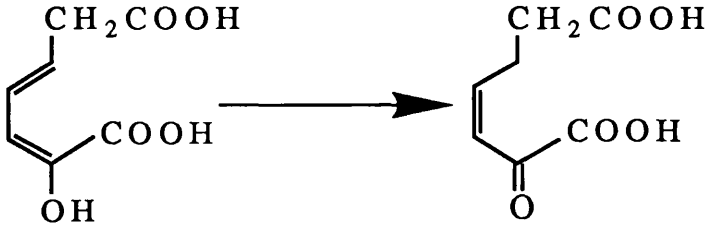


CHM **COHED** **HHDD**

The activity of COHED decarboxylase was measured by following the fall in absorbance at 300nm resulting from the conversion of CHM to COHED as COHED was removed from an equilibrium mixture by decarboxylation to give HHDD. An equilibrium mixture of CHM and COHED was generated rapidly in the cuvette by the addition of 0.6 units of partially purified CHM isomerase (kindly supplied by D.Roper) to a solution of CHM containing 0.1μmol of CHM in 1ml of 0.1 M sodium phosphate pH 7.5. The reaction was initiated by the addition of 5μmol of magnesium chloride and 5-50μl of extract. A molar extinction coefficient of 20,000 $M^{-1}.cm^{-1}$ was used for CHM at 300nm.

Assay for HHDD isomerase

(Jenkins, 1987; Roper, 1990)



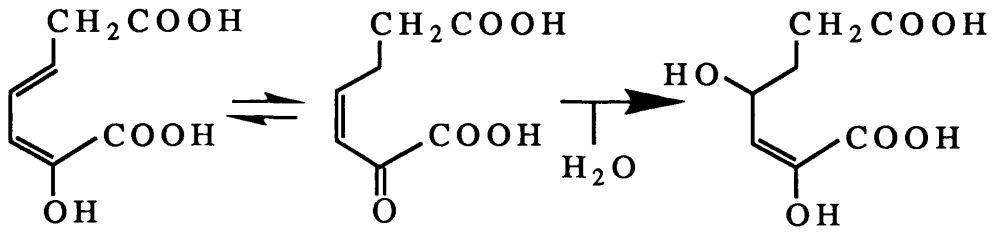
HHDD **OHED**

The activity of HHDD isomerase was measured through the fall in absorbance at 276nm resulting from the conversion of HHDD to OHED. In order to reduce the rapid spontaneous isomerisation of HHDD reported in aqueous solution (Jenkins, 1987) D_2O was employed as solvent since this was shown to reduce the spontaneous isomerisation

rate about 10-fold. In order to ensure adequate buffering of the reaction the basic assay method was modified according to Roper (1990). Hence the assay mixture consisted of 40µl of 2M potassium phosphate buffer pH 7.5 in D_2O, 0.05µmol HHDD and D_2O to a final volume of 0.8ml. The spontaneous rate of isomerisation was measured for 1 minute before 5-20µl of extract were added. A molar extinction coefficient of 18,600 $M^{-1}.cm^{-1}$ was used for HHDD.

Assay for OHED hydratase

(Jenkins, 1987)

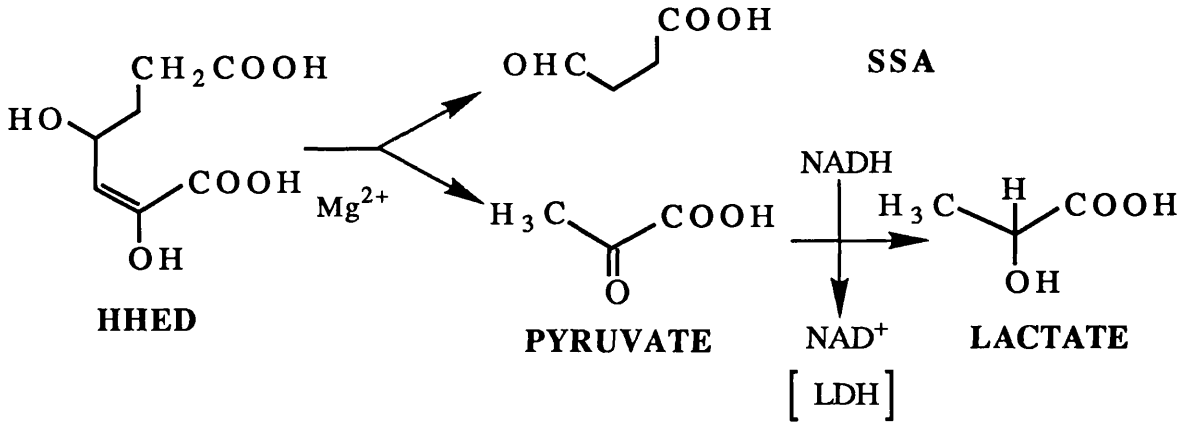


HHDD **OHED** **HHED**

The activity of OHED hydratase was measured as the fall in absorbance at 276nm resulting from the conversion of HHDD to OHED as OHED was removed from an equilibrium mixture of HHDD/OHED. The equilibrium mixture was generated through the rapid spontaneous isomerisation of HHDD in aqueous solution, 0.1µmol of HHDD was allowed to isomerise for a few minutes in 0.1 M sodium phosphate buffer pH 7.5. When the decrease in absorbance at 276nm had ceased 5µmol of magnesium chloride were added and the reaction started by the addition of 5-50µl of extract. A molar extinction coefficient of 18,600 $M^{-1}.cm^{-1}$ was assumed for HHDD.

Assay for HHED aldolase

(Jenkins, 1987)



The activity of HHED aldolase was measured in a coupled assay system through the decrease in absorbance at 340nm accompanying the oxidation of NADH as pyruvate formed from the cleavage of HHED was reduced by lactate dehydrogenase. The reaction mixture comprised of 5µmol of magnesium chloride, 0.15µmol NADH, 4 units of lactate dehydrogenase and an excess of HHED in 0.1M sodium phosphate buffer pH 7.5. A molar extinction coefficient of 6,220 $M^{-1}.cm^{-1}$ was assumed for NADH.

Preparation of plasmid DNA

Plasmids were prepared using an alkaline-lysis method based on that described by Maniatis *et al* (1981). In general 5ml cultures were used for the screening of plasmids of interest and to establish basic stocks. Where the requirement for plasmid DNA was greater 50ml cultures were processed.

Cells were harvested from an appropriate volume of an overnight Luria broth-antibiotic culture by centrifugation. The supernatant was carefully removed and the pellet resuspended in 0.015 volumes of ice-cold GTE solution (50mM glucose; 25mM Tris-Cl pH 8.0; 10mM EDTA), the mixture was allowed to stand at room temperature for 5 minutes. Next 0.03 volumes of alkaline lysis solution (1% SDS in 0.2M NaOH) was added and the contents mixed with three quick inversions. The mixture was then incubated on ice for 5 minutes and then 0.023 volumes of ice-cold potassium acetate solution (3M with respect to potassium; 5M with respect to acetate) added. The tube was inverted and vortexed for 1 second before being incubated on ice for a further 5 minutes. Protein and precipitated chromosomal DNA were removed by centrifugation for 5 minutes in a microfuge or for 20 minutes at 15,000g depending on the scale of the preparation. The supernatant was then transferred to a fresh tube and was extracted twice with phenol/chloroform/isoamyl alcohol (IAA) (25:24:1) equilibrated with 1M Tris-Cl pH 7.5, to remove any remaining protein. The aqueous phase was then extracted once with chloroform/IAA (24:1) to remove residual phenol. Nucleic acids were then precipitated by the addition of two volumes of ethanol. For small-scale preparations this involved incubation for two minutes at room temperature followed by centrifugation in a microfuge for 5 minutes. For larger scale preparations incubation was for 20 minutes and the precipitate was collected by centrifugation for 15 minutes at 10,000g. The precipitates were washed with ice-cold 70% ethanol and the drained pellets dried under vacuum for 10-15 minutes. The pellet was then dissolved in TE buffer (10mM Tris-Cl pH 8.0; 1mM EDTA) containing 20μg ml^{-1} Ribonuclease A. The DNA thus obtained was suitable for transformation or restriction endonuclease analysis.

Where purer plasmid DNA was required for nucleotide sequencing or exonuclease III digestion, the protocol above was modified to give higher quality DNA. The modifications followed were based on those described by Kraft *et al* (1988). Following the precipitation with potassium acetate in the alkaline lysis procedure the supernatant was treated with ribonuclease A added to a final concentration of 50 μg ml^{-1} and incubated for 30-60 minutes at 37°C prior to phenol extraction. The subsequent precipitation of DNA involved incubation on dry ice for 30 minutes rather than the shorter room temperature precipitation of the standard alkaline lysis procedure. The washed and dried DNA was resuspended in 320 μl of TE buffer and was then re-precipitated to remove residual protein by the addition of 80μl of 4M NaCl and 400μl of 13% polyethylene glycol (PEG) 8000 with incubation on ice for 60 minutes. The DNA was pelleted by centrifugation at 12,000g for 10 minutes, the pellet was washed with 70% ethanol and then dried before being dissolved in sterile distilled water.

Growth and isolation of M13 replicative form (RF)

For the isolation of M13 RF DNA a single phage plaque was transferred with a sterile toothpick into 10 ml of 2YT medium which contained freshly sub-cultured *E.coli* JM101 at an OD_{680} of 0.1. The cells were incubated with vigorous shaking for 2 hours at 37°C before a further 10 ml of freshly grown JM101 was added. Incubation was then continued for a further 3 hours at 37°C. The bacterial cells were harvested by centrifugation at 6,000g for 10 minutes. The replicative form was then prepared from cells using the basic alkaline lysis procedure described above.

Preparation of single-stranded M13 DNA

E.coli JM101 cells, grown overnight on 2YT, were sub-cultured by dilution into fresh medium. The cells were inoculated with a single tooth-picked plaque and grown for 7 hours at 37°C with vigorous shaking. The cells were pelleted by centrifugation at 6000g for 10 minutes. The supernatant was very carefully transferred to a fresh tube so as to avoid contamination with the pelleted bacterial cells. Single-stranded DNA was then precipitated by the addition of 0.1

volumes of 5M sodium chloride and 0.1 volumes of a 40% solution of PEG 6000 followed by incubation on ice for 1 hour. The precipitated material was pelleted by centrifugation at 10,000g for 10 minutes. The pellet was resuspended in 600µl of TE and the DNA precipitated again by incubation at room temperature for 30 minutes following addition of 60µl of 5M sodium chloride and 60µl of 40% PEG 6000. The resulting pellet obtained by further centrifugation was resuspended in 500µl of TE and was then extracted once with neutral phenol, twice with phenol/ chloroform and once with chloroform. The DNA was finally precipitated with two volumes of absolute ethanol and 0.1 volumes of 3M sodium acetate and incubation on dry-ice for 30 minutes and the precipitate collected by centrifugation at 10,000g for 10 minutes. The resulting pellet was washed with 70% ethanol and was finally resuspended in sterile distilled water.

Preparation of chromosomal DNA

Chromosomal DNA was prepared according to the method of Chow *et al* (1977). Cells were grown overnight with vigorous shaking in 200ml of Luria broth in a 1 litre flask to ensure good aeration. Cells were pelleted by centrifugation at 10,000g for 5 minutes and were then carefully washed twice with 10mM Tris-Cl pH 7.9; 1mM EDTA; 1M NaCl. The washed cell pellet was then resuspended in 25ml of the wash buffer and 5ml of lysozyme solution (10mg/ml) was added and the tube incubated at 37°C for 10 minutes. An equal volume of a 2% sarkosyl solution (in 10mM Tris-Cl pH 7.9; 1mM EDTA; 0.1M NaCl) was added with gentle mixing and the tube incubated at 42°C for 1 hour. Proteinase K, activated by self-digestion, was then added to a final concentration of 0.1 mg/ml and the mixture incubated overnight at 42°C. The lysate was then gently transferred to a 250ml conical flask and an equal volume of phenol (equilibrated with 0.5M Tris-Cl pH 8.0) added. The contents of the flask were mixed very gently by swirling in a room temperature shaker at slow speed for 30 minutes. The layers were separated by centrifugation at 12,000g at 15°C for 20 minutes. The aqueous layer was taken off using a snipped-off plastic Pasteur pipette and the extraction repeated twice. The volume of the final aqueous phase was determined using a sterile measuring cylinder and it was then transferred to a beaker on ice. Sodium acetate (3M, pH4.8) was added dropwise, with gentle mixing, to the

beaker to a final concentration of 0.3M. Two volumes of absolute ethanol were carefully poured down the side of the beaker whilst the interphase was gently mixed by swirling. The DNA which formed at this interphase was spooled onto a glass pipette and then unwound into 4ml of 10mM Tris-Cl pH 7.4; 1mM EDTA. The DNA was allowed to redissolve at 4°C for 4 days before being quantified and examined on a 0.5% agarose gel.

Restriction enzyme digestion

Digestions with restriction endonucleases were carried out according to the instructions of the manufacturer using the buffers provided. Enzymes were purchased from Bethesda Research Laboratories or Pharmacia-LKB Ltd. Routine digests of plasmids included 200-500ng of DNA in a total volume of 15µl with 5-10 units of enzyme added. Digestion times ranged from 1-3 hours for plasmids to overnight for complete chromosomal digests.

Phosphatase treatment of DNA

In order to minimise the number of recircularised vector molecules in ligation reactions the linearised plasmid DNA was treated with calf intestinal phosphatase (CIP) to remove terminal phosphate groups. One unit of phosphatase was added to the digested DNA. Reactions were incubated at 37°C for 30 minutes in the buffer recommended by the manufacturer. The CIP was removed by one extraction with phenol/chloroform and one with chloroform and the DNA precipitated on dry ice with two volumes of ethanol and 0.3M sodium acetate.

Agarose gel electrophoresis

Restriction enzyme digests were analysed by submarine gel electro - phoresis through agarose gels. Gels of 0.5-1.0% agarose prepared with TAE gel buffer (40mM Tris-acetate, 1mM EDTA (TAE buffer) plus ethidium bromide at 0.5µg ml^{-1}) were used according to the sizes of fragment expected. Gels were run at constant voltage (5-10 V cm^{-1}) in TAE buffer. Samples were prepared by the addition of 0.15 volumes of type II loading buffer (0.25% bromophenol blue (w/v), 0.25% xylene cyanol (w/v), 30% glycerol (v/v)) (Maniatis *et al* ,1982). Migration of sample was monitored by observation of the dye fronts.

Gels were usually run until the bromophenol blue front had reached the end of the gel when they were visualized under U.V light. Gels were photographed using short wave U.V illumination with T-Max 100 film (Kodak). Gels were calibrated using lambda phage DNA digested with *Hin* dIII or *Bst* EII as size markers.

Recovery of DNA fragments from agarose gels

Fragments generated by restriction endonuclease digestion were isolated from gels by use of the method described by Perbal (1987). Digests were run out on 0.8-1.0% low melting point agarose gels and the desired band(s) located by illumination with long wavelength ultra-violet light. The required fragments were excised with a sterile scalpel blade along with the minimum of extraneous agarose and placed in a 1.5ml eppendorf tube. The gel slice was collected at the bottom of the tube by brief centrifugation and 100µl of TE solution added. The tube was then incubated for 20 minutes at 65°C to melt the agarose. An equal volume of Tris-saturated phenol was added and the tube vortexed vigorously for 5 minutes. The phases were separated by a 5 minute spin and the aqueous layer transferred to a fresh tube. The organic phase was then re-extracted with 0.5 volumes of TE and the aqueous layer combined with the first sample. The pooled aqueous phases were then extracted with an excess of water-saturated ether to remove contaminating phenol. After addition of the ether the tube was vortexed briefly and the phases allowed to separate. The ether layer was removed with a pipette and the final traces eliminated by gently bubbling air through the residual layer with a Gilson pipette. The DNA was then precipitated by mixing the aqueous layer with two volumes of chilled absolute ethanol and adding 3M sodium acetate to a final concentration of 0.3M and allowing the tube to stand in dry-ice for 30-60 minutes. The DNA was recovered with a 10 minute spin in a micro-centrifuge and the pellet washed with 0.5 ml of 70% ethanol (kept at -20°C). The pellet was drained and then dried for 10-15 minutes in a vacuum desiccator. Pellets were routinely dissolved in 20-30µl of sterile water and 5µl was run on an agarose gel to check successful recovery.

Ligation of DNA

Vector and insert DNA were routinely mixed in a molar ratio of 1:2 in a total volume of 10-20µl and ligated in a reaction containing 1 unit of T4 DNA ligase (BRL) using the supplied buffer. Reactions were incubated at room temperature for 2-16 hours, blunt end ligations were always incubated overnight. Generally half of the ligation reaction was used for transformations of competent cells.

Treatment of digested plasmids with mung bean nuclease

Mung bean nuclease was used in order to remove the overhangs from non-compatible restriction enzymes so as to allow blunt end ligation of the two sites. Following digestion the restriction enzymes were removed by ethanol precipitation of the DNA. The DNA (2-5µg) was resuspended in 40µl of sterile distilled water. Mung bean nuclease reaction buffer was added to give final component concentrations of : 30mM sodium acetate, 50mM sodium chloride, 1mM zinc chloride, 5% glycerol. Mung bean nuclease (10-20 units) was added and the mixture incubated at 37°C for 20 minutes. Fragments were then purified using low melting point agarose gels.

Transformation of bacterial cells with plasmid DNA

Competent cells were freshly prepared as required by the method of Kushner (1978). The required strain was grown overnight in 5ml LB cultures and then diluted one-hundred fold into fresh LB. Cells were then grown at 37° C to an OD_{680} of approximately 0.5. Aliquots of cells (1.5ml) were harvested by centrifugation for 30 seconds and the drained pellets resuspended in 0.5 ml of sterile MOPS A solution (10mM MOPS pH7.0, 10mM RbCl). The cells were then pelleted with a further 20 second spin and resuspended in 0.5ml of MOPS B (100mM MOPS pH6.5, 10mM RbCl, 50mM $CaCl_2$) and left on ice for 60-90 minutes. Cells were collected by a 10 second spin and resuspended in 150µl of MOPS B with dimethyl sulphoxide included to 0.2% (v/v). DNA was added (10-200ng) and the mixture left on ice for 1 hour. The cells were heat-shocked at 55°C for 30 seconds and chilled on ice for

1 minute before 1ml of pre-warmed LB was added. The cells were then allowed to recover at 37°C for 1 hour before 100µl was spread on a pre-warmed LB-antibiotic plate. Plates were then incubated overnight at 37°C.

Transfection of bacterial cells with bacteriophage DNA

The strain used for transfections with bacteriophage M13 RF DNA was *E.coli* JM101. Competent cells were prepared from 10ml 2YT cultures which had been inoculated with 0.1 ml of an overnight culture and then grown up for 3-4 hours at 37°C. The cells were harvested by centrifugation for 5 minutes at 6000g and 4°C and were then resuspended in 2.4ml of 50mM $CaCl_2$ before being kept on ice for 20 minutes. The cells were pelleted by centrifugation for 5 minutes at 6000g, the supernatant was discarded and the pellet resuspended in 800µl of 50mM $CaCl_2$. Competent cells (300µl) were then mixed with the bacteriophage DNA (20-50ng) and the mixture was left on ice for 30 minutes. The cells were then subjected to a heat shock at 42°C for 2 minutes before being mixed with 200µl of fresh JM101 culture. The cells were then added to 3ml of Top-agar which had been kept at 42°C, the mixture vortexed briefly and then poured onto hard agar plates. The plates were then incubated at 37°C overnight.

Unidirectional deletion of DNA using ExonucleaseIII

Deletions were generated using a kit purchased from Pharmacia-LKB which employs the methodology described by Henikoff (1984). This method is based on the controlled digestion of DNA with exonuclease III. This enzyme is a 3'- exonuclease which is active only on double-stranded DNA : blunt and 5'-overhanging ends are susceptible to digestion whilst 3'-overhangs of at least 3 bases are resistant. Therefore by the appropriate use of restriction sites in a plasmid it is possible to generate a susceptible end adjacent to the target DNA and a protected end and since subsequent deletion can proceed only from the susceptible end unidirectional deletion results. For pNDR22 the *Bam* HI site of the polylinker was used to generate the susceptible end whilst the *Kpn* I site was used to protect vector functions and subsequent deletion of the insert DNA is from *Sph* I to *Hin* dIII.

For the reactions 5µg of pNDR22 prepared using the extended-alkaline lysis procedure was digested to completion with *Bam* HI and *Kpn* I. The restriction enzymes were inactivated by heating at 70°C for 10 minutes. The concentration of NaCl was then adjusted to 75mM by mixing 20µl of the digested DNA ([NaCl] = 100mM) with 20µl of 2x Exonuclease III([NaCl] = 50mM) buffer prepared according to the manufacturer's instructions. The tube was equilibrated at 30°C for 3 minutes and at this point a 2µl sample corresponding to the timepoint t=0 was removed into a tube containing 3µl of S1 nuclease/buffer mix, which was then kept on ice. Exonuclease III (1µl) was added to the reaction tube and the reaction allowed to proceed. Samples (2µl) were removed at 3 minute intervals and these were also mixed carefully with 3µl of S1 nuclease/buffer mix before being transferred to ice. When all of the required timed samples had been withdrawn all of the tubes were simultaneously incubated at room temperature for 30 minutes to allow the S1 nuclease to remove the single-stranded regions produced by the exonuclease III digestion. The reactions were then terminated by the addition of 1µl of Stop Solution and incubation at 65°C for 10 minutes. The extent of deletion at each point was then monitored by electrophoresis of 3µl samples of material from each timepoint mixed with 3µl of gel loading buffer on a 1% agarose gel. Recircularization of the deleted material was then achieved by incubation of the remaining 3µl from each timepoint with 17µl of a ligation mix prepared according to the manufacturer's instructions by mixing 85µl of 5x Ligase Mix with 85µl of 25% PEG and 195µl of sterile distilled water. Ligations were allowed to proceed overnight and the material used to transform *E.coli* 5K using the standard MOPS-RbCl method. The resulting recombinant plasmids were then screened by restriction analysis.

Transfer of DNA to nylon membranes

DNA was transferred from agarose gel to nylon membrane using a capillary transfer method (based on Southern, 1975). The gel was first soaked in 0.25M HCl for 15 minutes at room temperature in order to cause breakage of the DNA through depurination, this treatment improves the efficiency of transfer of large fragments. The DNA was then denatured by incubation of the gel with a solution of 1.5M NaCl; 0.5M NaOH for 30 minutes with gentle agitation. The

denaturing solution was then removed and replaced with neutralizing solution (1.5M NaCl; 0.5M Tris-Cl, pH 7.2; 1mM EDTA), the gel was then incubated for a further 30 minutes. The gel was removed to a fresh container and rinsed with distilled water. The gel was then placed on a piece of Whatman 3MM paper so that there were no air bubbles. The sheet of paper rested on a glass bridge over a plastic box with its ends projecting into the box such that when the container was filled the paper would act as a wick. A piece of Hybond-N membrane which had been cut to the same size as the gel and wetted by immersion in 2xSSC (SSC is 0.15M NaCl: 0.015M sodium citrate) was then carefully placed onto the surface of the gel. Any air bubbles between the membrane and the gel were eliminated by the application of gentle pressure with a pipette. Several layers of 3MM paper were placed on top of the membrane and a piece of Saranwrap film which had been cut to have an aperture the same size as the gel placed over the stack in order to reduce capillary movement of liquid bypassing the gel. A stack of paper towels approximately10cm high was placed in contact with the gel/membrane sandwich and on top of this was put a glass plate and an 500g weight. The reservoir was then filled with 20xSSC to immerse the ends of the wick and the transfer allowed to proceed for 16 hours with fresh paper towels being added after one hour. After transfer the blotting aparatus was disassembled and the membrane washed in 2xSSC to remove any adhering agarose. The membrane was then air-dried before being wrapped in Saranwrap film and exposed for 15 seconds to short wavelength U.V light to fix the DNA to the Hybond.

Radioabelling of restriction fragments for use as hybridisation probes

Restriction fragments were radiolabelled using [α-^{32}P] dCTP with an oligolabelling kit from Pharmacia-LKB, using the method of Feinberg and Vogelstein (1983). This method involves denaturation of the DNA to be labelled and then mixing with hexadeoxyribonucleotides of random sequence. These random hexamers anneal to random sites on the DNA and so serve as primers for DNA synthesis by the Klenow Fragment of *E.coli* DNA polymerase I with radiolabelling being achieved by the inclusion of labelled nucleotide.

Restriction fragments to be labelled were recovered from low melting point agarose gels and 50ng in a volume of 3µl was denatured by heating at 95°C for 3 min and the DNA chilled immediately on ice. The denatured fragment was then added to a mixture comprising 10µl of Reagent mix (Pharmacia), 5µl (50µCi) of [α^{32}P] dCTP (Amersham) and 31µl of sterile water. One µl of the provided Klenow fragment was added and the reaction was allowed to proceed at 37°C for 3 hours. The effectiveness of labelling was evaluated by paper chromatography of 0.5 µl of the reaction mixture on DE81 paper using 0.3M ammonium formate as the solvent. In this system labelled polynucleotide remains at the origin whereas unincorporated radioactive nucleotide migrates with the solvent front. After running the chromatogram was dried and wrapped in Saranwrap and autoradiographed using Fuji x-ray film. The film was developed after 10 minutes and showed that labelling had been highly efficient with essentially full incorporation of the label.The probe was denatured by heating at 95°C for 3 minutes before use.

Probing of Hybond-N filters

Filters were pre-hybridised and hybridised using a method based on that of Church and Gilbert (1984). Pre-hybridisation was carried out for 20 minutes at 65°C in sealed hybridisation chambers using 15ml of de-gassed pre-hybridisation solution (1mM EDTA; 0.5M Na_2HPO_4 pH 7.2; 7 %SDS). Hybridisation was carried out under similar conditions using pre-hybridisation solution to which the denatured probe had been added. Hybridisation was allowed to proceed overnight at 65°C with gentle shaking. The hybridisation solution was then removed and the filter subjected to three five minute washes at 65°C with 40mM Na_2HPO_4; 1% SDS followed by one 10 minute wash with 0.1x SSC; 0.01% SDS. The filter was then air-dried and covered with Saranwrap before autoradiography.

Nucleotide sequencing

Plasmid and M13 phage DNA was sequenced using the chain-termination method of Sanger *et al* (1977) using a modified T7 DNA polymerase (Tabor and Richardson ,1987). Sequencing kits (Sequenase version 2.0) were purchased from United States Biochemical and radio-isotope was purchased from Amersham

International. The procedure may be divided into five discrete stages. 1. Preparation of single stranded DNA template, for M13 this may be achieved directly, for double stranded plasmid DNA denaturation is required. 2. An oligonucleotide primer is annealed to the template. 3. A labelling reaction is performed using limiting concentrations of dNTPs including radio-labelled dATP. This continues to complete incorporation of label into DNA chains with lengths randomly distributed from several to several hundred nucleotides. 4. A termination reaction is performed at higher dNTP concentrations with a dideoxynucleotide triphosphate included. DNA synthesis proceeds until all DNA chains are terminated with a dideoxynucleotide. 5. Reactions are stopped and the products separated using high-resolution denaturing polyacrylamide gel electrophoresis.

Denaturation of plasmid DNA

Approximately 5µg of plasmid DNA in a total volume of 20µl was denatured by the addition of 2µl of a freshly prepared solution of 2M NaOH; 2mM EDTA and incubation at room temperature for 10 minutes. The mixture was then neutralised by the sequential addition, on ice, of 8µl of 1M Tris-Cl pH 4.5 and 3µl of 3M sodium acetate. The DNA was precipitated by the addition of 75µl of ethanol (-20°C) and incubation on dry ice for 10 minutes before the pellet was collected by centrifugation for 10 minutes. The drained pellet was washed with 200µl of ice-cold 70% ethanol and then dried under vacuum for 10 minutes before being resuspended in 7µl of sterile water.

Sequencing reactions

Sequencing reactions were performed using 5µg of denatured plasmid DNA or 1µg of M13 based constructs, in each case the DNA was dissolved in a total of 7µl of sterile water. The DNA was mixed with 2µl of reaction buffer and 1µl of the required primer ($5ng.ml^{-1}$) added. The primer was annealed by heating the mixture at 65°C for 2 minutes and then allowing slow cooling to room temperature over 30 minutes. During this time 2.5µl aliquots of A, C, G and T termination mixes were placed in separate tubes which were pre-warmed to 37°C. The annealed DNA solution was collected at the bottom of the tube by brief centrifugation and to this was added 1µl of 0.1M dithiothreitol (DTT), 2µl of diluted labelling mix, 0.7µl of [α-^{35}S] dATP (

approximately 3.6μCi) and 2μl of Sequenase version 2.0 which had been diluted 8-fold with ice cold dilution buffer immediately before addition. The contents of the tube were mixed and incubated for 4 minutes at room temperature. At the end of this period 3.5μl of the labelling reaction mixture was transferred to each of the termination mix tubes and the termination reactions were allowed to proceed at 37°C for a further 5 minutes. Reactions were stopped by the addition of 5μl of Stop solution (95% formamide; 20mM EDTA; 0.05% bromophenol blue; 0.05% xylene cyanol) to each tube. Reactions were then stored at -20°C until required.

Electrophoresis and autoradiography

Sequencing gels were 6% acrylamide-7M urea made up in 0.09M Tris-borate;EDTA (2mM) pH 8.3 (TBE) buffer. Gels were pre-run at 1400-1500V with TBE in the upper and lower reservoirs for 30-45 minutes before samples were loaded. An aluminium plate to act as a heat-sink was attached to the gel plates to minimise smiling of the gels during electrophoresis caused by uneven heat-distribution. In order to increase the amount of sequence data obtained two or three separate loadings of sample were run with 3μl being applied at each loading. The second sample was applied when the xylene cyanol front had migrated two-thirds of the way down the gel with the third sample (where required) being applied when the second xylene cyanol front had reached a similar position. Gels were then run until the final bromophenol blue dye front had reached the bottom of the gel. Following electrophoresis the gel was immersed in a 10% methanol /10% glacial acetic acid solution for 10 minutes before being transferred to Whatman 3MM paper. Gels were then dried down onto the paper under vacuum at 80°C for 90 minutes. Autoradiography was caried out overnight at room temperature using Fuji RX x-ray film.

Polyacrylamide gel electrophoresis

Protein samples were routinely electrophoresed on polyacrylamide gels using a Mini-Protean system from BioRad. Denaturing (SDS) gels were prepared with a 12% acrylamide; 0.1 %SDS resolving gel prepared in 375mM Tris-Cl pH 8.8 and were polymerised by the addition of 0.1% ammonium persulphate and 0.05% TEMED. Stacking

gels were prepared at 5.1% acrylamide in 125mM Tris-Cl pH 6.8 but were otherwise identical. For non-denaturing gels the same protocol was used but the SDS was omitted. Samples were prepared for denaturing gel electrophoresis by heating 2-40μg of protein for 2 minutes at 95°C with 0.3 volumes of sample buffer. The sample buffer consisted of 2% SDS; 0.1% bromophenol blue; 10% glycerol; 100mM DTT in 50mM Tris-Cl pH 6.8. For non-denaturing gels the sample buffer contained no SDS and the samples were not heated before loading on the gel. Gels were generally run for 45 minutes at a constant voltage of 200 V in a buffer which was 25mM with respect to Tris-base-250mM glycine (pH 8.3) which contained 0.1% SDS for denaturing gels. The protein bands were stained with Coomassie blue R250 solution (comprising 0.5% [w/v] Coomassie blue R250 in 45% [v/v] methanol; 10% acetic acid) for 0.5-3 hours. Gels were then destained in a solution of 7.5% [v/v] acetic acid; 5% methanol for 2-16 hours. Gels were calibrated using SDS-7 markers from Sigma, the relative molecular masses of the markers are as follows: bovine serum albumin (66,000); ovalbumin (45,000); glyceraldehyde-3-phosphate dehydrogenase (36,000); carbonic anhydrase (29,000); trypsinogen (24,000); trypsin inhibitor (20,100) and bovine lactoglobulin (14,100).

Amino-terminal amino acid sequencing

The amino terminus of purified 2,5-dihydroxybenzoate-1,2-dioxygenase was sequenced by Dr. K.Lilley and Miss E.Cavanagh using an Applied Biosystems 470A gas-phase sequencer. A purified fraction of the enzyme was run on a 12% polyacrylamide-SDS mini-gel and then electroblotted onto a polyvinylidine difluoride (PVDF) membrane using 50mM glycine-50mM Tris pH 10 as the transfer buffer (Matsudaira, 1987). The blot was stained with Coomassie blue R-250 and the protein band excised and loaded into the sequencer without polybrene.

Fast protein liquid chromatography (FPLC)

A Pharmacia FPLC system was used in protein purification procedures in accordance with the manufacturer's instructions. Purifications were performed at room temperature. The separation techniques employed

were anion exchange with the Mono Q HR 5/5 column and hydrophobic interaction chromatography using a Phenylsuperose HR 5/5 column. The buffers used were selected from the manufacturer's list of approved buffers.

Oligonucleotide synthesis

Oligonucleotides were synthesised by Mrs. D.Langton with an Applied Biosystems 380B DNA synthesiser using cyanoethyl phosphoamidite chemistry. The concentration of oligonucleotides was determined the A_{260} of a diluted oligonucleotide solution using the conversion constant A_{260} of 1.0 is equivalent to a concentration of 33µg/ml.

CHAPTER 3

Isolation and analysis of cloned 3-hydroxybenzoate (*mhb*) catabolic genes from *Klebsiella pneumoniae*

Key to enzymes:

MhbM= 3-HB monooxygenase
MhbD= 2,5-DHB dioxygenase
MhbI= maleylpyruvate isomerase
MhbH= fumarylpyruvate hydrolase

Isolation and analysis of cloned 3-hydroxybenzoate (*m*-HB) catabolic genes (*mhb*) from *Klebsiella pneumoniae*

Klebsiella pneumoniae has been shown to degrade 3-hydroxybenzoate (3-HB) via the gentisate pathway (Jones and Cooper,1990). Therefore, the basic biochemistry of the pathway has already been characterised but little is known about the genetic organisation of the pathway except that clustering of the genes was indicated by bacteriophage P1 transduction experiments which showed two of the genes namely *mhbM* encoding 3-hydroxybenzoate-6-mono-oxygenase and *mhbI* encoding maleylpyruvate isomerase were 100% co-transducible In order to analyse the fine genetic detail of the pathway it was decided to clone the genes involved. This chapter describes how an analysis of an existing partial clone was used to determine a cloning strategy and how its application allowed the entire pathway to be obtained on a single fragment.

A genomic clone was isolated (S.Parrott and R.A Cooper, pers. comm) by screening a genomic library of *K pneumoniae* DNA with an oligonucleotide probe representing a portion of the amino terminal sequence of purified maleylpyruvate isomerase. This clone designated pSP01 was characterised as containing a 9 Kbp *Hin*dIII fragment in pBR322. Cell-free extracts of *E.coli* 5K harbouring pSP01 expressed both the MhbM (3-hydroxybenzoate-6-monooxygenase) and MhbI (maleylpyruvate isomerase) gene products but did not express those of MhbD encoding 2,5-dihydroxybenzoate-1,2-dioxygenase or MhbH encoding fumarylpyruvate hydrolase. Additionally it was noted that, in contrast to wild-type *K.pneumoniae*, the expression of the catabolic genes was not induced when the cells were grown in the presence of 3-hydroxybenzoate suggesting the absence of a complete regulatory system.

Analysis of pSP01

Properties of cells harbouring pSP01

It had been observed previously that growth of *E.coli* 5K (pSP01) on glycerol in the presence of 3-HB had led to a brown colouration which had been presumed to result from the partial metabolism of 3-HB

leading to the accumulation of the ring fission substrate 2,5-DHB which cannot be metabolised because of the absence of the 2,5-DHB dioxygenase (MhbD) and so oxidises to give melanin-like substances. This conversion could be monitored by spectrophotometric scans of culture supernatants and was characterised by a shift from a peak absorbance of 290nm (3-HB) to a peak at 322nm co-incident with the absorbance maximum of 2,5-DHB. Examination of these scans suggested that complete conversion of 5mM 3-HB to 2,5-DHB occurred on prolonged incubation. The culture supernatant at the end of incubation also appeared to substitute for authentic 2,5-DHB when used in an assay for the 2,5-DHB dioxygenase with cell-free extracts prepared from wild-type *K.pneumoniae* thus providing confirmatory evidence for the identification of the major component as 2,5-DHB. As expected *K. pneumoniae* cells did not appear to accumulate 2,5-DHB during exposure to 3-HB since they possess the entire catabolic sequence and so are able to completely dissimilate 3-HB. No conversion of 3-HB was seen with *E.coli* 5K containing pBR322 showing that this ability is a feature of the cloned DNA and is indeed a consequence of partial metabolism. Another aspect of this property is that it shows that *E.coli* 5K harbouring pSP01 possesses a system allowing entry of 3-HB into the cell. This has direct relevance because it suggested that a clone including the other catabolic functions for 3-HB metabolism might allow growth of *E.coli* on the aromatic substrate. This situation would be in contrast to that observed with genomic clones of the *hpc* genes from *E.coli* C (Jenkins, 1987) and *K.pneumoniae* (Fawcett, 1989) where the presence of all genes required for the conversion of homoprotocatechuate (HPC) to central metabolites does not allow *E.coli* 5K to grow on HPC. In that case it has been hypothesized that entry of the aromatic substrate into the cell may be the limiting factor preventing growth.

Molecular analysis of pSP01

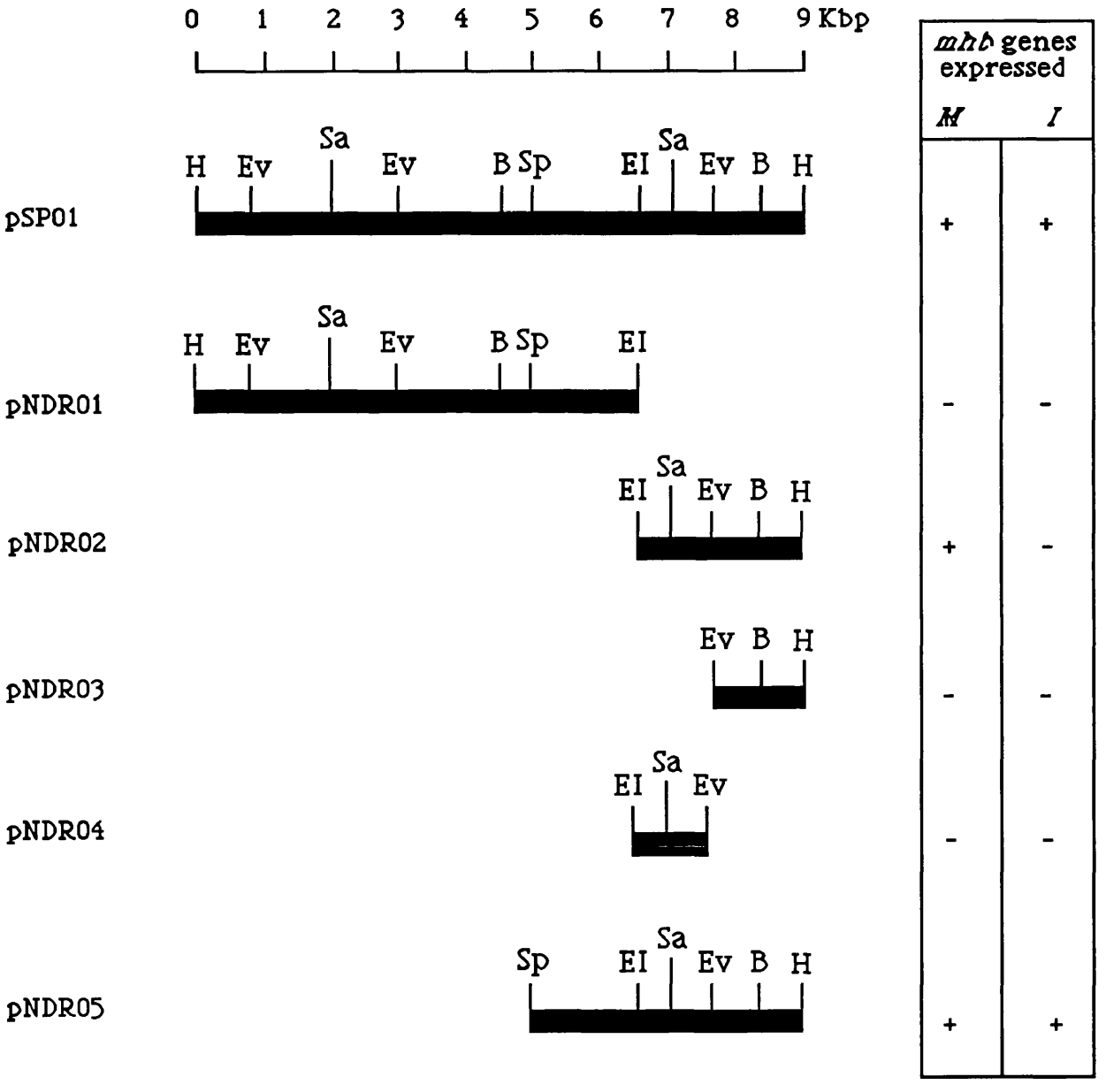
It was decided that the first stage in the molecular analysis of the *mhb* genes of *K.pneumoniae* would be a more detailed characterisation of the existing partial genomic clone pSP01. This would be of major assistance in the development of a scheme which would allow the

cloning of the remaining *mhb* genes, assuming that the remaining genes of the pathway would be clustered in the flanking DNA.
The first stage in the characterisation of this plasmid was the determination of a restriction map of the insert which would identify restriction sites which might be of value in localising the *mhb* genes. This was achieved by a combination of single and double restriction enzyme digestions. The restriction map is shown in Figure3.1.
The first sub-clones prepared were constructed using the *Eco*RI site located at co-ordinate 6.6 of the map. Digestion of pSP01 with *Eco*RI and *Hin*dIII cut the insert of *Klebsiella* DNA into two fragments of 6.6 Kbp and 2.4 Kbp. These fragments were then isolated by gel purification from a 1.0% LMP agarose gel and were ligated separately into pUC18 which had been cut with *Eco*RI and *Hin*dIII. The ligation mix was used to transform *E.coli* 5K and the ampicillin-resistant transformants were screened by restriction digestion of plasmid DNA. One sub-clone containing the 6.6 Kbp *Eco*RI-*Hin*dIII fragment was designated pNDR01 and another containing the 2.4 Kbp fragment was designated pNDR02. When *E.coli* 5K harbouring pNDR01 and pNDR02 was screened for the ability to convert 3-HB to 2,5-DHB *in vivo* it became apparent that pNDR02 was able to mediate the conversion whilst pNDR01 was not. This suggested that the *mhbM* gene was located on the DNA lying beyond position 6.6 Kbp of the map. This test was not capable of identifying which fragment the *mhbI* gene was located on so cell-free extracts were prepared. Enzyme assays revealed, as expected, MhbM activity associated with pNDR02 but it was not possible to detect the presence of MhbI in either extract suggesting that cleavage at the *Eco*RI may disrupt the *mhbI* gene.
The examination of the sub-clones pNDR01 and pNDR02 immediately allowed the relative positions of the *mhbM* and *mhbI* genes to be determined.
An attempt was made to locate more precisely the position of the *mhbM* gene by utilisation of the *Eco*RV restriction site lying at co-ordinate 7.7 of the map of pSP01. This site lies almost centrally in the insert DNA of pNDR02 and by digestion with *Eco*RV in combination with either *Eco*RI or *Hin*dIII it would be possible to generate deletants containing each half of the insert from pNDR02. This procedure would involve treatment with mung bean nuclease following digestion in order to blunt the *Hin*dIII or *Eco*RI to allow direct ligation . In this

way the sub-clone pNDR03, which includes DNA from co-ordinate 7.7 to 9.0 was generated by digestion with *Hin*dIII and *Eco*RV and the sub-clone pNDR04 including DNA from map position 6.6 to 7.7 was produced by digestion with *Eco*RV and *Eco*RI. It was shown that neither *E.coli* 5K (pNDR03) nor *E.coli* 5K (pNDR04) was able to mediate the *in vivo* conversion of 3-HB as had *E.coli* 5K harbouring either pSP01 or pNDR02. This could result either from the disruption of a plasmid encoded uptake gene or from the disruption of the *mhbM* gene. Enzyme assays of cell-free extracts showed that neither pNDR03 nor pNDR04 contained an expressible *mhbM* gene favouring the suggestion that the *Eco*RV site might lie within this gene.

Although the relative positions of the *mhbM* and *mhbI* genes had been determined it seemed appropriate to undertake further sub-cloning experiments to identify the boundary of the *mhbI* gene. The identification of restriction enzyme sites which would allow the retention of all of the catabolic functions of pSP01 would be helpful if the ultimate goal of cloning was to obtain the entire catabolic sequence on a single restriction fragment. Examination of the restriction map identified the site for *Sph*I at map position 5.0 as a likely candidate. Therefore, the 4.0Kbp *Hin*dIII-*Sph*I site including the DNA from map position 5.0-9.0 was isolated by gel purification and ligated into appropriately digested pUC19. The plasmids from transformants of *E.coli* 5K were screened by restriction enzyme digestion and one containing the desired fragment designated pNDR05. Enzyme assays of cell-free extracts prepared from this construct indicated, as expected, the presence of the 3-HB monooxygenase. This analysis also detected the presence of the maleylpyruvate isomerase showing that it was possible to remove the cloned DNA from map position 0 to at least 5.0 and still retain the catabolic activities associated with pSP01.The sub-cloning experiments are summarised in Figure 3.1. Additionally it was decided to attempt to obtain information regarding the likely direction of transcription of the *mhb* genes on pNDR05. In order to do this pNDR05 was introduced into *E.coli* NM522 which contains a *lacI*^{q} mutation which allows regulation of the expression of the *lac-UV5* promoter present in the pUC vectors. In this way it is possible to grow cells containing pUC-based constructs in the presence and absence of IPTG, the gratuitous inducer of the *lac*

Figure 3.1
Summary of sub-cloning of the plasmid pSP01.



Key to enzymes B=*Bam*HI EI=*Eco*RI Ev=*Eco*RV H=*Hin*dIII
Sa=*Sal*I Sp=*Sph*I

system, and then to evaluate the effect on the expression levels of the cloned genes. If the *mhb* genes were transcribed in the same direction as that directed by the *lac* promoter then the expectation would be that the presence of IPTG with its consequent increase in *lac* promoter activity would lead to increased *mhb* gene expression. *E.coli* NM522 (pNDR05) was grown on glycerol in the presence and absence of 0.3mM IPTG and the enzyme activities present in cell-free extracts were measured. The results of this analysis are shown in Table 3.1. It is apparent from this experiment that the expression of both the *mhbM* and *mhbI* genes from pNDR05 is increased 6-7 fold when IPTG is present. This suggests that the *mhb* genes are orientated such that they are expressed in the same direction that the *lac* promoter fires, that is, in pUC19, from *Hin*dIII to *Sph*I.

The analysis of pSP01 had localised the *mhb* genes to one end of the insert DNA and , assuming that the remaining genes were clustered on the chromosome of *K. pneumoniae*, this indicated that attempts at cloning should focus on the DNA lying beyond position 9.0. A restriction enzyme had been identified whose pattern of digestion allowed the retention of all of the catabolic functions present on the partial clone and the probable direction of transcription of the *mhb* genes had been determined.

Table 3.1

Expression of activities associated with *E.coli* 5K (pSP01) grown on 20mM glycerol in the presence and absence of 0.3mM IPTG. The ratio is expressed in terms of :Activity (+ IPTG) : Activity (- IPTG).

	Enzyme activities measured (nmole/min/mg protein)	
STRAIN/INDUCER	**MhbM**	**MhbI**
NM522(pNDR05) **+IPTG**	438	309
NM522 (pNDR05) **NONE**	60	52
RATIO	**7.3 : 1**	**5.9: 1**

Key to enzymes:

MhbM= 3-HB monooxygenase
MhbI= maleylpyruvate isomerase

Isolation of clones encoding the complete 3-HB pathway

Identification of most suitable enzyme for genomic library construction

The plasmid pSP01 represented only a partial clone of the *mhb* genes with the *mhbD* and *mhbH* genes being absent. The fact that expression of the activities associated with pSP01 was not inducible suggested that the system which gave regulated expression in *K.pneumoniae* was also absent. Therefore further analysis of the organisation of the *mhb* genes would require the cloning of the functions absent from pSP01.
The *mhb* genes present on pSP01 had been localised by sub-cloning to one end of the cloned DNA. If the assumption were made that the remaining genes were clustered this would suggest that there would be a high probability that the remaining genes could be isolated if the DNA abutting this end were cloned. If it is assumed that the average gene has a size of 1.0 Kbp this would suggest that around 3.0 Kbp of the flanking DNA would be required to encode the missing functions. It was necessary therefore to identify restriction enzymes which would cleave the chromosomal DNA to give rise to fragments which would include sufficient flanking DNA to allow the remaining genes to be cloned *en bloc* This was desirable because it would simplify the subsequent analysis. The ideal case would be to find a restriction enzyme whose pattern of digestion would additionally allow retention of the activities of pSP01.
In order to determine the restriction enzyme that would be most appropriate for the construction of a genomic library it was decided to conduct Southern blot analysis of various digests of chromosomal DNA with a probe known to include *mhb* sequence. It was therefore decided to prepare and label the 2.4 Kbp *Eco*RI-*Hin*dIII fragment of pNDR02 which included the intact *mhbM* gene. Aliquots (5.5µg) of *K. pneumoniae* chromosomal DNA digested overnight with a range of restriction endonucleases were run out on a 0.8% agarose gel before being transferred to a Hybond-N nylon membrane. A positive control of pSP01 DNA (400ng) digested with *Hin*dIII was also included on the agarose gel. The probe had been prepared by labelling the gel purified

restriction fragment with [$\alpha-^{32}P$] dCTP using the random hexanucleotide priming method of Feinberg and Vogelstein (1983). The membrane was treated according to the method of Church and Gilbert (1984) and allowed to hybridise with the probe overnight at 65°C. The membrane was then subjected to a high stringency wash regime, consisting of three 5 minute washes with 40mM Na_2HPO_4, 0.1% SDS and one 10 minute wash with 0.1 X SSC, 0.01% SDS at 65°C before being autoradiographed overnight at room temperature. The result of the autoradiography is displayed in Figure 3.2 along with a photograph of the agarose gel prior to DNA transfer.
Examination of the autoradiograph showed a very strong signal from the positive control plasmid DNA and clear hybridisation of the probe to the *K.pneumoniae* chromosomal DNA. Close scrutiny of the pattern of hybridising fragments in the various digests showed that the restriction enzyme *Sph*I gave rise to a fragment of approximately 11 Kbp. Examination of the restriction map of the insert of pSP01 suggested that this 11 Kbp would include 4.0 Kbp of DNA present in the existing clone and therefore approximately 7 Kbp of the flanking DNA. It seemed that this would be sufficient additional material for there to be a high probability of isolating the remaining *mhb* genes , assuming that they were tightly clustered. The additional benefit of using this enzyme that had been demonstrated in the sub-cloning of pSP01 was that it would allow the catalytic activities of pSP01 to be retained. This enzyme thus appeared to meet the conditions which would maximise the likelihood of isolating the supposed *mhb* operon intact. Therefore the conclusion was that *Sph*I would be used for the construction of the genomic library.

Construction of genomic library of *Klebsiella pneumoniae* DNA

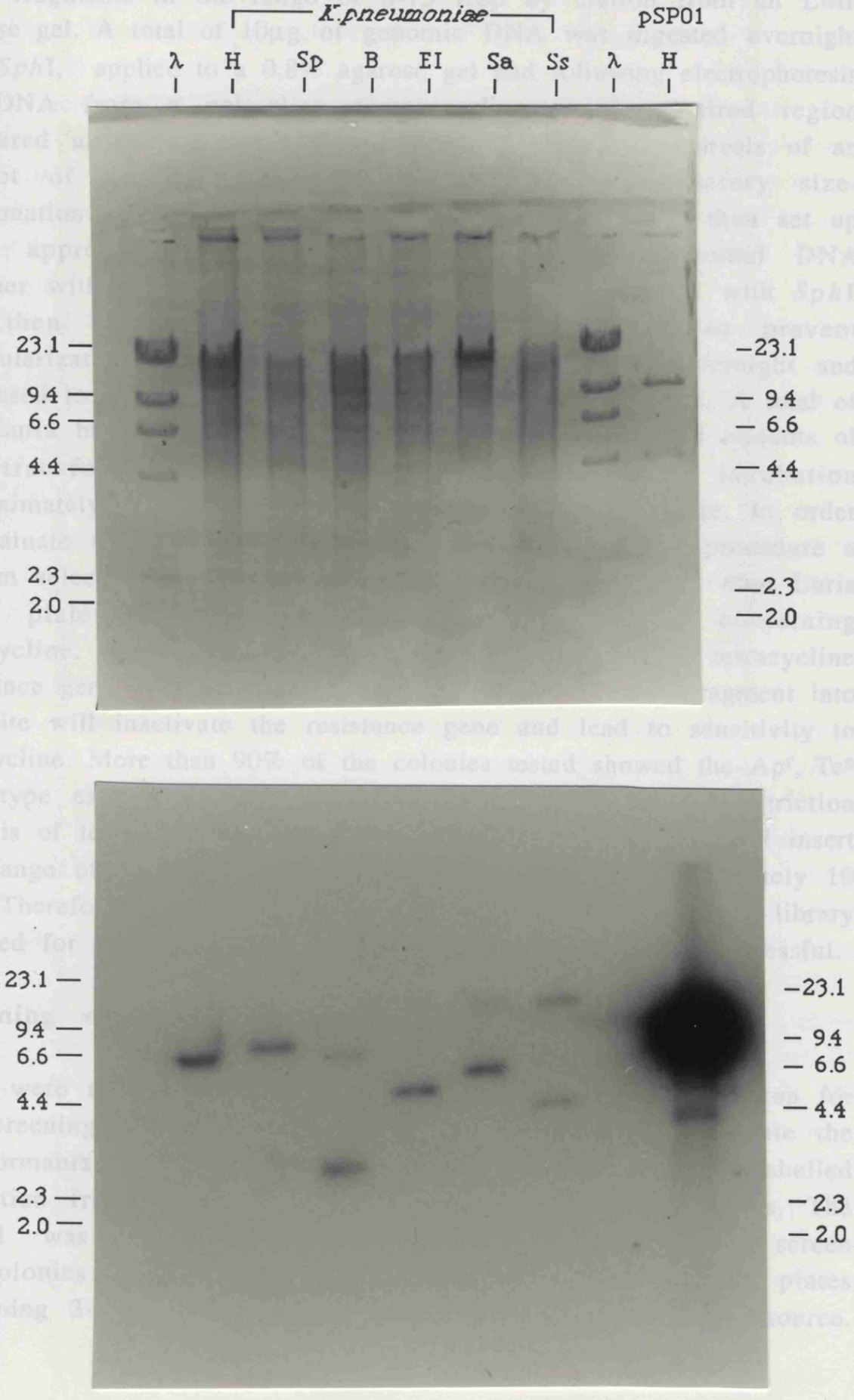
The analysis described in the previous section had not only identified a restriction enzyme suitable for use in the cloning of the *mhb* genes but had also suggested what the size of the desired fragment would be. This therefore afforded the opportunity of size-selecting the DNA which would be used for library construction. This would result in the production of an enriched library and would reduce the screening requirement of the library. The hybridising *Sph*I fragment had an

Figure 3.2

Southern blot analysis of *K.pneumoniae* chromosomal DNA digested wi various restriction endonucleases. 5.5µg samples were run out on a 0.8% agaro gel and transferred to Hybond-N membrane before being probed wi radiolabelled 2.3 Kbp *Eco*RI/*Hin*dIII fragment from pSP01. *Hin*dIII digeste pSP01 was also included to act as a positive control Following the washi procedure described in Chapter 2 the autoradiograph was exposed for 24hrs room temperature.

Key to enzymes : H=*Hin*dIII; Sp=*Sph*I; B=*Bam*HI; EI=*Eco*RI; Sa=*Sal*I; Ss=*Sst*I

λ- denotes *Hin*dIII digested phage λ used as DNA molecular weight marker


K. pneumoniae
pSP01
λ H Sp B EI Sa Ss λ H
23.1
9.4
6.6
4.4
2.3
2.0
23.1
9.4
6.6
4.4
2.3
2.0
23.1
9.4
6.6
4.4
2.3
2.0
23.1
9.4
6.6
4.4
2.3
2.0

apparent size of 11 Kbp and it was decided to isolate chromosomal DNA fragments in the range of 8-15 Kbp by elution from an LMP agarose gel. A total of 10μg of genomic DNA was digested overnight with *Sph*I, applied to a 0.8% agarose gel and following electrophoresis the DNA from a gel slice corresponding to the desired region recovered using a phenol-extraction procedure. Electrophoresis of an aliquot of the recovered DNA showed that satisfactory size-fractionation had been achieved. A ligation reaction was then set up using approximately 500ng of size-selected chromosomal DNA together with 100ng of pBR322 which had been digested with *Sph*I and then treated with calf-intestinal phosphatase to prevent recircularization. The ligation was allowed to proceed overnight and then used to transform *E.coli* 5K using standard procedures. A total of five Luria broth-ampicillin plates were spread with 100μl aliquots of the transformation solution. Following overnight incubation approximately 200-300 colonies were visible on each plate. In order to evaluate the quality of the library generated by this procedure a random selection of colonies from one plate were patched onto Luria broth plates containing separately ampicillin and containing tetracycline. The *Sph*I site of pBR322 lies within the tetracycline resistance gene and hence the insertion of a cloned DNA fragment into this site will inactivate the resistance gene and lead to sensitivity to tetracycline. More than 90% of the colonies tested showed the Ap^r, Tc^s phenotype expected of recombinant plasmids. Furthermore restriction analysis of ten colonies showing this phenotype indicated a total insert size range of 6-16 Kbp with a mean insert size of approximately 10 Kbp. Therefore, it appeared that the construction of a genomic library enriched for fragments in the desired size range had been successful.

Screening of the genomic library

There were two distinct approaches which could have been taken for the screening of the genomic library. One was to replica plate the transformants onto nitrocellulose and then to use the radiolabelled restriction fragment probe in colony hybridisation experiments. The second was a somewhat more direct approach which was to screen the colonies for the ability to grow on minimal medium plates containing 3-hydroxybenzoate as the sole carbon and energy source.

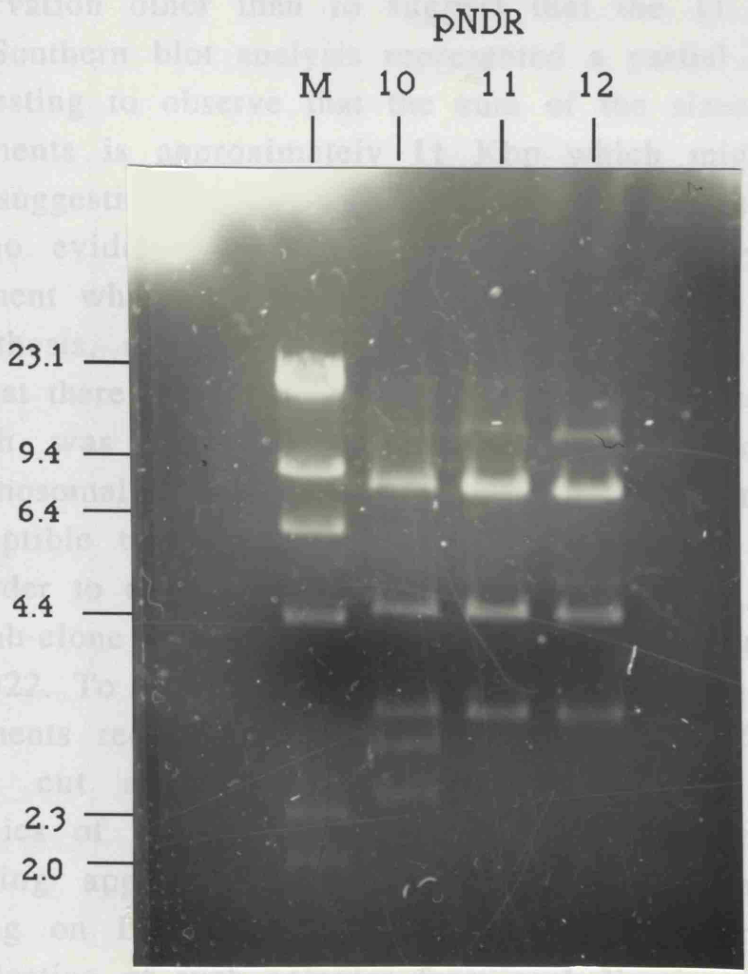
This latter alternative was more speculative in the sense that it would only give a positive result if the entire pathway for 3-HB catabolism were present and would not pick up incomplete clones even though they might contain some activities absent from pSP01. It was felt that the properties of *E.coli* 5K (pSP01) , particularly the demonstrable ability to take up and partially metabolise significant quantities of 3-HB, when allied with the considerable amount of flanking DNA that would be cloned using *Sph*I were such that it would be worthwhile to attempt using the direct-growth test approach on at least a portion of the library.

Colonies from two out of the five transformation plates were therefore replica-plated onto fresh Luria broth-ampicillin plates and onto minimal medium plates containing 5mM 3-HB which also contained the growth requirements of *E.coli* 5K. After overnight incubation a total of five colonies showing strong growth on the 3-HB plates were observed. This seemed to show that the cloning strategy had been effective and that clones including the entire pathway had been obtained. In order to verify this it was decided to patch material from each colony onto fresh 3-HB, Luria broth-ampicillin and Luria broth tetracycline plates to show that the phenotypes expected were reproducible. Whilst three of the colonies tested showed exactly the expected properties being 3-HB$^+$, Ap^r, Tc^s the other two colonies showed much weaker growth on 3-HB and less clear tetracycline sensitivity and were consequently eliminated from further study.

Plasmid DNA was prepared from each of the three remaining candidate clones and this was subjected to restriction analysis. From the Southern blot analysis it was predicted that digestion with *Sph*I would give rise to an insert fragment of approximately 11 Kbp in addition to a fragment of 4.3 Kbp corresponding to the pBR322 vector. The result of such a digest is shown in Figure 3.3 and it is clear that none of the three candidate clones designated pNDR 10,11 and 12 respectively exhibited a fragment of the expected size. Instead all three contained non-vector fragments of around 8 Kbp and 3.3 Kbp and pNDR10 contained further fragments of 3.0 and 2.5 Kbp. The presence of additional fragments might have resulted from a partial digestion of the chromosomal DNA or through adventitious incorporation during ligation and so is not particularly unusual but what was very surprising was that the predicted 11 Kbp fragment

Figure 3.3

Screening of clones allowing growth of *E.coli* 5K on 3-hydroxybenzoat
Plasmids were digested with *Sph*I and run out on a 0.8% agarose gel.


pNDR
M
10
11
12
23.1
9.4
6.4
4.4
2.3
2.0

was absent. There would seem to be no ready explanation for this observation other than to suggest that the 11 Kbp fragment seen in the Southern blot analysis represented a partial digestion product, it is interesting to observe that the sum of the sizes of the common insert fragments is approximately 11 Kbp which might be seen to support this suggestion. The problem with this idea is that there appears to be no evidence on the autoradiograph of any hybridisation to the fragment which would have resulted from total digestion . A possible hypothesis, albeit without experimental support, to explain this might be that there was an *Sph*I site present in the genome of *K.pneumoniae* which was in some way protected from digestion in prepared chromosomal DNA which as a result of replication in *E.coli* became susceptible to cleavage.

In order to clarify the exact nature of the *mhb* clones it was decided to sub-clone the 8.0 Kbp and 3.3 Kbp fragments from pNDR11 into pBR322. To this end pNDR11 was digested with *Sph*I and the required fragments recovered from a 0.8% LMP gel before being ligated with *Sph*I cut and phosphatase treated pBR322. Ampicillin resistant colonies of *E.coli* 5K were obtained from each ligation. Colonies showing appropriate antibiotic resistances were then identified by testing on Luria broth-ampicillin and Luria broth-tetracycline plates. A selection of such colonies from each ligation were then patched onto minimal medium plates containing 5mM 3-HB to test for the ability of each sub-cloned fragment to allow growth. The expectation was that it would be highly unlikely that the presence of the 3.3 Kbp fragment would allow growth since its coding capacity would seem to be insufficient for the expression of the entire pathway. Furthermore it had been shown that cloning of the entire operon would require at least the 4.0 Kbp of DNA bounded by the *Sph*I site of pSP01. Therefore the important test would be whether the colonies containing the larger fragment would be able to utilise 3-HB as a carbon source. It was clear from the examination of the 3-HB plates that the presence of a plasmid containing the 8.0 Kbp *Sph*I fragment was indeed able to allow *E.coli* 5K to grow on the aromatic carbon source. The plasmid DNA from one of these colonies was prepared and was subsequently designated pNDR20. It appeared therefore that the plasmid pNDR20

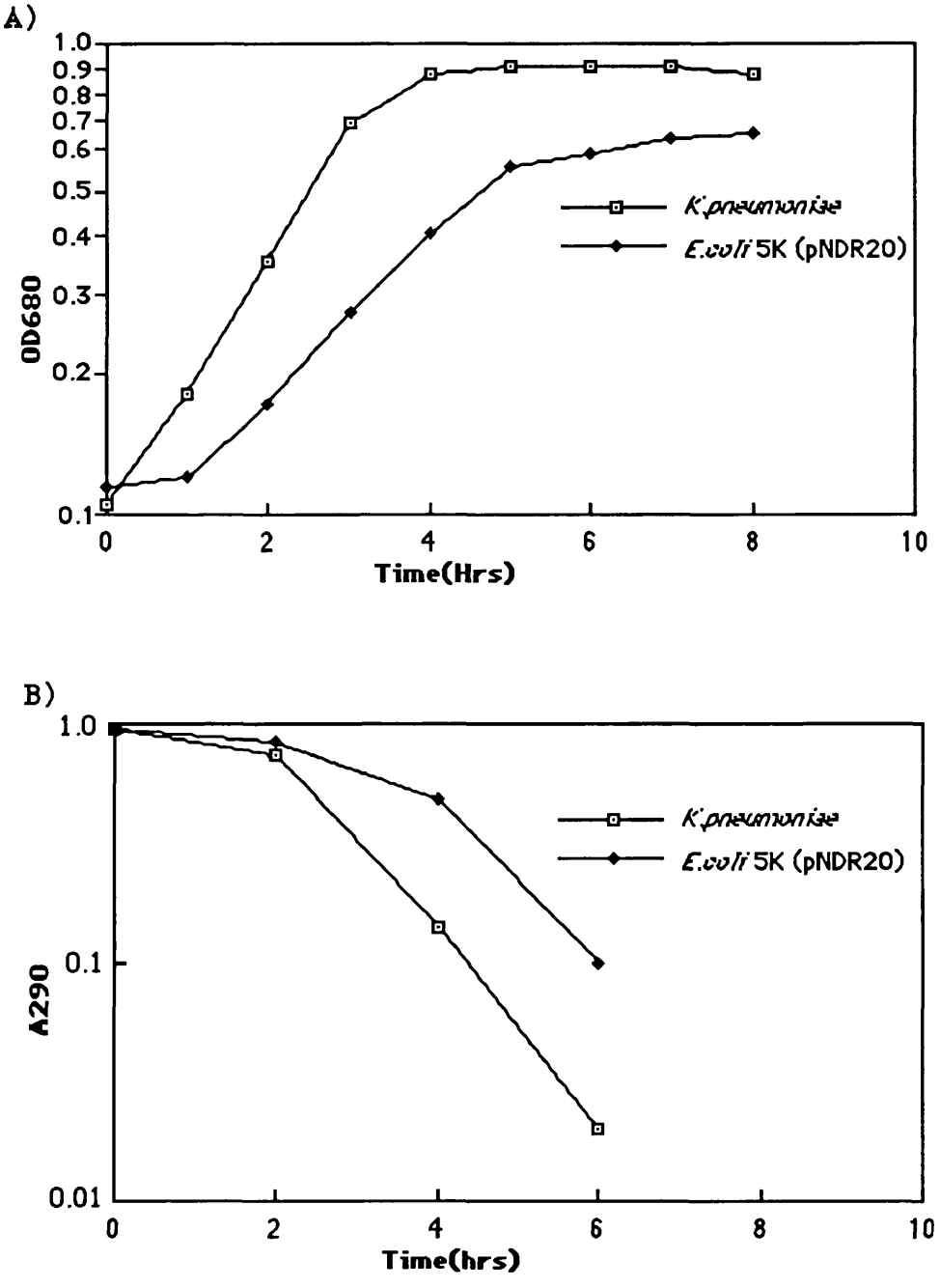
represented an artificially constructed catabolic plasmid since its presence was sufficient to give a new metabolic capability.
The next stage of the analysis was firstly to characterise the properties of pNDR20 in both a physiological sense through the analysis of the behaviour of cells harbouring this plasmid and at a more molecular level.

Properties of cells harbouring pNDR20

It had been shown during the cloning procedure that the presence of pNDR20 in *E.coli* 5K allowed growth on 3-HB plates. The next stage in the analysis was therefore to be a consideration of the behaviour of *E.coli* 5K (pNDR20) in liquid culture. This was to be in terms of both the growth rate and the pattern of substrate utilization and also the specific enzyme activities of pathway components. This information could then be examined in conjunction with the properties of *K.pneumoniae* to see how the systems compared.
Growth experiments were conducted by the sub-culturing of overnight Luria broth cultures of *E.coli* 5K (pNDR20) and *K.pneumoniae* M5a1 into flasks of minimal media containing 5mM 3-HB. The overnight culture of *E.coli* 5K (pNDR20) was grown in the presence of ampicillin to select for plasmid retention, however upon sub-culture ampicillin was not included so that selection was solely based on the ability to metabolise 3-HB. For the *E.coli* 5K cells the growth medium incorporated the specific strain growth requirements. Cells were initially sub-cultured to an OD_{680} of approximately 0.1 and samples were removed at hourly intervals for measurement of the OD_{680}. In order to obtain information regarding the pattern of substrate utilization growth medium supernatants prepared by brief centrifugation of cell samples were diluted and the A_{290}, which served as an indicator of the 3-HB concentration, determined. The results of this analysis are depicted in Figure 3.4.
The growth rates were measured through the determination of the doubling time of the cells during the exponential phase of growth. For the *K.pneumoniae* M5a1 this was calculated to be 72 minutes and for the *E.coli* 5K (pNDR20) it was 118 minutes. It seems therefore that the growth rate of the clone is only about 50% slower than that of *K.pneumoniae* which , given the fact that the clone represents a

Figure 3.4

Comparison of a) growth and b) substrate utilisation characteristics of *K.pneumoniae* M5a1 and *E.coli* 5K (pNDR20) grown on 5mM 3-hydroxybenzoate (3-HB) plotted logarithmically. Substrate utilisation is recorded through the reduction in the Absorbance at 290nm associated with metabolism of 3-HB, the values quoted represent absorbances of supernatants diluted ten-fold. After 6hrs no substrate was detected in the medium.



completely synthetic situation, suggested that the cloned genes are capable of a reasonable level of expression in the foreign background. A similar pattern was observed for the substrate utilization where the fall in the 3-HB concentration was somewhat faster for the 'parent' organism than for the clone suggesting more efficient utilization in the former. Again, however, the clone appears to be able to function at a level which is broadly comparable to *K. pneumoniae* suggesting that there were no pronounced problems associated with the expression of the cloned genes in the foreign background.

In order to evaluate more directly the expression of the individual *mhb* genes cell-free extracts were prepared from *E.coli* 5K (pNDR20) and *K.pneumoniae* that were harvested in late exponential phase. Enzyme assays were performed and the results are recorded in Table 3.2. It is apparent that, rather surprisingly, the specific activities recorded for *E.coli* 5K (pNDR20) are actually higher than those for *K.pneumoniae* . The activities of MhbD are very similar but the levels of the other enzymes are 1.5-2 fold higher in the clone. This observation is somewhat surprising in view of the fact that the growth rate was more rapid for *K.pneumoniae* than for the clone and thus there appears to be no direct correlation between growth rate and enzyme activities measured during exponential growth. What this means is that factors other than the raw enzyme levels are involved in determining the growth rate. It might be for example that the presence of plasmid, particularly one present in multiple copies, in the cell imposes a general metabolic load which restricts the rate of growth. However it was noticeable that the utilization of 3-HB was slower with the clone than with *K.pneumoniae* suggesting a specific limitation of the 3-HB catabolic system. One area which could restrict the growth rate and rate of substrate utilization would be the 3-HB uptake system which might be less efficient in the clone. If this were the case then increased activities of the other enzymes would not compensate for reduced entry of substrate.

The enzyme activities measured for the clone were as stated somewhat higher than those of *K.pneumoniae* , this might have been expected as the cloned *mhb* genes were present at a higher copy number than on the chromosome of *K.pneumoniae* , indeed much higher levels of expresssion might have been expected. Thus, it might

Table 3.2

Comparison of the activities of the enzymes for 3-hydroxybenzoate catabolism in cells of *K.pneumoniae* and of *E.coli* 5K harbouring the clone pNDR20. Cells were grown on 5mM 3-hydroxybenzoate.

	SPECIFIC ACTIVITIES nmole/min/mg protein			
STRAIN	**MhbM**	**MhbD**	**MhbI**	**MhbH**
Klebsiella pneumoniae M5al	168	1150	237	600
Escherichia coli 5K (pNDR20)	295	1421	419	1009

Key to enzymes:

MhbM= 3-HB monooxygenase
MhbD= 2,5-DHB dioxygenase
MhbI= maleylpyruvate isomerase
MhbH= fumarylpyruvate hydrolase

Table 3.3

Enzyme activities of 3-HB grown *Klebsiella pneumoniae* containing pNDR20 and pBR322 to show expression levels of the cloned genes in the natural background.

	SPECIFIC ACTIVITIES nmole/min/mg/protein			
STRAIN	**MhbM**	**MhbD**	**MhbI**	**MhbH**
K.pneumoniae M5a1 (pBR322)	170	1210	761	438
K.pneumoniae M5a1 (pNDR20)	461	3436	2136	1992

Key to enzymes:

MhbM= 3-HB monooxygenase
MhbD= 2,5-DHB dioxygenase
MhbI= maleylpyruvate isomerase
MhbH= fumarylpyruvate hydrolase

be possible that the cloned genes were in fact being poorly expressed in *E.coli* but this was being masked by the presence of multiple copies of the genes. Alternatively it was possible that the cloned genes were being expressed at a level appropriate for growth of the cells and that very high expression might be deleterious for the cell. In order to clarify this it was decided to introduce pNDR20 into *K.pneumoniae* to see how the cloned genes would function in the natural background. It was necessary to have *K.pneumoniae* transformed with pBR322 as a control so that the background levels of the *mhb* genes might be measured. This was also required for growth rate measurements because it would be necessary to include antibiotic in the medium to select for the maintenance of the plasmid. This was not required when *E.coli* was the host as growth on 3-HB would not be possible without the plasmid and this would serve as as selective pressure for its retention. *K.pneumoniae* is able to grow on 3-HB and so a different means of selection had to be applied.

The growth of *K.pneumoniae* transformed with pBR322 and with pNDR20 on 5mM 3-HB in the presence of 200μg/ml ampicillin was monitored and respective doubling times of 90 and 85 minutes were determined. This showed that the presence of the cloned genes did not cause any great increase in the growth rate. Enzyme assays were performed on cell-free extracts prepared from *K.pneumoniae* containing each plasmid. The cells had been grown on 5mM 3-HB and were harvested in the late logarithmic phase. The results of this analysis are shown in Table 3.3.

It seems clear that the activities of *K.pneumoniae* (pNDR20) are considerably higher than those of *K.pneumoniae* (pBR322) but that the net activities, that is those which are directly attributable to the clone are little different from those measured with pNDR20 in *E.coli* 5K. The exception to this is the MhbI gene product which is expressed at a considerably higher level , although it is apparent that its basal expression in *K.pneumoniae* is much higher than that recorded in Table 3.2, the reason for this anomaly is not clear. The overall conclusion is that most of the genes are expressed in *K.pneumoniae* at a level no more than two-fold that in *E.coli* suggesting that there are no particular problems with expression in the foreign background. It is also apparent that the growth rates of *K.pneumoniae* harbouring each of pBR322 and pNDR20 are not closely correlated with the

Table 3.4

Expression of the enzymes of 3-hydroxybenzoate catabolism in *E.coli* 5K (pNDR20) grown on 20mM glycerol +/- 5mM 3-HB with ampicillin to test for inducibility of the cloned genes.

	SPECIFIC ACTIVITIES nmole/min/mg/protein			
GROWTH CONDITION	**MhbM**	**MhbD**	**MhbI**	**MhbH**
GLYCEROL	15	82	26	85
GLYCEROL + 3-HB	137	344	366	798

Key to enzymes:

MhbM= 3-HB monooxygenase
MhbD= 2,5-DHB dioxygenase
MhbI= maleylpyruvate isomerase
MhbH= fumarylpyruvate hydrolase

specific activities measured for the individual enzymes of the pathway.

It had been shown (Jones and Cooper, 1990) that the expression of the *mhb* genes in *K.pneumoniae* was regulated being induced from a low basal level during growth in the presence of 3-HB. Therefore it was of considerable interest to determine whether the *mhb* genes present on pNDR20 were inducible. If this were the case then it would mean that in addition to the structural genes of the pathway that the gene(s) involved in the regulatory system had also been cloned. The presence of the regulatory system on a cloned DNA fragment might then facilitate the characterisation of the mode of regulation. In order to address the question of whether expression of the cloned *mhb* genes was inducible *E.coli* 5K was grown on 20mM glycerol in the presence and absence of 5mM 3-HB and ampicillin. Cell-free extracts were prepared and enzyme assays were performed. The results of this analysis are recorded in Table 3.4.

It is evident that the expression of pNDR20 is indeed regulated in *E.coli* 5K with the activities of the *mhb* genes in the cells grown in the presence of 3-HB being 4-15 times higher than those of cells grown on glycerol alone. This demonstrated that the system involved in regulation of the expression of the *mhb* genes was present intact on the clone. A further analysis of aspects of the regulatory system are discussed in a later chapter.

Summary

The work presented in this chapter shows how an analysis of the existing partial clone pSP01 was applied to determine a strategy for the cloning of the entire pathway for the catabolism of 3-HB. The execution of this scheme resulted in the isolation of clones which allowed *E.coli* 5K to acquire a novel catabolic function, that is growth on 3-HB. The nature of these initial clones was more complex than expected so a simpler construct designated pNDR20 was isolated which encoded the functions on an 8.0 Kbp *Sph*I fragment. The properties of cells harbouring this plasmid were examined and it was shown that *E.coli* cells containing pNDR20 were able to grow at a rate only 50% slower than wild-type *Klebsiella pneumoniae*. The expression of the *mhb* genes in a heterologous background was shown to be comparable to that in the natural background. It was then

shown that the expression of the cloned genes was inducible and that overall the properties of the cloned genes mirrored those of the natural system. The plasmid pNDR20 may therefore be viewed as a fully functional artificial catabolic plasmid albeit not self transmissable. Further examination of the cloned *mhb* genes is recorded in the following chapters.

CHAPTER 4

Further analysis of pNDR20

Further analysis of pNDR20

The isolation of pNDR20 is described in the preceding chapter along with a description of some of the basic properties of the plasmid. This chapter will focus on the attempts at further characterisation of the plasmid particularly with respect to the determination of the organisation of the *mhb* genes and also some aspects of the regulatory system.

Determination of the order of the *mhb* genes

The gene order of a portion of the 3-HB degradative pathway had already been determined during the characterisation of pSP01. Therefore the main requirement was to determine the relative positions of the *mhbD* and *mhbH* genes and the postulated regulator gene designated *mhbR*. It will be seen shortly that the relative positions of the two structural genes were determined quite readily but that the localisation of the regulatory gene was more problematic.

Determination of the order of the *mhbD* and *mhbH* genes

The analysis of pSP01 summarised in Figure 3.1 had focused upon the DNA running from the *Sph*I site at map position 5.0 to the *Hin*dIII site at co-ordinate 9.0. The 8.0 Kbp *Sph* I fragment present in pNDR20 includes DNA from map position 5.0 running through to a position designated 13.0 by extension of the map of pSP01. The obvious sub-cloning experiment would involve the isolation of the DNA novel to pNDR20 separate from that present on pSP01. This could be readily achieved by digestion of the insert of pNDR20 with *Hin*dIII to generate one segment of DNA identical to that present in pNDR05 (prepared from pSP01) and one novel fragment. If it were assumed that the *mhb* genes were clustered without significant intervening DNA then there would seem to be a high probability that the *Hin*dIII site would lie within one of the *mhb* genes. It would be expected that the disruption of this gene at this site would mean that the associated activity would be expressed by neither of the *Sph*I/ *Hin*dIII fragments but that all the other activities would be represented on one of the fragments. Thus by consideration of the activities present

on the novel fragment in combination with knowledge of those associated with pNDR05 positioning of the *mhb* genes would be straightforward. The problem that existed was that digestion of pNDR20 with *Hin*dIII and *Sph*I would give rise to three fragments with very similar sizes thereby complicating the sub-cloning procedure. Therefore it was decided to prepare a slightly larger sub-clone containing some of the DNA present on pSP01 as well as the 4.0 Kbp of novel DNA by utilising the *Bam*HI site at map position 8.4. If, however, the analysis of this larger sub-clone was not able to resolve the gene order directly then its construction would greatly facilitate the preparation of the preferred sub-clone.

The first DNA to be sub-cloned therefore was the 4.6 Kbp fragment delineated by the *Bam*HI site at position 8.4 and the *Sph*I fragment at 13.0. This fragment therefore contained all of the DNA novel to pNDR20 as well as a small segment common to both pNDR20 and pSP01. The fragment was isolated from a 1% LMP agarose gel of pNDR20 DNA that had been digested to completion with *Sph*I and *Bam*HI. The fragment was then ligated into appropriately digested pUC19 and *E.coli* 5K transformed with the ligation mixture with selection for ampicillin resistance. Plasmid DNA was prepared from a number of colonies and was then analysed by restriction enzyme digestion. A plasmid containing the desired fragment was isolated and this was designated pNDR21. Enzyme assays of cell-free extracts prepared from *E.coli* 5K (pNDR21) indicated the presence of both the MhbD (2,5-DHB dioxygenase) and MhbH (fumarylpyruvate hydrolase) gene products. It was therefore clear that the construction of pNDR21 had not allowed the relative positions of *mhbD* and *mhbH* genes to be determined and further sub-cloning would be required.

The production of pNDR21 had, as expected, provided a pathway for the preparation of a sub-clone containing the 4.0 Kbp *Sph*I/ *Hin*dIII fragment representing the DNA from position 9.0-13.0. Digestion of pNDR21 with *Sph*I and *Hin*dIII gave rise to fragments which were readily separable on agarose gels and it was possible to isolate the desired 4.0Kbp fragment from a LMP agarose gel. The eluted DNA was then ligated with appropriately digested pUC18 and the mixture used to transform *E.coli* 5K. A plasmid containing the desired fragment was identified by restriction analysis and this was designated pNDR22. When cell-free extracts were prepared from *E.coli* 5K (pNDR22) the

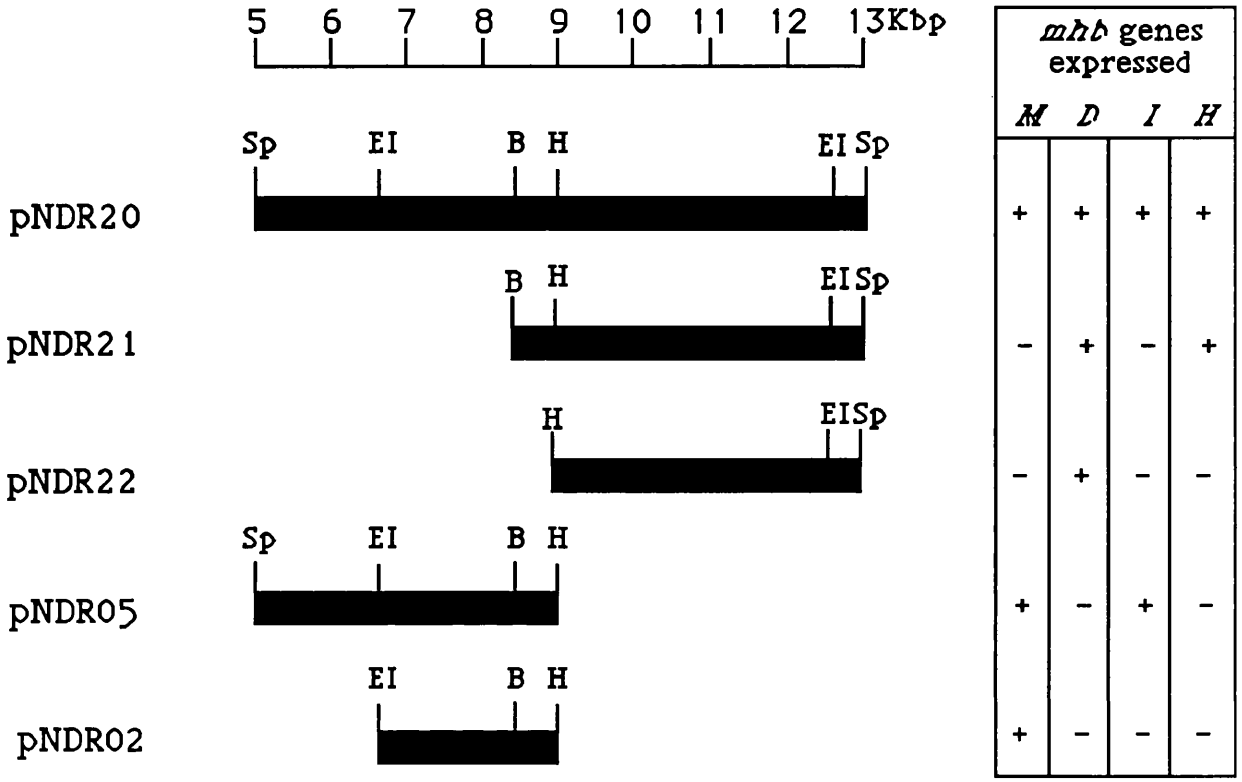
activity of the MhbD (2,5-DHB dioxygenase) gene product was detectable but that of the MhbH (fumarylpyruvate hydrolase) protein was absent. This suggested that the hypothesis that the *Hin*dIII site at position 9.0 lay within one of the *mhb* genes was correct and that the gene involved was *mhbH*. It was therefore evident that it was the *mhbH* gene which immediately bordered the genes present on pSP01 and that the *mhbD* gene occurred further out towards the *Sph*I site at position 13.0. The sub-cloning experiments are summarised in Figure 4.1.

Regulation of the 3-HB degradative pathway

It was shown in Chapter 3 that the *mhb* genes present on the chromosome of *K.pneumoniae* and on the complete genomic clone pNDR20 were induced by growth in the presence of 3-HB. That is that 3-HB appeared to serve as an inducer of the 3-HB pathway. Evidence has been presented (Jones, 1985; Jones and Cooper, 1990) which suggests that 3-HB is not actually an inducer of the pathway *in vivo* but that a downstream intermediate serves this role. This view was based upon experiments with a mutant strain of *K.pneumoniae* designated OH101 which is defective for the 3-HB mono-oxygenase (encoded by the *mhbM* gene) which catalyses the hydroxylation of 3-HB. *K.pneumoniae* OH101 is not defective for any of the other *mhb* genes and will grow on 2,5-DHB. When this strain was grown in the presence of 3-HB no induction of *mhb* gene expression was observed. It therefore seemed that a mutation blocking further metabolism of 3-HB prevented induction thereby implying that 3-HB was not an inducer *per se*. If *K.pneumoniae* OH101 was grown in the presence of 2,5-DHB, the compound normally generated by hydroxylation of 3-HB, then increased *mhb* gene expression was observed. The phenomenon of 2,5-DHB induced expression had similarly been observed with wild-type *K.pneumoniae* M5a1 cells. These findings appeared to confirm the hypothesis that induction was mediated by a downstream pathway intermediate since the production, or presence, of 2,5-DHB was required for the induction process. It was not possible to specifically identify the intermediate required for induction. The fact that increased *mhb* gene expression was apparent when cells of a

Figure 4.1

Sub-cloning of pNDR20 to determine the order of the genes encoding the enzymes of 3-HB catabolism.



Key to enzymes : B=*Bam*HI E=*Eco*RI H=*Hin*dIII Sp=*Sph*I

further mutant strain *K.pneumoniae* 214/I (*mhbM*, *mhbI*) were grown in the presence of 2,5-DHB suggested that fumarylpyruvate, which cannot be produced by this strain, was not the inducer. This left only 2,5-DHB and maleylpyruvate as potentially inducing intermediates.
Dissection of the regulation of the cloned *mhb* genes would require that the inducibility of *mhb* gene expression on particular sub-clones could be evaluated. A problem area that emerged during the early stages of this analysis related to the choice of the inducer. The most obvious candidate for use was 2,5-DHB. However, several of the sub-clones to be examined for the presence of the regulatory gene (e.g pNDR21/22) did not express a functional maleylpyruvate isomerase. When this activity is absent but 2,5-DHB dioxygenase is present the 2,5-DHB can be metabolised only as far as maleylpyruvate. Previous studies have suggested that maleylpyruvate is toxic to cells and that accumulation has very severe effects on cell-growth (Jones and Cooper, 1990). Consequently it was perhaps not surprising when it was found that cultures of *E.coli* 5K harbouring plasmids such as pNDR21/22 could not be grown in the presence of 2,5-DHB although cells harbouring pNDR20 could grow under these conditions. This meant that it was not possible to test for the presence of the *mhbR* gene on such constructs using 2,5-DHB as inducer. It was apparent that alternative approaches would be required which would allow testing for the presence of the regulatory genes. Two immediate possibilities were considered . The first of these was to screen a number of different aromatic compounds for their ability to serve as inducers of the *mhb* genes with the added proviso that they should not be metabolised to any toxic compound. The second alternative would be to generate constructs in which the activity of the *mhbI* gene was present so that the toxic maleylpyruvate could be removed. A problem with this second approach was that the gene order determined from sub-cloning of pSP01 and pNDR20 placed the *mhbI* gene at the extreme edge of the operon and the need to retain this segment of DNA would severely curtail the available sub-cloning options. Therefore in the first instance it was decided to screen a range of potential inducer analogues.

Identification of inducer analogues

There were two main criteria involved in the identification of inducer analogues. These were that the compounds should be sufficiently like the true inducers to be effective and that they should not prove toxic either directly or via subsequent metabolism. It was decided in the first instance to use *E.coli* 5K(pNDR20) for the screening procedure since this clone was known to show inducible expression. Once a compound had been identified as a general inducer then possible toxicity problems would become evident when that compound was used to test the inducibility of sub-clones in the attempt to locate the *mhbR* gene.

The initial screening of candidate inducer analogues was designed to focus on compounds having structural similarity to 2,5-DHB rather than those similar to maleylpyruvate because of the relative availabilities of suitable compounds. The choice of candidate inducer analogues was refined further by consideration of data on the ability of various aromatic compounds to serve as competitors with 2,5-DHB for entry into the cell via the 3-HB uptake system (Jones, 1985). In the first instance three compounds which had shown competition with 2,5-DHB were selected these were 2-hydroxybenzoate (salicylate), 2,3-dihydroxybenzoate and 3,4-dihydroxybenzoate (protocatechuate). An additional compound, 5-methoxy-2-hydroxybenzoate (5-methoxy-salicylate) which had not been tested as a competitor for uptake was also included in the initial screening. This compound is very similar to 2,5-DHB having similar substituents at both the 2- and 5- positions of the aromatic ring.

The activities of *E.coli* 5K (pNDR20) were measured from cells grown on glycerol in the presence and absence of each aromatic compound, added to a concentration of 5mM, and the ability of each compound to induce the expression of the *mhb* genes evaluated. The results of this analysis are summarised in Table 4.1. It appeared that two of the compounds tested, 2-hydroxybenzoate and 5-methoxy-2-hydroxybenzoate were able to induce *mhb* gene expression whilst the remaining compounds appeared to be non-inducers. Interestingly an investigation of the pattern of expression of the *mhbD* gene from *Pseudomonas testosteroni* has shown that growth in the presence of 5-methoxy-2-hydroxybenzoate, though not 2-hydroxybenzoate, also

leads to induction of the *mhb* genes (Harpel and Lipscomb, 1990a). The next stage of the analysis was to attempt to determine whether there would be any toxicity problems associated with the use of the candidate inducer analogues in conjunction with sub-clones containing the *mhb* genes . This was tested by growing *E.coli* 5K (pNDR22), which had been shown to experience toxic effects when grown in the presence of 2,5-DHB, on glycerol with and without each of the salicylate based compounds. This investigation showed that neither compound had severe effects upon the growth of these cells and that they therefore fulfilled the two major criteria for inducer analogues being both inducing and non-toxic. It was then proposed to use the inducer analogues to attempt to determine whether expression of the *mhb* genes was regulated on any of the sub-clones.

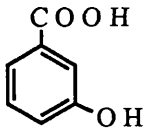
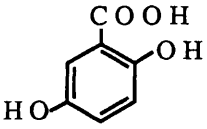
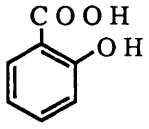
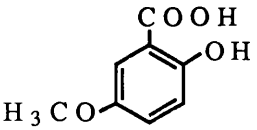
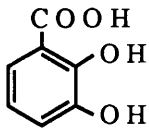
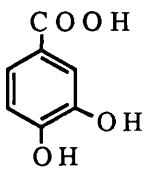
Extracts were prepared from cells of *E.coli* 5K (pNDR22) and also *E.coli* 5K (pNDR21) grown on glycerol in the presence and absence of each of the salicylate-based compounds . Comparison of the activities revealed that the expression of the *mhb* genes present on each clone was indeed increased when the inducer analogues were present, the information is summarised in Table 4.1. This suggested that the *mhbR* gene was in fact present on these sub-clones.

Localisation of the *mhbR* gene

The use of inducer analogues had allowed the presence of the regulatory system to be detected on pNDR22. The next stage in the analysis would therefore be to localise the *mhbR* gene and by doing so complete the determination of the *mhb* gene order. It was decided that the best approach for the localisation of the *mhbR* gene would be via the generation of deletants of pNDR22. This was the smallest construct retaining an enzyme activity which would allow testing for induction. In the first instance it was decided to generate unidirectional deletions from the *Sph*I end of this sub-clone using the exonuclease III based deletion system (Henikoff, 1984). A system involving unidirectional deletion was preferred because of the greater flexibility of this system in terms of its ability to generate deletions of various sizes in a single experiment. This flexibility was important because it was necessary to retain the activity of the *mhbD* gene in order to assay for the presence of the *mhbR* gene.

Table 4.1

Screening of compounds as inducers of *mhb* gene expression. Tests were performed sequentially with a satisfactory outcome from a test required before proceeding onto the next stage.

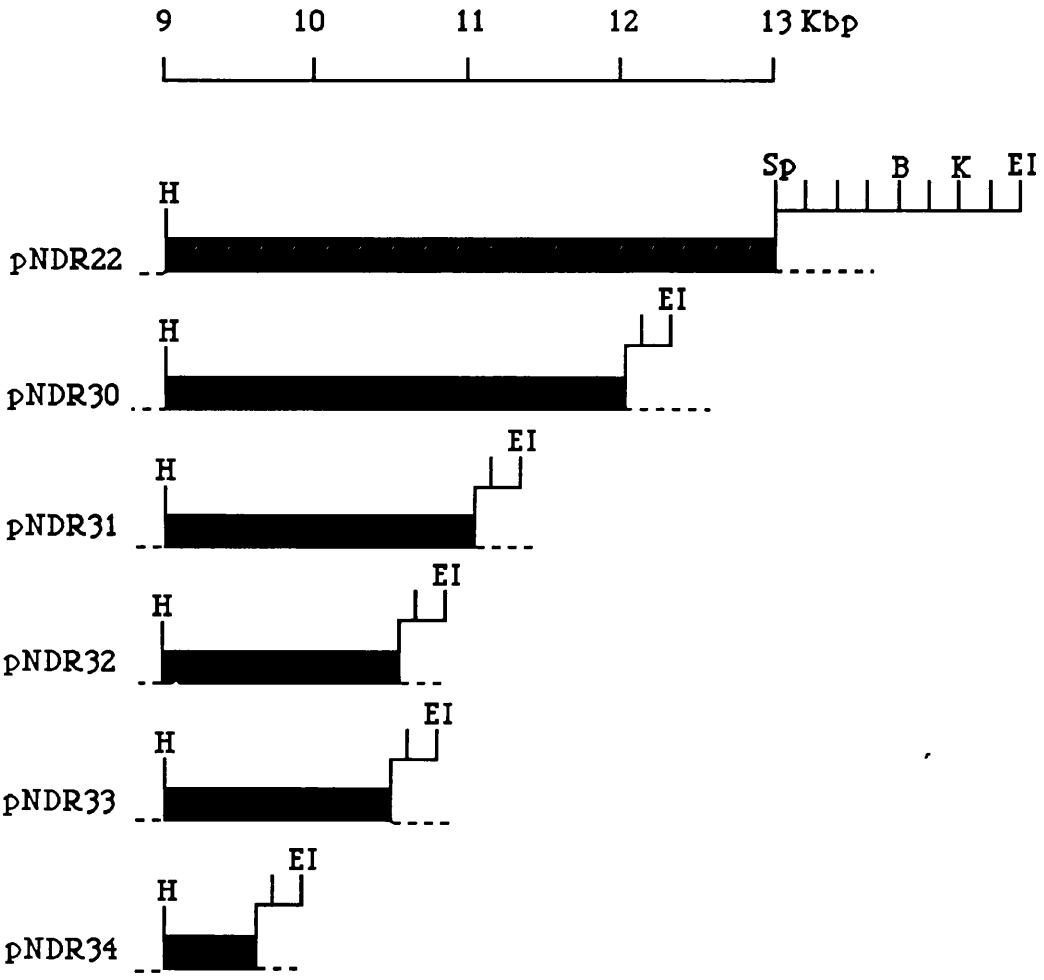
COMPOUND	Induction of pNDR20	Growth inhibition	Induction of pNDR21/22
 3-Hydroxybenzoate	**YES**	**NO**	NO
 2,5-Dihydroxybenzoate	**YES**	YES	
 2-Hydroxybenzoate	**YES**	**NO**	**YES**
 5-Methoxy-2-hydroxybenzoate	**YES**	**NO**	**YES**
 2,3-Dihydroxybenzoate	NO		
 3,4-Dihydroxybenzoate	NO		

Exonuclease III reactions were performed as described in Chapter 2. In this instance pNDR22 was digested with *Bam*HI so as to render the cloned insert susceptible to exonuclease III digestion and with *Kpn*I to ensure protection of the vector. In order to maximise the chances of being able to localise the *mhb* genes present on pNDR22 it was decided to generate a wide spectrum of deletions ranging from a few hundred base pairs right up to 3 Kbp, the latter representing deletion of 75% of the insert DNA. The properties expected from the deletants over this range would depend upon the organisation of the *mhb* genes. If, for example, the *mhbR* gene was located such that it would be encountered in the deletion series before the *mhbD* gene then expectation would be that deletants might initially retain regulated expression of *mhbD* progressing then to non-regulated expression followed ultimately by loss of the *mhbD* gene. Samples from each time-point of the deletion procedure were ligated overnight and this material was used to transform *E.coli* 5K. Plasmids from random colonies at a variety of time-points were screened and five constructs encompassing a range of deletion selected for further analysis. These constructs were designated pNDR30-34 in ascending order of the amount of DNA deleted and are represented in Figure 4.2.

Initially these constructs were screened for the expression of the *mhbD* gene. It emerged from enzyme assays of cell-free extracts prepared from *E.coli* 5K harbouring each of these plasmids that neither pNDR33 nor pNDR34 contained detectable levels of the 2,5-DHB dioxygenase (MhbD). Conversely the levels of MhbD were very high in each of the other three constructs, indeed with the plasmid pNDR32 MhbD accounted for up to 15% of soluble cell protein (this is considered in more detail in Chapter 5). The specific activities measured are recorded in Table 4.2. These results appeared to locate the *mhbD* gene to around map position 10.5. The fact that the extent of expression increased with increasing deletion up to the point where MhbD activity was abolished suggested that the *mhbD* gene was transcribed in the same direction as the firing of the *lac* promoter. In this case towards the *Hin*dIII site. It appeared from the measurement of activities of cells grown in the presence and absence of 5mM 2-hydroxybenzoate that the expression of the *mhbD* gene in the deletants tested was constitutive with no evidence of induction.

Figure 4.2

Deletants generated by treatment of pNDR22 with exonuclease III in order to determine the relative position of the *mhbD* and *mhbR* genes.



mhb genes expressed	
D	*R*
YES	YES
YES	NO
YES	NO
YES	NO
NO	–
NO	–

Key to restriction sites: B= *Bam*HI; EI= *Eco*RI; H= *Hin*dIII; K= *Kpn*I; Sp= *Sph*I

This loss of regulation suggested that the *mhbR* gene lay at a higher map co-ordinate than the *mhbD* gene, that is closer to the *Sph*I site of pNDR22.
This finding appeared to conclude the analysis of the organisation of the *mhb* genes and this order is shown in Figure 4.3.

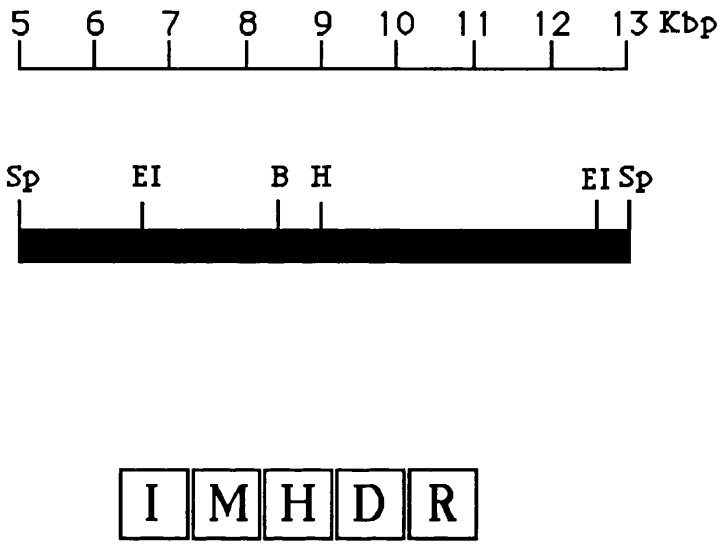
Further consideration of the regulation of cloned *mhb* genes

The evidence obtained from the analysis of the deletion derivatives of pNDR22 in *E.coli* 5K suggested that loss of the *mhbR* gene gave rise to a pattern of expression which might be described as NON-INDUCIBLE/HIGH. This pattern, characterised by an elevated level of non-induced basal expression, is typical of systems which are negatively regulated. In these systems expression is repressed when the inducer is absent and the expression is increased when the presence of the inducer causes derepression. Loss of the repressor gene, for example by deletion or mutation leaves the system in a state of permanent derepression with gene expression switched on.

Whilst the analysis of the patterns of expression of the pNDR22 based deletants in *E.coli* 5K allowed the position of the *mhbR* gene to be determined this particular background is not ideal for the assignment of the mode of regulation. This is because the deletants of pNDR22 are based upon the pUC vector which includes the strong *lac*-UV5 promoter. The complication that is introduced is that expression of cloned genes might be directed from the non-repressed *lac* promoter rather than the endogenous *mhb* promoter. In this case the true pattern of expression observed might be masked by the effects of the external promoter. Therefore the more proper approach is to consider the expression levels in a background where the influence of the external promoter is reduced. This may be achieved simply by the introduction of the plasmids into a strain where the activity of the *lac* promoter is diminished because of the presence of the $LacI^q$ product, a more potent repressor of *lac* gene expression than the normal LacI product. For this reason *E.coli* NM522 was transformed with pNDR22 and with pNDR30-32 and the activities of the 2,5-DHB dioxygenase (MhbD) measured in cells grown in the presence and absence of 5mM

Figure 4.3

Model for the organisation of the *mhb* genes in *Klebsiella pneumoniae* M5a1. For purposes of clarity only selected restriction sites are shown. The individual *mhb* genes are represented schematically and are not drawn to scale.



Key to enzymes: B=*Bam*HI; EI= *Eco*RI; H= *Hin*dIII; Sp=*Sph*I

Table 4.2

Activities of 2,5-DHB dioxygenase (MhbD) in *E.coli* strains harbouring pNDR22 and deletants prepared by exonucleaseIII treatment of this plasmid. Activities were measured in the presence and absence of an inducer (5mM) to test for the presence of the MhbR product.

	Sp.activity of MhbD (nmol/min/mg) HOST=5K		Sp.activity of MhbD (nmol/min/mg) HOST=NM522	
PLASMID	+2-HB	-2-HB	+2-HB	-2-HB
pNDR22	2030	290	3310	360
pNDR30	5200	7200	800	490
pNDR31	8300	11700	540	465
pNDR32	17900	15900	690	555
pNDR33	0	0	0	0
pNDR34	0	0	0	0
pNDR932			345	380

2-hydroxybenzoate. The results of this series of experiments are shown in Table 4.2.

It seems from these results that the expression of the *mhbD* gene is, as expected, inducible with pNDR22 but non-inducible with each of the deletants. However, it is important to note that the levels of expression in each of the deletants is much lower than that resulting when the plasmids were present in *E.coli* 5K. The activities are in fact comparable to those representing the basal, non-induced, levels of expression of pNDR22. The pattern of expression of the deletants in NM522 appears to be best described as NON-INDUCIBLE/LOW. This is characteristic of systems which are positively regulated, that is where the basal level of expression is increased by an activator in the presence of an inducer. If the activator is not produced then expression remains at the non-induced level. It thus appeared that the *mhb* genes were positively regulated and hence that the MhbR gene product was an activator rather than a repressor.

A matter of concern at this point was that the properties of the deletants of pNDR22 in *E.coli* 5K and in NM522, that is high level expression in 5K and low level expression in NM522, might result simply from differential activity of the *lac* promoter in the two hosts. It was possible to envisage a case where deletion of the endogenous promoter sequence would leave expression solely dependent on the activity of the *lac* promoter. Thus when the activity of the *lac* promoter was high, for example when 5K was the host, expression would be high with the converse being true for the other host. Therefore it was felt necessary to obtain evidence suggesting that the natural promoter remained on the deletants and was responsible for the expression seen. In order to do this the 1.7 Kbp insert fragment from pNDR32 was prepared by digestion with *Eco*RI and *Hin*dIII followed by elution from an LMP gel slice. This fragment was then ligated with appropriately digested pUC19 to give rise to a plasmid designated pNDR932 which contains the insert fragment orientated in the opposite direction to pNDR32 with respect to the *lac* promoter. If the expression of pNDR32 was simply the consequence of *lac* promoter activity then the expectation would be that pNDR932 would not show detectable 2,5-DHB dioxygenase (MhbD) activity because the promoter would be in the wrong orientation to drive expression. If however a *mhb* promoter was present on the 1.7 Kbp *Eco*RI/ *Hin*dIII fragment

then MhbD activity should be detectable in cells harbouring pNDR932. As a consequence of the fact that in pNDR932 transcription from the *lac* promoter would be counter to that of the *mhbD* gene which could result in a pronounced reduction in, or even elimination of, *mhbD* gene expression it was decided to use *E.coli* NM522 as the host because *lac* promoter activity in this host would be reduced. Enzyme assays of cell-free extracts of *E.coli* NM522 (pNDR932) recorded in Table 4.2 indicated the presence of the MhbD product which suggested that a *mhb* promoter had been retained. It seemed that the low activities seen in NM522 containing pNDR32 were caused by the low activity of the endogenous promoter and hence the regulation of the *mhb* pathway appeared to be under positive control.

Summary

In this Chapter the complete order of the *mhb* genes was determined. The relative positions of the *mhbD* and *mhbH* genes were demonstrated by straightforward sub-cloning experiments and which produced pNDR21 and 22. The location of the *mhbD* gene was refined by the construction of a deletion series from pNDR22 and the results of analysis of the deletants used to infer the probable direction of transcription. This aspect is considered further in the following chapter. It was not possible to assay for the presence of the *mhbR* gene on these sub-clones using 2,5-DHB because of problems resulting from the partial metabolism of 2,5-DHB, an inducer of the pathway, which caused the accumulation of a toxic intermediate maleylpyruvate. In order to allow assays for the presence of the *mhbR* gene a small-scale survey of prospective inducer analogues was undertaken. This identified two compounds, 2-hydroxybenzoate and 5-methoxy-2-hydroxybenzoate neither of which are growth substrates for *K.pneumoniae*, which were able to induce expression of the *mhb* genes and which did not give rise to toxicity problems with sub-clones. When the effects of the inducer analogues on the expression from pNDR21 and pNDR22 were tested it was apparent that induction occurred and thus that the *mhbR* gene was present. This localised the *mhbR* gene to a region comprising 4.0 Kbp of cloned DNA but its position relative to the *mhbD* genes was not known.

The *mhbR* gene was further localised by the generation of a unidirectional deletion series from pNDR22. This analysis revealed

that this gene was located to the side of the main *mhb* gene block closest to the *Sph*I site at map position 13.0 and this therefore completed the ordering of the genes.
Experiments in which the expression of deletants of pNDR21 were measured in a background in which the influence of the external, vector-based, *lac* promoter was minimised indicated a pattern of activity best described as non-inducible/low. It was therefore inferred that the *mhb* genes were positively regulated and thus that the MhbR produced was an activator.

CHAPTER 5

Purification of 2,5-dihydroxybenzoate-1,2-dioxygenase and nucleotide sequencing of its gene (*mhbD*)

Purification of 2,5-dihydroxybenzoate-1,2-dioxygenase and nucleotide sequencing of its gene (*mhbD*)

The work described in this Chapter focuses on the purification of 2,5-DHB dioxygenase (MhbD) which permitted the determination of its first 19 N-terminal residues. This information was used in the analysis of nucleotide sequence obtained from sub-clones of the *mhb* genes to locate precisely the *mhbD* gene. Once the reading frame had been identified a sequencing strategy was devised which allowed the complete sequence of the *mhbD* gene to be determined.
The purification of 2,5-dihydroxybenzoate dioxygenase has been reported on two separate occasions. The enzyme was first purified (Crawford *et al*, 1975) from an organism originally classified as *Moraxella osloensis*, although the isolate was subsequently reclassified as a non-motile species of *Bacillus*. The second report involved the purification of the enzyme from *Pseudomonas testosteroni* and *P.acidovorans*, the two divergent species of the *acidovorans* group of *Pseudomonas* (Harpel and Lipscomb,1990a).

Purification of 2,5-DHB dioxygenase

In Chapter 4 experiments were described which led to the determination of the order of the *mhb* genes. A part of this analysis had involved the generation of derivatives of the plasmid pNDR22 using a unidirectional deletion procedure. It was observed that deletions of 2.5 Kbp or more resulted in total loss of expression of the *mhbD* gene (encoding 2,5-DHB-1,2-dioxygenase). However, when introduced into *E.coli* 5K, a number of plasmids which had less extensive deletions (pNDR30-32) were observed to contain greatly elevated levels of 2,5-DHB dioxygenase (MhbD) activity, indicating overexpression of the *mhbD* gene. The highest levels of activity were associated with the plasmid pNDR32 which retains some 1.7 Kbp of the cloned DNA and preliminary experiments suggested that the 2,5-DHB dioxygenase, which was estimated to have a sub-unit molecular mass of 40,500, represented up to 15% of soluble cellular protein. It was expected that this degree of overexpression would facilitate the purification of the protein such that its N-terminal amino acid sequence could be determined.

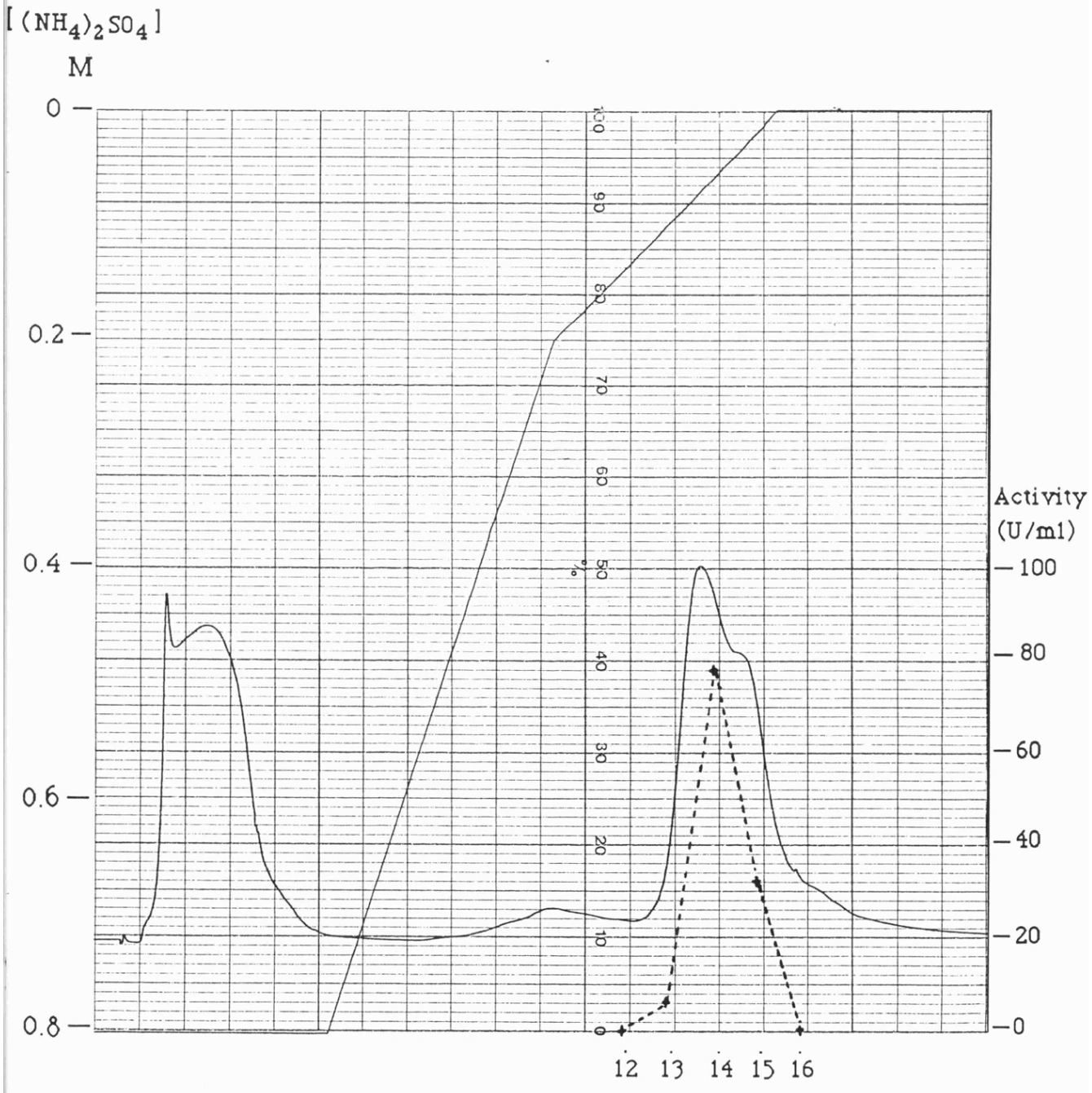
Cells of *E.coli* 5K harbouring the plasmid pNDR32 from an overnight culture grown on Luria broth-ampicillin were sub-cultured into glycerol minimal medium containing ampicillin at a concentration of 25μg/ml to a starting OD_{680} of 0.05 IPTG was added to a concentration of 0.3mM to ensure full induction of the *lac* promoter on the plasmid. The cells were grown at 30°C with vigorous shaking until they reached the mid-exponential phase of growth. The cells were then harvested by centrifugation at 8,000g with the cell pellet being resuspended in 0.02 volumes of 50 mM sodium phosphate buffer pH 7.5. A cell-free extract was prepared by sonication for 45 seconds at a peak amplitude of 7μm followed by centrifugation at 8,000g for 15 minutes at 4°C to remove cellular debris. A small quantity of the extract was removed and ferrous sulphate added to 1mM to activate the 2,5-DHB dioxygenase prior to assay. The reason for the withdrawal of a small sample of material for activation rather than direct treatment of all of the material was that preliminary experiments suggested that precipitates were formed with time when ferrous sulphate was added which interfered with further purification. This activation scheme was applied for all enzyme assays throughout the purification protocol which was carried out without added $FeSO_4$.

The first stage of the purification procedure involved fractionation of the crude extract with ammonium sulphate. Preliminary experiments had showed that at an ammonium sulphate concentration of 20% very little of the enzyme was precipitated but that increasing the ammonium sulphate concentration to 40% resulted in the precipitation of the vast majority of the enzyme activity. Therefore it seemed appropriate to prepare a 20-40% ammonium sulphate precipitate as part of the purification. Similar precipitation regimes were used for the purification of the 2,5-DHB dioxygenases from *P.testosteroni* (25-40%) and from *P.acidovorans* (33-40%) (Harpel and Lipscomb, 1990a) although the enzyme from *M.osloensis* was not precipitated until 70% saturation was achieved. The fraction representing the material precipitating at 20-40% saturation was prepared in the following way: finely ground ammonium sulphate was slowly added to crude extract in the ratio of 0.106g of solid per ml of extract, maintained at 4°C with gentle stirring for 15 minutes and the precipitated material representing the 0-20% fraction separated via centrifugation of the mixture for 20 minutes at 10,000g and 4°C. The ammonium sulphate

concentration of the supernatant was then increased to 40% by the careful addition of a further 0.113g of solid per ml of extract. The precipitate was collected by a further centrifugation step and the resulting pellet redissolved in 50mM sodium phosphate buffer pH 7.5. The redissolved 20-40% $(NH_4)_2SO_4$ fraction was then mixed with a solution of 0.8 M $(NH_4)_2SO_4$ in 50mM sodium phosphate buffer pH7.5 in the ratio 5 : 6 (fraction : buffer) and was then applied to a Phenylsuperose 5/5 hydrophobic-interaction FPLC column. The enzyme was eluted using a non-linear decreasing gradient of 0.8 -> 0.0 M $(NH_4)_2SO_4$ at a flow rate of 0.5 ml/min. The concentration was decreased from 0.8-> 0.2M $(NH_4)_2SO_4$ over 5ml with the reduction from 0.2->0.0 M occurring over a further 5ml. The elution profile from this column is shown in Figure 5.1. Enzyme assays of the column fractions revealed that the 2,5-DHB dioxygenase eluted at approximately 0.06 M $(NH_4)_2SO_4$. The activities associated with the fractions across the 2,5-DHB dioxygenase peak are recorded in Table 5.1. Samples from across this peak from each of two phenylsuperose column runs were then separated on SDS-polyacrylamide gels. This allowed the extent of purification to be monitored visually and the suitability of individual fractions for amino terminal sequencing to be assessed. This gel indicated that the peak fractions appeared to contain a protein band, occurring in the expected position, which accounted for approximately 90% of the total protein. Whilst this type of sample would normally be regarded as suitable for sequence analysis following transfer onto a PVDF membrane it was decided to include a further column step. This was because when a previous sample of apparently similar quality had been blotted two bands had been apparent on the membrane. As it was not known why an apparent single protein band had appeared to resolve into two bands in a subsequent electrophoretic separation further fractionation seemed appropriate. Therefore, material from the two peak fractions was pooled and diluted with 4 volumes of 0.5M Tris-Cl buffer pH 7.5 to reduce the concentration of $(NH_4)_2SO_4$before being applied to a Mono-Q 5/5 anion exchange column. The material was eluted with a linear gradient of 0-> 1.0M NaCl (in Tris-Cl buffer) over 10ml. The elution profile is shown in Figure 5.2. It was apparent that the major peak eluted form this column at approximately 0.27M NaCl although there appeared to be a very minor peak trailing at around 0.3M NaCl.

Figure 5.1

Elution profile from Phenylsuperose 5/5 column using a decreasing non-linear gradient in which the concentration of ammonium sulphate in 50mM sodium phosphate buffer pH 7.5 was first reduced from 0.8M ->0.2M over 5ml and then from 0.2-> 0.0 M over a further 5ml The elution profile of the 2,5-DHB dioxygenase is superimposed on the protein trace.



Sample	UNITS APPLIED µmol/min	UNITS RECOVERED µmol/min	RECOVERY %	[PROTEIN] mg/ml	SPECIFIC ACTIVITY µmol/min/mg	PURIFICATION (FOLD)
CRUDE EXTRACT	623	---	100	4.1	24	--
20-40% $(NH_4)_2SO_4$ CUT	623	516	83	1.5	100	4.1
PHENYL-SUPEROSE FRACTIONS						
1.14	\	49.4 \		0.45	109	
1.15	251	63.3--133.3	57	0.5	127	5.4
1.16	/	20.6 /		0.3	68	
2.13	\	6.7 \		0.1	67	
2.14	246	78.1--139.6	54	0.6	130	5.4
2.15	/	31.7 /		0.4	79	
MONO-Q FRACTIONS						
8	\	12.1 \		0.3	40	1.7
	120	--18.4	15			
9	/	6.3		0.2	32	1.3

Table 5.1

Record of the purification of 2,5-DHB dioxygenase from *K.pneumoniae*. Recoveries reflect total activity recovered from a given step whilst purifications are cumulative and do not reflect each step.

Figure 5.2

Elution profile from MonoQ 5/5 column using a linear gradient of 0.0-> 1M NaCl in 50mM Tris-Cl buffer pH 7.5 over 10ml at a flow rate of 1.0 ml/min . The elution profile of the 2,5-DHB dioxygenase activity is superimposed on the protein trace.

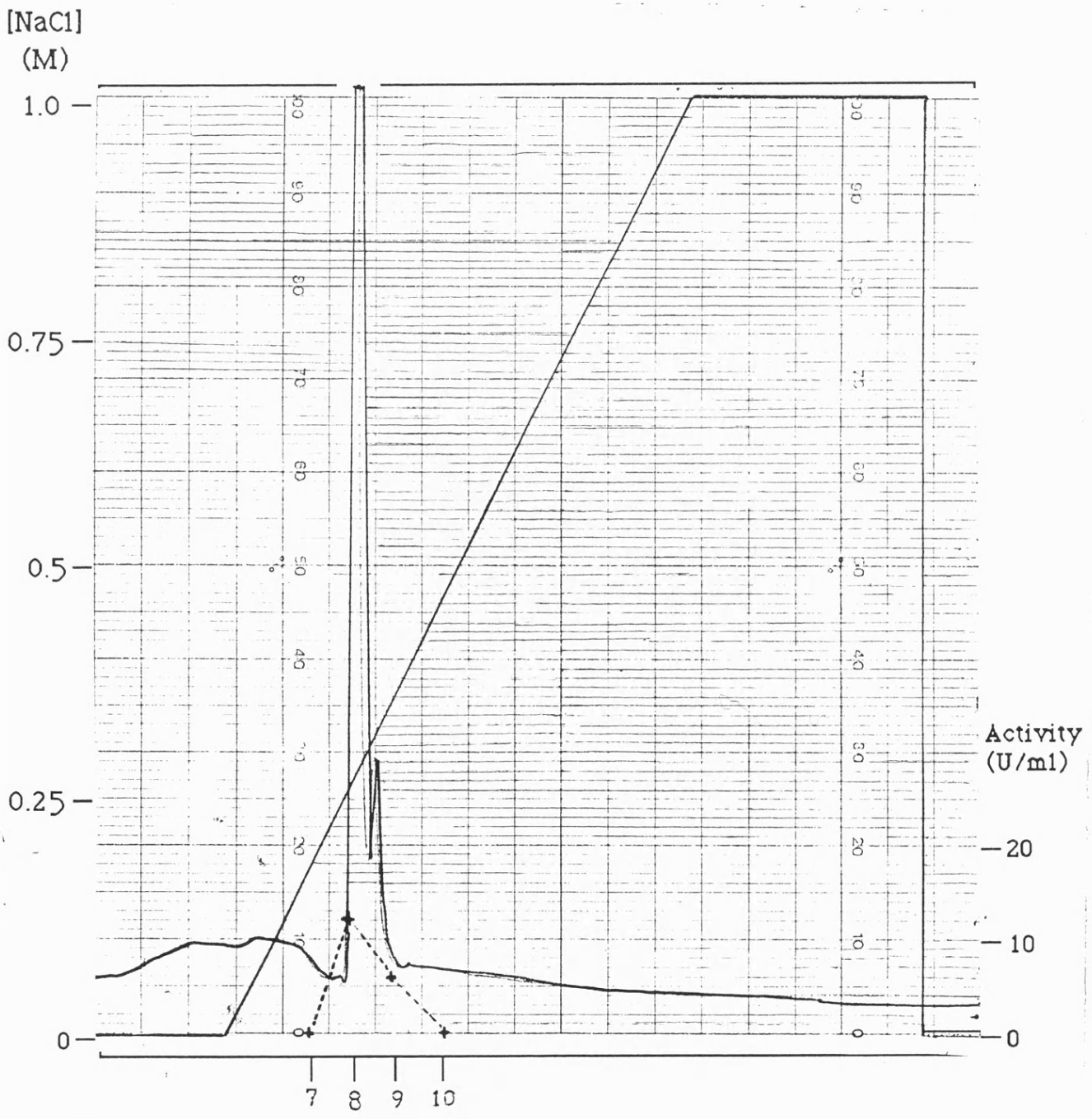
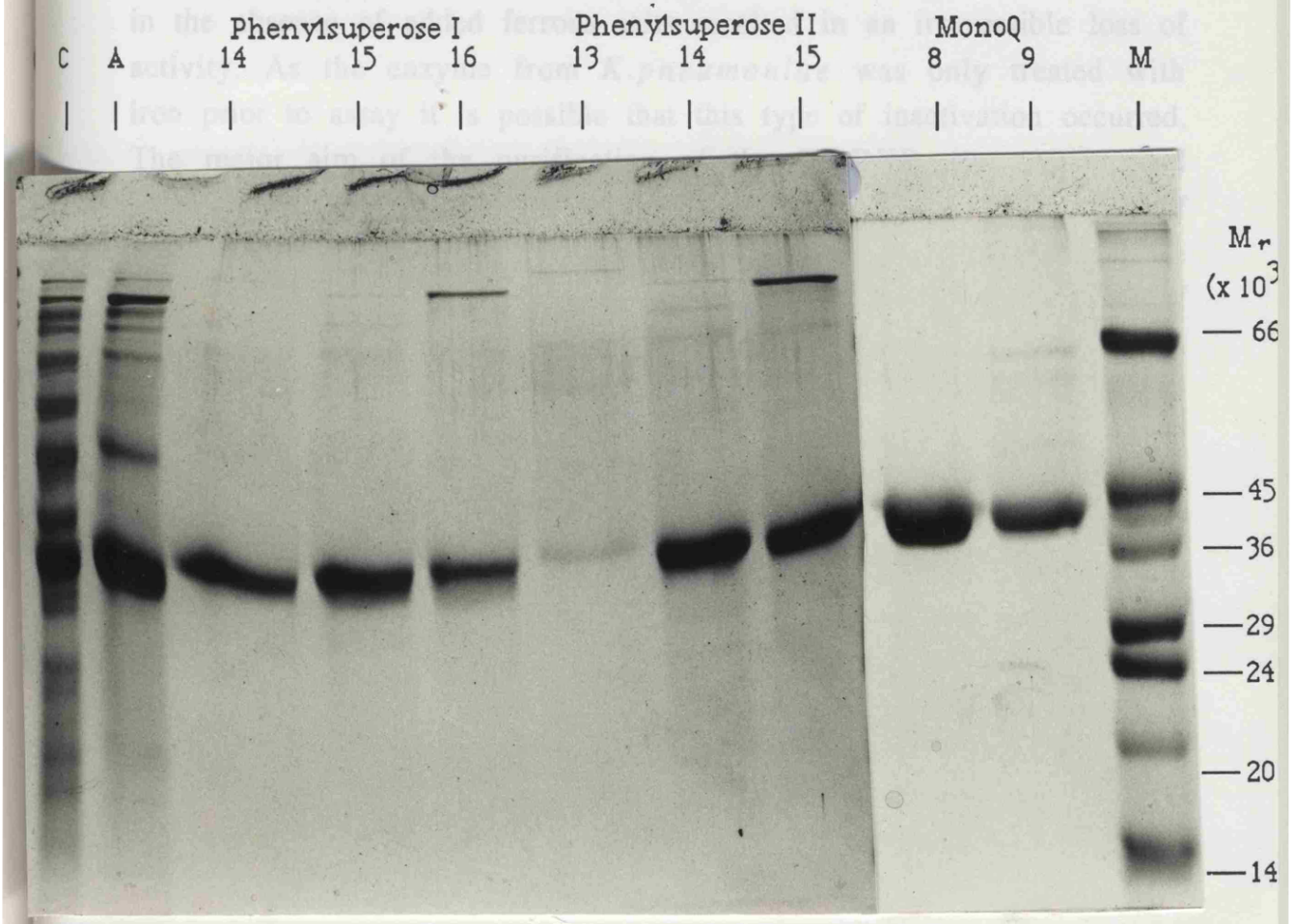


Figure 5.3

Photograph showing the progression of the purification of 2,5-DHB dioxygenase. Samples were run out on denaturing 12% polyacrylamide gels. 10 µg samples of crude extract (C) and 20-40% ammonium sulphate fractions (A) were loaded. Equivalent volumes of sample were loaded for each of the column fractions with the amount of protein present in the peak fraction also set at 10µg.



Enzyme assays of the fractions revealed that all of the 2,5-DHB dioxygenase activity appeared to be located in the fractions corresponding to the major peak. The results of the electrophoretic analysis of the MonoQ fractions along with that of the other samples marking the progression of the purification are shown in Figure 5.3.
It was apparent however that the employment of this column approach had resulted in a very significant loss of activity (Table 5.1). The precise explanation for the loss of activity on this column is not clear. It has been reported (Harpel and Lipscomb, 1990a) that maintenance of the equivalent protein from two *Pseudomonas* strains in the absence of added ferrous salts resulted in an irreversible loss of activity. As the enzyme from *K.pneumoniae* was only treated with iron prior to assay it is possible that this type of inactivation occurred. The major aim of the purification of the 2,5-DHB dioxygenase had been to obtain N-terminal amino acid sequence. Despite the loss of activity it was still possible to identify the fraction containing the desired protein. Therefore a sample containing approximately 40μg of protein was then supplied to the protein sequencing facility at the University of Leicester. The sequence analysis was conducted by Dr.K.Lilley and Miss.E.Cavanagh as described in Chapter 2. Unambiguous sequence was obtained for the first 19 residues and the sequence obtained is shown in Figure 5.4. This sequence was quite different from that found at the N-terminus of the α-fragment of β-galactosidase. This eliminated the possibility that the protein purified was a fusion protein formed from the α-peptide and the 2,5-DHB dioxygenase.

Figure 5.4

N-terminal amino acid sequence of the purified 2,5-DHB dioxygenase from *K.pneumoniae*

1.	Ser
2.	Gln
3.	Ser
4.	Thr
5.	Thr
6.	Glu
7.	Ala
8.	Asn
9.	Asn
10.	Gly
11.	Arg
12.	Gln
13.	Gln
14.	Phe
15.	Tyr
16.	Gln
17.	His
18.	Ile
19.	Ser

Nucleotide sequencing of the *mhbD* gene

There is a general requirement in any sequencing project for an ability to be able to recognise the sequence corresponding to the gene of interest within a particular segment of DNA. In this instance identification of the *mhbD* gene would be possible because N-terminal amino acid sequence had been obtained from the purified MhbD gene product. The occurrence of the stretch of nucleotides corresponding to this protein sequence would then define the open reading frame.
It had been shown that the *mhbD* gene was present intact on pNDR32 which includes some 1.7 Kbp of cloned DNA but that further deletions removing as little as 200 bp, for example in pNDR33, led to a loss of expression of this gene. As it was believed that the *mhbD* gene was transcribed towards the *Hin*dIII site it seemed probable that the inactivation of the *mhbD* gene in pNDR33 resulted from deletion of DNA representing the N-terminal portion of the protein. This therefore suggested that if sequencing was carried out from the deleted end of pNDR32 towards the *Hin*dIII site the sequence corresponding to the amino-terminus would be located within the first 200 bp.
Plasmid sequencing of pNDR32 using Reverse Primer (RP) did indeed identify a segment of DNA whose predicted amino acid sequence matched perfectly that of the amino-terminus of the purified protein. The start of the gene was thus mapped to a position around map co-ordinate 10.5. (Figure 4.2) With the start-point of the *mhbD* coding sequence defined and the knowledge that 1.1-1.2 Kbp of DNA would be required to encode the *mhbD* gene it was clear that sequencing should focus on the region running from map position 10.5-> 9.3.
The sequencing of the amino-terminal region was extended through analysis of pNDR33, which was predicted to have lost some DNA from the amino-terminus of the protein. In fact, it transpired that the amount of coding sequence deleted in this construct was remarkably low with only two residues being lost. Expression of the protein may be prevented by the absence of the necessary translational signals or through a frame shift accompanying fusion with the α-peptide sequence. Through the sequencing of this plasmid it was possible to obtain sequence information running from the third codon through

the *Pvu*II site at map-position 10.2, this is some 280 nucleotides into the sequence.

In order to extend the sequence further a plasmid, designated pNDR40 was prepared by ligation of the central 500 bp *Pvu*II fragment from pNDR32 into pUC19. Sequence contiguous with that obtained from pNDR33 was obtained by sequencing pNDR40 with the Universal primer whilst sequence from the opposite strand, extending for approximately 275 bp from map position 9.7, was obtained using the pUC reverse primer. This meant that the entire 500 bp *Pvu*II fragment had been sequenced on one or other strand with a small central region which had been sequenced on both strands.

Calculations based upon the known start point of the *mhbD* gene and its expected size determined that the plasmid pNDR34 would be expected to include approximately 300bp of the carboxyl terminus of the gene. Analysis of the sequence obtained from this construct revealed an area showing complementarity to the sequence of pNDR40 obtained using the Reverse Primer. The regions of overlap identified meant that it was possible to add the sequence data obtained from pNDR34 to that previously obtained to allow coverage from the start point of the coding sequence right through to a region close to the expected end of the gene. The sequence obtained thus far related to one or other strand of the DNA with little data corroborated by sequencing of the opposite strand. Therefore the next phase of the sequencing was to obtain full double stranded sequence.

The efficiency of plasmid sequencing had proved somewhat variable so it was decided that future constructs should be based on bacteriophage M13 vectors. The intention was that constructs which would allow the sequencing of each strand would be produced and that gaps in the sequence would be filled using specific primers based upon the existing nucleotide data. A construct designated mp933 was prepared by ligation of the 1.5 Kbp *Eco*RI/ *Hin*dIII fragment from pNDR33 into M13 mp19 and a construct designated mp832 prepared by ligation of the 1.7 Kbp *Eco*RI/ *Hin*dIII fragment from pNDR32 into M13 mp18.

Nucleotide sequencing of mp933 using the primer NR02 allowed coding strand sequence to be obtained which extended just beyond the *Pvu*II site which lay some 780 nucleotides into the coding sequence at around map position 10.7. This was thus the sequence

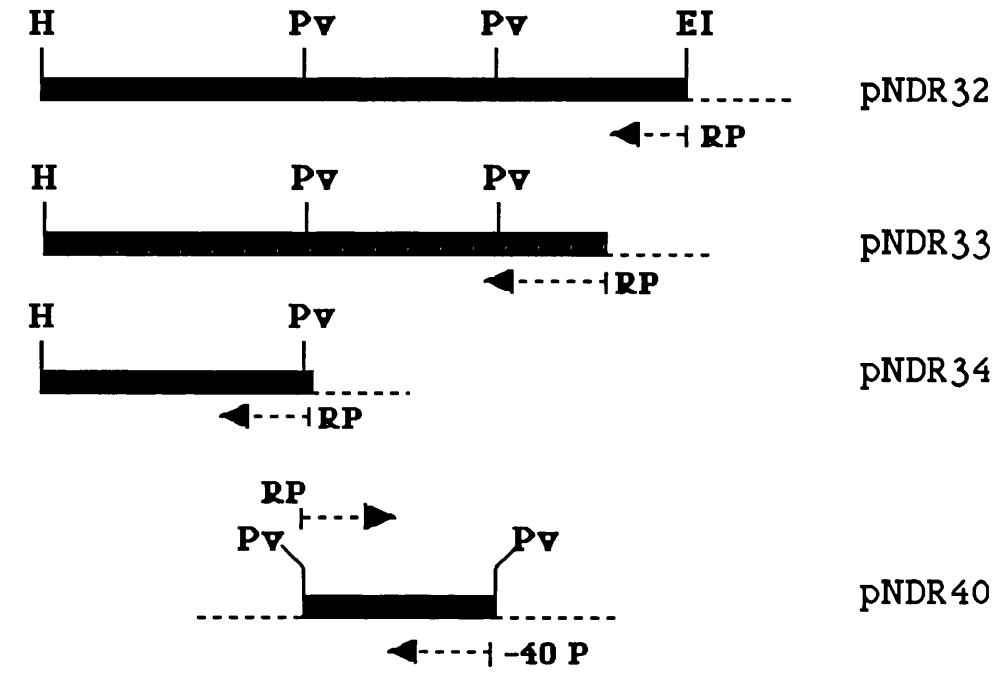
representing the complement of the sequence obtained from pNDR40 with the reverse primer. The coding strand sequence in the region encoding, and extending beyond, the carboxyl terminus of the MhbD gene product was obtained through the use of the primer designated NR04.
Nucleotide sequencing of the non-coding strand was performed using mp832 as the template. The use of the primer NR03 allowed the sequence in the central portion of the gene, specifically in the region from around map position 9.9 to 10.25, to be determined. This sequence was thus complementary to that obtained by sequencing pNDR40 with the Universal primer. The primer NR05 whose sequence had been derived from data extending beyond the likely end of the coding sequence was then used to sequence back across the carboxyl terminus to provide information complementary to that from the sequencing of mp933 with NR04. Finally the non-coding strand in the region representing the amino terminus was sequenced using the primer NR06, this provided information which was used to confirm the sequence obtained from pNDR32.
The sequencing strategy is summarised in Figure 5.5.

The sequence of the *mhbD* gene

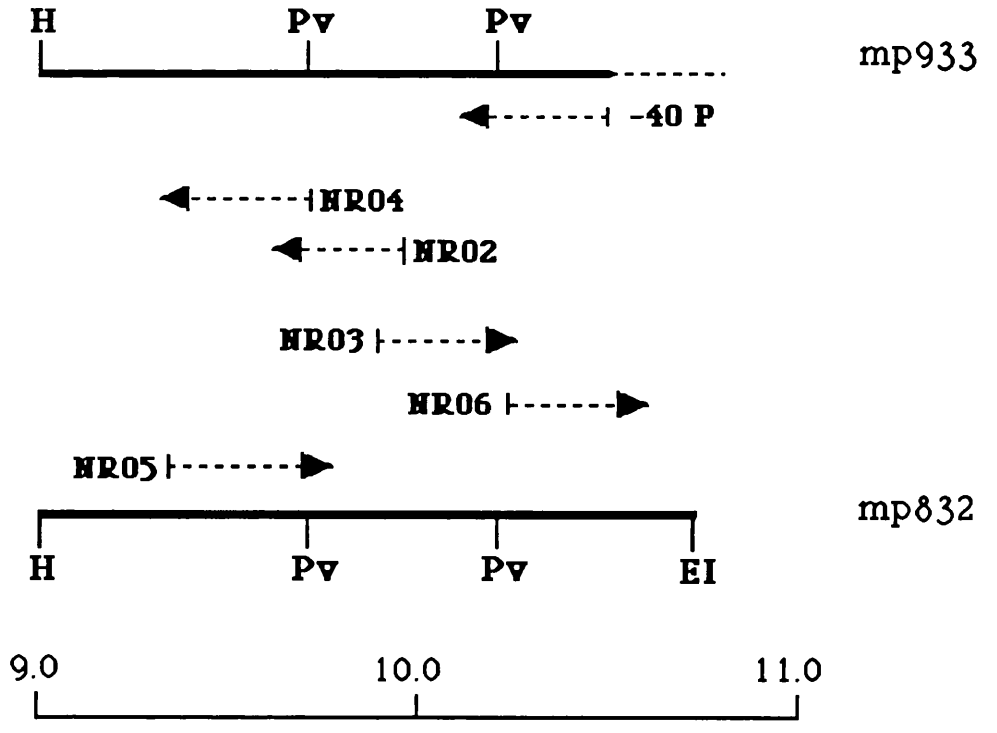
The complete nucleotide sequence of the *mhbD* gene is represented in Figure 5.6. The reading frame of the *mhbD* gene appeared to comprise some 1029 nucleotides including the initiating ATG codon. This sequence encodes 342 amino acids (the initiating methionine is absent from the mature protein) and the MhbD protein has a predicted relative molecular mass for the sub-unit of 38,651. This predicted molecular mass is in reasonable accord with the 40,500 estimated for the purified protein. The predicted amino acid composition of the sub-unit is shown in Table 5.2. A sequence showing similarity to the consensus Ribosome Binding Site (RBS) (Shine and Dalgarno, 1975) was found 7 nucleotides upstream of the initiator methionine and this is the proposed translational initiation site. The reading frame is terminated by a TAA (Ochre) codon which is closely followed by a second in frame termination codon. It is interesting to note that just downstream of this is a candidate RBS sequence which precedes an ATG sequence by approximately 7 nucleotides. Given the gene order

Figure 5.5 Strategy for sequencing of the *mhbD* gene. Arrows are indicative of the amount of sequence obtained with the stated oligonucleotide.

Plasmid constructs



M13 constructs



```
GAATAAATAAGGAGAACGCTTATGTCCCAGTCCACCACGGAAGCAAATAACGGCCGTCAG
        -RBS-         MetSerGlnSerThrThrGluAlaAsnAsnGlyArgGln      13
CAGTTTTACCAGCATATTTCCGGGCAAAACCTGACGCCGCTTTGGGAATCGCTGCACCAT  99
GlnPheTyrGlnHisIleSerGlyGlnAsnLeuThrProLeuTrpGluSerLeuHisHis      33
CTGGTGCCGAAAACGCCGAACGCCACCTGCGCGCCGGCTTACTGGAATTATCAGGAGATT 159
LeuValProLysThrProAsnAlaThrCysAlaProAlaTyrTrpAsnTyrGlnGluIle      53
CGCCCGCTGCTGGAAAGCGGCAAGCTGATTGGCGCGAAAGAGGCGATACGCGCGTGCTGG 219
ArgProLeuLeuGluSerGlyLysLeuIleGlyAlaLysGluAlaIleArgAlaCysTrp      73
TGCTGGAAAACCCGGCGCTGCGCGGCAGTCGTCGATTACCTCCTCGCTGTACGCGGGTTT 279
CysTrpLysThrArgArgCysAlaAlaValValAspTyrLeuLeuAlaValArgGlyPhe      93
GCAGCTGATTATGCCCGGCGAAGTGGCGACCGAGCCATCGGCACAACCAGTCGGCGCTCG 339
AlaAlaAspTyrAlaArgArgSerGlyAspArgAlaIleGlyThrThrSerArgArgSer     113
CGCTTTGTCGTGGAAGGCGAAGGGGCGTTTACCGCCGTTGACGGCGAACGGACCGCCATG 399
ArgPheValValGluGlyGluGlyAlaPheThrAlaValAspGlyGluArgThrAlaMet     133
CGCGCCGGCGACTTTATTCTGACGCCGCAGTGGCGCTGGCACGATCACGGCAACCCCGGC 459
ArgAlaGlyAspPheIleLeuThrProGlnTrpArgTrpHisAspHisGlyAsnProGly     153
AACGAACCGGTTATCTGGCTCGACGGCCTCGATCTGCCGCTGGTGAACTATCTGGGCTGC 519
AsnGluProValIleTrpLeuAspGlyLeuAspLeuProLeuValAsnTyrLeuGlyCys     173
GGCTTCGCGGAGGACTATCCGCAAGATCAGCAGCCGGTGACCCGCAAAGAGGGGGATTAT 579
GlyPheAlaGluAspTyrProGlnAspGlnGlnProValThrArgLysGluGlyAspTyr     193
CTGCCGCGCTACGCCGCCAACATGCTGCCGCTGCGCCATCAGTCCGGTAACTCTTCGCCC 639
LeuProArgTyrAlaAlaAsnMetLeuProLeuArgHisGlnSerGlyAsnSerSerPro     213
ATCTTTAACTATCGCTACGACCGCAGCCGCGAAGCGCTGCACGATCTGACGCGGATGGGC 699
 IlePheAsnTyrArgTyrAspArgSerArgGluAlaLeuHisAspLeuThrArgMetGly    233
GATGCCGACGAGTGGGATGGCTATAAGATGCGCTACGTCAACCCGGTCACCGGCGGCTAC 759
AspAlaAspGluTrpAspGlyTyrLysMetArgTyrValAsnProValThrGlyGlyTyr     253
CCGATGCCGTCGATGGGCGCCTTCCTGCAGCTGTTACCGAAAGGTTTACCTCGCGAGCGG 819
ProMetProSerMetGlyAlaPheLeuGlnLeuLeuProLysGlyLeuProArgGluArg     273
CGAAAACCACCGACAGTACCGTCTATCACGTGGTGGAAGGCAGCGGCCAGGTCACCATTG 879
ArgLysProProThrValProSerIleThrTrpTrpLysAlaAlaAlaArgSerProLeu     293
GCGAACAGACCTTCGCTTTTCAGGCAAAAGATATCTTCGTGGTGCCGACCTGGCACGCCG 939
AlaAsnArgProSerLeuPheArgGlnLysIleSerSerTrpCysArgProGlyThrPro     313
TCTCTTTTATCTCCGCCGAAGATAGCGTGTTATTCAGCTTTTCGGACCGTCCCGTGCAGG 999
SerLeuLeuSerProProCysIleAlaCysTyrSerAlaPheArgThrValProCysArg     333
AAGCGCTCGGCCTGTTCCGCGAAGCGCGTTATTAAAAATTAAGAATAAGGAATCGAATCA 1059
LysArgSerAlaCysThrAlaLysArgValIleLysAsnEND    -RBS-       M
```

```
TGACTCAATACGTTTTTGCACCCCAGGCTCCGATTAGCGTTCCGGTTGTCGGCAGCGATG 1149
etThrGlnTyrValPheAlaProGlnAlaProIleSerValProValValGlySerAsp
```

Figure 5.6

Complete nucleotide sequence of the *mhbD* gene along with the predicted amino acid sequence of the protein. The figures in normal type refer to the nucleotide position starting from the ATG codon. Figures in bold number the amino acid residues including the initiating methionine although this is not present in the mature protein. RBS denotes sequences showing similarity to the consensus ribosome binding site. Also shown is the translation of the downstream sequence which may represent part of the *mhbH* gene.

Table 5.2

Determined amino acid composition of 2,5-DHB dioxygenase from *K.pneumoniae* M5a1 deduced from nucleotide sequence of its gene.

AMINO ACID	RESIDUES PER SUB-UNIT
ALA	32
ARG	34
ASN	13
ASP	16
CYS	9
GLN	13
GLU	14
GLY	25
HIS	7
ILE	11
LEU	28
LYS	13
MET	6
PHE	10
PRO	29
SER	23
THR	18
TRP	11
TYR	15
VAL	15
CALCULATED R.M.M	**38,651**

previously postulated it is possible that this could indicate the start of the *mhbH* gene although it must be emphasised that the confirmation of this hypothesis would require determination of the amino-terminal sequence of the MhbH protein.
With nucleotide sequence information available for the *mhbD* gene it was possible to begin searches which would establish whether there was similarity between this sequence and those of other dioxygenases. Whole database searches using the FASTA software available on the University of Leicester VAX of both nucleotide and protein sequence repositories (EMBL and NBRF) respectively failed to identify any strong similarities between *mhbD*/MhbD and any other dioxygenases or other aromatic catabolic enzymes. Further comparisons between a range of individual dioxygenases and the *mhbD*/MhbD sequences using the BESTFIT algorithm failed to find any evidence of any locally strong similarity between sequences.

Summary

In this chapter the purification of 2,5-DHB dioxygenase has been described. This was achieved via a three step procedure featuring an ammonium sulphate fractionation followed by two FPLC column steps. The second column was included as a final clean-up step but unfortunately the material recovered from this column had lost a great deal of activity. In the context of any investigation of the protein chemistry this would have been of great concern. However, the main reason for the purification of the protein had always been to allow the N-terminal amino acid sequence to be determined so as to facilitate the nucleotide sequencing of the *mhbD* gene. This was reflected in the fact that the purification was performed on a small scale commensurate with the requirement for the ultimate production of the sub-mg quantities required for N-terminal sequencing. What has become clear from this purification is that it would not be suitable for any investigation in which the enzymology of the dioxygenase were to be studied. During the purification of the corresponding enzymes from two *Pseudomonas sp* (Harpel and Lipscomb, 1990a) it was observed that the maintenance of protein fractions in the absence of ferrous salts led to a slow but irreversible inactivation of 2,5-DHB dioxygenase. This was not possible using the activation methodologies

employed in this study as precipitates were formed when samples were mixed with ferrous sulphate. This was incompatible with the use of FPLC columns. However, it is possible that the use of alternative iron salts such as the ferrous ammonium sulphate used by Harpel and Lipscomb (1990a) might not give rise to precipitates. In this case it might be possible to stabilise the enzyme activity and obtain a more active final product. Given the degree of overexpression of the 2,5-DHB dioxygenase by plasmids described in this chapter then it should be possible to scale up the purification procedure to prepare ample material for the charaterisation of the enzyme from *K.pneumoniae*.

The nucleotide sequencing of the *mhbD* gene from *K.pneumoniae* described in this chapter represents the first analysis of any dioxygenase gene of this type. This is because no other cloning of gentisate pathway genes has been recorded. As such there are no immediate comparisons that can be made with equivalent systems from other sources. The N-terminal sequences of the 2,5-DHB dioxygenases from the two *Pseudomonas* strains are known so it is likely that these genes will be cloned at some point in the future. The N-terminal sequences of the two *Pseudomonas* enzymes show no sequence similarity to each other nor apparently to the sequence from *K.pneumoniae*. This is shown in Figure 5.7. Initial comparisons with other dioxygenase sequences failed to find any strong similarities. The precise implications of this are not absolutely clear. Given the speculation by Harayama *et al* (1989) that the extra- and intra-diol dioxygenases fall into discrete families it is perhaps to be expected that 2,5-DHB dioxygenase should not be particularly like either. However before hypothesising that the distal extradiol dioxygenases may be part of a distinct sub-family one should perhaps remember that the model of Harayama *et al* has not proved universally appropriate (Noda *et al*, 1989; Roper and Cooper, 1990).

Figure 5.7

Amino-terminal amino acid sequences of three purified 2,5-DHB dioxygenases from a) *K.pneumoniae*
b) *P.testosteroni*
c) *P.acidovorans*

a)

SerGlnSerThrThrGluAlaAsnAsnGlyArgGlnPheTyrGlnHisIleSer

b)

SerLeuValGlnAlaGlnProSerProValLysLeuThrAlaArg

c)

MetGlnGluLeuGlyArgLeuGluAspLeuProGlnAspGluLeuThrArgAsnAsnLeuValProLeu

CHAPTER 6

Analysis of cloned *hpc* genes from *Klebsiella pneumoniae*

KEY TO GENES

hpcB= HPC dioxygenase

hpcC= CHMS dehydrogenase

hpcD= CHM isomerase

hpcE= COHED decarboxylase

hpcF= HHDD isomerase

hpcG= OHED hydratase

hpcH= HHED aldolase

Analysis of cloned *hpc* genes from *Klebsiella pneumoniae*

The organisation of the genes involved in the catabolism of homoprotocatechuate (HPC) by *Escherichia coli* C has been determined through the analysis of genomic clones containing *hpc* genes (Jenkins and Cooper, 1988; Fawcett, 1989; Roper, 1990). Genomic clones which include the corresponding genes from the organism *Klebsiella pneumoniae* M5a1 have become available. This chapter describes the analysis which allowed the order of the genes from *K.pneumoniae* to be determined permitting comparison with the organisation deduced for *E.coli* C.

Two discrete types of genomic clone including genes involved in the catabolism of HPC by *K.pneumoniae* were available for analysis, the pTF100/102 plasmids of Fawcett (1989) and a plasmid designated pJF1 (J.Fernandez and R.A Cooper, pers.comm). In each case the clones were identified as a result of their hybridisation to oligonucleotide probes corresponding to portions of the amino terminus of the purified CHMS dehydrogenase protein (HpcC) from *K.pneumoniae* (Fawcett *et al*, 1989). The first clones (Fawcett, 1989) were identified in a genomic library constructed by the ligation of size-selected *Hin*dIII fragments of *K.pneumoniae* into the vector pBR322. Two clones showing discrete restriction patterns although each containing a 14 Kbp insert fragment were thus identified and designated pTF100 and pTF102. These plasmids were shown to be able to suppress the *hpcC* mutation of the strain *E.coli* C221 to allow growth on HPC. Assays of cell-free extracts of *E.coli* 5K (which is naturally devoid of *hpc* genes) harbouring either pTF100 or pTF102 demonstrated the presence of each of the *hpc* genes involved in the metabolism of HPC through to pyruvate and succinic semialdehyde. It was also shown that the expression of the *hpc* genes on these plasmids was induced by growth in the presence of 4-HPA or HPC. However *E.coli* 5K (pTF100/102) was unable to grow on HPC suggesting the absence of a critical metabolic function such as HPC uptake (Fawcett, 1989). The plasmid pJF1 had been constructed by the ligation of *K.pneumoniae* DNA digested with *Bam*HI into the vector pBR328. This plasmid also proved able to suppress the *hpcC* mutation of *E.coli* C221 but assays of cell-free extracts of *E.coli* 5K harbouring pJF1 suggested the presence of only a sub-set of the activities present on pTF100/102.

In this Chapter experiments will be described which helped to elucidate the order of the *hpc* genes from *K.pneumoniae* so that a comparison could be made with the organisation previously determined for *E.coli* C.

Preliminary analysis of genomic clones

It was decided that the approach to be employed for the determination of the *hpc* gene order would involve the sub-cloning of specific restriction fragments and the analysis of the enzyme activities present in extracts of *E.coli* 5K harbouring these constructs. A basic requirement of this approach is that restriction map information should be available for the primary clones.

pTF100/102

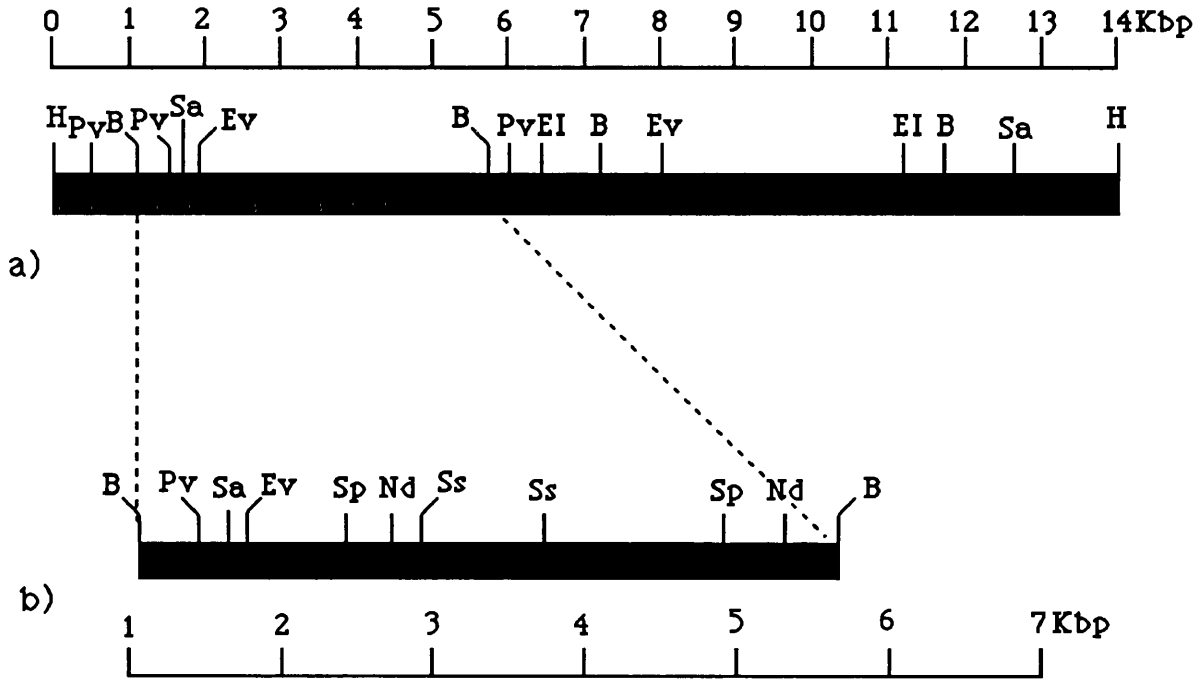
It had been shown that both pTF100 and pTF102 had *Hin*dIII inserts of approximately 14 Kbp but that there were differences in the digestion patterns observed with other restriction enzymes (Fawcett, 1989). This had raised questions about the possible relationship between the two plasmids. Fawcett considered a number of possible explanations including DNA rearrangement but suggested that the most probable explanation was that the two plasmids did indeed contain identical inserts but orientated oppositely in the vector. This suggested that the differences occurring on restriction digestion arose because of the different relative positions of vector-based sites with respect to those of the insert. In order to address this question two discrete approaches were employed. The first of these involved digestion of the plasmids with various restriction enzymes in combination with the enzyme *Hin*dIII. The rationale behind this was that *Hin*dIII digestion would effectively excise the insert from the vector and thus that digestion with the other enzyme would generate patterns representing digestion of insert alone (and vector alone). The results of this analysis suggested that concurrent digestion with *Hin*dIII did indeed lead to the elimination of the differences in the restriction patterns. This therefore seemed to confirm the reversed orientation hypothesis. The second approach was a more direct approach which involved the digestion of pTF100 with *Hin*dIII to

Figure 6.1

Restriction maps of inserts carrying *hpc* genes from *K.pneumoniae*

a) The 14Kbp *Hin*dIII fragment from pTF100/102

b) The 4.6 *Bam*HI fragment prepared from pJF1



Key to enzymes B=*Bam*HI EI=*Eco*RI Ev=*Eco*RV H=*Hin*dIII Nd=*Nde*I

Pv=*Pvu*II Sa=*Sal*I Sp=*Sph*I Ss=*Sst*I

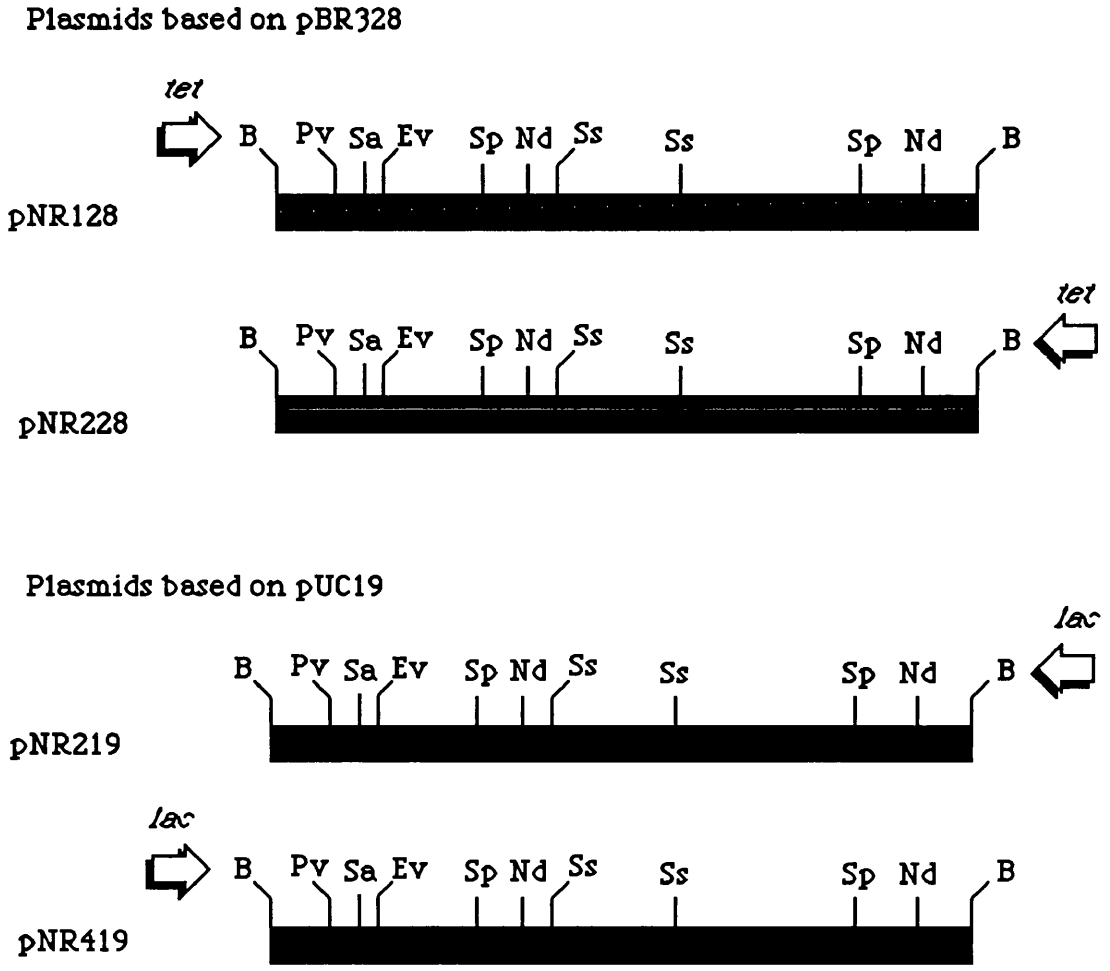
liberate the insert fragment. The digestion mixture was then religated and following transformation the resulting plasmids were subjected to restriction analysis. The expectation, assuming the reversed orientation hypothesis, was that it would be possible to obtain plasmids showing the restriction pattern of pTF102 as well as that of pTF100 in cases where religation had involved inversion of the insert. The observation of plasmids having the restriction pattern of pTF102 therefore confirmed the hypothesis. The fact that pTF100 and 102 contained the same insert in opposite orientations in the vector simplified restriction mapping because it allowed easy identification of those fragments including portions of the vector DNA. Restriction mapping involved a range of single and double digestions the results of which were combined to derive the map shown in Figure 6.1a.

pJF1

Southern blot analysis of various chromosomal DNAs digested with *Bam*HI had indicated that the hybridising fragment from *K.pneumoniae* had a size of approximately 4.6 Kbp (Fawcett *et al*, 1989). It was therefore expected that digestion of pJF1 with *Bam*HI would give rise to a fragment of this size in addition to one of 4.9 Kbp (representing the vector). However digestion of pJF1 with *Bam*HI gave rise to three fragments, those described above and an additional fragment of about 2.5 Kbp. It was not clear whether this fragment represented DNA contiguous with the 4.5 Kbp fragment or whether it was an extraneous fragment that had been incorporated adventitiously during ligation. When the digestion pattern of pJF1 was compared with that of pTF100 it became apparent that this 2.5 Kbp *Bam*HI fragment was not present on the larger plasmid. This tended to support the view that this fragment was non-contiguous but it was possible that this fragment might not be present on pTF100 because it might extend beyond the *Hin*dIII site marking the boundary of the insert of pTF100. If this were the case then digestion with *Hin*dIII and *Bam*HI would be expected to generate equivalent 1.1Kbp (representing DNA from map position 0.0 to position 1.1) fragments from each plasmid The fact that no fragment of this size was seen with pJF1 seemed to show that the 2.5 Kbp fragment was probably

Figure 6.2

Sub-cloning of the 4.6Kbp *Bam*HI fragment from pJF1 into pBR328 and pUC19 to give rise to constructs containing the fragment in both possible orientations.



Key to restriction sites: B= *Bam*HI; Ev= *Eco*RV; Nd= *Nde*I; Sa= *Sal*I; Sp= *Sph*I; Ss=*Sst*I

adventitious and that all subsequent analysis should focus on the 4.6 Kbp fragment. It was decided that the mapping of this fragment would be facilitated by sub-cloning. Therefore the 4.6 Kbp fragment was gel purified and ligated with pBR328 and also with pUC19. Ampicillin resistant transformants were screened and plasmids containing the insert in each orientation in pBR328 were isolated and designated pNR128 and pNR228. Similarly plasmids were isolated containing the 4.6 Kbp *Bam*HI fragment in each orientation in pUC19, these were designated pNR219 and pNR419. A number of restriction sites occurring within this insert were then mapped by analysis of the digestion patterns of the various constructs. A map of the insert is depicted in Figure 6.1b. The pJF1 derivatives pNR128/228 and pNR219/419 are shown in Figure 6.2.

Sub-cloning experiments to determine the *hpc* gene order

Derivatives of pTF100/102

The aim of the initial sub-cloning experiments was to localise the *hpc* gene cluster on the 14 Kbp *Hin*dIII fragment. Southern blot analysis (Fawcett, 1989) had already identified the restriction fragments which hybridised to a probe corresponding to the 5' end of the *hpcC* gene. Examination of this data in conjunction with the restriction map deduced in this study allowed the approximate position of the *hpcC* gene to be determined. The gene appeared to lie within the region bounded by the *Eco*RV site at map position 1.9 and the *Bam*HI site at map position 5.7. This therefore placed the *hpc* gene cluster towards the left-hand end of the insert as represented in Figure 6.1a. In order to confirm this it was decided to prepare a sub-clone which retained the left hand end of the parental clone. Therefore pTF100 was digested with *Eco*RI and *Hin*dIII and the 6.4 Kbp fragment isolated by extraction from an LMP agarose gel. This fragment was then ligated with appropriately digested pUC18. The screening of the transformants resulted in the isolation of a plasmid containing the desired fragment. This plasmid was designated pNR18. Enzyme assays were performed on cell-free extracts of *E.coli* 5K (pNR18) grown on Luria broth +/- 4-HPA. These indicated the presence of the activities

associated with the HPC dioxygenase, CHMS dehydrogenase, CHM isomerase, OHED hydratase and HHED aldolase (HpcB, C, D, G and H respectively but the activities of the COHED decarboxylase and HHDD isomerase (HpcE and F) were absent. The fact that expression, in contrast to the parent clone pTF100, was not regulated also suggested the absence of the HpcR product which was part of the regulatory system. This information immediately suggested that the *hpcE* and *F* genes and also the *hpcR* gene were located (or extended) beyond the *Eco*RI site at map position 6.4 . This location was quite different from that originally proposed for the *hpc* genes from *E.coli* C (Jenkins, 1987) where the *hpcR* gene was positioned next to the *hpcH* gene. However, further analysis of the *E.coli* C genes (Roper, 1990) suggested that the *hpcR* gene was in fact located in an equivalent position to that determined for the *hpc* genes from *K.pneumoniae*.

An attempt was made to confirm the presence of the *hpcE* and *F* genes on the neighbouring 4.7 Kbp *Eco*RI fragment by introducing this into pUC18 generating a plasmid designated pNR17. However, no activities were detectable in extracts of *E.coli* 5K harbouring this plasmid suggesting that cleavage at the *Eco*RI site at map position 6.4 prevents expression of both activities. In the absence of any measurable enzyme activity it was not possible to confirm the presence of the *hpcR* gene on this fragment.

The plasmid pNR18 had been prepared in such a way that DNA was deleted from one end only of the 14 Kbp *Hin*dIII insert of PTF100/102. It had been possible to detect a number of activities on pNR18 and in order to elucidate further the gene order it was decided to prepare a sub-clone of pTF100/102 where DNA was deleted from the end previously retained intact. Therefore the 6.0 Kbp *Eco*R V fragment representing the DNA from map position 1.8 to position 7.8 was isolated and ligated into pUC18 which had been digested with *Sma*I to generate a ligatable blunt end. Screening of the transformants by restriction analysis of plasmid DNA identified a plasmid designated pNR16 which contained the desired fragment. This construct includes some 1.4 Kbp of DNA not present on the *Eco*RI-*Hin*dIII fragment of pNR18 but has lost approximately 1.8 Kbp of DNA at the left-hand end of the fragment. Enzyme assays were then performed on cell-free extracts of *E.coli* 5K (pNR16). It appeared from these measurements that the deletion of DNA from map position 0 to 1.8 had caused the

loss of the activities associated with the OHED hydratase (HpcG) and HHED aldolase (HpcH) . This immediately located these activities at the left-hand end of the *Hin*dIII insert but it was not possible at this point to determine the relative positions of these genes. The inclusion of the DNA from map position 6.4 to 7.8 appeared to allow the expression of the activities formally associated with the COHED decarboxylase (HpcE) and HHDD isomerase (HpcF). This observation is clearly compatible with the location suggested for these genes on the basis of the analysis of pNR18. It was not possible to determine the relative positions of these genes. The analysis of the *hpc* genes from *E.coli* C has , however, suggested that the assignment of separate genes for COHED decarboxylase and HHDD isomerase activities is not correct (Roper, 1990). It was shown that it was possible to purify a single polypeptide species which possessed both activities and subsequent nucleotide sequence analysis identified a single open reading frame. If the same situation pertained in *K.pneumoniae* the expectation would be that it would not be possible to separate the two activities. The fact that the expression of the *hpc* genes on pNR16 was not regulated suggested that the complete *hpcR* gene was absent from this construct just as it had been from pNR18.

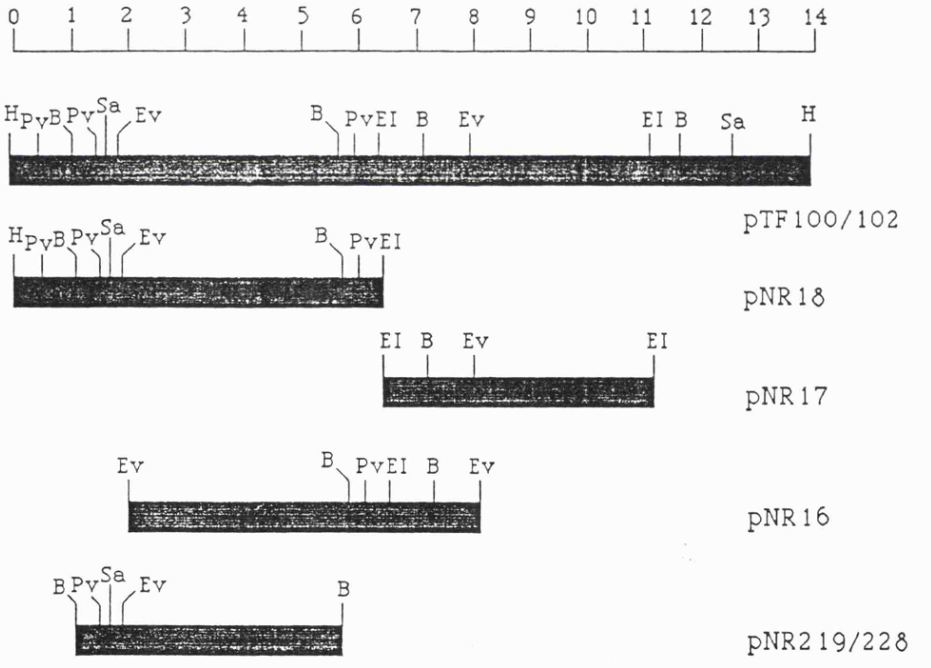
The relative positions of the *hpcG* and *H* genes were determined when sub-clones were prepared which included the 4.6 Kbp *Bam*HI present as part of the independently isolated genomic clone pJF1. The analysis of this clone is described in more detail in the following section but for the determination of this part of the gene order the derivatives of pJF1 described previously (pNR128/228 and pNR219/419) are considered as sub-clones of pTF100/102. When cell-free extracts of *E.coli* 5K (pNR228) were assayed for the enzymes of the HPC pathway it was possible to detect the activities associated with the HPC dioxygenase, CHMS dehydrogenase, CHM isomerase, OHED hydratase (HpcB, C, D and G) but it was not possible to detect any other activities. The integration of this data with that obtained from the analysis of pNR16 placed the *hpcG* gene to the right of the *hpcH* gene which appeared to be the gene located closest to the left hand end of the insert from pTF100/102.

The sub-cloning of pTF100/102 is summarised in Figure 6.3.

Figure 6.3

The sub-cloning of the 14 Kbp *Hin*dIII fragment from pTF100/102. Activities associated with individual sub-clones are marked present (+) or absent (-). The (?) associated with pNR17 indicates that the presence or absence of the *hpcR* gene cannot be assayed owing to the absence of any measurable catabolic function.

Key to enzymes : B= *Bam*HI; EI= *Eco*RI; Ev= *Eco*RV; H= *Hin*dIII; Pv= *Pvu*II; Sa= *Sal*I



hpc GENES EXPRESSED							
B	*C*	*D*	*E*	*F*	*G*	*H*	*R*
+	+	+	+	+	+	+	+
+	+	+	-	-	+	+	-
-	-	-	-	-	-	-	?
+	+	+	+	+	-	-	-
+	+	+	-	-	+	-	-

Derivatives of pJF1

There were four enzyme activities detected on the 4.6 Kbp *Bam*H I fragment which had been sub-cloned from pJF1 to give pNR128/228 and pNR219/419. These sub-clones were analysed to derive the restriction map shown in Figure 6.1b and are represented individually in Figure 6.2. The availability of derivatives differing only in the orientation of the insert fragment with respect to vector-based promoters, the *tet* promoter for the pBR328 derivatives and the *lac* promoter for the pUC-based constructs, afforded an opportunity to determine the direction of transcription of these genes. Cell-free extracts were prepared from *E.coli* 5K harbouring each of these plasmids and enzyme assays performed. These measurements are recorded in Table 6.1. It is apparent that the levels of expression of each gene are higher from pNR228 than from pNR128 and higher from pNR219 than pNR419 and thus that the vector promoters have a considerable effect on *hpc* gene expression. If the orientations of the vector promoters are considered it is apparent that each of the *hpc* genes present on this fragment is transcribed from right to left as drawn in Figure 6.1b . Only one activity, that associated with the *hpcG* gene, had been localised through the sub-cloning of fragments from pTF100/102. The remaining three genes (*hpcBCD*) had been present or absent *en bloc* on the sub-clones so far produced so it was not possible to determine their order. A number of restriction sites within the 4.6 Kbp *Bam*HI fragment had been mapped and it was hoped that it would be possible to apply this information to construct a series of sub-clones which would allow the ordering of these genes.

It was decided initially to prepare a sub-clone in which the insert DNA to the left of the *Sst*I site at map position 3.7 was deleted. This would remove a total of 2.6 Kbp of the insert DNA leaving 2.0 Kbp of cloned DNA. The *Sst*I site within the polylinker sequence of pNR219 was orientated such that digestion with this enzyme would liberate two fragments. One of these was a fragment of 4.7 Kbp which included the entire 2.7 Kbp of vector DNA together with the desired 2.0 Kbp piece of the cloned DNA. This fragment was isolated from an LMP gel and was then religated to give rise to a plasmid designated pNR259. Enzyme assays of cell-free extracts of *E.coli* NM522 (pNR259) grown

Table 6.1

Enzyme activities associated with presence of plasmids containing the 4.6Kbp *Bam*HI fragment from pJF1 in opposing orientations in pBR328 and pUC19 in *E.coli* 5K cells grown on glycerol.

	Expression of *hpc* genes (nmoles/min/mg protein)			
Construct	***B***	***C***	***D***	***G***
pNR128	250	15	75	ND
pNR228	1200	40	565	60
pNR219	3095	310	555	150
pNR419	35	13	30	ND

KEY TO GENES

hpcB= HPC dioxygenase

hpcC= CHMS dehydrogenase

hpcD= CHM isomerase

hpcG= OHED hydratase

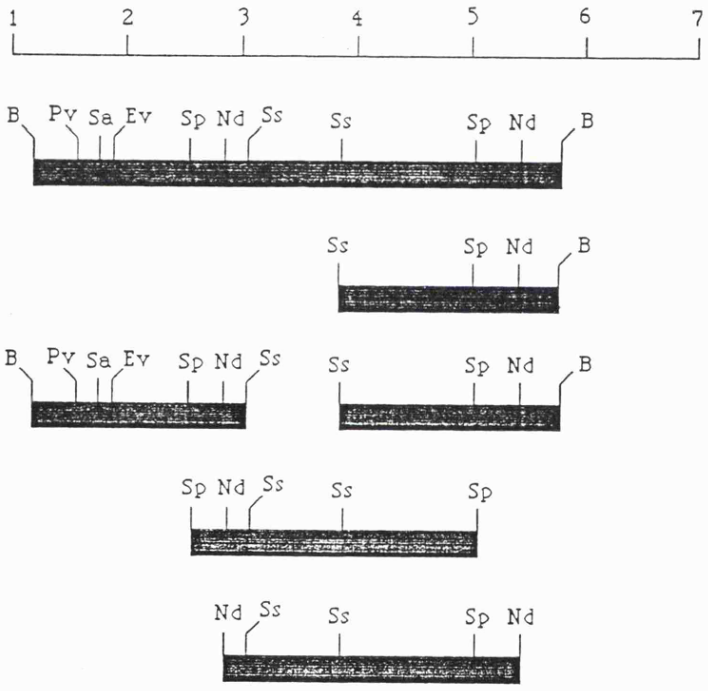
on Luria broth with IPTG added indicated the presence of only a single activity, that of the CHMS dehydrogenase (HpcC). This immediately localised the *hpcC* gene, which has a predicted size of approximately 1.4 Kbp, estimated from the sub-unit relative molecular mass of the purified protein (Fawcett *et al*, 1989), to a 2.0 Kbp region of the cloned DNA. It was also apparent that the other three genes present on pNR219 lay at least partially outside this region.

A second *Sst*I sub-clone was generated by the treatment of pNR228 with *Sst*I. The aim of this procedure was to delete the central 0.8 Kbp *Sst*I fragment. This was possible with a pBR328-based construct because this vector has no *Sst*I site and so the only *Sst*I sites present were those delineating this fragment. It was hoped that any intact gene(s) lying in front of the *Sst*I site at map position 2.9 would be expressed along with the *hpcC* gene which would be retained in this construct. This would therefore identify the genes located towards each end of the *Bam*HI insert through the presence of their products in cell-free extracts and those genes located centrally by the absence of their products. The digestion with *Sst*I was performed and the 8.7 Kbp fragment isolated from an LMP agarose gel. Religation and the screening of the resulting transformants allowed the isolation of a plasmid, designated pNR438, which showed the expected restriction pattern. Cell-free extracts from *E.coli* 5K (pNR438) showed the presence of the activities associated with the CHMS dehydrogenase (HpcC) and OHED hydratase (HpcG) products but those of the HPC dioxygenase (HpcB) and CHM isomerase (HpcD) products were absent. As the *hpcG* gene had been absent from pNR259 but was present on pNR438 this suggested that this gene must be located in front of the *Sst*I site at map position 2.9.

These experiments suggested that the *hpcB* and *hpcD* genes would be located in the central region of the 4.6 Kbp *Bam*HI insert. In order to confirm this a sub-clone designated pNR289 was prepared by ligation of the 2.6 Kbp *Sph*I fragment , which spans this central region, into pUC19. Cell-free extracts of *E.coli* 5K (pNR289) were found to contain activities associated with HPC dioxygenase (HpcB) and CHM isomerase (HpcD) but those associated with CHMS dehydrogenase (HpcC) and OHED hydratase (HpcG) were absent. This result was clearly compatible with a model in which the *hpcB* and *D* genes were located

Figure 6.4

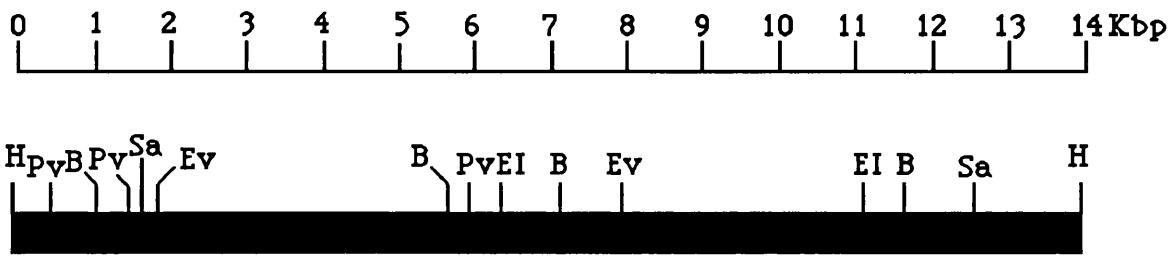
The sub-cloning of the 4.6Kbp *Bam*HI fragment prepared from the plasmid pJF1. Key to enzymes : B= *Bam*HI; Ev= *Eco*RV; Nd= *Nde*I; Pv= *Pvu*II; Sa= *Sal*I; Sp= *Sph*I; Ss= *Sst*I

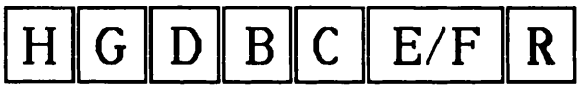


hpc GENES EXPRESSED			
B	*C*	*D*	*G*
+	+	+	+
-	+	-	-
-	+	-	+
+	-	+	-
+	+	-	-

Figure 6.5

Order deduced for the *hpc* genes of *K.pneumoniae* M5a1. Genes are located to their approximate map locations but are not drawn to scale. The designation *hpcE/F* is used to illustrate the fact that the precise status of these genes is unclear.





Key to enzymes : B=*Bam*HI EI=*Eco*RI

Ev=*Eco*RV H=*Hin*dIII

Pv=*Pvu*II Sa=*Sal*I

together between the *hpcC* and *G* genes whose positions had already been determined.

The key question to be addressed at this point related to the determination of the relative positions of the *hpcB* and *D* genes. Sub-clones prepared up to this point had contained either both or neither activity. It was decided to sub-clone the 2.5 Kbp *Nde*I fragment which included the DNA from map position 2.75 to 5.25. This fragment contained some 1.0 Kbp of DNA which had not been represented on pNR259 which might be sufficient to encode one or other of the activities absent from pNR259. The plasmid identified by the screening of the colonies resulting from the ligation of the gel-purified *Nde*I fragment with pUC 19 was designated pNR299. Enzyme assays of *E.coli* 5K (pNR299) indicated the presence of the HPC dioxygenase (HpcB) andCHMS dehydrogenase (HpcC) but not the CHM isomerase (HpcD). This strongly suggested that the *Nde*I site at position 2.75 lay within the *hpcD* gene with the *hpcB* gene lying intact to the right of this site as drawn. Interestingly nucleotide sequencing of the *hpcD* gene from *E.coli* C (Roper and Cooper, 1990a) has identified an *Nde*I site lying almost centrally in the coding sequence so it may be that this site is a common feature of the two systems.

The sub-cloning of the 4.6 Kbp *Bam*HI fragment is summarised in Figure 6.4 with the deduced gene order.shown in Figure 6.5.

Discussion

In this chapter two discrete genomic clones pTF100/102 and pJF1 were available for analysis. A recent report (Martin *et al*, 1991) has described the isolation of a further clone which includes a number of the *hpc* genes. This latter clone was obtained through the screening of a library constructed by the ligation of *K.pneumoniae* M5a1 DNA partially digested with *Bam*HI into the vector pBR328. The screening procedure employed by Martin *et al* in this report was almost identical to that employed during the isolation of pJF1 except that the oligonucleotide probe was derived from a different portion of the 5' end of the *hpcC* gene. The insert present in this clone was reported to have a size of 7.0 Kbp, which is rather surprising in view of the fact that no such fragment was apparent in pTF100/102 and that Southern blot analysis of genomic DNA had suggested that the hybridising

*Bam*HI fragment from *K.pneumoniae* had a size of 4.6 Kbp. It is possible that 7.0 Kbp fragment is a product of partial digest but in view of the fact that the report states that as part of the analysis the fragment was reversed, which would require complete digestion with *Bam*HI, it is somewhat surprising that this did not become apparent. It was reported that a total of five activities were detectable on the clone rather than the four detected on pJF1 and its derivatives. As no detailed analysis of this clone was reported and given the fact that the insert size of this clone does not fit expectation it is not possible at this stage to make meaningful comparisons between this clone and those examined in this study. The order, deduced from the experiments described in this chapter, for the *hpc* genes from *K.pneumoniae* is identical to that of the corresponding genes from *E.coli* C. This conservation of gene order provides evidence for a common origin of the pathway in the two organisms. This view was suggested by the previous finding of strong homologies between the N-terminal amino acid sequences of the CHMS dehydrogenase (HpcC) and the OHED hydratase (HpcG) proteins (Fawcett *et al*, 1989; R.A Cooper, pers. comm) in the two organisms. At this point the major question outstanding is whether the COHED decarboxylase and HHDD isomerase activities are combined in a single protein in *K.pneumoniae*. Given the similarities of gene order and specific protein homologies between *K.pneumoniae* and *E.coli* C the finding of a dual function protein in the latter would seem to provide a strong precedent but is, of course, far from definitive.

CHAPTER 7

Summary and future prospects

Summary and future prospects

The information presented in this thesis has investigated the organisation of two aromatic catabolic systems in *Klebsiella pneumoniae*. In Chapter 3 a cloning strategy was devised and executed which resulted in the isolation of all of the *mhb* catabolic genes on a single restriction fragment. The presence of this plasmid in cells of *E.coli* conferred a new metabolic function allowing 3-HB to be utilised as sole carbon and energy source.

The arrangement of the *mhb* genes was elucidated through the preparation of sub-clones from this novel genomic clone and from a pre-existing partial clone. Whilst the ordering of the structural genes of the pathway was relatively straightforward a more involved methodology was involved in the localisation of the *mhbR* gene. This was necessary because several sub-clones were able to partially metabolise the presumed *in vivo* inducer 2,5-dihydroxybenzoate to produce the toxic compound maleylpyruvate. Therefore it was not possible to assay for the presence of regulatory systems on these clones using 2,5-dihydroxybenzoate as inducer. Screening of a range of candidate inducer analogues identified two compounds capable of giving induction whilst being non-toxic. The use of these analogues allowed the position of the *mhbR* gene to be located and also enabled experiments to be performed which suggested that the *mhb* genes were under a system of positive control. These investigations have only really started the dissection of the regulatory system and this may represent an appropriate topic area for future study. For example whilst it has been shown that certain compounds are able to act as inducer analogues no in-depth study of the structural elements required for interaction with the regulatory system has been performed. This type of analysis has been of interest in other systems (e.g the TOL *meta*-pathway) particularly when attempts were made to expand the substrate range of a pathway. On a more molecular level the examination of the regulatory system might include some characterisation of the *mhbR* gene. This might well involve the determination of the nucleotide sequence of this gene although a necessary prelude to this would be the purification of the regulatory protein. This in turn would require the development of a suitable assay system which would allow specific identification of this product.

Here one might envisage the development of a gel-shift assay methodology. This would be dependent on some preliminary characterisation of the *mhb* operator/promoter region in order to provide an appropriate target sequence. This would of course be a very useful exercise in its own right. The expectation would be that there would be a promoter region in front of the *mhbD* gene which appears to be the first gene transcribed in the operon.

The purification of the 2,5-DHB dioxygenase was satisfactory in the context of allowing N-terminal sequencing which defined the start of the *mhbD* gene. However the loss of enzyme activity during this procedure was somewhat disappointing as it virtually precluded any kinetic or mechanistic analysis of the purified protein. The analysis of the corresponding proteins from *P.testosteroni* and *P.acidovorans* (Harpel and Lipscomb,1990a,b) has yielded a substantial amount of data which would be very useful for the purpose of comparison with the *K.pneumoniae* enzyme. It is possible that the existing purification protocol for the enzyme from *K.pneumoniae* could be modified to include some of the key features involved in the preparation of the enzymes from *Pseudomonas*. The most obvious change would be the use of buffers containing ferrous salts and possibly cysteine throughout the purification. This may result in a more active preparation because it might avoid the apparently irreversible inactivation occurring with the use of iron-free buffers. The use of ferrous sulphate in the activation procedure described in this thesis led to problems with the formation of precipitates which interfered with subsequent purification steps. It would therefore be necessary to identify salts which could maintain activity without compromising subsequent manipulations.

Whilst neither the 3-hydroxybenzoate mono-oxygenase nor the fumarylpyruvate hydrolase has been overexpressed to the same degree as the 2,5-DHB dioxygenase it is likely that this might be achieved through further manipulation of the genomic clones. If this were achieved then purification of these proteins would be greatly facilitated. Each of these enzymes has been purified from other sources and so this might aid in the design of purification schemes as well as providing data which could be used to compare with that from the *K.pneumoniae* enzymes. Basic characterisation of these enzymes would be useful as it would help to fill in some of the fine detail

regarding the architecture of the *mhb* gene cluster particularly if combined with further nucleotide sequencing.

The analysis of the *hpc* genes from *K.pneumoniae* has been significantly less extensive than that of the *mhb* genes. The sub-cloning of the available genomic clones has determined the basic gene order but there remain several outstanding questions. Perhaps the most notable of these is whether the *hpcE* and *hpcF* genes are discrete or whether the activities suggested by this nomenclature are combined in a single gene product as has been reported for *E.coli* C. In order to determine this it would be necessary to carry out a protein purification and show that the activities co-purified ultimately to a single polypeptide species. The purification would be facilitated by the overexpression of these genes .However this has not been observed with any of the *K.pneumoniae hpc* genes expressed in an *E.coli* background. With the basic organisation of the hpc genes having been defined for *K.pneumoniae* it might be of interest to compare the nucleotide sequences of these genes with their equivalents from *E.coli* C. This would be most readily achieved for the *hpcC* and *hpcG* genes whose N-terminal sequences are already known thus eliminating the requirement for further protein purification. The N-terminal sequences of the proteins from the two different sources have been shown to have a high degree of homology and it would be interesting to determine the overall homology as a means of investigating the evolutionary origins of catabolic pathways.

REFERENCES

REFERENCES

Abril, M-A, Michan, C, Timmis, K.N and Ramos, J.L (1989) Regulator and enzyme specificities of the TOL plasmid-encoded upper pathway for degradation of aromatic hydrocarbons and expansion of the substrate range of the pathway. J. Bacteriol. **171** : 6782-6790

Adachi, K., Takeda, Y, Seno, S and Kita, H. (1964). Metabolism of *p*-hydroxyphenylacetic acid in *Pseudomonas ovalis*. Biochim. Biophys. Acta . **93:** 483-493.

Allewell, N (1989) Evolving to dissimilate hydrocarbons. Trends in Bioch. Sciences **14** : 473-474

Arciero, D.M, and Lipscomb, J.D (1986) Binding of ^{17}O labelled substrate and inhibitors to protocatechuate 4,5-dioxygenase-nitrosyl complex. Evidence for direct substrate binding to the active site Fe^{2+} of extradiol dioxygenases. J. Biol. Chem. **261** : 2170-2178

Assinder, S.J, and Williams, P.A (1988) Comparison of the meta pathway operons on NAH plasmid pWW60-22 and TOL plasmid pWW53-4 and its evolutionary significance. J. Gen. Microbiol. **134** : 2769-2778

Burlage, R.S , Hooper, S.W, and Sayler, G.S (1990) The TOL (pWW0) catabolic plasmid. Appl. Environ. Microbiol. **55** : 1323-1328

Burlingame, R, and Chapman, P.J (1983) Catabolism of phenylpropionic acid and its 3-hydroxy derivative by *Escherichia coli*. J. Bacteriol. **155** : 113-121

Chatfield , L.K and Williams, P.A (1986) Naturally occurring TOL plasmids in Pseudomonas strains carry either two homologous or non-homologous catechol 2,3-oxygenase genes. J. Bacteriol. **168** : 878-885

Chow, L.T, Kahmann, R, and Kemp, D (1977) Electron microscopic characterization of DNAs of non- defective deletion mutants of bacteriophage Mu. J. Mol. Biol. **155** : 113-121

Church, G.M and Gilbert, W. (1984) Genomic sequencing. Proc. Natl. Acad. Sci. U.S.A **81** : 1991-1995

Cooper, R.A and Skinner, M.A (1980) Catabolism of 3- and 4-hydroxyphenylacetate by the 3,4-dihydroxy-phenylacetate pathway in *Escherichia coli* J. Bacteriol. **143** : 302-306

Cooper, R.A, Jones, D.C.N and Parrott, S (1985) Isolation and mapping of *Escherichia coli* K12 mutants defective in phenylacetate degradation. J. Gen. Microbiol. **131** : 2753-2757

Crawford, R.L, Hutton, S.W and Chapman, P.J (1975) Purification and properties of gentisate 1,2-dioxygenase from *Moraxella osloensis*. J. Bacteriol. **121** : 794-799

Crawford, R.L, Olson, P.E, and Frick, T.D (1979) Catabolism of 5-chlorosalicylate by a *Bacillus* isolated from the Mississippi river. Appl. Environ. Microbiol. **38** : 379-384

Dagley, S. (1975) A biochemical approach to some problems of environmental pollution. Essays in Biochemistry **11** : 81-138

Deschamps, A.M, Richard, C, and Lebeault, J-M (1983) Bacteriology and nutrition of environmental strains of *Klebsiella pneumoniae* involved in wood and bark decay. Ann. Microbiol. (Paris) **134A** : 189-196

Dixon, R. (1986) The *xylABC* promoter from the *Pseudomonas putida* TOL plasmid is activated by nitrogen regulatory genes in *Escherichia coli*. Mol. Gen. Genet. **203** : 129-136

Donnelly, M.I and Cooper, R.A (1981) Succinic semi-aldehyde dehydrogenases of *Escherichia coli* -their role in the degradation of *p*-hydroxyphenylacetate and g-aminobutyrate. Eur.J. Biochem. **113**: 556-561

Doten, R.C and Ornston, L.N (1987) Protocatechuate is not metabolized via catechol in *Enterobacter aerogenes*. J. Bacteriol **169** : 5827-5830

Doten, R.C, Ngai, K-L, Mitchell, D.J, and Ornston, L.N (1987) Cloning and genetic organisation of the *pca* gene cluster from *Acinetobacter calcoaceticus*. J. Bacteriol. **169** : 3168-3174

Duggleby, C.J and Williams, P.A (1986) Purification and some properties of the 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase (2-hydroxymuconic semialdehyde hydrolase) encoded by the TOL plasmid pWW0 from *Pseudomonas putida* mt-2. J. Gen. Microbiol. **132**: 717-726

Farrell, R, and Chakrabarty, A.M (1979) Degradative plasmids : molecular nature and mode of evolution. In Plasmids of medical, environmental and commercial importance (Ed. K.N Timmis and A. Puhler). Elsevier/ North-Holland Biomedical Press. Amsterdam pp97-109

Fawcett, T (1989) Ph.D Thesis. University of Leicester

Fawcett, T, Garrido-Pertierra, A and Cooper, R.A (1989) 5-Carboxymethyl-2-hydroxymuconic semialdehyde dehydrogenases of *Escherichia coli* C and *Klebsiella pneumoniae* M5a1 show very high N-terminal sequence homology. FEMS Microbiol.Lett. **57** : 307-312

Feinberg, A.P, and Vogelstein, B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132** : 6-13

Ferrer, E and Cooper, R.A (1988) Studies with a cloned *Escherichia coli* C 2-oxo-hept-3-ene-1,7-dioate hydratase gene. FEMS. Microbiol. Lett. **52** : 155-160

Franklin, F.C.H, Bagdasarian, M, Bagdasarian, M.M and Timmis, K.N (1981) Molecular and functional analysis of TOL plasmid pWWO from *Pseudomonas putida* and cloning of the entire regulated aromatic ring *meta* cleavage pathway. Proc. Natl. Acad. Sci. U.S.A **78** : 7458-7462

Furukawa, K, Arimura, N, and Miyazaki, T. (1987) Nucleotide sequence of the 2,3-dihydroxybiphenyl dioxygenase gene of *Pseudomonas pseudoalcaligenes*. J. Bacteriol. **169** : 427-429

Garrido-Pertierra, A. and Cooper, R.A (1981) Identification and purification of distinct isomerase and decarboxylase enzymes involved in the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli*. Eur.J.Biochem. **117** : 581-584

Gornall, A.C, Bardawill, C.S, and David, M.M (1949) Determination of serum proteins by means of the biuret reaction. J. Biol. Chem. **177** : 751-766

Gough, J.A and Murray, N.E (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol **166**: 1-19

Grant, D.J.W and Patel, J.C (1969) The non-oxidative decarboxylation of *p*-hydroxybenzoic acid, gentisic acid, protocatechuic acid and gallic acid by *Klebsiella aerogenes* (*Aerobacter aerogenes*). Antonie van Leuwenhoek **35** : 325-343

Groseclose, E.E, and Ribbons, D.W (1973) 3-Hydroxybenzoate 6-hydroxylase from *Pseudomonas aeruginosa*. Biochim. Biophys. Res. Commun **55** : 897-903

Hagedorn, S.R (1980) Doctoral thesis. University of Minnesota

Harayama, S and Rekik,M (1988) Bacterial aromatic ring-cleavage enzymes are classed into two different gene families. J. Biol. Chem. **264** : 15328-15333

Harayama, S. and Rekik, M. (1990) The *meta* cleavage operon of TOL degradative plasmid pWW0 comprises 13 genes.
Mol. Gen. Genet.**221** : 113-120

Harayama, S, Leppik, R.A, Rekik, M, Mermod, N, Lehrbach, P.R, Reineke, W, and Timmis, K.N (1986) Gene order of the TOL catabolic plasmid upper pathway operon and oxidation of both toluene and benzyl alcohol by the *xylA* product. J. Bacteriol. **167** : 455-461

Harayama, S, Rekik, M, Wasserfallen, A. and Bairoch, A. (1987) Evolutionary relationships between catabolic pathways for aromatics : Conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. Mol. Gen. Genet. **210** : 241-247

Harayama, S, Rekik, M, Ngai, K-L, and Ornston, L.N (1989) Physically associated enzymes produce and metabolize 2-hydroxy-2,4-dienoate, a chemically unstable intermediate formed in catechol metabolism via *meta* cleavage in *Pseudomonas putida*. J. Bacteriol **171** : 6251-6258

Hareland, W.A, Crawford, R.L, Chapman, P.J and Dagley, S (1975) Metabolic function and properties of 4-HPA-1-hydroxylase from *Pseudomonas acidovorans*. J. Bacteriol **121**: 275-285

Harpel, M.R. and Lipscomb, J.D (1990a) Gentisate 1,2-dioxygenase from *Pseudomonas* . Purification, characterization and comparison of the enzymes from *Pseudomonas testosteroni* and *Pseudomonas acidovorans*. J. Biol. Chem **265** : 6301-6311

Harpel, M.R. and Lipscomb, J.D (1990b) Gentisate 1,2-dioxygenase from *Pseudomonas* . Substrate coordination to active site Fe^{2+} and mechanism of turnover. J. Biol. Chem. **265** : 22187-22196

Hartnett, C, Neidle, E.L, Ngai, K-L, and Ornston, L.N (1990) DNA sequences of genes encoding *Acinetobacter calcoaceticus* protocatechuate 3,4-dioxygenase : evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence. J. Bacteriol.**177** : 956-966

Henikoff, S. (1987) Unidirectional digestion with Exonuclease III in DNA sequence analysis. Meth. in Enzymol **155** : 156-165

Hubacek, J and Glover, S.W (1970) Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli* C. J. Mol. Biol. **50**: 111-127

Inouye, S, Nakazawa, A, and Nakazawa, T (1981) Molecular cloning of TOL genes *xylB* and *xylE* in *Escherichia coli*. J. Bacteriol. **145** : 1137-1141

Jenkins, J.R (1987) Ph.D Thesis. University of Leicester

Jenkins, J.R. and Cooper, R.A. (1988) Molecular cloning, expression, and analysis of the genes of the homoprotocatechuate catabolic pathway of *Escherichia coli* C. J. Bacteriol. **170** : 5317-5324

Jones, D.C.N (1985) Ph.D Thesis. University of Leicester.

Jones, D.C.N. and Cooper, R.A (1990) Catabolism of 3-hydroxy-benzoate by the gentisate pathway in *Klebsiella pneumoniae* M5a1. Arch. Microbiol. **154** : 489-485

Kamp, P.F, and Chakrabarty, A.M (1979) Plasmids specifying *p*-chlorobiphenyl degradation in enteric bacteria.In Plasmids of medical, environmental and commercial importance (Ed. K.N Timmis and A. Puhler). Elsevier/ North-Holland Biomedical Press. Amsterdam

Kirk, T.K and Farrell, R.L (1987) Enzymatic "Combustion" : the microbial degradation of lignin. Ann. Rev. Microbiol. **41** : 465-505

Kita, H, Kamimoto, M, & Senoh, H (1965) Crystallisation and some properties of 3,4-dihydroxyphenylacetate 2,3-dioxygenase. Biochim. Biophys. Res. Commun. **18**: 66-70.

Kraft, R, Tardiff, J, Krauter, K.S and Leinwant, L.A (1988) Using mini-prepped DNA for sequencing double-stranded templates with Sequenase. Biotechniques **6** : 544-547

Kushner, S.R (1978) An improved method for the transformation of *Escherichia coli* with ColE1-derived plamids. In Genetic Engineering pp 17-23 ed. Boyer, H.B and Nicosia, S. Elsevier/North-Holland Publishing Co. Amsterdam.

Lack, L (1959) The enzymatic oxidation of gentisic acid. Biochim. Biophys. Acta **34** : 117-123

Lack, L. (1961) Enzymatic *cis -trans* isomerization of maleylpyruvate. J. Biol. Chem. **236** : 2835-2840

Layne, E (1959) Spectrophotometric and turbidimetric methods for measuring proteins. Methods in Enzymol. **3** : 447-454

Maniatis, T, Fritsch, E.F, and Sambrook, J. (1982) Molecular Cloning : A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory

Martin, M, Gibello, A, Fernandez, J, Ferrer, E, and Garrido-Pertierra, A. (1991) Catabolism of 3- and 4-hydroxyphenylacetic acid by *Klebsiella pneumoniae*. J. Gen. Microbiol. **132** : 621-628

Matsudaira, P (1989) Sequence form picomole quantiites of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262** : 10035-10038

Miller, J.H (1972) Ed. In: Experiments in molecular genetics. Cold Spring Harbor, New York

Nakai, C, Hori, K, Kagamiyama, H, Nakazawa, T and Nozaki, M (1983) Purification, sub-unit structure, and partial amino acid sequence of metapyrocatechase. J. Biol. Chem. **258**: 2916-2922

Nakazawa, T, Inouye, S, and Nakazawa, A (1980) Physical and functional mapping of RP4-TOL plasmid recombinants : analysis of insertion and deletion mutants. J. Bacteriol. **144** : 222-231

Neidle, E.L, Shapiro, M.K, and Ornston, L.N (1987) Cloning and expression in *Escherichia coli* of *Acinetobacter calcoaceticus* genes for benzoate degradation. J. Bacteriol. **169** : 5496-5503

Neidle, E.L, Hartnett, C, Bonitz, S, and Ornston L.N (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, Hartnett, C, and Ornston, L.N (1989) Characterization of *Acinetobacter calcoaceticus catM*, a repressor gene homologous in sequence to transcriptional activator genes. J. Bacteriol. **171** : 5410-5421

Noda, Y, Nishikawa, S, Shiozuka, K-I, Kadokura, H, Nakajima, H, Yoda, K, Katayama, Y, Morohoshi, N, Haraguchi, T, and Yamasaki, M. (1990) Molecular cloning of the protocatechuate 4,5-dioxygenase genes of *Pseudomonas paucimobilis* . J. Bacteriol. **172** : 2704-2709

Parrott, S, Jones, S, and Cooper, R.A (1987) 2-Phenylethylamine catabolism by *Escherichia coli* K12. J. Gen .Microbiol. **133** : 347-351

Patel, J.C and Grant, D.J.W (1969) The formation of phenol in the degradation of *p*-hydroxybenzoic acid by *Klebsiella aerogenes* (*Aerobacter aerogenes*). Antonie van Leuwenhoek **35** : 53-64

Perbal, B (1988) A practical guide to molecular cloning. pp 366-374. Wiley, New York.

Ramos, J.L, Mermod, N and Timmis, K.N (1987) Regulatory circuits controlling transcription of TOL plasmid operon encoding *meta* - cleavage pathway for degradation of alkylbenzoates by *Pseudomonas*. Mol.Microbiol. **1** : 293-300

Ramos, J.L, Wasserfallen, A, Rose, K, and Timmis, K.N (1987) Redesigning metabolic routes : manipulation of TOL plasmid pathway for catabolism of alkylbenzoates. Science **235** : 593-596

Reineke, W and Knackmuss, H-J (1978) 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 Knackmuss, H-J (1979) Construction of haloaromatic-utilizing bacteria. Nature (London) **277** : 385-386

Ribbons, D.W. & Eaton, R.W. (1982). Chemical transformations of aromatic hydrocarbons that support growth of microorganisms. p59-84. In Biodegradation and detoxification of enviromental pollutants. ed. A.M. Chakarbarty. CRC press, Boca Raton, Fla.

Roper, D.I (1990) Ph.D Thesis. University of Leicester.

Roper, D.I, and Cooper,R.A (1990a) Purification, some properties and nucleotide sequence of 5-carboxymethyl-2-hydroxymuconate isomerase of *Escherichia coli* C. FEBS. Lett. **266** : 63-66

Roper, D.I and Cooper, R.A (1990b) Subcloning and nucleotide sequence of the 3,4-dihydroxyphenylacetate (homoprotocatechuate) dioxygenase gene from *Escherichia coli* C. FEBS Lett. **275** : 53-57

Sanger, F, Nicklen, S, and Coulson, A.R (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A **74** : 5463-5467

Shanley, M.S, Neidle, E.L, Parales, R.E, and Ornston, L.N (1986) Cloning and expression of *Acinetobacter calcoaceticus catBCDE* genes in *Pseudomonas putida* and *Escherichia coli*. J. Bacteriol. **165** : 557-563

Shine, J, and Dalgarno, L (1975) Determinant of cistron specificity in bacterial ribosomes. Nature **254** : 34-38

Soberon, X, Covarrubias, L, and Bolivar, F (1980) Construction and characterisation of new cloning vehicles IV. Deletion derivatives of pBR322 and pBR325. Gene **9** : 287-305

Southern, E.M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98** : 503-508

Sparnins, V.L, Chapman, P.J, and Dagley,S (1974) Bacterial degradation of 4-hydroxyphenylacetic acid and homoprotocatechuic acid. J. Bacteriol. **120** : 159-167

Spoelstra, S.F (1978) Degradation of tyrosine in anaerobically stored piggery wastes and in pig feces. Appl. and Environ.Microbiol. **36** : 631-638

Stanier, R.Y, and Ornston, L.N (1973) The β-ketoadipate pathway. Adv.Microb.Physiol. **9** : 89-151

Tabor, S, and Richardson, C.C (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. U.S.A **84** : 4767-4771

Taira, K, Hayase, N, Arimura, N, Yamashita, S, Miyazaki, T, and Furukawa, K. (1988) Cloning and nucleotide sequence of the 2,3-dihydroxybiphenyl dioxygenase gene from the PCB-degrading strain of *Pseudomonas paucimobilis* Q1. Biochemistry **27** : 3990-3996

Wang, L.H, Hamzah, R.Y, Yu,Y, and Tu, S.C (1987) *Pseudomonas cepacia* 3-hydroxybenzoate-6-hydroxylase : induction, purification and characterization. Biochemistry **26** : 1099-1104

Wheelis, M.L, Palleroni, N.J, and Stanier, R.Y (1967) The matabolism of aromatic acids by *Pseudomonas testosteroni* and *P. acidovorans*. Arch. Mikrobiol. **59** : 302-314

Williams, P.A and Murray, K (1974) Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida* (*arvilla* mt-2) evidence for the existence of a TOL plasmid. J. Bacteriol. **120** : 416-423

Worsey, M.J and Williams, P.A (1975) Metabolism of toluene and xylenes by *Pseudomonas putida* (arvilla) mt-2 : evidence for a new function of the TOL plasmid. J. Bacteriol **124** : 7-13

Yanisch-Perron , C, Vieira, J, and Messing, J. (1985) Improved M13 cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene **33** : 103-119

Yano, K and Arima, K (1958) Metabolism of aromatic compounds by bacteria, II *m*-Hydroxybenzoic acid hydroxylase A and B; 5-dehydroshikimic acid, a precursor of protocatechuic acid , a new pathway from salicylic acid to gentisic acid. J. Gen. Appl. Microbiol. **4** : 241-258

Yeh, W.K, and Ornston, L.N (1980) Origins of metabolic diversity : substitution of homologous sequences into genes for enzymes with different catalytic activities. Proc. Natl. Acad. Sci USA **77** : 5365-5369

Yeh , W.K, Fletcher, P, and Ornston, L.N (1980) Evolutionary divergence of coselected β-ketoadipate enol-lactone hydrolases in *Acinetobacter calcoaceticus*. J. Biol. Chem. **255** : 6342-6346