MOLECULAR ANALYSIS OF AN AROMATIC DEGRADATIVE PATHWAY: STUDIES ON THE GENES AND ENZYMES FOR HOMOPROTOCATECHUATE DEGRADATION FROM *Escherichia coli* C.

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

DAVID IAN ROPER OCTOBER 1990
MOLECULAR ANALYSIS OF AN AROMATIC DEGRADATIVE PATHWAY: STUDIES ON THE GENES AND ENZYMES FOR HOMOPROTocatechuate Degradation FROM Escherichia coli C.

DAVID IAN ROPER

The homoprotocatechuate degradative pathway of Escherichia coli C contains two isomerisation reactions with chemically similar intermediates, which were thought to be catalysed by distinct but genetically linked enzymes. The possibility that these two isomerases may have arisen from the same ancestral precursor, given their similar substrate structures and close physical location of their genes, was investigated by nucleotide sequencing and purification of the individual enzymes. The purified proteins are of very different subunit molecular weight and have been shown by kinetic measurements to be specific for their respective substrates. The two isomerisation events are separated by an enzyme catalysed decarboxylation and it has been shown that the second isomerisation and the decarboxylation reactions are distinct activities of the same protein subunit. Comparison of the amino acid sequence of the latter with that of the first isomerase of the pathway, reveals a very low level of similarity, suggesting that the two enzymes are unlikely to be derived from a common ancestor.

Subcloning of the rest of hpc gene cluster has revealed that the gene order and direction of transcription are not as previously reported. All the genes for the pathway enzymes are transcribed in the same direction and are subject to negative regulation by a protein which appears to be transcribed in the opposite direction. A putative operator site to which the regulatory protein could bind has been located in the same region as a mapped promoter for the pathway genes, which also contains a binding site for the catabolite activator protein.

Pairwise comparison of the amino acid sequences of the rest of the Hpc enzymes have shown only low levels of similarity. However, the single dehydrogenase enzyme of the pathway is similar to isoforms of human aldehyde dehydrogenase.

The subcloning procedures used in this study have enabled high level expression of several of the pathway enzymes. This has enabled preliminary crystallographic analysis of the isomerase and decarboxylase/isomerase enzymes.
For my parents.
ACKNOWLEDGEMENTS

Although the work detailed in this thesis was a solo effort, many people in Leicester have made a great contribution in terms of their advice, encouragement and friendship. I would first like to thank Dr. R.A. Cooper for his advice and guidance over the past three years, for giving me the chance to start a scientific career and for putting up with my spelling. Lab 108 has seen a large turnover of people since I arrived: Suzanne Parrott, Jadi Sujadi, Amando Garrido-Pertierra, Neil Robson, Ayse Hacisalihoglu and Tony Fawcett (even though he broke my rib!) have all played a part. There are a great number of people in the department who should not go unmentioned, who have offered advice and help render my problems more soluble in alcohol. The DNA gyrase boys: Andy Bates, Richard Reece, Paul Hallett, Porky, Alison Grimshaw, and Tony Maxwell have been the main repository of molecular biological advice. Numerous supporting roles have been played by Dale Wigley, Jerramy Derrick, Bill Primrose, Dave Holmes, Magda Zalacian, Clive Bagshaw, and all the people in 109. A special thanks goes to Chris Thomas who made it possible to transfer this thesis from Amstrad to Apple Macintosh. Many thanks must go to the technical staff, especially Derrick Gamble whose cheery manner and unending ability to be helpful make him a great unsung hero. A special thanks go to Nigel Silman for teaching me how to sail. A number of people outside the department have also made life more bearable by making Hartopp Road a good place to live including Emma Whitham, Kathryn Lilley, Mac, puss-cat and particularly Mike Oliver and Lynne Warwick who have remained friends even though I managed to destroy their kitchen ceiling. There are numerous others who have all helped make my stay in Leicester memorable and I thank them. With luck, I can now answer the question which seems to be asked every time I go home: yes, it is time I got a job!
STATEMENT.

This thesis, submitted for the degree of Doctor of Philosophy entitled: Molecular analysis of an aromatic pathway: Studies on the genes and enzymes for homoprotocatechuic degradation from *Escherichia coli* C, is based upon work conducted by the author in the Department of Biochemistry during the period between October 1987 and September 1990. All the work recorded in this thesis is original unless otherwise stated. I acknowledge the work carried out by Dr. D.B. Wigley in chapter 6 who grew all the crystals of CHM isomerase and carried out the X-ray crystallographic analysis. None of this work has been submitted for another degree in this or any other University.

Signed [Signature]  
Date 27-09-90
Abbreviations

A<sub>680</sub> Absorbance at 680nm
Amp Ampicillin
bps base pairs
CAP Catabolite Activator Protein
CHM 5-carboxymethyl-2-hydroxymuconate
CHMS 5-carboxymethyl-2-hydroxymuconate semialdehyde
COHED 5-carboxymethyl-2-oxo-hex-3-ene-1,6-dioate
Da Dalton
EDTA ethylenediaminetetra acetic acid
ExoIII Exonuclease III
FPLC Fast protein Liquid Chromatography
HHDD 2-hydroxyhepta-2,4-diene-1,7-dioate
HHED 2,4-dihydrox-hepta-2-ene-1,7-dioate
HPC homoprotocatechuate (3,4 - dihydroxyphenylacetate)
IPTG isopropyl-β-D-thiogalactoside
IR Inverted Repeat
Kbp Kilobase pairs
KDa KiloDaltons
λ wavelength
MOPs 3-[N-moropolino] propanesulphonic acid
ORF Open Reading Frame
PAGE Polyacrylamide Gel Electrophoresis
PEG Polyethylene glycol
RNase Ribonuclease
SDS Sodium dodecylsulphate
TEMED  N,N,N',N'-tetramethylethylenediamine
TRIS     Tris hydroxymethylaminomethane
Table of Contents

CHAPTER 1. INTRODUCTION. 1
1.1. AROMATIC CATABOLISM. 2
1.2. RING CLEAVAGE OF THE AROMATIC NUCLEUS. 5
1.3. ORGANISATION OF AROMATIC DEGRADATIVE PATHWAYS. 7
1.4. REGULATION OF GENE EXPRESSION IN PROKARYOTES
   a. GLOBAL REGULATORY MECHANISMS. 9
   b. TRANSCRIPTIONAL REGULATORY MECHANISMS 13
   c. REGULATION AT THE LEVEL OF TRANSCRIPT ELONGATION 16
   d. TRANSLATIONAL REGULATION. 18
1.5. EVOLUTION OF DEGRADATIVE PATHWAYS. 19
   b. CONSERVATION OF PATHWAYS IN SEVERAL SPECIES. 22
1.6. DETAILED ANALYSIS OF KNOWN AROMATIC DEGRADATIVE PATHWAYS. 25
   a. The TOL pathway 25
   b. The β-ketoadipate pathway. 30
   c. Crystallographic analysis of enzymes of the β-ketoadipate pathway 33
1.7. HOMOPROTOCATECHUATE (HPC) DEGRADATION.
   a. The HPC pathway 35
   b. Analysis of the cloned E.coli C hpc genes. 41
1.8. AIMS. 48

CHAPTER 2. METHODS AND MATERIALS 49
2.1. BIOCHEMICAL METHODS. 50
2.1.1. BACTERIAL STRAINS. 50
2.1.2. PLASMIDS.
   a. Plasmid vectors
   b. Genomic clones containing hpc genes of E.coli C. 51
   c. Constructs created in this study. 51
   d. Other constructs used 51
2.1.3. GROWTH MEDIA AND CONDITIONS. 55
2.1.4. PREPARATION OF CELL FREE EXTRACTS. 56
2.1.5. ESTIMATION OF PROTEIN CONCENTRATIONS. 56
2.1.6. PREPARATION OF INTERMEDIATES OF THE HPC PATHWAY. 57
2.1.6. PREPARATION OF INTERMEDIATES OF THE HPC PATHWAY. 57
2.1.7. ENZYME ASSAYS. 59
2.1.8. SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS. (SDS-PAGE) 64
2.1.9. AMINO TERMINAL AMINO ACID SEQUENCING. 65
2.1.10. FAST PROTEIN LIQUID CHROMATOGRAPHY (FPLC). 66
2.1.11. OLIGONUCLEOTIDE SYNTHESIS. 66
2.1.12. CRYSTALLOGRAPHIC EXPERIMENTS. 67
2.2. GENETIC PROCEDURES. 67
2.2.1. PREPARATION OF PLASMID DNA. 67
2.2.2. AGAROSE GEL ELECTROPHORESIS. 70
2.2.3. RESTRICTION ENZYME DIGESTION. 71
2.2.4. PHOSPHATASE TREATMENT OF DNA. 71
2.2.5. ISOLATION OF DNA FRAGMENTS FROM AGAROSE GELS. 72
2.2.6. LIGATION OF DNA. 74
2.2.7. TRANSFORMATION OF BACTERIA WITH PLASMID DNA. 74
2.2.8. EXONUCLEASE III (ExoIII) DIGESTIONS. 75
2.2.9. DIDEOXYNUCLEOTIDE SEQUENCING. 78
2.2.10. TRANSCRIPT MAPPING STUDIES. 82

CHAPTER 3. ORDERING THE hpc GENES 85
3.1. INTRODUCTION. 86
3.2.1. SUBCLONING OF THE 3.8 Kbp SalI-BamH1 FRAGMENT of pJJ801. 86
3.2.2. ANALYSIS OF ENZYME ACTIVITIES ASSOCIATED WITH pDR1835 AND pDR1935. 89
3.2.3. SUBCLONING OF THE 3.0 Kbp BamH1-SalI FRAGMENT OF pJJ801. 91
3.2.4. ANALYSIS OF ENZYME ACTIVITIES ASSOCIATED WITH pDR1830 AND pDR1930. 93
3.2.5. REVISION OF THE GENE ORDER. 97
3.2.6. APPARENT INDUCIBILITY OF hpcB AND hpcD IN pDR1930. 99
3.2.7. DIRECTION OF TRANSCRIPTION OF THE hpc GENES ENCODED ON pJJ801. 101
3.3. CONSTRUCTS REQUIRED TO DETERMINE THE GENE
ORDER OF hpcD AND hpcB. 102

3.4. ORDER OF THE GENES ENCODING HHDD ISOMERASE AND COHED DECARBOXYLASE. 107

3.5.1. INVESTIGATING THE POSITION OF THE GENES ENCODING HHED ALDOLASE AND THE hpc REGULATORY PROTEIN. 112

3.5.2. OBTAINING EXPRESSION OF THE GENES ENCODED ON F2 AND CONFIRMING THE hpcGH GENE ORDER. 116

3.5.3. PROBING THE REGION ENCODING THE Hpc REGULATOR GENE (hpcR). 120

3.6. ORDER OF THE hpc GENES. 124

CHAPTER 4. ENZYME PURIFICATION 125

4.1.1. INTRODUCTION. 126

4.1.2. PURIFICATION OF CHM ISOMERASE. 126

4.1.3. AMINO TERMINAL AMINO ACID SEQUENCING OF CHM ISOMERASE. 130

4.1.4. Km DETERMINATION FOR CHM ISOMERASE. 132

4.2.1. PURIFICATION OF HPC DIOXYGENASE. 136

4.2.2. pH OPTIMUM FOR HPC DIOXYGENASE. 144

4.2.3. Km FOR HPC DIOXYGENASE. 144

4.2.4. AMINO TERMINAL AMINO ACID SEQUENCE FOR HPC DIOXYGENASE. 148

4.3.1. PURIFICATION OF HHDD ISOMERASE/COHED DECARBOXYLASE. 150

4.3.2. AMINO TERMINAL AMINO ACID SEQUENCE FOR COHED DECARBOXYLASE / HHDD ISOMERASE. 159

4.3.4. KINETIC ANALYSIS OF COHED DECARBOXYLASE/HHDD ISOMERASE. 161

CHAPTER 5. NUCLEOTIDE SEQUENCING OF THE hpc GENES, TRANSCRIPT MAPPING STUDIES AND SIMILARITY SEARCHES 166

5.1.1. INTRODUCTION. 167

5.2.1. DELETION SUBCLONING OF pDR1930. 167

5.2.2. GENE ORDER DEDUCED FROM THE pDR93 DELETION SERIES. 170

5.3. NUCLEOTIDE SEQUENCING OF hpcD. 170
5.4. AMINO ACID COMPOSITION OF THE CHM ISOMERASE 174
5.5. NUCLEOTIDE SEQUENCING OF hpcB. 176
5.6. AMINO ACID COMPOSITION OF THE HPC DIOXYGENASE SUBUNIT AND SUBUNIT MOLECULAR WEIGHT. 180
5.7. DELETION SUBCLONING OF pDR1835. 182
5.8. NUCLEOTIDE SEQUENCING OF hpcEF. 185
5.9. AMINO ACID COMPOSITION OF COHED DECARBOXYLASE/HHDD ISOMERASE AND SUBUNIT MOLECULAR WEIGHT. 190
5.10. NUCLEOTIDE SEQUENCING OF hpcC. 192
5.11. AMINO ACID COMPOSITION OF THE CHMS DEHYDROGENASE SUBUNIT AND SUBUNIT MOLECULAR WEIGHT. 194
5.12. NUCLEOTIDE SEQUENCING OF PART OF hpcG. 200
5.13. NUCLEOTIDE SEQUENCING OF THE REGION 3' hpcR. 204
5.14. DOUBLE STRAND NUCLEOTIDE SEQUENCING OF hpcR. 206
5.14.3. AMINO ACID COMPOSITION OF THE Hpc REPRESSOR AND SUBUNIT MOLECULAR WEIGHT. 207
5.15. NUCLEOTIDE SEQUENCING FROM THE HindIII END OF pDR1865. 211
5.16. TRANSCRIPT MAPPING STUDIES. 211
5.16.1. TRANSCRIPT MAPPING UPSTREAM OF hpcEF. 213
5.17.1. TRANSCRIPTIONAL TERMINATION 5' TO hpcR. 217
5.17.2. CATABOLITE ACTIVATOR PROTEIN (CAP) BINDING SITE. 217
5.18. OPERATOR SITE IN THE REGION BETWEEN hpcR and hpcEF. 219
5.19. SEQUENCE COMPARISON STUDIES. 222
5.19.1. DATABASE SEARCHES. 222
a. HPC DIOXYGENASE. 222
b. CHMS DEHYDROGENASE. 227
c. CHM ISOMERASE, COHED DECARBOXYLASE/HHDD ISOMERASE, OHED HYDRATASE AND Hpc REPRESSOR PROTEIN. 230
5.19.2. SEQUENCE COMPARISONS BETWEEN THE Hpc ENZYMES. 230
CHAPTER 6. CRYSTALLISATION OF CHM ISOMERASE AND HHDD ISOMERASE/COHED DEACRBOXYLASE

6.1. INTRODUCTION. 233
6.2. CRYSTALLISATION OF CHM ISOMERASE. 233
6.3. CRYSTALLISATION OF COHED DECARBOXYLASE/HHDD ISOMERASE. 244

CHAPTER 7. DISCUSSION AND FUTURE DIRECTIONS 246
7.1. SYNOPSIS 247
7.2. SUMMARY OF THE RESULTS. 247
7.3. FUTURE WORK. 250

APPENDICES:
1. Oligonucleotides used in this study
2. Nucleotide sequence of the hpc gene cluster obtained during this study.
3. Publications
CHAPTER 1.

INTRODUCTION
1.1. AROMATIC CATABOLISM.

Studies of the ways in which the carbon cycle operate are becoming increasingly important in the modern world. The increased awareness in how man is changing his environment must surely promote a growing interest in how nature responds to environmental changes. A major part of the carbon cycle hinges on the fixation of carbon dioxide (CO$_2$) by plants, and it has been estimated that some $1.5 \times 10^{10}$ tons of carbon in the form of CO$_2$ are fixed annually as wood. This carbon is not removed from the carbon cycle however, since it eventually returns due to the breakdown of plant material by animals and bacteria, or increasingly through man's usage. The bulk of plant biomass is in the form of cellulose and lignin. Since microorganisms are solely responsible for initial steps in breakdown of these compounds, it must follow that their role in the carbon cycle is profound. Thus a greater knowledge of the processes involved in degradative events, will lead to a better global understanding of how the carbon cycle operates and how man can influence it.

Currently there is a great divide in the knowledge of how cellulose breakdown occurs compared to that of lignin. Cellulose is a linear polymer of D-glucose in β-(1-4) linkage, and comparatively a great deal is known about the degradation of the basic structural unit (i.e. glucose) by microorganisms. This is in stark contrast to the information available on lignin degradation, where the structural unit is based on the coupling of three aromatic alcohols: $p$-coumaryl, coniferyl and sinapyl alcohol (see fig 1.1). The key to understanding how microbial breakdown of lignin occurs is in the study of the aromatic catabolism. The major structural unit of these three aromatic alcohols is the benzene nucleus, a relatively stable chemical structure.
Fig.1.1.


The proposed structure of beech wood lignin. There are 25 different C₉ units of which several can, to some extent, be replaced by the three dimeric units in brackets.
Fig. 1.1.

Representation of the structure of lignin.
The study of aromatic catabolism does not end with lignin however, since the aromatic nucleus also occurs in a variety of other naturally occurring compounds such as toluene and phenol found in coal and tar and in the aromatic amino acids. Moreover, the occurrence of man made aromatic chemicals in the environment is becoming more evident and perhaps unsurprisingly some microorganisms have evolved to degrade these novel compounds. An example of this is the ability of a *Flavobacterium* species (Okada et al 1983) and a *Pseudomonas* species (Kazuo et al 1989) to degrade 6-aminohexanoate cyclic-dimer, which is found in waste water from the production of nylon. The ability to degrade novel man-made aromatic compounds, is not generally found and a great number remain recalcitrant thus giving rise to problems of pollution. A better understanding of how aromatic catabolism occurs and is controlled, may eventually lead to several possibilities for man: the design of novel biodegradable chemicals, the "design" of specialised organisms to effectively remove previously recalcitrant chemicals, and a more global understanding of how we affect the environment.
1.2 RING CLEAVAGE OF THE AROMATIC NUCLEUS.

The degradation of lignin is in general initiated by the extracellular enzymes produced by fungi, a subject beyond the scope of this thesis. Breakdown of the lignin structure liberating the individual aromatic components is followed by a variety of processes culminating in the cleavage of the aromatic ring allowing degradation to more simple non-aromatic chemicals. The ring-fission event is generally preceded by substitution reactions which render the aromatic nucleus susceptible to cleavage. The aromatic nucleus is destabilised by the addition of hydroxyl groups by the action of hydroxylases (mono-oxygenases or mixed function oxygenases) or by the addition of two ortho hydroxyl groups in a single step by a dioxygenase (dihydroxylase). The dihydroxy aromatic nucleus is then a substrate for ring cleavage either between the two hydroxyls (ortho -cleavage) or to the side of one of the hydroxyls (meta -cleavage). In cases where the substituted hydroxyls are para to each other, ring fission occurs between one of the hydroxyls and a side chain. (see Fig.1.2.) Once the aromatic nucleus has been opened the structure can then serve as a substrate for other reactions leading to conversion to central metabolites. The cleavage of the aromatic nucleus in either the ortho or para configuration serves as a means of classifying the various different degradative pathways. Since the aromatic nucleus is relatively stable the ring cleavage step represents a committal step in metabolism.
Fig. 1.2.
Ring fission of the aromatic nucleus

*ortho* and *meta* fission of the catechol by different dioxygenases.

*meta* fission of gentisate
Fig.1.2.

ortho and meta fission of the aromatic nucleus.

![Diagram of aromatic fission](image_url)
As such there is a great deal of interest in oxygenase and dioxygenase enzymes especially those involved with the degradation of halogenated aromatic rings. In this context a great deal of research has been concentrated in identifying oxygenase enzymes with relaxed substrate specificity, which could bring about the degradation of novel aromatic chemicals (Ngai & Ornston 1988, Ghosal et al 1985).

1.3. ORGANISATION OF AROMATIC DEGRADATIVE PATHWAYS: THE CHANELLING OF RING-FISSION PRODUCTS INTO CENTRAL PATHWAYS.

There are a vast range of aromatic chemicals which are degraded in the environment. However, it appears nature has been economical in the use of mechanisms to bring about such degradation. As different microbial species have been studied it has become apparent that there are a small number of central pathways by which a range of chemicals are channelled. A large range of aromatic chemicals converge on one of a small range of ring-cleavage substrates such as catechol, gentisate and protocatechuate or their derivatives (Ribbons and Eaton 1982). This type of mechanism must surely represent a more economical use of the cell’s resources including a reduced genetic load and simplification of the regulatory mechanisms required to allow growth on a range of different compounds. An example of such channelling is depicted in Fig.1.3. involving the ortho cleavage β-ketoadipate pathway found in a variety of organisms but principally studied in Acinetobacter calcoaceticus and Pseudomonas putida. The meta-cleavage, TOL pathway represents one of the other major degradative pathways and is responsible for the dissimilation of catechol, toluene, m-xylene and p-xylene. Both pathways are discussed in more detail in later sections.
Fig. 1.3.

The β-keto adipate pathway taken from Shanley et al (1986).

A variety of other aromatic compounds than those shown can be channelled into the pathway. They included halocatechols which can enter at the level of the β-keto adipate. Other compounds included quinic acid and shikimic acid which enter via protocatechuate. The catechol side of the pathway is the entry point for mandelate and tryptophan (Wheelis & Stanier 1970). The two branches of the pathway converge on the intermediate β-keto adipate enol-lactone. Genetic designations for the enzymes are shown. A single hydrolase (pcaD) and a single transferase (pcaE) are synthesised in Pseudomonas putida.
Fig.1.3.

Channeling of aromatic compounds towards the β-ketoadipate pathway
1.4 REGULATION OF GENE EXPRESSION IN PROKARYOTES:

a. GLOBAL REGULATORY MECHANISMS.

An understanding of the ways in which gene expression is regulated in prokaryotes is fundamental to a study of aromatic catabolism. Moreover, if aromatic degradative systems are to be used by man as a means of controlling pollution for example, a proper understanding of the control mechanisms involved in gene expression will be essential. It seems clear that there are several global regulatory systems employed by bacteria to control carbon and nitrogen utilisation and the heat-shock response, for example. The various genes and operons involved in these responses are generally controlled as a regulon i.e. they are controlled by a common regulatory element. "Enteric bacteria are able to grow in a minimal medium containing one of a large variety of organic compounds as a sole source of carbon and energy an one of a more limited variety of organic or inorganic compounds as a sole nitrogen source; judging from the rates of growth in such minimal media, glucose is their preferred source of carbon and ammonia is their preferred source of nitrogen." (Magasanik & Neidhardt 1987). The special position of glucose and ammonia in the metabolism of such cells is reflected in the fact that the enzymes required for their metabolism are always present in the cell regardless of the growth medium. This is not true of the enzymes required for growth on other carbon substrates, which in general are present at much reduced levels, if present at all. Control of expression of a variety of genes under these conditions is said to be mediated by the global carbon-energy (Cer) regulon for glucose and by the global nitrogen (Ntr) regulon for nitrogen. The best understood is the Cer regulon. When growth occurs on glucose the enzymes required for growth on other carbon substrates are present at very low levels. The genes involved are said to be subject to the glucose effect. The glucose effect has three components, including the exclusion of inducers by a
mechanism linked to glucose, whereby uptake systems for those inducers are inactivated. The other two mechanisms are the severe repression of enzyme synthesis for about one generation and catabolite repression mediated by a requirement for the presence of the Catabolite Activator Protein (CAP)-cAMP complex for transcription activation. During conditions of glucose limitation, intracellular cAMP levels rise in the cell. In such conditions the CAP-cAMP complex forms and activates the transcription of a number of catabolic operons including gal, ara, mal and lac by binding upstream of the RNA polymerase binding site in the promoter region of these operons. In the case of lac, the presence of the cAMP-CAP complex facilitates binding of RNA polymerase to an extent that only 2% of the maximum level of transcription is seen in the absence of the complex. Similar effects are seen for other operons subject to the glucose effect. Thus the binding of the cAMP-CAP complex signals a requirement for substrates other than glucose to be used as a carbon source.

The global nitrogen regulon was first characterised by increased expression from a number of operons, when ammonia levels become restricted (Magasanik 1985). These nitrogen regulated operons encode products that initially facilitate the assimilation of low concentrations of ammonia by means of the glutamine synthetase-glutamate synthase pathway and then the utilisation of alternative sources of nitrogen. Part of this response involves the transcription of two regulatory genes; ntrA and ntrC which in turn activate the transcription of glnA which encodes glutamine synthetase. It has been shown that the ntrA gene product (also called glnF & rpoN) is an RNA polymerase σ factor (Hirschman et al 1985). This particular σ factor is required for the activation of transcription from promoters which have a -24 (GG) -12 (GC) recognition sequence. These promoters lack the -35 and -10 regions characteristic of σ70 promoters. Although originally found in Ntr operons, this type of promoters has now been found in a variety of other operons which
appear unconnected to global nitrogen regulation (Thony & Hennecke 1989). For example the ntrA gene product has been shown to be required for the expression of diverse genes including the Flagellin gene in *Pseudomonas aeruginosa* (Totten *et al* 1990), the expression of the hydB involved in hydrogen metabolism in *E.coli* (Sankar & Shanmugan 1986) and diverse metabolic functions including C₄-dicarboxylate transport in *Rhizobium meliloti* (Ronson *et al* 1987). The general features of these types of promoters are (i) the conserved -24(GG)/-12(GC) consensus sequence, (ii) its recognition by σ⁶⁴ RNA polymerase sigma factor and (iii) the requirement for a transcriptional regulatory protein to activate the transcription of the associated genes. In short, the expression from all -24(GG)/-12(GC) consensus promoters known to date is positively controlled (Thony & Hennecke 1989). Perhaps more significantly for this study, the ntrA gene product has been shown to be involved in the transcription of TOL plasmid xylS regulatory gene and the structural genes of the so called, upper pathway (Kohler *et al* 1989). The regulatory system for the TOL plasmid genes is shown in Fig.1.4. Briefly, the TOL pathway enzymes are encoded in two operons. The first (upper) operon encodes the three enzymes that oxidise toluenes and xylenes to benzoates and toluates. The second (lower) operon encodes nine enzymes that complete the degradative metabolism to give tricarboxylic acid cycle intermediates. Transcription from the upper operon is positively regulated by the product of xylR which is activated in the presence of toluene or its alcohol derivatives. The lower operon is positively regulated by the product of xylS which is activated in the presence of the effectors: benzoates and toluates. The active xylR protein induces the xylS promoter which results in transcription of the lower operon enzymes even in the absence of the xylS effectors (Mermod *et al* 1987). Expression of xylR is auto regulated.
Fig. 1.4.

Regulation of the TOL pathway.

All of the catabolic genes encoded by the TOL plasmid of \textit{P.putida} are clustered into two operons. The first operon (upper operon) encoded three enzymes which oxidise toluene and xylenes to benzoates and toluates whereas the second lower operon (meta operon) encodes nine enzymes which transform benzoates and toluates to tricarboxylic acid cycle intermediates. Transcription from the upper-operon promoter (Pu) is positively regulated (+) by the XylR protein (●) which is activated in the presence of toluene, xylenes or their alcohol derivatives. Expression from the lower (meta) pathway promoter (Pm) requires the XylS protein (O) and its effectors, benzoates or toluates. The XylR protein, in its activated form also induces the \textit{xylS} gene promoter (Ps) resulting in a hyper-production of the XylS protein, which in turn induces the meta operon even in the absence of XylS effector molecules. Expression of \textit{xylR} is autoregulated (−).
Fig. 1.4.

Representation of the regulation of the catabolic operons of the TOL plasmid from Kohler et al. 1989.
Transcription from the \textit{xylS} gene and the upper operon promoters requires the \textit{ntrA} factor for transcription (Dixon 1986 & Ramos \textit{et al} 1987, respectively), such that their promoters have the -24, -12 motif. In contrast the transcription of the lower operon genes and the \textit{xylR} gene does not require the \textit{ntrA} \(\sigma\) factor. In summary then, the involvement of various sigma factors enabling transcriptional initiation at particular sites not recognised by other sigma factors, may represent another level at which global regulatory mechanisms act.

b. TRANSCRIPTIONAL REGULATORY MECHANISMS

Regulation of gene expression at the level of transcription represents one of the most economical ways to exert control and many examples are found. In general protein factors are implicated in affecting transcriptional activation. Such proteins are categorised into either positive or negative regulators of gene expression. Positive regulators bind to their target DNA and increase the ability of RNA polymerase to bind to a promoter site. Negative regulators bind to their target DNA and interfere with the RNA polymerase-promoter interaction preventing transcriptional initiation.

The operon model of gene regulation, as proposed by Jacob and Monod (1961) envisaged a simple model for the organisation of gene expression in a regulated system in which an intermediate in gene expression mRNA would play a role. Jacob and Monod defined the operon as a unit of transcriptional activity and suggested a repressor-operator mode of control which was applied to the \textit{lac} system of \textit{E.coli} and the bacteriophage lambda. The \textit{lac} system consists of three structural genes for lactose metabolism, (\textit{lacZ}, \textit{lacY} and \textit{lacA}) whose expression is controlled by the product of the \textit{lacI} gene. The \textit{lacI} repressor was proposed to bind to a control region (or operator) in the absence of lactose which prevented the expression of the structural genes for lactose
metabolism. The presence of lactose in the growth medium prevented the repressor from interacting with the operator site allowing the transcription of the structural genes. As such the product of lacI, the lacI repressor is a negative regulator of gene expression. As discussed in section 1.4.a., the lac operon is part of the Cer regulon and is therefore subject to positive control by the cAMP-CAP complex as well as negative control by lacI.

The proposal by Jacob and Monod is now known to be an oversimplification since initiation control signals have been found within, downstream and far upstream of the genes they regulate. This is a more widespread phenomenon in eukaryotes but bacterial examples are known. The lac system is still a model for study today, and recent reviews by Gralla (1989a, 1989b) describe an updated version of the mechanism of repression in this system, which is much more complex than originally proposed. At least three binding sites for the lac repressor occur in vitro, one of which was the classically defined operator site just upstream of the promoter site, the second a very weak site 93 bps upstream of the operator and the third was located 402 bps downstream within the lacZ gene. However, even in the light of this and other studies, the exact mechanism how lac repression is still incomplete and remains a topic of great research. The concept of a single promoter site and a single operator site controlling the expression of a single transcript as exemplified in the original Jacob and Monod model has been largely replaced with more wide ranging ideas as other systems are studied. Today we know that the control mechanism elaborated for the lac system is only one of many different ways in which bacterial gene expression is controlled. The arabinose operon of E.coli is an example of how positive control of transcription operates. The operon is responsible for the controlled synthesis of three enzymes required to convert arabinose to xylulose-5-phosphate. The araC gene directs the synthesis of the araC regulatory protein, which in the absence of
arabinose, binds to an operator site and represses the initiation of transcription of the three structural genes for arabinose metabolism. In the presence of arabinose the araC binds to the operator site and facilitates the binding of RNA polymerase to the promoter site. Most of the interest in the AraC protein has been generated by its positive-acting function when arabinose is present. In the case of the arabinose operon, the AraC protein can function either as a repressor or as a transcriptional activator. The araC gene and the araBAD operon have their own oppositely orientated promoters. The formation of a DNA loop by the AraC protein bound to two distinct sites; O2 and I, which are separated by 211 bps is necessary for the repression seen in the absence of arabinose (Lee & Schleif 1989, Gralla 1989a). In vivo, the I site is always occupied by AraC, a configuration required for transcriptional activation. The role of loop formation is apparently to prevent AraC from activating transcription by preventing a required confirmational change, and by preventing interaction with a second copy of AraC which may bind to another site during induction. During induction the loop is thought to be broken allowing RNA polymerase to interact with the araBAD promoter.

As already described in section 1.4.a. the control of expression of the TOL pathway genes is largely thought to be at the level of transcription where two regulatory proteins, the products of xylS and xylR act in a positive sense i.e. they are required for transcriptional activation. This also seems to be true generally in the control of expression of the β-ketoadipate pathway genes. The β-ketoadipate pathway is another major route for the degradation of aromatic compounds, principally protocatechuate and catechol, and has been extensively studied in P.putida and Acinetobacter calcoaceticus. The pathway is discussed in more detail later but has two branches for the degradation of protocatechuate and catechol. Genetic experiments indicate that the genes of the protocatechuate and catechol arms form two linkage groups and that the
regulation of gene expression is primarily by induction by positive acting regulatory proteins (Ornston & Yeh 1981, Allewell 1989). In particular the catR gene product of \textit{P. putida} has been shown to be a positive regulator of the \textit{catBC} operon and bears sequence homology to a group of prokaryotic regulatory proteins designated as the LysR family (Rothmel \textit{et al} 1990).

c. REGULATION AT THE LEVEL OF TRANSCRIPT ELONGATION AND TERMINATION.

This is a very large topic to which a great deal of study has been directed. Although regulation of gene expression at the level of transcription is thought to be the main level of control, there are many examples of regulation of expression during later events. Although elongation of the mRNA transcript seems to be an extremely stable process, under certain conditions temporary or permanent interruptions can occur. Studies on the elongation of \textit{lacZ} have shown that pausing of the transcript can occur upstream of particularly GC rich regions, presumably due to the greater degree of hydrogen bonding in such base pairs necessitating larger changes in free energy during the translation events (Yarger & Hippel 1987).

It has been known for some time that the generation of self complementary sequences in the mRNA transcript can induce pausing of RNA polymerase. The classical example of this effect is in the regulation of expression from amino acid biosynthetic operons, such as the \textit{trp} operon (for review see Landick \& Yanofsky 1987). The \textit{trp} operon encodes the enzymes required for the synthesis of tryptophan. The transcriptional start site in the \textit{trp} operon is separated from \textit{trpE}, the first structural gene in the operon by a leader sequence of around 160 bps. This region is responsible for the transcriptional termination: attenuation control. Attenuation provides a mechanism of controlling expression of the \textit{trp} genes depending on the
requirement for tryptophan within the cell, specifically in response to changes in the level of charged tRNA^{Trp}. The leader sequence within the mRNA transcript encodes several codons for tryptophan so that when tryptophan levels are high the mRNA is translated quickly by ribosomes and the portion of the transcript that has not yet been translated, folds itself into a particular stem loop structure that causes termination of transcription. However, when tryptophan is scarce, ribosomes attached to the growing mRNA transcript are stalled allowing the rest of the mRNA molecule to adopt a conformation that prevents premature termination and a full length mRNA transcript of the five trp genes is produced (Yanofsky & Crawford 1987). This type of regulation is also seen in operons responsible for phenylalanine and histidine biosynthesis.

Transcriptional termination at attenuator sites is a rho-independent event that is spontaneous unless affected by ribosomal pausing as outlined above. In contrast, however, there are some systems that are regulated by specific antitermination proteins. One of the best studied systems is the rho-dependent early events that occur upon infection of E.coli by bacteriophage λ. Transcription from the λ PI and Pr promoters during the early stages of infection is terminated after 1000 bps and 500 bps respectively by the rho protein of E.coli. The mechanism involves two sequential steps: rho-independent stalling of RNA polymerase followed by rho mediated release of the RNA transcript. The PI transcript encodes the N antitermination protein. The action of the N protein prevents rho from terminating transcription of the PI transcript after 1000 bps. As such the transcript is extended into the delayed-early region of the λ genome which encodes recombination and package functions. Several rho-dependent chain terminating regions have been identified in λ and E.coli DNA. Comparisons of these sequences have yielded no obvious similarities but mutations in the rho gene itself cause RNA polymerase to read through from one operon to the next so it seems clear the
rho plays an important role in terminating transcription. It is clear that the amount of a particular mRNA species in the cell will depend on both the rate of synthesis and degradation of that mRNA. In general bacterial mRNAs have short life spans in the region of a few minutes. Thus the finely tuned mechanisms for the control of mRNA initiation plus the rapid degradation of mRNA molecules, allows very efficient and flexible control of metabolism.

There are conflicting reports on the role which ribosomes play on the stability of mRNA. Har-El et al (1979) showed that the presence of ribosomes can decrease the half life of mRNA. However, a conflicting report of Yates and Nomura (1981) showed that in the absence of ribosomes the rate of mRNA degradation was increased. It has been suggested in this context that there is a nuclease activity associated with the ribosomes themselves, so that under certain conditions the presence of ribosomes enhances degradation.

d. TRANSLATIONAL REGULATION.

The efficiency of the binding interaction between mRNA and ribosome can affect the rate at which that mRNA is translated. Specifically, mRNA molecules have a short region just upstream of the translational start signal which binds to a complementary region of 16S rRNA at its 3' end. The region found in the mRNA is called the Ribosome Binding site (RBS) or Shine-Dalgarno sequence (1974). The strength of the interaction between the RBS and the 16S rRNA sequence can drastically effect the efficiency with which the downstream mRNA is translated. Other signals not yet understood, probably play a part in regulation at this level as some sequences complementary to the 3' end of 16S rRNA do not act as sites for translational initiation. In addition, instances of proteins effecting translational initiation have been documented (Gold & Stromo 1987). Genes expressed in a 1:1 ratio are often separated by short intergenic regions so that the open reading frames are translationally
coupled as has been shown with the \textit{galT} and \textit{galK} genes (Schumperli et al 1982). In contrast, genes which are translated at different rates may be regulated by the efficiency of individual ribosome binding sites. The translation of the mRNA may also be regulated by the codon usage employed in that message. Highly expressed genes contain very few codons which are recognised by minor isoaccepting species of tRNA while poorly expressed genes tend to have a higher proportion of these rarely used codons (Konisburg & Godson 1983). Evidence to show that translational efficiency is affected by codon usage has yet to be demonstrated.

1.5. \textbf{EVOLUTION OF DEGRADATIVE PATHWAYS.}

a. \textbf{MECHANISMS OF EVOLUTION}

Since there is no fossil record for comparison of present day bacterial species to ancestral relatives, the establishment of evolutionary links has been a topic of great debate. With the advent of molecular biological techniques, such studies have become more feasible since comparisons can be made at the level of nucleotide and amino acid sequence. The study of metabolic pathways in this regard has shed some light onto the processes involved.

One of the earliest theories to account for the evolution of metabolic pathways was the retrograde evolution proposal of Horowitz (1945) in which the development of a metabolic pathway was proposed to occur in terms of microevolutionary steps. Horowitz envisaged an ancient environment where a simple organism reproduced itself at the "expense of prefabricated organic molecules in the environment". With time those molecules may have become depleted in the environment limiting further multiplication of the organism. If a structurally-related compound to that initially used was available, then a variant of the organism could develop which would possess the ability to convert the related compound to that originally used. Horowitz proposed this
hypothesis in terms of biosynthetic pathways, but the argument is equally valid for dissimilarity pathways. The Horowitz hypothesis was further developed in 1951 by Lewis who proposed a two-step mechanism by which retrograde evolution could occur. The gene required to encode the new metabolic function could arise by duplication of an existing gene followed by its mutation to new functions, while the first gene retained its original function. Such a theory would seem to be borne out by the study of methionine biosynthesis in \textit{E.coli} and threonine biosynthesis in \textit{Bacillus subtilis}. Cystathionine \( \gamma \)-synthase and \( \beta \)-cystathionase are encoded by the \textit{metB} and \textit{metC} genes located at 88 and 65 minutes on the \textit{E.coli} chromosome, respectively (Belfaiza \textit{et al} 1986) The deduced amino acid sequence of \( \beta \)-cystathionase shows extensive sequence homology with Cystathionine \( \gamma \)-synthase which catalyses the proceeding step in the biosynthesis of methionine, suggesting that the \textit{metB} and \textit{metC} evolved from a common ancestral gene. Cloning and nucleotide sequencing of the \textit{B.subtilis} genes for threonine synthase (\textit{thrC}) and homoserine kinase (\textit{thrB}) showed the gene order to be different to that of the corresponding \textit{E.coli} genes implying independent construction of the \textit{thr} operon in the two species. However, analysis of the predicted amino acid sequences encoded by \textit{thrB} and \textit{thrC} in the same species and between species, revealed extensive homologies between the enzymes. The observed sequence homology which correlates with similarities in the catalytic mechanisms of the enzymes, was taken as an indication that these proteins have evolved from a common ancestral precursor (Parsot 1986).

Although a theory based on retrograde evolution may explain some instances of pathway formation, it would seem impossible to advocate it in all circumstances. This theme was taken up by Dagley (1975) who pointed out that not all the reactions that occur in a pathway, involved similar reactions. In aromatic degradative pathways for example, it is common to find
dehydrogenation and ring-fission reactions which would seem to have unrelated catalytic mechanisms. From these types of arguments other theories of gene evolution were put forward. Riley and Anilionis (1978) proposed that new genetic material may be produced by chromosome duplication. They hypothesized that the \textit{E.coli} chromosome has undergone two duplicative events and that the additional DNA was then available for mutation to produce novel catalytic activities. Thus new catalytic activities could be utilised by the recruitment of genes from different parts of the chromosome to give the required metabolic activity after appropriate mutation. This type of process has been shown to occur for \textit{E.coli} K12 with mutations in \textit{lacZ} encoding β-galactosidase (Cambell \textit{et al} 1973). Mutants of K12 with a \textit{lacZ} mutation were grown under intense selection for growth on lactose until a mutant able to grow on lactose was selected. The mutant strain produced a form of β-galactosidase which had different kinetic, immunological and sedimentary characteristics to the \textit{lacZ} enzyme. The gene encoding the new activity was mapped to minute 59 on the \textit{E.coli} chromosome whereas \textit{lacZ} maps to minute 10.

The two mechanisms outlined above are not the only way in which an organism could obtain the genes necessary to degrade a novel compound. There is abundant evidence that in the laboratory at least, genetic transfer and exchange can occur by a variety of mechanisms: plasmid-mediated conjugation, phage infection and transduction, conjugative transposition of mobile genetic elements and transformation by unassociated DNA (Ochman & Wilson 1987). The degree to which these factors play a role in the transfer of new catalytic activities in the environment is uncertain. However, it should be noted that several metabolic pathways such as the TOL pathway, are found to occur predominantly on naturally occurring plasmids. Thus there is a potential
for transfer of individual genes or entire gene clusters between species by these types of mechanisms outlined above.

b. CONSERVATION OF PATHWAYS IN SEVERAL SPECIES.

The analysis of metabolic pathways offers a system in which evolutionary differences can be probed, since it is possible to investigate differences in the same pathway from different species. The TOL pathway has been a major area of study, not least because of its importance as a major degradative pathway, but also because several naturally-occurring variants of the TOL plasmid have been identified. Keil et al.(1985) described a TOL plasmid pWW53 from P.putida MT53 that differs significantly in its restriction endonuclease digestion pattern and incompatibility group from the archetypal TOL-plasmid pWW0. However, on closer examination of the region encoding the meta-pathway genes, both the gene order and restriction map of the pWW53 genes was very similar to that of the pWW0 genes suggesting a high degree of conservation of the catabolic operons in the two plasmids. A similar story was true when the naturally-occurring plasmid pWW60-22 encoding genes of a meta-pathway for the degradation of salicylate was compared to the TOL pathway genes (Assinder and Williams 1988).

The β-ketoadipate pathway of P.putida and A.calcoaceticus has been a model system for the study of evolutionary homologies in degradative pathways. Although the two species are widely divergent there would appear to be a high degree of homology between the individual enzymes. Pairwise comparisons of the γ-carboxymuconolactone decarboxylase and muconolactone isomerases from the two species show approximately 50% sequence identity in their amino termini (Yeh et al 1980, Yeh et al 1978). Such a high degree of conservation is not seen in all the pathway enzymes, however. In A.calcoaceticus two forms of enol lactone hydrolase are encoded whose
expression is dependent on which branch of the pathway is induced. Although comparison of the genes for the *A. calcoaceticus* hydrolases shows wide divergence, comparison of the amino terminal sequences of the corresponding single enzyme isoform from *P. putida* provides convincing evidence that all three hydrolases evolved from a common ancestor. Convergent evolutionary forces appear to have played little part in the development of the hydrolases. As such, the homologies seen are due to a common ancestor and not selection of the lactone binding site (Ornston & Yeh 1981). Perhaps a more interesting comparison can be made between the decarboxylases and isomerases from the two species. The interatomic electron rearrangements that occur upon decarboxylation and isomerisation in this pathway are analogous, so it is possible that selection of a common active site led to the evolution of the different enzymes from a common ancestor. Four pairwise comparisons of decarboxylase and isomerase sequences can be made and reveal a level of common identity comparable to that found between the hydrolase sequences. Hence it appears that these proteins have diverged widely from a common ancestor.

This is not the end of the story, since comparison of the isomerase and hydrolase shows significant homologies. Optimal isomerase/hydrolase identities appear in different alignments of the sequences. Given the generally wide divergence of sequence when the decarboxylase is taken into account, this finding is taken as indicating that multiple genetic events were required to create the sum of the observed isomerase/hydrolase homologies (Ornston and Yeh 1981). More recently the nucleotide sequences of the protocatechuate 3,4-dioxygenase enzymes from both *A. calcoaceticus* and *P. putida* have become available (Harnett *et al* 1990). This enzyme has two subunits which appear to share a common ancestry in both species. The ring fission event in the catechol branch of the β-ketoadipate pathway is catalysed by catechol 1,2-
oxygenase which is a single subunit type enzyme. The catechol oxygenase amino acid sequence resembles the iron binding β subunit of protocatechuic dioxygenase more than the α subunit, with conservation being particularly high in regions encoding certain tyrosyl and histidyl residues that have been implicated in iron binding (Ohlendorf et al. 1989). In some regions of the sequence alignment between the oxygenase subunits the DNA sequences appear to be conserved at a level beyond that which may have been demanded at the level of the protein. In other regions divergence of the aligned sequence appears to have been achieved by the substitution of DNA sequence from one genetic segment to another. This is proposed to have been achieved by exchange of sequence during slippage of complementary DNA sequences. Mismatch repair between such slipped strands could then contribute to the short stretches of conserved sequence seen today (Hartnett et al. 1990, Neidle et al. 1988).

The amino acid sequences of several ring-fission enzymes from both ortho and meta pathways indifferent species are now known. This prompted Harayama & Rekik (1988) to propose that dioxygenases that catalyse the cleavage of the aromatic ring, are classified into two different gene families: intradiol enzymes which cleave the ring between two hydroxyls and extradiol enzymes which cleave to the side of one of the hydroxylated carbons. Although the chemical structures of the substrates for both intradiol and extradiol oxygenases are similar, and in some cases identical, (eg. catechol can degraded by either an ortho or meta pathway.) the reaction mechanisms are completely different between these types of enzymes. Extradiol enzymes are composed of a single polypeptide species and contain ferrous iron as a prosthetic group, whereas intradiol enzymes are composed of either one or two subunits and contain ferric iron at their catalytic site. Moreover a statistical comparison of similarities between the amino acid sequences of intradiol and
extradiol dioxygenases, "did not show any significant global or localised similarity." However, this generalised conclusion is contradicted by a recent paper describing the cloning of the protocatechuate 4,5-dioxygenase genes of *Pseudomonas paucimobilis* (Noda *et al* 1990). Protocatechuate 4,5-dioxygenase catalyses the extradiol cleavage of protocatechuate, but does not have any sequence similarities with the extradiol dioxygenases identified by Harayama & Rekik (1988). Comparison to the intradiol dioxygenases, which have a similar subunit composition, also showed no significant degree of homology.

The theme of conservation of amino acid sequences between enzymes involved in similar reactions is not restricted to degradative pathways. For example, Woods *et al.* (1988) have reported considerable degrees of sequence homology between enzymes involved in trans-elimination reactions involving fumarate. Aspartase, fumarase (class II) and argininosuccinase are homotetramers that catalyse such reactions to give fumarate. Comparison of their amino acid sequences indicates that five regions in all the enzymes are conserved, suggesting that the enzymes have arisen from a common evolutionary origin.

1.6 **DETAILED ANALYSIS OF KNOWN AROMATIC DEGRADATIVE PATHWAYS.**

a. The TOL pathway.

The majority of molecular studies of aromatic degradative pathways have involved the plasmid borne genes of the TOL pathway. The location of such genes on naturally occurring plasmids has facilitated such study since it has proved easier to isolated the genes concerned on restriction fragments that could be further manipulated. The ability of *Pseudomonas* to dissimilate benzoate via catechol using an ortho fission pathway has been known for some time (Stanier 1947). Phenol and methyl substituted benzene is degraded via
benzoate using a meta-cleavage pathway indicating that there are two pathways for benzoate metabolism. Although benzoate is a growth substrate for strains carrying the TOL plasmid, it supports a faster rate of growth when utilised by the chromosomally encoded β-ketoadipate pathway. This pathway is found in all *P. putida* and *P. aeruginosa* strains but is not expressed when the TOL plasmid is present. Growth on benzoate provides a nutritional selection for faster growing strains with a nonfunctional plasmid encoded meta-pathway. The benzoate induced loss of the meta-pathway genes is characterised by either loss of the entire TOL plasmid or more commonly by the precise excision of a 39 Kbp segment leaving a residual 78 Kbp plasmid. This residual plasmid does not express any of the meta-pathway genes suggesting that they are located in the excised 39 Kbp segment (Jeenes & Williams 1981). The plasmid nature of the meta pathway was demonstrated in mating experiments with *Pseudomonas* mutants by Williams & Murray (1974) and Wong & Dunn (1974). These studies confirmed the benzoate induced loss of the meta pathway genes and also demonstrated the transmissibility of the pathway genes. Subsequently the plasmid, now designated TOL was isolated and was shown to specify the meta pathway enzymes and a toluate dioxygenase required to convert p- or m-toluates to methylcatechols. Other aromatic chemicals have been found to be substrates including m- and p-xylenes (Worsley and Williams 1975), 3-ethyltoluate, and 1,2,4-trimethylbenzene (Kunz and Chapman 1981). Extensive cloning of the TOL genes has helped elucidate the architecture of the gene cluster. There are two gene clusters, one specifying the three genes of the upper pathway enzymes, the other containing the eight genes for the lower pathway (see fig.1.4.). Restriction mapping of the archetypal TOL plasmid pWW0 (Downing and Broda 1979) followed by cloning and transposon mutagenesis (Franklin et al 1981) determined that the two gene clusters were separated by 14 Kbp. Inouye
et al (1981) cloned the lower pathway xylDEGF genes and determined their gene order. It was also demonstrated that the product of the regulatory gene xylS was required in addition to the lower pathway precursors m-toluate or benzoate, for expression of this gene cluster, indicating a positive transcriptional control mechanism. The number of genes and the gene order was extended to xylDLEGFJIH with xylK encoding the last enzyme of the pathway probably between xylJ and xylII, by Harayama et al 1984. This gene order takes into account the genes required to encode the enzymes which participate in the branching of the lower pathway, where the ring-fission semialdehyde can be further metabolised via one of two routes (Fig.1.6.a.).

The gene order of the upper pathway genes was found to be xylCAB by Harayama et al 1986. The expression of the upper pathway genes was also found to be positively controlled by the action of the xylR gene product in association with the pathway inducers toluate and its alcohol derivatives. In general, there seems to be no logical correlation between the gene order and the order in which the individual reactions occur. When the regulation of expression of the genes (as discussed earlier) is taken into account it seems that control of the metabolism of substrates for the pathway is achieved by coordinated control of gene clusters and not by controlling the synthesis of the first enzyme of the pathway followed by the other pathway enzymes in order.

Although the TOL pathway in a variety of forms, is found encoded on naturally-occurring plasmids, Sinclair et al (1986) reported a chromosomal location of the genes in one P.putida strain. Moreover, the chromosomal genes were shown to be able to move from the chromosome to a naturally occurring plasmid in a transposon like manner. Such a property may yield yet another evolutionary dimension to the development of such degradative pathways.
The xylene/toluene degradative pathway encoded by the pWW0 TOL plasmid taken from Frantz & Chakrabarty (1986).

xyl op1 represents the first gene cluster regulated as an operon who’s products comprise the upper pathway. xyl op2 represents the second gene cluster regulated as an operon, which gives rise to the lower pathway. The two operons are separated by approximately 14 Kbp on the TOL plasmid. Enzyme abbreviations: XO, xylene oxygenase; BADH, benzyl alcohol dehydrogenase; BZDH, benzaldehyde dehydrogenase; TO, toluate dioxygenase; DHCDH, dihydroxycyclohexadiene carboxylate dehydrogenase; C23O, catechol 2,3 dioxygenase; HMSN, hydroxymuconic semialdehyde hydrolase; HMSC, hydroxymuconic semialdehyde dehydrogenase; 4-OI, 4-oxalocrotonate isomerase; 4-OD, 4-oxalocrotonate decarboxylase; OEH, 2-oxopent-4-enoate hydratase; HOA, 2-oxo-4-hydroxy-pentonate aldolase. Initial compounds: R,R’=H, toluene; R=H, R’=CH₃, m-xylene; R=CH₃, R’=H, p-xylene. xylA to xylZ, genes encoding the pathway enzymes.
The xylene/toluene degradative pathway encoded by the pWW0 TOL plasmid. From Frantz and Chakrabarty (1986).
Many *Pseudomonas* species utilise naphthalene as a sole source of carbon through the formation of salicylate as an intermediate (Frantz & Chakrabarty 1986). Other species can utilise salicylate but not naphthalene, generally by *meta* pathway enzymes encoded on naturally-occurring plasmids such as SAL1 (Yen *et al* 1983) and pMWD-1 (Zunniga *et al* 1981). The NAH7 (Dunn & Gunsalus 1973) and pWW60-1 (Cane & Williams 1986) plasmids encoded additional enzymes to degrade naphthalene via salicylate. The genes encoding the two pathways on NAH7 are arranged in a similar manner to the TOL genes i.e. in two clusters separated by 14 Kbp (Yen & Gunsalus 1982). There are several other similarities between the TOL pathway and the pWW60-1 encoded pathway, both are positively regulated and the restriction maps of the catechol 2,3-oxygenase genes are similar and it has been speculated that the two pathways share related *meta*-pathway DNA (Cane & Williams 1986). Harayama *et al* (1987) reported that the gene order of the NAH7 plasmid *meta*-pathway genes is identical to the order of the isofunctional genes of the TOL plasmid pWW0. The nucleotide sequence of *nahH* encoding catechol 2,3 oxygenase from the naphthalene degradative pathway was found to be homologous to the isofunctional *xylE* gene of the TOL pathway. The homology between the NAH7 and TOL plasmids ends upstream of the ribosome binding site for *nahH* and *xylE* but continues downstream of the genes. This suggests that the catechol oxidative enzymes of both plasmids are derived from a common ancestor which was transferred as a discrete segment of DNA between the plasmids. A more detailed analysis of the naphthalene degradative pathways may shed light on how these long catabolic pathways have developed.
b. The β-ketoadipate pathway.

Catabolic degradation via ortho fission of the aromatic ring has mainly been studied in the chromosomally encoded β-ketoadipate pathway. The β-ketoadipate pathway is a ortho-ring fission pathway for a variety of compounds which converge on catechol or protocatechuate. The pathway has two arms as depicted in Fig.1.5.a. which converge after the formation of enol-lactone structures. Although this pathway has been conserved among the various organisms studied, the regulatory mechanisms governing the expression of the pathway enzymes have generally diverged widely. The regulation of the pathway genes in *P. putida* and *P. aeruginosa* is similar. The catechol branch of the pathway is induced by cis, cis muconate which is the product of the action of catechol 2,3-oxygenase on catechol. Cis, cis muconate is further metabolised by muconate lactonising enzyme and muconolactone isomerase which are coordinately induced by muconate. The protocatechuate arm of the pathway is controlled slightly differently. Protocatechuate 3,4- dioxygenase is induced by protocatechuate or *p*-hydroxybenzoate (Zylstra et al. 1989). The remaining enzymes of the protocatechuate arm are coordinately induced by β-ketoadipate. This is shown digramatically in Fig.1.6.b.

The first published report of chromosomally encoded genes for an aromatic pathway being cloned was for the *catBCDE* cluster from the β-ketoadipate pathway of *A. calcoaceticus* (Shanley et al. 1986). This segment was known to encode only a part of the genes required and was used to probe a gene library for flanking regions of DNA which might encode the other genes. A region encoding *catA* was isolated in this way by Neidle & Ornston 1987. Subsequently the genes required for the conversion of benzoate to catechol (*benABCD*) were cloned from *A. calcoaceticus* (Neidle et al. 1987).
Fig.1.6.b.

The β-ketoadipate pathway and its regulation in *P. aeruginosa* and *P. putida*. Brackets indicate coordinately induced enzymes.

In contrast to *P. putida*, *A. calcoaceticus* elaborates type isofunction enol lactone hydrolases depending on which arm of the pathway is induced. Enol lactone hydrolase I is coordinately induced with carboxymuconolactone decarboxylase by the metabolic precursor protocatechuate. Enol lactone hydrolase II is coordinately induced with muconolactone isomerase by the intermediate *cis, cis* muconate.
Fig. 1.6.b.

\( \beta \)-ketoadipate pathway regulation in *P. aeruginosa* and *P. putida*. Taken from Zystra et al. 1989.
The *ben* genes were found to be clustered on the *A.calcoaceticus* chromosome in a region near to the independently regulated *cat* genes. In a 16 Kbp region all the genes required for the complete degradation of benzoate were found, controlled in three transcriptional units. Such an arrangement is termed supraoperonic clustering and is a feature also found in *Pseudomonads* (Holloway & Morgan 1986).

Six *pca* genes encode the enzymes that convert protocatechuate to acetyl-CoA. Cloning of the genes for this arm of the pathway from *Pseudomonas cepacia* (Zylstra et al 1989) and from *P.putida* (Hughes et al 1988) has been reported. The gene order of the *pca* genes cloned from *P.putida* is different to that of the genes from *A.calcoaceticus* (Doten et al 1987). In *P.putida* β-keto-adipate induces the expression of four of the *pca* structural genes. The genes cloned from this organism by Hughes *et al* (1988) were clustered into an operon with the order *pcaBDC*. In addition the *pcaE* gene was found to lie more than 15 Kbp away from the *pcaBDC* operon. The positive regulator (in response to β-keto-adipate) encoded by *pcaR* was found to be located 15Kbp upstream from the *pcaBDC* cluster. In contrast, the gene order elucidated by Shanley *et al* (1986) for the *A.calcoaceticus* genes, which was extended to include the genes coding for protocatechuate 3,4-dioxygenase by Harnett *et al* (1990) was found to be *pcaEFDBCHG*. A considerable degree of gene shuffling must therefore have occurred in the *pca* genes after the divergence of *P.putida* and *A.calcoaceticus*.

There are differences in the regulatory patterns between species and between arms of the pathway as already discussed. But as more of the cloned genes are analysed several areas of interest come to light. It has been reported by Allewell (1989) that the evolutionary homologue of the gene encoding a transcriptional activator in one species (*catR* gene of *P.putida*) encodes a
repressor in another strain (\textit{catM} in \textit{A. calcoaceticus}). Stanier and Ornston (1973) point out that differences in regulatory mechanism tend to be associated with differences in the organisation of the genes in operons such that selection occurs first at the level of catalysis and then at the level of regulation. The discovery of a negative regulator controlling the expression of an aromatic degradative pathway is unusual since most of the systems studied to date (TOL, \(\beta\)-ketoadipate, NAH) seem to be regulated by positive acting regulators exclusively. The only other published report of a negatively controlled system is the homoprotocatechuate pathway of \textit{Escherichia coli} C (Jenkins & Cooper 1988; see section 1.7.

c. Crystallographic analysis of enzymes of the \(\beta\)-ketoadipate pathway.

Analysis of the \(\beta\)-ketoadipate pathway has been taken a step further in the last decade by the determination of crystal structures for several of the enzymes. This offers an opportunity to see how homologies and differences at the level of amino acid and nucleotide sequence influence protein structure. The enzymes involved are muconate lactonizing enzyme and muconolactone isomerase from \textit{P. putida}, protocatechuate 3,4-dioxygenase from \textit{P. aeruginosa} and dienelactone hydrolase from \textit{Pseudomonas sp} B13. Of the four enzymes the hydrolase seems to have the least in common with the other three enzymes. The crystal structure of the hydrolase (Pathak \textit{et al} 1988) shows the active site to consist of a catalytic triad of cysteine, aspartate and histidine, resembling those of the serine and cysteine proteases. There is also a reported low level of topological similarity to subtilisin. The other three enzymes have complex oligomeric structures. The structure of the muconate lactonizing enzyme (Goldman \textit{et al} 1987) shows it to be constructed of three domains, the central one of which bears similarity to a common feature reported in six other
enzymes including triose phosphate isomerase and pyruvate kinase. The active site of the muconate lactonizing enzyme which contains manganese was located to a cleft between the amino terminal of the enzyme and one of the barrel domains. Superimposing the structure of triose phosphate isomerase shows that different parts of the two enzymes carry the homologous active site residues. This implies that the two enzymes do not have a common ancestor and are linked by convergent rather than divergent evolution. Such a conclusion is arguable however, since the generally accepted view of common protein folds, is that they arose by divergent evolution from a common ancestor (Schultz 1980).

The isomerase crystal structure (Katti et al 1989) shows the subunits to adopt a relatively simple tertiary structure consisting of two α-helices and an antiparallel twisted β-sheet of four strands which ends in an extended arm. The putative active site is thought to be formed in a hydrophobic core formed by the fold of the α-helices and β-sheets. The subunits of the isomerase are thought to associate to form two stacked pentamers with the active sites occurring between subunits.

The dioxygenase enzyme also has a complex ternary structure (Ohlendorf et al 1988) consisting of 12 dimers of the α and β subunits which assemble to form a truncated tetrahedron. As with the isomerase, the active site is thought to occur between the subunits with the β subunit carrying the iron residue linked to specific tyrosines and histidines. These residues are also conserved in sequences homologues of protocatechuate 3,4-dioxygenase, which include catechol oxygenase from the other arm of the β-ketoadipoate and the isofunctional chlorocatechol oxygenase (Hartnett et al 1990).
1.7 HOMOPROTOCATECHUATE (HPC) DEGRADATION.

a. The HPC pathway.

The ability of an enteric organism to degrade aromatic compounds is unsurprising given the fact that aromatic amino acids such as phenylalanine and tyrosine occur in the intestinal-faecal environment (Spoelstra 1978). Parrot et al (1987) showed that \textit{E.coli} K12 can grow on 2-phenylethylamine as a sole carbon and energy source. In addition \textit{E.coli} has been shown to grow on phenylpropionate and 3-hydroxyphenylpropionate (Burlingame & Chapman 1983) as well as phenylacetate (Cooper \textit{et al.} 1985).

The metabolism of 4-hydroxyphenylacetic acid (4-HPA) by a species of \textit{Acinetobacter} and two species of \textit{Pseudomonas} was first described by Sparnins \textit{et al} (1974). The metabolism of 4-HPA proceeded through the dihydroxylated ring-fission substrate homoprotocatechuate (HPC; 3,4-dihydroxyphenylacetate) to yield pyruvate and succinate via a \textit{meta} fission pathway. Sparnins \textit{et al} found the metabolism of 3-HPA to occur via homogentisate in the \textit{Acinetobacter} species, however. HPC was identified by Adachi \textit{et al} (1964) as an intermediate in 4-HPA metabolism in \textit{Pseudomonas ovalis}. The following year HPC dioxygenase was crystallised from that organism by Kita \textit{et al} (1965). The product of the ring fission reaction was shown to be 5-carboxymethyl-2-hydroxymuconate semialdehyde (CHMS) in this study which was confirmed by Sparnins \textit{et al} (1974) who 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 along the pathway to pyruvate. A pathway for the dissimulation of HPC to pyruvate and succinate was proposed by Dagley (1975) which included an NAD\textsuperscript{+} dehydrogenation step followed by three keto-enol isomerisation steps (See Fig.1.7.1.). The first of these isomerisations was confirmed by the work of Garrido-Pertierra & Cooper (1981).
The possible intermediates in the catabolism of HPC as proposed by Dagley (1975).

I. HPC, homoprotocatechuate
   a. dioxygenase

II. CHMS, 5-carboxymethyl-2-hydroxymuconate semialdehyde
    b. dehydrogenase

III. CHM, 5-carboxymethyl-2-hydroxymuconate
     c. isomerase

IV. COHED, 5-carboxymethyl-2-oxo-hex-3-ene-1,6-dioate
    d. decarboxylase

V. HHDD, 2-hydroxyhepta-2,4-diene-1,7-dioate
   e. isomerase

VI. OHED, 2-oxo-hepta-3-ene-1,7-dioate
    f. hydratase

VII. HHED, 2,4-dihydroxy-hepta-2-ene-1,7-dioate
     g. isomerase

VIII. HKP, 4-hydroxy-2-ketopimilate
      h. aldolase
Fig. 1.71.
Pathway for the dissimilation of HPC as proposed by Dagley (1975)

\[ \text{Pathway diagram as described in the text.} \]
This metabolic route remained unchanged until analysis of the cloned genes for HPC degradation from *E.coli C* (Jenkins 1987). The availability of clones which could be used to convert HPC to selected intermediates of the pathway showed there were only two isomerisation steps before aldol fission occurred to give pyruvate and succinic semialdehyde. The substrate for the aldol fission reaction was shown to be 2,4-dihydroxy-hepta-2-ene-1,7-dioate (HHED) by NMR analysis of the isolated intermediate. The analysis also showed that HHED is the enol form of HKP. HKP was chemically synthesised by Jenkins and was shown to have the same λ-max (260nm) as the biologically produced HHED and served as a substrate for the aldolase enzyme. With this information a revised diagram of the pathway was produced and is presented in Fig. 1.7.2.

There are unmistakable similarities between the order of reactions and chemical intermediates which occur in the dehydrogenative route of the meta-cleavage TOL pathway and the HPC pathway. A recent paper by Harayama *et al.* (1989) identified several of the chemical intermediates in the dissimilation of catechol by this pathway. The pathway includes a single isomerisation step which converts the enol form of 4-oxalocrotonate to the keto form (Fig. 1.7.3.). This step which was originally thought to be a tautomerisation (Sala-Trepat & Evans 1971). However, NMR analysis of the isolated intermediates by Harayama *et al.*, established that the reaction proceeds via an isomerisation and is therefore analogous to the isomerisation steps proposed originally by Dagley for the HPC pathway. Interestingly, there is no second isomerisation step after the decarboxylation, which is therefore different to the sequence of reactions in the HPC pathway proposed by Jenkins & Cooper (1988). Otherwise the nature of the chemical intermediate in the two pathways are very similar.
Fig. 1.7.2.

The intermediates in the HPC pathway as proposed by Jenkins (1987):

HPC, homoprotocatechuate;
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;
SSA, succinic semialdehyde
Fig.1.7.2.

Pathway for the degradation of HPC (Jenkins 1987)
The *meta*-cleavage pathway for catechol degradation as proposed by Harayama *et al* (1989).

Possible chemical structures of each intermediate are shown in the boxes. Enzyme abbreviations: C23O, catechol 2,3-dioxygenase; HMSD, 2-hydroxymuconate semialdehyde dehydrogenase; HMSH, 2-hydroxymuconate semialdehyde hydrolase; 4OI, 4-oxalocrotonate isomerase; 4OD, 4-oxalocronate decarboxylase; OEH, 2-oxopent-4-enoate hydratase; HOA, 4-hydroxy-2-oxovalerate aldolase. For compounds in which R1 and R2=H,

1=catechol,
2=hydroxymuconate semialdehyde,
3a=2-hydroxyhexa-2,4-diene-1,6-dioate
(enol form of 4-oxalocrotonate)
3b=2-oxohex-4-ene-1,6-dioate,
3c=2-oxohex-3-ene-1,6-dioate
4a=2-hydroxypent-2,4-dieneoate
4b=2-oxopent-4-enoate
4c=2-oxopent-3-enoate
5a=1,4-dihydroxypent-2-enoate
5b=4-hydroxy-2-oxovalerate
6=pyruvate
7=acetaldehyde
8=formate

The pathway was shown by Harayama *et al* (1989) to proceed via compounds 3c and 4a. However, it has not been shown whether compound 5a or 5b is the product of the hydratase.
Fig. 1.7.3.
The meta-cleavage pathway for catechol degradation according to Harayama et al (1989).

R₂ - CH₂CHO + CH₃COOH
The ability of *E. coli C* to grow on 3-HPA and 4-HPA was reported by Skinner and Cooper (1980). The genes required for the degradation of HPC to pyruvate and succinate semialdehyde (SSA) along with a gene encoding a regulatory protein were cloned from *E. coli C* by Jenkins (1987). These cloned genes form the basis for Fawcett's work on similarities between the *E. coli* and *Klebsiella pneumoniae* HPC genes and also form the basis for this work. The HPC pathway of *E. coli* represents the longest and most studied aromatic degradative pathway in *E. coli*.

Studies of induction of *hpc* genes in *P. ovalis* (Adachi *et al* 1964) and in *P. putida* (Barbour and Bayly 1977) showed that HPA induces both the HPA hydroxylase and the HPC catabolic enzymes whereas HPC induces only the HPC catabolic enzymes. From this data the genes were thought to be organised in at least two regulatory groups. The *hpa* and *hpc* genes of *E. coli C* were shown by Skinner (1981) to be over 95% co-transducible in P1 transduction experiments, suggesting that the two regulatory groups were very closely linked. The set of genes responsible for the degradation of HPA can be considered in terms of at least three regulatory groups:

**Group 1:** 3-HPA and 4-HPA uptake and hydroxylation. Consisting of *hpaA* (permease), *hpoB* (hydroxylase) and *hpaR* (regulatory protein)

**Group 2:** HPC to pyruvate catabolic group. This group of genes, designated *hpcB* through *hpcH* are responsible for the complete dissimilation of HPC to pyruvate and succinic semialdehyde. This group of genes are inducible by both HPA and HPC: as HPC can be transported into the cell and induces the gene for its own degradation this group represents a complete pathway.
Group 3: Succinic semialdehyde dehydrogenase. The oxidation of SSA to succinate is achieved by an NAD-linked dehydrogenase encoded by the sad gene, which is induced by SSA (Donnelly & Cooper 1981, Marek & Henson 1988).

Skinner (1981) proposed from a study of mutants of the 4-HPA pathway, that the product of a single regulatory gene controlled both 4-HPA and HPC degradative genes. A study of the cloned genes from E.coli C by Jenkins (1987) showed that in the absence of the regulatory gene expression of the HPC genes was constitutive. Thus Jenkins proposed that the expression of these genes was negatively controlled in contrast to the majority of other aromatic degradative pathways that are positively controlled.

b.Analysis of the cloned E.coli C Hpc genes.

E.coli strains B,C and W are known to degrade 4-HPA. The presence of the HPC catabolic enzymes in E.coli K12 was tested by Skinner (1981) and they were undetectable. This result was confirmed by hybridisation studies against chromosomal DNA using synthetic oligonucleotide probes for hpcG (Ferrer & Cooper 1988) and hpcC (Fawcett et al 1989). Similar results were obtained by Jenkins (1987) using the cloned E.coli C DNA as a probe. E.coli K12 strains do carry a copy of the sad gene encoding the NAD-linked succinic semialdehyde dehydrogenase and it has been suggested (Marek & Henson 1988) that perhaps sad is a vestigial gene that was retained when K12 strains lost their ability to use 4-HPA as a growth substrate. There is no known requirement for the sad gene in K12 strains other than it may be a selective advantage in allowing the removal of succinic semialdehyde which is toxic to E.coli. Derivatives of K12 therefore make good hosts for study of the E.coli C cloned genes since they are a genetically blank background. The most commonly used strain for the study of these genes has been E.coli 5K which is an hsdR derivative allowing efficient
transformation of DNA from the naturally hsdM E.coli C strain. Recently the HPC catabolic genes have been introduced into 5K by P1 transduction to produce a derivative of 5K which constitutively expresses the hpc genes and is able to grow on 4-HPA as a sole carbon source (R.A. Cooper: pers comm). This strain has been useful in assessing the location of the cloned hpcR gene.

Jenkins isolated three genomic subclones which carry all or part of the genes required to degrade HPC (see Fig.1.7.4.). E.coli C chromosomal DNA was digested in partial digests with Sau3A1 to produce fragments of various sizes which were cloned into BamH1 digested pBR322 or pBR328. Two of the genomic subclones pJJ200 and pJJ210 which were constructed in pBR322 carry all of the hpc catabolic genes as well as the hpcR regulator gene. Expression of the cloned hpc catabolic genes required the presence of the inducer 4-HPA during growth of the transformed bacteria. The third clone, pJJ801 carried the genes required to complete the first five reactions of the HPC pathway. Both the OHED hydratase and HHED aldolase could not be detected in cell extracts made from cultures harbouring pJJ801.

In addition expression of the genes from pJJ801 was found to be constitutive. The constitutive nature of expression from pJJ801 was investigated by testing their expression in a derivative of E.coli C: E.coli C221 which does not express an active CHMS dehydrogenase (hpcC). The presence of 4-HPA is required for expression in these cultures indicating that pJJ801 carries a region to which a host encoded-regulatory protein can bind in trans.

pJJ801 was constructed using pBR328 as the vector and constitutively expresses hpcBCDEF. Both pJJ200 and pJJ210 were constructed using pBR322 as the vector and express all the Hpc pathway enzymes in a regulated manner.
Fig. 1.7.4.

The three geneomic clones isolated by Jenkins (1987)
These experiments led Jenkins to propose a negative model for the regulation of the hpc genes. Further experiments by Jenkins (1987) implicated the role of cAMP-CRP as a positive element in the regulation system. The HPC catabolic enzymes of both E.coli C and E.coli 5K (pJJ210) were shown to be subject to glucose repression which could be partially relieved by the addition of 5mM cAMP to the growth medium.

Jenkins used a variety of restriction sites and fragments created by partial Sau3A1 digestions to position the cloned genes with respect to each other and thus assign a gene order. The catabolic genes were arranged in two blocks hpcBCDEF and hpcGH whose expression was controlled by a pair of centrally arranged divergent promoters such that the gene blocks were transcribed in different directions. The arrangement of transcriptional units in this way is not an uncommon form of gene organisation (Beck & Warren 1988). The regulatory gene hpcR was placed downstream of the hpcGH gene cluster. A model for the organisation of the hpc genes was proposed on this basis and is detailed in Fig.1.7.5.

Investigations of the gene order and direction of transcription of the cloned hpc genes by Fawcett (1989) highlighted notable differences to the model proposed by Jenkins. Firstly the gene order did not follow the order of the reactions as proposed by Jenkins. The direction of transcription was found to be the same for all the catabolic genes and was from right to left as drawn. Jenkins identified a PvuII site to lie within the hpcF gene since subclones which ended in that site did not apparently produce an active HHDD isomerase enzyme. In an experiment which duplicated the cloning experiments of Jenkins, it was found that an active HHDD isomerase was produced from such a restriction fragment. In addition experiments to order hpcE and hpcF which were known to occur in this region were unsuccessful.
Fig. 1.7.5.

Model for the organisation of the Hpc pathway according to Jenkins (1987).

The genes are arranged in two operons: *hpcGHR* and *hpcBCDEF*, with transcription originating from two divergent promoters. The two operons were thought to be negatively regulated with 3,4-DHPA (HPC) as the inducer. The intermediates in the pathway are: 3,4-DHPA (HPC), 3,4-dihydroxyphenylacetate, (homoprotocatechuate)

CHMS, 5-carboxymethyl-2-hydroxymuconate semialdehyde;
CHM, 5-carboxymethyl-2-hydroxymuconate;
COHED, 5-carboxymethyl-2-oxo-hex-3-ene-1,6-dioate;
HHDD, 2-hydroxhepta-2,4-diene-1,7-dioate;
OHDH, 2-oxo-hepta-3-ene-1,7-dioate;
HHED, 2,4-dihydroxy-hepta-2-ene-1,7-dioate;
SSA, succinic semialdehyde.
Fig. 1.7.5.

Model of the organisation of the $hpc$ genes as proposed by Jenkins (1987)
At the same time experiments conducted in this study showed that \textit{hpcD} encoding CHM isomerase was not associated with \textit{hpc(EF)} as was originally thought. Fawcett independently confirmed this result and proposed a new gene order on the basis of the apparent inducibility of \textit{hpcB} encoding HPC dioxygenase during growth in minimal media cultures. This finding was taken as an indication that there were at least two transcriptional units on pJJ801. Fawcett conducted Northern blotting experiments using \textit{E.coli} C mRNA and mRNA produced by \textit{E.coli} 5K(pJJ200) probed with the whole \textit{BamH1} fragment of pJJ200 and detected three transcripts of lengths 4.5, 2.7 and 1.6 Kbp in size. The two larger hybridising bands were thought to be due to the structural gene transcripts while the smaller band was due to a transcript from \textit{hpcR}.

The position of the 5' regions of \textit{hpcC} and \textit{hpcG} were confirmed by Southern blotting and nucleotide sequencing. Using this information the gene order was assigned by Fawcett as \textit{hpc(EF)CD, hpcBGH} with transcription of the two operons in the same direction. This is summarised in Fig.1.7.6.
Fig.1.7.6.

Model for the organisation and expression of the *E.coli* C *hpc* genes according to Fawcett (1989).

I.- In the presence of glucose in the medium: The regulatory gene, *hpcR*, is transcribed and the product of that gene binds to the two operator sites of the *hpc* operon and prevents transcription of the structural genes.

II- In the absence of glucose and the absence of an inducer of the *hpc* genes: Transcription of the *hpcR* gene and the binding of the regulatory protein to the operator site occurs as in I.

III- In the absence of glucose, but in the presence of a specific inducer of the *hpc* genes: The *hpcR* gene is transcribed and a mature regulatory protein is produced. However, the inducer in some way interferes with the binding of the regulatory protein to the operator site; thus the two blocks of structural genes *hpcFECD* and *hpcBGH* are transcribed and mature proteins produced.
Model for the order of the *hpc* genes according to Fawcett (1989)
1.8 AIMS.

Two enzymes of the HPC pathway: CHM isomerase and HHDD isomerase, catalyse the isomerisation of structurally similar intermediates of the pathway but are apparently specific for their own substrates. It is possible that they may share some degree of sequence homology, given their structurally similar substrates and the close physical location of their genes. The two isomerisation events are separated in the pathway by an enzyme-catalysed decarboxylation reaction of the intermediate COHED. The electron rearrangements that occur upon decarboxylation are similar to those that occur upon isomerisation. An similar analogy has been drawn for the decarboxylase and isomerase enzymes of the β-ketoadipoate pathway.

These possibilities were to be studied by nucleotide sequencing of the genes. This was supplemented by purification of the individual enzymes followed by amino terminal sequencing to positively identify the open reading frames encoding them. An order of priority was set with regard to the nucleotide sequencing, such that the genes specifying the isomerases (\textit{hpcF} and \textit{hpcD}) were to be sequenced first. Sequencing of \textit{hpcE} encoding COHED decarboxylase was then to follow along with any of the other \textit{hpc} genes that were available.

Such a project would initially require more precise information on the position of the \textit{hpc} genes to facilitate the nucleotide sequencing strategy. As a consequence of these studies, it was possible to purify large quantities of several of the proteins which has resulted in preliminary crystallographic analysis of them.
CHAPTER 2.

METHODS AND MATERIALS.
2.1. BIOCHEMICAL METHODS.

2.1.1. BACTERIAL STRAINS.

<table>
<thead>
<tr>
<th>STRAIN:</th>
<th>GENOTYPE:</th>
<th>SOURCE:</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td>Prototroph</td>
<td>Lab stock</td>
</tr>
<tr>
<td><em>E.coli C221</em></td>
<td><em>hpcC, recA</em></td>
<td>Jenkins, 1987</td>
</tr>
</tbody>
</table>
| *E.coli 5K*        | *F*, *supE44, tonA21*
|                    | *hsdR, rpsL, thr-1, leu-B6, thi-1, λ* | Lab stock (Huback and Glover, 1970.) |
| *E.coli 5K (hpa+)* | As above except *hpa+* | R.A.Cooper pers com |
| *E.coli DH5α*      | *F-*, *lacZM15, endA1*
|                    | *recA1, hsdR17(ri^m), supE44, thi, λ, gyrA96*
|                    | *relA1?, (lacZYA-argF), U169* | Lab stock (GIBCO-BRL) |
| *E.coli NM522*     | *supE, thi, hsd5, (lac-proAB),[F', proAB lacIqZM15]* | Lab stock |
| *E.coli JM101*     | *thi,(lac-proAB),[F', traD36, proAB, lacIqZM15]* | Lab stock |
| *E.coli JM105*     | *thi, strA, endA, shcB15*
|                    | *hsdR4, (lac-proAB),[F', traD36, proAB, lacIqZM15]* | Lab stock |
| *E.coli JM109*     | *recA1, endA1, thi, gyrA96*
|                    | *hsdR17, sup44, relA1, lac-proAB),[F',traD36, proAB, lacIqZM15]* | Lab stock |

2.1.2. PLASMIDS.

a. Plasmid vectors:

<table>
<thead>
<tr>
<th>VECTOR</th>
<th>GENES ENCODED</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBR322</td>
<td><em>ApR, TcR</em></td>
<td>Soberon et al., 1980</td>
</tr>
<tr>
<td>pBR328</td>
<td>*ApR, TcR, CmR</td>
<td>Soberon et al., 1980</td>
</tr>
<tr>
<td>pUC18/pUC19</td>
<td><em>ApR, lacZ</em></td>
<td>Yanisch-Perron et al., 1985</td>
</tr>
<tr>
<td>pKK223-3</td>
<td><em>ApR</em></td>
<td>Brosius &amp; Holy 1984</td>
</tr>
</tbody>
</table>
b. Genomic clones containing \textit{hpc} genes of \textit{E.coli C}.

<table>
<thead>
<tr>
<th>CONSTRUCT:</th>
<th>VECTOR:</th>
<th>GENES ENCODED:</th>
<th>SOURCE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJJ200</td>
<td>pBR322</td>
<td>\textit{hpcBCDEFGHR}</td>
<td>Jenkins, 1987</td>
</tr>
<tr>
<td>pJJ210</td>
<td>pBR322</td>
<td>\textit{hpcBCDEFGHR}</td>
<td>Jenkins, 1987</td>
</tr>
<tr>
<td>pJJ801</td>
<td>pBR328</td>
<td>\textit{hpcBCDEF}</td>
<td>Jenkins, 1987</td>
</tr>
</tbody>
</table>

c. Constructs created in this study.

<table>
<thead>
<tr>
<th>CONSTRUCT</th>
<th>VECTOR</th>
<th>DESCRIPTION</th>
<th>\textit{hpc GENES}</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDR1830</td>
<td>pUC18</td>
<td>3.0 Kbp \textit{BamH1-Sal1}</td>
<td>none, hpcB, hpcD</td>
</tr>
<tr>
<td>pDR1930</td>
<td>pUC19</td>
<td>3.0 Kbp \textit{BamH1-Sal1}</td>
<td>hpcE</td>
</tr>
<tr>
<td>pDR1835</td>
<td>pUC18</td>
<td>3.8 Kbp \textit{BamH1-Sal1}</td>
<td>none</td>
</tr>
<tr>
<td>pDR1935</td>
<td>pUC19</td>
<td>3.8 Kbp \textit{BamH1-Sal1}</td>
<td>hpcB</td>
</tr>
<tr>
<td>pDR337</td>
<td>pUC19</td>
<td>2.3 Kbp \textit{Nde1-Sal1}</td>
<td>hpcB</td>
</tr>
<tr>
<td>pDR334</td>
<td>pUC19</td>
<td>1.5 Kbp \textit{Kpn1-Sal1}</td>
<td>none</td>
</tr>
<tr>
<td>pDR14</td>
<td>pKK223-3</td>
<td>2.5 Kbp \textit{SalI-EcoRI}</td>
<td>hpcEF</td>
</tr>
<tr>
<td>pDRF1</td>
<td>pBR322</td>
<td>3.3 Kbp \textit{EcoRI-SalI}</td>
<td>hpcEF, hpcR</td>
</tr>
<tr>
<td>pDRF1'</td>
<td>pBR322</td>
<td>2.4 Kbp \textit{BamH1-SalI}</td>
<td>hpcE</td>
</tr>
<tr>
<td>pDRF2</td>
<td>pUC18</td>
<td>6.8 Kbp \textit{EcoRI-EcoRI}</td>
<td>none</td>
</tr>
<tr>
<td>pDR1865</td>
<td>pUC18</td>
<td>6.5 Kbp \textit{EcoRI-HindIII}</td>
<td>hpcBDGH</td>
</tr>
<tr>
<td>pDR336</td>
<td>pUC18</td>
<td>3.6 Kbp \textit{EcoRI-b'}</td>
<td>hpcBDG</td>
</tr>
<tr>
<td>pDR332</td>
<td>pUC18</td>
<td>3.2 Kbp \textit{EcoRI-b'}</td>
<td>hpcBDG</td>
</tr>
<tr>
<td>pDR1905</td>
<td>pUC19</td>
<td>0.5 Kbp \textit{EcoRI-SalI}</td>
<td>none</td>
</tr>
<tr>
<td>pDR1937</td>
<td>pUC19</td>
<td>2.7 Kbp \textit{HindIII-BamH1}</td>
<td>hpcEF</td>
</tr>
<tr>
<td>pDR1903</td>
<td>pUC19</td>
<td>0.3 Kbp \textit{BamH1-b'}</td>
<td>none</td>
</tr>
</tbody>
</table>

(b' = blunt end, no restriction site)

(d. Other constructs used

<table>
<thead>
<tr>
<th>CONSTRUCT</th>
<th>VECTOR</th>
<th>DESCRIPTION</th>
<th>\textit{hpc GENES}</th>
<th>SOURCE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJJ003</td>
<td>pBR322</td>
<td>4.0 Kbp \textit{EcoRI-BamH1}</td>
<td>\textit{hpcBCD}</td>
<td>Jenkins (1987)</td>
</tr>
<tr>
<td>pTF802</td>
<td>pUC18</td>
<td>2.5 Kbp \textit{BamH1-EcoRI}</td>
<td>\textit{hpcBD}</td>
<td>Fawcett (1989)</td>
</tr>
<tr>
<td>pTF830</td>
<td>pUC18</td>
<td>0.3 Kbp \textit{EcoRV-EcoRV}</td>
<td>none</td>
<td>Fawcett (1989)</td>
</tr>
</tbody>
</table>
DELETION SERIES SUBCLONES:
pDR93 deletion series
(Deletion series from the \textit{Sal}1 site of pDR1930)

<table>
<thead>
<tr>
<th>Subclone</th>
<th>Insert size (Kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9301</td>
<td>2.4</td>
</tr>
<tr>
<td>9302</td>
<td>2.25</td>
</tr>
<tr>
<td>9303</td>
<td>2.2</td>
</tr>
<tr>
<td>9304</td>
<td>2.05</td>
</tr>
<tr>
<td>9305</td>
<td>1.95</td>
</tr>
<tr>
<td>9306</td>
<td>1.9</td>
</tr>
<tr>
<td>9307</td>
<td>1.88</td>
</tr>
<tr>
<td>9308</td>
<td>1.70</td>
</tr>
<tr>
<td>9309</td>
<td>1.65</td>
</tr>
<tr>
<td>9310</td>
<td>1.62</td>
</tr>
<tr>
<td>9311</td>
<td>1.4</td>
</tr>
<tr>
<td>9312</td>
<td>1.35</td>
</tr>
<tr>
<td>9313</td>
<td>0.7</td>
</tr>
<tr>
<td>9314</td>
<td>0.8</td>
</tr>
<tr>
<td>9315</td>
<td>1.25</td>
</tr>
<tr>
<td>9316</td>
<td>0.9</td>
</tr>
<tr>
<td>9317</td>
<td>1.1</td>
</tr>
</tbody>
</table>

pDR335 deletion series
(Deletion series from the \textit{Bam}H1 site of pDR1835)

<table>
<thead>
<tr>
<th>Subclone</th>
<th>Insert size (Kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>83501</td>
<td>2.50</td>
</tr>
<tr>
<td>83502</td>
<td>2.30</td>
</tr>
<tr>
<td>83503</td>
<td>2.02</td>
</tr>
<tr>
<td>83504</td>
<td>1.84</td>
</tr>
<tr>
<td>83505</td>
<td>1.80</td>
</tr>
<tr>
<td>83506</td>
<td>2.02</td>
</tr>
<tr>
<td>83507</td>
<td>1.83</td>
</tr>
<tr>
<td>83508</td>
<td>1.46</td>
</tr>
<tr>
<td>83509</td>
<td>1.35</td>
</tr>
<tr>
<td>83510</td>
<td>0.81</td>
</tr>
<tr>
<td>83511</td>
<td>0.79</td>
</tr>
<tr>
<td>83512</td>
<td>0.78</td>
</tr>
<tr>
<td>83513</td>
<td>0.80</td>
</tr>
<tr>
<td>83514</td>
<td>0.82</td>
</tr>
<tr>
<td>83515</td>
<td>0.39</td>
</tr>
<tr>
<td>83516</td>
<td>0.41</td>
</tr>
<tr>
<td>35819</td>
<td>2.10</td>
</tr>
</tbody>
</table>

pDR358 deletion series
(Deletion series from the \textit{Sal}1 site of pDR1835)

<table>
<thead>
<tr>
<th>Subclone</th>
<th>Insert size (Kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35801/02</td>
<td>3.40</td>
</tr>
<tr>
<td>35803/4</td>
<td>3.35</td>
</tr>
<tr>
<td>35805</td>
<td>3.30</td>
</tr>
<tr>
<td>35806</td>
<td>3.20</td>
</tr>
<tr>
<td>35807</td>
<td>3.10</td>
</tr>
<tr>
<td>35808</td>
<td>3.05</td>
</tr>
<tr>
<td>35809</td>
<td>3.00</td>
</tr>
<tr>
<td>35810</td>
<td>3.10</td>
</tr>
<tr>
<td>35811</td>
<td>2.85</td>
</tr>
<tr>
<td>35812</td>
<td>2.80</td>
</tr>
<tr>
<td>35813</td>
<td>2.75</td>
</tr>
<tr>
<td>35814</td>
<td>2.00</td>
</tr>
<tr>
<td>35815</td>
<td>2.70</td>
</tr>
<tr>
<td>35816</td>
<td>3.60</td>
</tr>
<tr>
<td>35817</td>
<td>3.50</td>
</tr>
<tr>
<td>35818</td>
<td>2.95</td>
</tr>
</tbody>
</table>
The plasmid vectors pUC18 and pUC19.
(Yanisch-Perron et al 1985)

The pUC18 and pUC19 vectors are derived from a pBR322 ampicillin resistance gene and origin of DNA replication, ligated to a portion of the \textit{lacZ} gene encoding the \(\alpha\) fragment of \(\beta\)-galactosidase. The amino terminal portion of this gene contains a cluster of restriction enzyme sites termed the polylinker or multiple cloning site as detailed. The polylinker is oppositely orientated in the two vectors with respect to the \textit{lacUV-5} promoter. In pUC18 the \textit{lacUV-5} promoter initiates transcription in the \textit{EcoRI} to \textit{HindIII} direction. In pUC19 the direction of transcription is in the \textit{HindIII} to \textit{EcoRI} direction. The presence of the polylinker region allows insertional inactivation of the \textit{lacZ} gene allowing blue/white selection of clones as discussed in the text.
Fig.2.1.a.
The plasmid vectors pUC18 and pUC19.
(Yanisch-Perron et al 1985)
The expression vector pKK223-3 contains the strong tac (trp/lac fusion) promoter which in appropriate host strains is subject to regulation by the lac repressor and can be induced by the presence of isopropyl-β-D-thiogalactoside (IPTG) in the growth medium. This vector has a pBR322 derived ampicillin resistance gene and origin of replication and contains a pUC8 derived polylinker immediately after the tac promoter. However, due to the presence of SalI and BamH1 restriction sites in the vector only a portion of the polylinker can be easily used to ligate with restriction fragments carrying genes of interest. The strong rrn ribosomal terminator is found downstream of the polylinker to stabilise the vector-host system by inhibiting transcription of the parent plasmid from the tac promoter.
Fig 2.1.b.

The plasmid vector pKK223-3.

(Brosius & Holy 1984)
2.1.3. GROWTH MEDIA AND CONDITIONS.

Bacteria were grown in either complex or minimal media. Complex media (Luria broth: LB) was as described by Miller (1972) at 37°C and minimal media was as described by Hareland et al (1975) at 30°C. Liquid cultures were incubated at the desired temperature in an orbital shaker at 200 r.p.m. Where appropriate liquid media was solidified by the incorporation of 1.6% (w/v) Bacto agar (Difco).

Carbon sources and supplementary amino acids were sterilised separately and added aseptically to the media to give final concentrations:

- HPA: 5mM
- Glycerol: 10mM
- Glucose: 10mM
- Thiamine: 80µg.ml⁻¹
- Leucine: 80µg.ml⁻¹

Minimal medium cultures were inoculated with 1/100 volume of an overnight culture grown in LB. Luria broth cultures larger than 50ml were inoculated with 1ml of an overnight culture taken from a single colony. Small scale 5ml or 10ml LB cultures were inoculated from a single colony picked with a sterile toothpick. Antibiotics were added as required to LB media to final concentrations of: ampicillin (100µg.ml⁻¹) or tetracycline (12.5µg.ml⁻¹). These concentrations were reduced to one quarter for minimal media cultures. Bacteria in minimal media were grown to an OD₆₈₀ of 0.8.ml⁻¹ before harvesting. Cultures grown in LB were routinely harvested after overnight growth when the OD₆₈₀ had reached approximately 3.0.ml⁻¹.

When blue/white selection was used to detect the presence of inserts in the multiple cloning site of pUC vector by inactivation of the α fragment of the
β-galactosidase gene, ampicillin (100μg.ml⁻¹) and isopropyl-β-D-thiogalactopyranoside (IPTG 1mM) were incorporated into the solid 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 diethyl formamide, spread onto the surface of the plates before the addition of the bacterial culture.

2.1.4. PREPARATION OF CELL FREE EXTRACTS.

Bacteria from liquid culture were harvested by centrifugation at 10,000 g for 10 minutes at 4°C, washed with 0.2 volumes of 0.1M sodium phosphate buffer pH 7.5 before being resuspended in 0.04 volumes of the same buffer. The cells were then disrupted by ultrasonication in an MSE 100 W ultrasonic disintegrator in 4ml samples maintained at 0°C with oscillations of 7μM peak to peak amplitude for 30 seconds. The remaining cells and cell wall material was then removed by centrifugation at 20,000 xg for 30 minutes at 4°C. When necessary the extracts were ultracentrifuged at 120,000 xg for 90 minutes to remove membrane particles.

2.1.5. ESTIMATION OF PROTEIN CONCENTRATIONS.

The protein concentration of crude or ultracentrifuged extracts was measured by the biuret method (Gornall et al 1949). The protein concentration of such extracts grown on LB was typically in the range 5-25 mg.ml and in the range of 2-6mg.ml for cells grown in minimal media. The protein concentration of more purified extracts was measured by the method of Lowery et al (1951). In both cases bovine serum albumin (Sigma) was used as a standard. Calculations of molar extinction coefficients for the purified enzymes are made on the basis of the molar extinction coefficients of tyrosine (ε=1100 M⁻¹.cm⁻¹) and trptophan (ε=5050 M⁻¹.cm⁻¹) in neutral aqueous solution.
2.1.6. **PREPARATION OF INTERMEDIATES OF THE HPC PATHWAY.**

The diverse series of clones produced by Jenkins (1987) enables the production of the intermediates of the HPC pathway for use in spectrophotometric assays. *E.coli.* 5K strains harbouring the appropriate cloned *hpc* genes were used in a modified system to that reported by Jenkins (1987) to produce milligram quantities of the HPC pathway intermediates. The method assumes that the HPC dioxygenase enzyme is the rate limiting step of the pathway, since this enzyme seems to be rapidly deactivated in spectrophotometric assay systems. To overcome any possible deficiency in the *meta-ring* cleavage step, the high expressing deletion subclone pDR9304 (which expresses the *hpcB* gene product to some 10% of total cell protein) was grown in addition to a clone which produces the other Hpc pathway enzymes required to produce the desired intermediate. For the production of CHMS, only pDR9304 was used. The clones used to prepare the various intermediates are as follows.

<table>
<thead>
<tr>
<th>Intermediate</th>
<th>Clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHMS</td>
<td>pDR9304 (<em>hpcBD</em>)</td>
</tr>
<tr>
<td>CHM</td>
<td>pJJ003 (<em>hpcEDC</em>)</td>
</tr>
<tr>
<td>HHDD</td>
<td>pJJ801 (<em>hpcBCDEF</em>)</td>
</tr>
</tbody>
</table>

Generally the method is as follows. Two 100ml LB cultures, one of 5K(pDR9304) and one of 5K with the required construct, were grown overnight and extracts prepared as previously described. The activity of the extracts prepared from clones carrying all the required activities was estimated by determining how much extract was required to convert completely 0.1μmol of HPC into CHMS. This step measures the activity due to the HPC dioxygenase only and assumes that the other enzymes in the extract will be present in sufficient stable quantities to enable complete conversion of HPC to the desired intermediate. Once the activity in terms of the amount of extract required to
convert a known quantity of HPC was established the total amount of HPC that could be converted by all the extract was calculated. Typically a volume of 20-50μl of extract per μmol of HPC was required. The required amount of HPC was then dissolved in 25ml of 0.1M sodium phosphate buffer pH 7.5. NAD+ was then added to the solution to a molar ratio of 1:1.3 in favour of NAD+ as in the method described by Jenkins (1987). The mixture was gently shaken in a flask and warmed to 37°C before the addition of the converting extract. The mixture was then allowed to incubate with gentle stirring until the yellow colour due to CHMS had disappeared to give a pale straw colour. Small quantities of 5k.(pDR9304) extract were added periodically to ensure complete conversion of HPC to CHMS. Generally the reaction was complete within 30 minutes. This was checked by taking a 100μl sample of the reaction mixture and testing whether any HPC was left in it, by adding some of the 5K.(pDR9304) extract. The presence of any unconverted HPC was signified by the formation of yellow CHMS.

The reaction was terminated by the addition of concentrated HCl (2ml) which serves the dual purpose of precipitating the protein of the reaction mixture and also converting the final reaction product to the protonated form. The acidified mixture was centrifuged at 20,000 xg to remove the precipitated protein. The remaining solution was then extracted three times with an equal volume of ether, before the organic phases were pooled and dried over anhydrous sodium sulphate. The extracted intermediate were obtained by evaporation to dryness in a rotary evaporator. The residue thus obtained was dissolved in 2ml of H₂O and freeze dried. When HHDD was prepared D₂O was used instead of H₂O due to the rapid spontaneous isomerisation of HHDD to OHED in aqueous solutions. A further one or two freeze drying steps were then performed to enable very pure samples of the intermediate to be obtained. The
intermediates could then be stored for several months as freeze dried powders
or as solutions at -20°C before use.

2.1.7. **ENZYME ASSAYS.**

Enzyme assays were carried out at 30°C in disposable plastic cuvettes
(λ > 320nm) or in matched quartz cuvettes (λ < 320nm) with a path length of 1
cm in a total volume of 1ml. HHDD isomerase assays were carried out in a
total volume of 0.8 ml to reduce the amount of D₂O used. All assays were
carried out in a Pye-Unicam SP1800 recording spectrophotometer unless
otherwise stated. One unit of enzyme activity is defined as the ability to convert
1μmol.min⁻¹.

**HPC 2,3 DIOXYGENASE (EC 1.13.11.15)**

(Cooper and Skinner, 1980)

\[
\begin{align*}
\text{HPC} & \rightarrow \text{CHMS} \\
\text{CH}_2\text{COOH} & \quad \text{CH}_2\text{COOH} \\
\text{OH} & \quad \text{CHO} \\
\text{O}_2 & \quad \text{COOH}
\end{align*}
\]

The meta ring cleavage of HPC to give CHMS by HPC dioxygenase was
monitored by the increase in absorbance at 380nm due to the formation of
CHMS. The reaction mixture contained 0.2μmol HPC in 0.1 M sodium
phosphate buffer pH 7.5 and was started by the addition of 5-50μl of extract. A
molar extinction coefficient of 31,800 dm³. mol⁻¹. cm⁻¹ was assumed for CHMS
at pH 7.5
CHMS DEHYDROGENASE.

(Cooper and Skinner, 1980)

\[
\begin{align*}
\text{CHMS} & \xrightarrow{\text{NAD}^+} \text{CHM} \\
\end{align*}
\]

CHMS dehydrogenase was assayed by measuring the decrease in absorbance at 380nm as CHMS is oxidised to CHM on addition of the NAD$^+$ dependent dehydrogenase. The reaction mixture contained 0.035µmol of CHMS and 5-50µl of extract in 1 ml of 0.1 M sodium phosphate buffer pH 7.5. The reaction was started by the addition of 0.2µmol of NAD$. A molar extinction coefficient of 31,800 dm$^3$. mol$^{-1}$. cm$^{-1}$ was assumed for CHMS at pH 7.5.

CHM ISOMERASE.

(Garrido-Pertierra and Cooper 1981)

\[
\begin{align*}
\text{CHM} & \leftrightarrow \text{COHED} \\
\end{align*}
\]

CHM isomerase activity was monitored by the decrease in absorbance at 300nm due to the loss of CHM. The assay mixture consisted of 0.05µmol of CHM in 0.1 M sodium phosphate buffer pH 7.5. The spontaneous rate of isomerisation was monitored for 1 minute before the addition of 5-50µl of
In previous studies an assay mixture for the measurement of COHED decarboxylase was formed by the slow spontaneous isomerisation of CHM, to an equilibrium mixture of CHM and COHED, before the decarboxylase could be measured. (Garrido-Pertierra and Cooper 1981) In this study large quantities of partially purified stable CHM isomerase were available, such that an equilibrium mixture of CHM and COHED could be formed in the cuvette by the action of excess CHM isomerase. This system is more advantageous than the method of Garrido-Pertierra and Cooper 1981 since an equilibrium concentration of COHED is maintained throughout the assay due to CHM isomerase being present in excess. The reaction mixture contained 0.1μmol of CHM in 0.1 M sodium phosphate buffer pH 7.5 to which approximately 0.6 units of CHM isomerase were added to rapidly produce an equilibrium mixture of CHM and COHED. MgCl₂ (5μmol) was added to the mixture before the addition of 5-50μl of extract. A molar extinction coefficient of 20,000 dm³ mol⁻¹ cm⁻¹ was used for CHM at 300nm.
The presence of a second isomerase enzyme in the Hpc pathway was demonstrated by Jenkins (1987). Jenkins was able to isolate a stable intermediate, HHDD which was shown to undergo rapid spontaneous isomerisation in aqueous solutions. A D$_2$O based system was employed to slow down the spontaneous rate of isomerisation enabling an enzyme catalysed rate to be measured. In the assay system developed by Jenkins (1987) 10µl of 0.1M sodium phosphate buffer pH 7.5 was included in the assay to provide some degree of buffering capacity. It has been found in this study that the pH of such a system is variable depending on individual preparations of HHDD, generally giving rise to low pH ( < pH 6 ) conditions in the assay. Accordingly a new assay system was devised. The assay mixture consisted of 40µl of 2.0M potassium phosphate buffer pH 7.5 in D$_2$O, 0.05µmol HHDD and D$_2$O to 0.8ml. The spontaneous rate of isomerisation (which appears to be subject to a primary kinetic isotope effect) was then measured for 40 seconds at 276nm before the addition of 5µl of extract. A molar extinction coefficient of 18,600 dm$^3$. mol$^{-1}$. cm$^{-1}$ was used for HHDD.
The rapid spontaneous isomerisation of HHDD in aqueous phosphate buffer was utilised to form an equilibrium mixture of HHDD and OHED in order to measure OHED hydratase activity. The reaction mixture consisted of 0.1M sodium phosphate buffer pH 7.5 and 0.1μmol of HHDD which was allowed over the period of a few minutes to form an equilibrium mixture. When the decrease in absorbance at 276nm had stopped 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 dm$^3$. mol$^{-1}$. cm$^{-1}$ was assumed for HHDD.
HHED aldolase activity is measured in a coupled assay system by following the decrease in absorbance at 340nm as NADH is oxidised in the presence of lactate dehydrogenase by the pyruvate formed from HHED. The reaction was assayed in 0.1M sodium phosphate buffer pH 8.0 containing 5 µmol magnesium chloride, 0.15 µmol NADH, 4 units of lactate dehydrogenase and an excess of HHED. A molar extinction coefficient of 6,220 dm$^3$. mol$^{-1}$. cm$^{-1}$ was used for NADH.

2.1.8. **SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS.** (SDS-PAGE)

High resolution SDS-PAGE was performed using an LKB vertical slab gel system utilising 7.5% -20% (w/v) gradient gels (Hames 1981). Samples of protein (5-50µg) were mixed with 0.33 volumes of sample buffer and heated to 100°C for one to two minutes before application to the gel. Sample buffer consisted of 0.18M Tris.Cl pH 6.8, 5.7 (w/v) SDS, 29% glycerol and 0.005 % (w/v) bromophenol blue). The gels were run at 4mA per track for approximately
2.5h. This type of system was also utilised when protein samples were electrophobotted for amino terminal amino acid sequencing. The proteins were stained with Coomasie blue R250 solution (50% [v/v], 10% acetic acid [v/v] and 0.5% [w/v] Coomasie blue R250 for 1 hour. Gels were then destained with three changes of a solution of 7.5% (v/v) acetic acid and 5.0% (v/v) methanol. A second Biorad mini gel (mini-Protean) system was utilised for analysis of fractions obtained during purification of individual enzymes. Electrophoresis using this system was carried out according to the manufacturers instructions using 12% acrylamide gels which were run at 200V constant voltage for 45 minutes. Gels were then stained for twenty minutes and destained for one hour with the solutions described above. The "SDS 7" molecular weight marker proteins purchased from Sigma were used to calibrate the gels. The molecular weights 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).

2.1.9. AMINO TERMINAL AMINO ACID SEQUENCING.

Purified enzymes were sequenced on an Applied Biosystems 470A gas-phase sequencer by Dr K.S.Lilley. A purified fraction of the enzyme was run on a 7.5%-20% SDS-polyacrylamide gel and electrophobotted onto a polyvinylidene difloride (PVDF) membrane using 50 mM glycine-50 mM Tris pH 10 as transfer buffer and stained with Coomassie blue R-250 10. (Matsudaira 1987). The stained protein band was excised and loaded into the sequencer without polybrene.
2.1.10. **FAST PROTEIN LIQUID CHROMATOGRAPHY (FPLC).**

The Pharmacia FPLC system was run at room temperature in accordance with the manufacturer's instructions. The columns used in the purification of enzymes in this study were: Mono Q HR 5/5 and 10/10 anion exchange columns, a Phenylsuperose HR 5/5 hydrophobic interaction column and a Superose 12 HR 10/30 gel filtration column. Native molecular weight estimations by gel filtration were made on a system comprising two Superose 12 columns connected in series. This system increased the resolution two fold over that given by a single column. In this case the proteins used to calibrate the system as recommended by Pharmacia were: immunoglobulin G (160,000), bovine serum albumin (66,000) β-lactoglobulin (35,000) and cytochrome C (12,400).

2.1.11. **OLIGONUCLEOTIDE SYNTHESIS.**

Oligonucleotides were synthesised by Mr J.Kyte, Dr K.S. Lilley or Mrs D Langton with an Applied Biosystems 380B DNA synthesiser using cyanoethyl phosphoramidate chemistry. Oligonucleotides were ethanol precipitated and redissolved in 10mM Tris.Cl 1mM EDTA pH 8.0 before use. A list of the oligonucleotides used in this study is given in appendix 1.
2.1.12. **CRYSTALLOGRAPHIC EXPERIMENTS.**

Crystals were grown from protein solutions in the region of 10mg.ml$^{-1}$ using hanging drop techniques under the supervision of Dr D.B.Wigley. Preliminary X-ray studies on crystalline CHM isomerase were carried out at the SERC synchrotron source at Daresbury and a full native data set was produced at the EMBO synchrotron source at Hamburg, FRG with the help of Dr K.Wilson. Data processing was carried out on the VAX mainframe computer in Leicester.

2.2. **GENETIC PROCEDURES.**

2.2.1. **PREPARATION OF PLASMID DNA.**

Several different procedures were employed for the preparation of plasmid DNA. Small scale preparations of plasmid DNA for use in restriction digests, transformation and general DNA manipulation were made by the alkaline lysis method of Ish-Horowitz as described in Maniatis *et al* 1981. Overnight cultures grown in the presence of the appropriate antibiotic were used as a source of plasmid DNA. 1.5ml of an overnight culture were harvested in eppendorf tubes by centrifugation in a bench top microcentrifuge for 2 minutes. The remaining supernatant was removed and the pellet resuspended in 100|il of ice cold 25mM Tris.Cl pH 8.0, 50mM glucose, 10mM EDTA and allowed to stand at room temperature for 5 minutes. The cells were then broken open by the addition of 200|il of alkaline SDS solution consisting of 1% SDS and 0.2M NaOH. The tubes were then inverted several times and placed on ice for 5 minutes. Chromosomal DNA was then preferentially precipitated by the addition of 150|il of ice cold potassium acetate solution ( 3M with respect to potassium and 5M with respect to acetate pH 5.0 ). The tubes were inverted and then vortexed for 1 second before being incubated on ice for 5
minutes. The precipitated chromosomal DNA was removed by centrifugation in a bench top microfuge for 5 minutes. The resultant supernatant was then transferred to a new eppendorf tube and extracted once with an equal volume of 1M Tris.Cl pH 7.5 buffered equilibrated phenol : chloroform : isoaamy alcohol (25:24:1) to remove the remaining protein. The mixture was centrifuged for 5 minutes after which the upper aqueous phase was removed to a fresh tube and extracted once with an equal volume of chloroform : isoaamy alcohol (24:1). After the phases were separated by centrifugation for two minutes the nucleic acids in the upper aqueous phase were precipitated by the addition of two volumes of isopropanol (propan-2-ol) upon incubation for 5 minutes at room temperature. The precipitate was collected by centrifugation for 5 minutes in a bench top microfuge. The pellets were then dried under vacuum for ten minutes and then dissolved in 40μl of 10mM Tris.Cl pH 8.0, 1 mM EDTA (TE) containing ribonuclease A (RNase) at a concentration of 20μg.ml⁻¹. Some larger scale preparations were made by directly scaling up this procedure substituting 15,000 g spins for those carried out in a bench top microfuge. In such cases the cells were initially pelleted by a 5000 xg spin for 10 minutes.

Highly purified DNA was prepared from CsCl gradients using dye-bouyant density equilibrium centrifugation as follows. The plasmid DNA obtained by a scaled up mini-prep procedure from a 400ml overnight culture, was RNase treated for 30 minutes at 37°C at an RNase concentration of 20μg.ml⁻¹ before phenol chloroform extraction and precipitation of the DNA with 100% ethanol on dry ice. The dried pellet was then dissolved in 4.1ml of TE to which 4.3g of CsCl, 218μl of ethidium bromide of concentration 10mg.ml⁻¹ was added and vortexed to ensure the CsCl had completely dissolved. The solution was then loaded into Beckman quick seal 5.2ml tubes and centrifuged overnight in a VTi 65.2 rotor in a Beckman L5-65 ultracentrifuge at 45,000 rpm at room temperature. After centrifugation, covalently closed circular plasmid
DNA was located under long wave UV light and removed from the gradient using a large gauge needle and syringe. The ethidium bromide was removed by ten extractions with NaCl saturated isopropanol, and the DNA precipitated with 2 volumes of -20°C ethanol in the presence of 0.3M sodium acetate and incubated on dry ice for 20 minutes.

When a large number of mini-preps were required a much shorter protocol was used which yielded DNA of the quality required for restriction digests and retransformation. This method is based on that of Serghini et al (1989). Small bacterial cultures are grown overnight and 1.5 ml used as a source of plasmid DNA. Cells were harvested by centrifugation for two minutes and then resuspended in 150μl of 10mM Tris.Cl pH 7.5, 100mM NaCl, 1mM EDTA (TNE) after removing the supernatant. An equal volume of buffer equilibrated phenol (as above) was then added and the mixture vortexed for one second. Prolonged vortexing causes the amount of chromosomal DNA fragments in the resulting preparations to increase. The solution was then centrifuged for 5 minutes to separate the phases and the upper aqueous phases removed to a clean tube. The nucleic acids are then precipitated with two volumes of ethanol taken straight from the freezer, on dry ice for 10 minutes. The pellets were then treated as in the alkaline lysis procedure.

Plasmid DNA for use in sequencing reactions was prepared by a modification of the method of Kraft et al (1987). Overnight LB cultures of 50ml were grown and the cells from 4.5 ml of bacterial culture pelleted per eppendorf tube. Typically 6 tubes per overnight culture were prepared at once. The cells are then broken open and chromosomal DNA preferentially precipitated as in the alkaline lysis method. Following removal of the supernatant to a fresh tube, DNase free RNase A was added to a final concentration of 50μg.ml⁻¹ and incubated at 37°C for 30 minutes. An equal
volume of Tris.Cl pH 7.5 buffer equilibrated phenol was then added, vortexed and the aqueous phase removed to a fresh tube after centrifugation for 5 minutes. The DNA was then precipitated on dry ice for 30 minutes with two volumes of ethanol straight from the freezer. The pellets are washed with 0.5ml of 70% ethanol after centrifugation for 5 minutes and dried under vacuum for 10 minutes. The remaining protein was then removed by the addition of 40μl of a solution containing 16.8μl H₂O, 3.2μl 5M NaCl and 20μl of 13% polyethylene glycol 8000 ( Sigma ). The tubes were incubated on ice for 5 minutes and then briefly vortexed to redissolve the pellets. The tubes were then incubated on ice for 30 minutes before centrifugation in a bench top microcentrifuge for ten minutes. The supernatants were then removed with a drawn out Pasteur pipette and the remaining pellets washed with 70% ethanol. The pellets were then dried under vacuum for 10 minutes before the DNA is redissolved in 40μl of TE per tube. The DNA from several tubes were pooled and the DNA concentration assessed by measuring the absorbance at 260nm of a sample. This procedure yields very high quality plasmid DNA which is pure enough to serve as template in DNA sequencing reactions and also as a substrate in exonuclease III reactions.

2.2.2. AGAROSE GEL ELECTROPHORESIS.

Restriction enzyme digests were analysed by submarine gel electrophoresis through agarose gels. Gels of 0.5% to 1.5% agarose were employed depending on the size of the expected fragments. The agarose was dissolved in TAE buffer (40mM Tris-acetate pH 8.0, 1mM EDTA) to which ethidium bromide at a concentration of 0.5μg.ml⁻¹ was included. Gels were also run in TAE buffer at constant voltage, generally in the range 5-10V.cm⁻¹. Before loading of the DNA sample 0.1 volumes of type II electrode dye mix (Maniatis 1981; 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol and
30% (w/v) glycerol), was added to enable the progress of the electrophoresis to be monitored. The gels were run until the bromophenol blue had reached the end of the gel, and then viewed under short wave UV light. Photographs were taken using TMAX-100 film (Kodak). Lambda phage DNA cut with HindIII or BstEII restriction enzymes were used to calibrate the gels after measurement of the migration distances of the individual fragments.

2.2.3. RESTRICTION ENZYME DIGESTION.

Restriction digests were carried out according to the manufacturer's instructions using the buffers provided. Restriction enzymes were purchased from Bethesda Research Laboratories (BRL) or from Pharmacia-LKB Ltd. Restriction digests contained between 200-500ng of plasmid DNA in a 15μl volume to which 0.5μl of the required enzyme was added, representing an average 5-10 fold excess number of units of enzyme to bring about complete digestion in 1 hour. Larger digestions were carried out in either 25μl or 60μl volumes and allowed to digest overnight. BstEII digestions were carried out at 60°C under mineral oil for 90 minutes.

2.2.4. PHOSPHATASE TREATMENT OF DNA.

To minimise the number of recircularised vector molecules in a ligation reaction, plasmid DNA was treated with Calf intestinal phosphatase (CIP) to remove terminal phosphate groups from the linearised plasmid. CIP was purchased from Pharmacia and 0.1 units added per μg of vector DNA. The CIP reaction was allowed to continue for 30 minutes at 37°C in the required buffer as per the manufacturer's instructions. The CIP was then removed by phenol/chloroform extraction and the DNA precipitated on dry ice with two volumes ethanol and 0.3M sodium acetate.
2.2.5. **ISOLATION OF DNA FRAGMENTS FROM AGAROSE GELS.**

Two different methods for extracting DNA fragments from agarose gels were used during the course of this study. The CETAB method (Maniatis 1981) used in the earlier stages of the project was replaced by the more efficient and faster phenol extraction method of Perbal (1988).

**CETAB method:**

Approximately 1 μg of plasmid DNA was digested and electrophoresed through a 1% low melting point agarose gel. A sterile scalpel blade was then used to cut the required fragment(s) from the gel taking as little excess agarose with the fragment as possible. The recovered gel slice was then placed in a 1.5 ml microfuge tube and centrifuged briefly to take it to the bottom of the tube. The tube was placed in a 65°C waterbath for 30 min to melt the agarose before the addition of one volume of butanol-saturated water and one volume of water saturated butanol. The mixture was vortexed vigorously for 2 min and the phases separated by centrifugation for 5 minutes. The upper phase, but none of the interphase, was transferred to a fresh tube and the original mixture re-extracted with 0.5 volumes of water saturated butanol. The two butanol layers were then pooled. The pooled layers were then extracted once with 0.25 volumes of 0.3 M sodium acetate pH 7.0 followed by centrifugation. The bottom phase was then carefully collected without taking any of the upper butanol layer. The aqueous layer was then extracted once with an equal volume of chloroform/isoamyl alcohol [24:1 (v/v)] and the DNA precipitated at -70°C with 2.5 volumes of ethanol in the presence of 0.3 M sodium acetate for thirty min. The precipitate was collected by centrifugation for 10 min in a microfuge, dried briefly in a vacuum desiccator and redissolved in water.
The solutions used were made as follows: An equal volume of butanol and water were vigorously shaken together and the phases allowed to separate. One gram of cetyl trimethyl ammonium bromide (CETAB) was dissolved in 100 ml of the upper (water saturated butanol) phase to which 100 ml of the lower (butanol saturated water) phase was added and the two mixed by shaking. The mixture was allowed to separate overnight and the phases put into separate bottles.

Phenol extraction method:

Between 1μg and 5μg of plasmid DNA was digested overnight and loaded onto a 1% low melting point agarose gel (BRL). After electrophoresis the required band was cut out of the gel as a slice and placed in an eppendorf tube to which TE buffer was added to a total volume, including the gel slice, of 0.5ml. The gel was then melted at 65°C for 5-10 minutes and vortexed briefly. 150μl of Tris.Cl pH 7.5 buffered phenol was then added and the mixture vortexed for two minutes. Once the phases had been separated by centrifugation for 5 minutes, the upper aqueous phase was removed to a fresh eppendorf tube and an equal volume of water saturated ether added. The mixture was vortexed briefly and centrifuged before the upper ether layer was removed. The residual ether was removed by either blowing air briefly through the aqueous phase or by leaving the tube open at room temperature for 5-10 minutes. The DNA was then precipitated by the addition of 2 volumes of ethanol and sodium acetate to a concentration of 0.3M, on dry ice for 30 minutes or overnight. The DNA pellet was then washed with 70% ethanol and dried under vacuum before redissolving the pellet in 50μl of TE. The recovery could then be assessed by running a small volume on an agarose gel.
2.2.6. **LIGATION OF DNA.**

Vector and insert DNA were mixed in a 1:2 molar ratio of ends and ligated in a reaction containing 1-5 units of T4 DNA ligase (BRL) in accordance with the manufacturer's instructions. The reaction was allowed to continue for between 2-16 hours at room temperature before the ligation mixture was used to transform competent cells.

2.2.7. **TRANSFORMATION OF BACTERIA WITH PLASMID DNA.**

(Kushner, 1978)

Competent cells were made when required each time by the morpholinepropanesulfonic acid (MOPS) -rubidium chloride method of Kushner. The required cell line was grown in small LB cultures (5-10ml) to an OD$_{680}$ of approximately 0.5 before 1.5ml aliquots were pelleted by centrifugation for 2 minutes. The pellet of cells was then washed in 0.5ml of sterile 10mM MOPS pH 7.0, 10mM RbCl and 50mM CaCl$_2$ and finally resuspended in 10mM MOPS pH 6.5 10mM RbCl and 50mM CaCl$_2$. The cells were then incubated on ice for 30-90 minutes before they were collected by a short spin and resuspended in 100μl of the MOPS pH6.5 buffer. Dimethyl sulphoxide (DMSO) was added to 0.2% (v/v) and the DNA added. In the case of a ligation reaction all the incubation was added to the competent cells, whereas for a simple retransformation approximately 100ng of DNA was added. The cells were then incubated on ice for a further hour before being heat shocked for 30 seconds at 55°C followed by incubation on ice for 2 minutes. Prewarmed LB was then added to a total of 1ml before the cells were allowed to recover for 1 hour at 37°C. Cells were then spread onto the appropriate LB+antibiotic agar plate and allowed to grow overnight at 37°C.
2.2.8. EXONUCLEASE III (ExoIII) DIGESTIONS.

ExoIII digestion were made using a Kit purchased from Pharmacia-LKB based on the techniques described by Henikoff (1984, 1987). The temperature used in the Exo III reaction was derived empirically, since the temperatures quoted in the manufacturer's protocol seemed to give approximately half the expected rate of digestion. The Exo III reactions were carried out at 35°C using 5μg of plasmid DNA. This technique relies on the substrate specificity of Exo III for free ends of double stranded DNA which are either blunt ended or have 5' overhangs. Exo III will not digest 3' overhangs or overhangs filled in with thionucleotides. For pDR1930 and pDR1835 the inserts are bounded by SalI and BamH1 restriction site which create 5' overhangs and are therefore susceptible to Exo III digestion allowing digestion of the cloned DNA. The SalI and BamH1 sites in pUC18 and pUC19 occur inbetween SphI, PstI and KpnI, SstI site respectively which produce 3' overhangs and are not therefore susceptible to Exo III. So by digesting the required subclone with a combination of 3' and 5' restriction endonucleases it is possible to produce a linearised molecule which will undergo Exo III digestion in one direction only. The temperature and time intervals required to give a set of subclones which differ in size by approximately 200 bps was found empirically before a set of deletion subclones was derived by this method. Generally Exo III reactions were set up so that ten time points of 4μl aliquots were taken at timed intervals to be mixed with 6μl of S1 nuclease/buffer mix provided with the kit. The S1 step was carried out for 30 mins at room temperature and terminated with 2μl of S1 stop mix before heat inactivation as described in the Pharmacia protocol. The size of the digestion products were checked by agarose gel electrophoresis on 0.8% gels at this point. Ligation reactions using the T4 DNA ligase and buffers provided were carried out for 2 hours at room temperature and then used to transform competent E.coli 5K.
Ten colonies per time point were analysed by mini-prep followed by
digestion with appropriate restriction endonucleases. Using this method
approximately one colony in six or seven would yield plasmid DNA which
retained the parental pUC plasmid DNA along with a deleted insert. As such a
large number of colonies needed to be analysed before a complete deletion
series was isolated. This is obviously not an ideal situation, but did yield 17
deletion subclones which were chosen to form the pDR93 deletion series.

However, the conditions described in the manufacturer's instructions
gave rise to an unacceptably low rate of recovery of desired deletion subclones
from any one time point on several separate occasions when a deletion series
was attempted from pDR1835. Therefore a number of different strategies were
employed to increase the recovery frequency.

Strategy 2: Replacement of the S1 nuclease with Mung Bean nuclease.
Approximately 10µg of pDR1835 was digested with BamHI and SacI overnight
at 37°C in a total of 50µl before use in the ExoIII reaction. The ExoIII reactions
were carried out at 35°C in a 100µl reaction with the required amount of Exo
III buffer and 0.3M NaCl. Fourteen aliquots of 7µl were taken at 2 min
intervals after an initial period of 8 mins and added to 15µl of mung bean
nuclease buffer (See over). Each time point was stored on dry ice until all had
been collected. All the time points were then incubated at 68°C for 15 mins to
inactivate the Exo III. Some 3 units of mung bean nuclease were then added to
each time point. The enzyme was diluted to a concentration of 3 units/µl just
before use, with 1 x mung bean dilution buffer (See over). Mung bean nuclease
digestions were carried out for 30 mins at 30°C and terminated with the
following:
4μl 20% SDS
10μl 1M Tris.Cl pH 9.5
20μl 8M LiCl
200μl sterile water
250μl Tris.Cl pH 7.5 buffer equilibrated
phenol/chloroform

Each tube was then vortexed, spun 5 mins in a microfuge and the upper aqueous layer removed for extraction with an equal volume of chloroform. The DNA was then precipitated with 1 ml of absolute ethanol (−20°C) and 180μl of 3.0M sodium acetate pH 4.5 on dry ice for 20 mins. The DNA was then pelleted by spinning in a microfuge for 10 mins before draining of the supernatants and washing with 250μl of 70% ethanol. The pellets were then dried under vacuum for 10 mins and redissolved in 15μl of sterile water before ligation overnight at 4°C.

<table>
<thead>
<tr>
<th>5x Mung bean nuclease buffer</th>
<th>1x Mung bean dilution buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>150 mM Na Acetate pH 5.0</td>
<td>10.0 mM Na Acetate pH 5.0</td>
</tr>
<tr>
<td>250 mM NaCl</td>
<td>0.1 mM Zn Acetate</td>
</tr>
<tr>
<td>5 mM ZnCl₂</td>
<td>1.0 mM cysteine</td>
</tr>
<tr>
<td>25% Glycerol</td>
<td>0.1% Triton X-100</td>
</tr>
<tr>
<td></td>
<td>50% Glycerol</td>
</tr>
</tbody>
</table>

Analysis of the colonies after transformation, showed that approximately one in every three mini-preps gave a deletion subclone. This protocol has made the isolation of a deletion series much easier, especially when coupled with the very rapid mini prep method of Serghini et al (1987).

Strategy 3: A methodology using a filling in reaction with the Klenow fragment of DNA polymerase 1 and dNTPs after the S1 step was used. In the
original strategy no filling in step was included. A series of Exo III reactions with 20μg of pDR1835 plasmid DNA in a 100μl ExoIII reaction was carried out as previously described (strategy one). The S1 step was carried out as described for 30 minutes and terminated by the addition of S1 stop solution provided with the kit. After the S1 step 2μl of Klenow mix were added and then incubated for 5 mins at 37°C. 2μl of a solution containing 2.5mM dNTP’s was then added and incubated for a further 5 mins at 37°C before the volume was raised to 200μl and phenol extracted in the same way as the protocol for the reactions utilising mung bean nuclease.

5x Klenow buffer: 100mM Tris.Cl pH 8.0
500mM MgCl₂
Klenow mix: 7μl Klenow buffer
9μl Klenow fragment (10 units/μl Pharmacia FPLC pure Klenow fragment)
19μl sterile water

This procedure increased the rate at which deletion subclones of were obtained to approximately one in three or four mini-preps analysed.

2.2.9. DIDEOXYNUCLEOTIDE SEQUENCING.

Plasmid DNA of the purity required for nucleotide sequencing was prepared as described in section 2.2.1. by the method of Kraft et al (1988). Nucleotide sequencing was carried out using T7 DNA polymerase from Pharmacia or the Sequenase version 2 enzyme of United States Biochemical. Buffers and reagents were provided in the kits from each manufacturer. Sequencing reactions were carried out with both unmodified nucleotide mixes and those incorporating c7dGTP to overcome G-G base pair compressions. When necessary dITP was used instead of dGTP to overcome particularly
Due to the high degree of overlap between deletion subclones of the same series, the accuracy of sequencing in any one region was significantly higher than could be achieved by a simple strategy using synthetic oligonucleotides as primers for dideoxynucleotide sequencing. DNA sequence analysis was carried out on Leicester University's Vaxcluster mainframe computer system utilising the University of Wisconsin's WIMP package. Open reading frames were identified by a combination of computer aided prediction and comparison of deduced amino acid sequence to that of the amino terminal sequence of individual purified proteins.

Sequencing procedure:

The solutions required are listed in table 2.2.9. Approximately 5μg of plasmid DNA in a total volume of 20μl was denatured by the addition of 2μl of 2M NaOH, 2mM EDTA which was made up freshly before each set of sequencing reactions. The denaturation reaction was carried out at room temperature for 10 minutes before the sequential addition on ice of 8μl of 1M Tris.Cl pH 4.5 and 3μl 3M sodium acetate followed by 75μl of ethanol (−20°C). The denatured DNA was then precipitated on dry ice for 10 minutes before the pellet was collected by centrifugation for 10 minutes. The supernatant was removed with a drawn out Pasteur pipette and the pellet washed in 200μl of 70% ethanol. The pellet was then dried under vacuum for 10 minutes before being redissolved in 10μl of sterile water to which 2μl of annealing buffer and 2μl of the required primer (5ng.μl⁻¹) were added. The annealing step was then carried out at 65°C for 2 minutes before being allowed to cool slowly to room temperature over 30 minutes. During this time 2.5μl of A,C,G and T termination mix are aliquoted into separate tubes and kept on ice until required. After the annealing step was complete the annealed template and primer solution was collected at the bottom of the tube by a brief spin. T7 DNA
polymerase was diluted with enzyme dilution buffer to a concentration of 2 units.µl⁻¹ and 2µl were added to the annealing mixture. Labelling mix (3µl) was then added to the mixture along with 1µl of [α-³⁵S] dATP (approximately 1000mCi/mmol; Amersham International plc) and incubated for 5 minutes at room temperature. During this time the tubes labelled A, C, G and T containing 2.5µl of the corresponding termination mixes were prewarmed in a water bath at 37°C. After the 5 minute labelling step 4.5µl of the labelling reaction was aliquoted into each of the 4 termination tubes and the termination reactions allowed to continue for 5 minutes at 37°C. The reactions were finally stopped by transferring the tubes to an ice bath and by the addition of 5µl of stop mix (90% formamide, 0.3% bromophenol blue, 0.3% xylene cyanol and 10mM EDTA pH 8.3). Sequencing reactions in which the sequence close to the primer was required were performed in exactly the same way except that 1µl of Mn buffer (Tabor and Richardson 1989) was included at the labelling step and this kind of sequencing reaction was only loaded once on the gel since the presence of Mn²⁺ in the sequencing reaction causes nucleotides to be incorporated in the first 150 bps only.
Table 2.2.9.

Composition of reagents for dideoxynucleotide sequencing

1. Annealing buffer:
   280mM Tris.Cl pH 7.5
   100mM MgCl₂
   350mM NaCl

2. Labelling mix:
   2.0μM dGTP
   2.0μM dCTP
   2.0μM dTTP

3. Termination mixes:
   Each of the four termination mixes consisted of:
   150μM dATP, 150μM dCTP, 150μM c'dGTP or dGTP and 150μM dTTP
   In addition individual mixes contained the following
   A mix: 15μM ddATP
   G mix: 15μM ddGTP
   C mix: 15μM ddCTP
   T mix: 15μM ddTTP
   Nucleotides were made up in buffered solutions containing 40mM Tris.Cl pH 7.5, 10mM MgCl₂ and 50mM NaCl

4. 20x Mn buffer:
   800mM MOPS pH 7.5
   100mM MnCl₂
   200mM sodium isocitrate
   Mixes were made up according to Tabor and Richardson (1987, 1989).
   c'dGTP: 7-deaza-2'-deoxyguanosine 5' triphosphate
   All nucleotides and nucleotides analogues were purchased from Pharmacia-LKB Ltd.

5. 10 X Tris-Borate-EDTA buffer (pH 8.3)
   121.1g Tris.L⁻¹
   53.0g Boric acid.L⁻¹
   3.7g EDTA.L⁻¹
Electrophoresis and autoradiography:

6% acrylamide-7M urea gels were made up in 1 x Tris-borate-EDTA pH 8.3 (TBE) buffer. Gels of 37cm length were run after prerunning at 1500-1700V for 45 minutes. 1 x TBE was used in the top and bottom tanks with an aluminium back plate to act as an efficient heat sink preventing "smiling" of the DNA bands during electrophoresis. Two loadings per sample were generally run using 4µl of the individual sequencing reactions. The first loading was run until the xylene cyanol marker band was two thirds of the distance down the gel. In cases were three loadings were applied, the first loading was run until the xylene cyanol reached the bottom of the gel before the second loading was applied. The third loading was then applied when the xylene cyanol from the second loading was two thirds the way down the gel. Following electrophoresis the gels were immersed in 10% methanol, 10% acetic acid for 15-20 minutes before being transferred to Whatman 3MM paper. The gels were then dried down on the paper under vacuum at 80°C for one hour. Autoradiography was normally carried out overnight with Fuji RX X-ray film at room temperature without intensifying screens.

2.2.10. TRANSCRIPT MAPPING STUDIES.

The first nucleotide of mRNA transcripts were mapped by primed synthesis using synthetic oligonucleotides and reverse transcriptase. The methodology used was as described by Boulnois (1987) with slight modifications.

Isolation of total RNA from E.coli.

The cells from a 30ml overnight Luria broth culture were harvested by centrifugation in 50ml sterile polypropylene centrifuge tubes and resuspended in 2ml of sterile water. 2ml of preheated lysis buffer: 1% SDS, 0.2M Tris.Cl pH
7.5, 0.2M NaCl, 40mM EDTA was then added before the mixture was allowed to stand in a boiling water bath for two minutes. An equal volume of 1M Tris pH 7.5 buffered phenol preheated to 64°C was then added. The mixture was vortexed for 2 minutes, allowed to stand for two minutes, and again vortexed for two minutes. The mixture was cooled rapidly by placing on wet ice before the phases were separated by centrifugation at 20,000 xg. The top aqueous phase was carefully removed to a fresh centrifuge tube and re-extracted with cold buffered phenol. The phases were again separated by centrifugation and the upper aqueous phase removed to a fresh tube and stored as an ethanol suspension at -20°C until required. The recovery was checked by running a small sample of the preparation on an agarose gel.

Primer extension.

An aliquot of RNA was precipitated in ethanol on dry ice before redissolving in 10µl of reverse transcriptase hybridization buffer: 40mM PIPES, pH 6.7, 0.4M NaCl, 1mM EDTA, 0.2% SDS. Approximately 10ng of the required oligonucleotide primer was then added. The mixture was incubated at 80°C for 5 minutes followed by a further 40 minutes at 63°C. The annealed mixture was precipitated by the addition of 35µl of ethanol followed by incubation on dry ice for 15 minutes. The pellet was then collected by centrifugation in a microfuge for 10 minutes, washed in 70% ethanol and dried in a vacuum desiccator. The pellet was redissolved in 10µl of water to which the following was added:

1.5µl 10x M-MuLV Reverse transcriptase buffer
0.5µl [35S]-dATP (approximately 1000mCi/mmol)
1.5µl 100µM dCTP, dGTP, dTTP mixture
1.5µl M-MuLV Reverse transcriptase (15 units)
10x M-MuLV Reverse transcriptase buffer:
500mM Tris.Cl pH 8.3,
80mM MgCl$_2$
100mM DTT

Cold nucleotides and Molony Murine Leukemia Virus (M-MuLV) Reverse transcriptase were purchased from Pharmacia-LKB Ltd. $^{35}$S-dATP (>1000mCi/mmol) was purchased from Amersham International Plc.

The reaction was started by incubation at 37°C for 30 minutes and terminated by the addition on ice of 4μl of sequencing stop mix (see section 2.2.9.). The mixture was heated in a boiling water bath for 1 minute before a 6μl sample was loaded onto a 6% acrylamide-7M urea sequencing gel. The reverse transcriptase extension product was run against a set of dideoxy sequencing tracks produced using the same oligonucleotide as used in the primer extension reaction. In this manner the extension product can be directly compared to the nucleotide sequence from which it was derived.
CHAPTER 3.
ORDERING THE *Hpc* GENES
3.1. INTRODUCTION.

According to Jenkins (1987) the genes encoding CHM isomerase, COHED decarboxylase and HHDD isomerase (encoding hpcD, hpcE and hpcF respectively) were to be found in a 3.8 Kbp SalI-BamH1 fragment present in all the genomic clones isolated (See Fig 3.1.). The genomic clone, pJJ801 was chosen as a source from which this 3.8 Kbp fragment could be obtained. As an initial step toward nucleotide sequencing of hpcD, hpcE and hpcF, this fragment was subcloned into the high expression and nucleotide sequencing vectors, pUC18 and pUC19. The two subclones produced were named pDR1835 and pDR1935 with respect to pUC18 and pUC19.

3.2.1. SUBCLONING OF THE 3.8 Kbp SalI-BamH1 FRAGMENT of pJJ801.

Some 2μg of pJJ801 was digested overnight in a 50μl digest containing SalI and BamH1. This digest was then applied to a 1% low melting point agarose gel and the 3.8 Kbp band extracted by the CETAB method. A similar quantity of pUC18 and pUC19 DNA was digested with the same enzymes in the presence of CIP and applied to a 1% low melting point agarose gel. After extraction, approximately 100μg of linearised plasmid was incubated with 200μg of the 3.8Kbp fragment in the presence of T4 DNA ligase at room temperature overnight. The ligation mixture was used to transform competent E.coli NM522 and the resultant colonies were screened by blue/white selection on Luria broth plates supplemented with ampicillin, x-gal and IPTG. Several white colonies were selected, grown in 5ml cultures and plasmid DNA analysed by restriction digestion. By this method, two clones were identified: pDR1835 and pDR1935 (See Fig.3.2.1)
Fig.3.1.

Representation of the gene order and direction of transcription of the *Hpc* genes according to Jenkins (1987).

The genomic clones pJJ200 and pJJ210 express all of the *hpc* pathway genes including the regulator gene: *hpcR*. The third genomic clone, pJJ801 constitutively expresses *hpcBCDEF* only. Jenkins (1987) placed a divergent operator/promoter region near the left hand end of pJJ801 as shown, such that the structural genes were transcribed in two transcriptional blocks: *hpcBCDEF* and *hpcGHR* as shown. Restriction sites are as follows: B=BamH1, E=EcoRI, S=SalI, P=PvuII, H=HindIII.
Fig. 3.1.

Representation of the gene order and direction of transcription of the $H_{pc}$ genes according to Jenkins (1987) and their relationship to the three genomic clones isolated.
Fig. 3.2.1.

Subclones derived from the 3.8 Kbp SalI-BamHI fragment of pJJ801

The 6.8 Kbp BamHI-BamHI genomic fragment of pJJ801 is shown for reference. Restriction sites are as follows: B = BamHI, Ev = EcoRV, Nd = NdeI, Kp = KpnI, E = EcoRI, Ss = SstII (SacII), S = SalI, Pv = PvuII, H = HindIII, Sp = SphI, Ps = PstI, Sm = SmaI, Sc = StI (SacI), Ss = SstII. The arrows denote the direction of transcription with respect to the lacUV-5 promoter of pUC in both pDR1835 and pDR1935. Restriction sites of the polylinker in each of the pUC derived plasmids are not drawn to scale, due to their close physical location, and are shown above the rest of the diagram. This convention is adopted for all subsequent diagrams.
Fig. 3.2.1.

Restriction maps of pDR1835 and pDR1935.
3.2.2. **ANALYSIS OF ENZYME ACTIVITIES ASSOCIATED WITH pDR1835 AND pDR1935.**

Competent *E.coli* 5K was transformed with pDR1835 and pDR1935, from which crude extracts were prepared from cells grown in 100ml minimal media cultures. These cultures were then used to inoculate fresh growth medium in the morning. The growth of the cultures was monitored with time, until the cell density had reached 0.8 OD$_{680}$ units. Inducible expression was tested by growing the cultures in the presence or absence of 1mM IPTG. Analysis of the enzyme activities associated with these extracts suggested that the gene order and direction of transcription was different to that previously described (Jenkins & Cooper 1988). The details of the enzyme activities associated with each subclone are detailed in Table 3.2.2.

This result was unexpected, as Jenkins (1987) had originally mapped *hpcD* encoding CHM isomerase to a 2.9 Kbp EcoR1-PvuII fragment of pJJ801 and therefore this fragment would be expected to give CHM isomerase activity. SDS-PAGE was used to analyse the crude extracts obtained from these cultures. Previous experiments in this study (not detailed) had established the conditions under which a partial purification of CHM isomerase could be achieved. A sample of this partially pure enzyme was also included in the SDS-PAGE analysis so that comparisons of the protein profiles of the extracts obtained in this experiment, could be made.
Table 3.2.2.

Activities of CHM isomerase, COHED decarboxylase and HHDD isomerase from various subclones.

<table>
<thead>
<tr>
<th>Subclone</th>
<th>IPTG</th>
<th>Protein (mg ml⁻¹)</th>
<th>CHM isomerase</th>
<th>COHED decarboxylase</th>
<th>HHDD isomerase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>SA</td>
<td>A</td>
</tr>
<tr>
<td>pDR1835</td>
<td>-</td>
<td>3.5</td>
<td>NDA</td>
<td>-</td>
<td>5.0</td>
</tr>
<tr>
<td>pDR1835</td>
<td>+</td>
<td>3.5</td>
<td>NDA</td>
<td>-</td>
<td>4.5</td>
</tr>
<tr>
<td>pDR1935</td>
<td>-</td>
<td>3.0</td>
<td>NDA</td>
<td>-</td>
<td>NDA</td>
</tr>
<tr>
<td>pDR1935</td>
<td>+</td>
<td>2.0</td>
<td>NDA</td>
<td>-</td>
<td>NDA</td>
</tr>
</tbody>
</table>

Activity (A)=μmol.min⁻¹.ml⁻¹
Specific Activity (SA)= mmol.min⁻¹.mg protein⁻¹
NDA= No Detectable Activity
Previous attempts to purify HHDD isomerase and COHED decarboxylase, indicated that both enzymes migrate as approximately 50 KDa bands on SDS-polyacrylamide gels. An increased number of bands, in this region of the gel was observed in the tracks corresponding to the extracts prepared from *E. coli* (pDR1835)+/IPTG. However, no band with the same mobility as the partially purified CHM isomerase, was seen in these extracts. The result of this SDS-PAGE analysis is shown in fig 3.2.2. To investigate these observation further, the 3.0 Kbp *Bam*H1-*Sal*1 fragment of pJJ801, believed to carry only *hpcB* encoding HPC dioxygenase (Jenkins 1987), was subcloned into pUC18 and pUC19.

3.2.3. **SUBCLONING OF THE 3.0 Kbp *Bam*H1-*Sal*1 FRAGMENT OF pJJ801.**

The 3.0 Kbp *Bam*H1-*Sal*1 fragment was isolated from pJJ801 by appropriate endonuclease digestion and recovery from a low melting point agarose gel. This fragment was ligated into pUC18 and pUC19 which had been cut to give complimentary cohesive ends and treated with CIP. The ligation mixture was used to transform competent *E. coli* NM522 and the resultant colonies were screened by blue/white selection on luria broth plates supplemented with ampicillin, IPTG and x-gal. Two subclones were isolated in this way: pDR1830 and pDR1930 (See Fig.3.2.3.)
Fig.3.2.2.
SDS-PAGE analysis of crude extracts cells carrying
pDR1835 and pDR1935.

Lanes
1. Marker proteins ranging in size (KDa) as shown.
2. pDR1835 - IPTG
3. Partially purified sample of CHM isomerase.
4. pDR1835 + IPTG
5. pDR1935 - IPTG
6. Partially purified sample of CHM isomerase.
7. pDR1935 + IPTG
8. Marker proteins ranging in size as shown.

The large arrow denotes the position of an increase number of bands in the pDR1835 extracts (lane 2 and lane 4) which are associated with the presence of HHDD isomerase and COHED decarboxylase in those extracts. No band in the 14 KDa region corresponding to CHM isomerase is seen in the pDR1835 extracts.
Fig. 3.2.2.
SDS-PAGE analysis of extracts made from cells transformed with either pDR1835 or pDR1935.

Lane: 1  2  3  4  5  6  7  8

M.P. (KDa)
66
45
36
29
24
20.1
14.2

CHM isomerase

CHM isomerase
3.2.4. **ANALYSIS OF ENZYME ACTIVITIES ASSOCIATED WITH pDR1830 AND pDR1930.**

Mini-preped plasmid DNA of the two clones was used to transform *E.coli* 5K and crude extracts were prepared from these cells grown in PJC minimal media, in the same way as detailed in section 3.2.2. The details of the enzyme activities associated with each subclone are detailed in Table 3.2.4. The finding of CHM isomerase activity associated with the 3.0 Kbp *BamH1-Sal1* fragment of pJJ801 was unexpected. SDS-PAGE was used to analyse the crude extracts obtained in the presence of pDR1830 and pDR1930. (See Fig. 3.2.4.) A band of some 33 KDa corresponding to the HPC dioxygenase (T.Fawcett pers comm.) was found in the pDR1930 extracts along with a band at 14KDa which migrated with the same mobility as the partially purified sample of CHM isomerase run alongside. These results confirmed that the gene order according to Jenkins was at least in part incorrect and required a different strategy to sequence the two isomerase genes than that originally proposed.
Subclones derived from the 3.0 Kbp *BamHI-SalI* fragment of pJJ801.

The 6.5 Kbp *BamHI-BamHI* genomic fragment of pJJ801 is shown for reference. Restriction sites are as follows: B= *BamHI*, Ev= *EcoRI*, Nd= *NdeI*, Kp= *KpnI*, E= *EcoRI*, Ss= *SstII* (*SacII*), S= *SalI*, Pv= *PvuII*, H= *HindIII*, Sp= *SphI*, Ps= *PstI*, Sm= *SmaI*, Sc= *SstI* (*SacI*). The arrow denotes the direction of transcription with respect to the *lacUV-5* promoter of pUC.
Fig. 3.2.3.
Restriction maps of pDR1830 and pDR1930.
Table 3.2.4.

Activities of HPC dioxygenase and CHM isomerase from various subclones

<table>
<thead>
<tr>
<th>Subclone</th>
<th>IPTG</th>
<th>Protein Concentration (mg.ml⁻¹)</th>
<th>CHM isomerase</th>
<th>HPC dioxygenase</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDR1830</td>
<td>-</td>
<td>3.8</td>
<td>NDA</td>
<td>NDA</td>
</tr>
<tr>
<td>pDR1830</td>
<td>+</td>
<td>3.6</td>
<td>NDA</td>
<td>NDA</td>
</tr>
<tr>
<td>pDR1930</td>
<td>-</td>
<td>3.6</td>
<td>5.4</td>
<td>8.7</td>
</tr>
<tr>
<td>pDR1930</td>
<td>+</td>
<td>3.5</td>
<td>5.6</td>
<td>7.3</td>
</tr>
</tbody>
</table>

Activity (A)= µmol.min⁻¹.ml⁻¹
Specific Activity (SA)= µmol.min⁻¹.mg protein⁻¹
NDA= No Detectable Activity
SDS-PAGE analysis of crude extracts from cells carrying pDR1830 and pDR1930.

Lanes 1. Marker proteins ranging in size as shown.
2. pDR1830 -IPTG
3. pDR1830 +IPTG
4. Partially purified sample of CHM isomerase.
5. pDR1930 -IPTG
6. pDR1930 +IPTG
7. Marker proteins ranging in size as shown.

HPC dioxygenase and CHM isomerase enzyme activity was found in the pDR1930 extracts +/ IPTG. A band which migrates with the same mobility as the partially purified sample of CHM isomerase occurs in the pDR1930 extracts as does a band of approximately 33 KDa which was thought as that due to the HPC dioxygenase enzyme (T. Fawcett pers comm). Neither band is found in the pDR1830 extracts.
Fig. 3.2.4.
SDS-PAGE analysis of extracts prepared from cells transformed with pDR1830 and pDR1930.
3.2.5. REVISION OF THE GENE ORDER.

The arrangement of the five complete hpc genes encoded upon pJJ801 from this study is now different to that first proposed by Jenkins (1987). Jenkins describes the gene order for pJJ801 as being hpcBCDEF. The experiments described above, clearly demonstrate that hpcD and hpcB are separated from hpcE and hpcF. However, at this stage no definite gene order could be assigned to this region of the Hpc operon. Experiments to resolve this question are described in section 3.3.

Fawcett (1989) proposed the most recent model for the order of the genes on pJJ801. This model is based upon several pieces of evidence. The 5' region of the hpcC encoding CHMS dehydrogenase has been localised to the 3.8 Kbp SalI-BamH1 fragment of pJJ801 by Southern blotting (Fawcett et al. 1989). Purified CHMS dehydrogenase was subjected to amino terminal amino acid sequencing by automated techniques. The first 34 amino acid residues of the protein were obtained in this way from which a degenerative 17mer oligonucleotide probe was constructed corresponding to residues 6-10 of the amino acid sequence. The oligonucleotide probe was then used to detect the 5' end of the hpcC gene in a southern blot experiment. This experiment in conjunction with enzyme activities detected in a variety of different subclones produced by Fawcett indicated that hpcC lies across the internal SalI and EcoR1 sites of pJJ801 with the 5' end of the gene to the right of the SalI site as pJJ801 is drawn (see Fig.3.2.3. for example). This is in the opposite direction to the Jenkins model (1987).

The 5' end of the OHED hydratase gene, hpcG was located in a similar way. A degenerative 17mer oligonucleotide probe, corresponding to the first six residues of the purified hydratase protein was constructed and used to probe against various restriction fragments of the genomic subclones pJJ801 and
pJJ200 in a Southern blot experiment. The 5' end of \textit{hpcG} was located in this way, to a 0.3 Kbp \textit{EcoRV-EcoRV} fragment which lies toward the left hand end of the cloned DNA, as pJJ801 is drawn (See Fig.3.2.3 for example). The genomic clone, pJJ801 does not express an active OHED hydratase enzyme, so that the left hand \textit{BamH}1 site as drawn, must occur within the open reading frame for the \textit{hpcG} gene, in the same way that the \textit{Sal1} and \textit{EcoR1} sites do for the \textit{hpcC} gene. The location of the 5' end of \textit{hpcG} gene was further corroborated by Fawcett upon nucleotide sequencing of this region of pJJ801. An open reading frame (ORF) corresponding exactly to the amino terminal amino acid sequence of the purified protein, was located between the two \textit{EcoRV} sites of pJJ801. The direction of transcription of the \textit{hpcG} gene was toward the left hand \textit{BamH}1 site of pJJ801 as drawn, with the ATG-start codon lying 4 nucleotides downstream of the most internal \textit{EcoRV} site. Fawcett attempted to order the five intact genes of pJJ801: \textit{hpcB}, \textit{hpcC}, \textit{hpcD}, \textit{hpcE} and \textit{hpcF}, given the information obtained as described above, in the following way. The deduced gene order of \textit{hpcB} and \textit{hpcD} was based upon the inducibility of \textit{hpcB} and constitutive expression of \textit{hpcD}. Such behaviour is displayed by subclones of pJJ801, when present in \textit{E.coli} JJ221 and \textit{E.coli} 5K which are regulated and unregulated background strains respectively. Fawcett constructed pTF306 which contained a 5.2 Kbp \textit{BamH}1-\textit{PvuII} of pJJ801 in pUC18 and expressed \textit{hpc(FE)C(DB)} constitutively in 5K. When present in the regulated background strain: JJ221 and grown in glycerol (ie non-inducing conditions) the subclone pTF306 expressed \textit{hpc(FE)CD} constitutively, but expressed \textit{hpcB} at a low levels as assessed by measurement of the specific activities for individual enzymes. This suggested that the \textit{PvuII} site lies either in a operator/promoter region or between the promoter and the translational start signal of either \textit{hpcE} or \textit{hpcF} gene. This would then mean that the expression of \textit{hpc(FE)CD} was under the control of the vector's \textit{lac UV-5}
promoter and is therefore not subject to transcriptional regulation by the product of the chromosomally encoded Hpc repressor gene. The low levels of HPC dioxygenase activity measured from pTF306 in *E.coli* C JJ221, suggest that the *hpcB* gene is regulated and therefore in a different regulatory group to *hpc(FE)CD*. This hypothesis was supported by the measurement of enzyme activities in *E.coli* C JJ221(pTF306) when grown on glycerol in the presence of the inducer: 4-HPA. In such conditions the levels of the enzymes encoded by *hpc(FE)CD* increase by approximately 1.5-fold according to Fawcett, whereas the level of HPC dioxygenase encoded by *hpcB* increases 9-fold. The presence of 4-HPA in the growth medium also induces expression of the chromosomal *hpc* genes. The contribution of the chromosomal gene expression was taken into account during the measurement of HPC dioxygenase and CHM isomerase activities in these experiments. Fawcett also subcloned a 2.4 Kbp *BamH1-EcoR1* fragment of pJJS01 into pBR322 to create pTF302 which expressed *hpcB* and *hpcD* only. When expression of *hpcD* and *hpcB* was tested in 5K and JJ221, the expression of *hpcB* appeared inducible, whereas that of *hpcD* appeared constitutive. On this basis Fawcett proposed that there are at least two transcriptional units in pJJS01, one of which lies 5' to *hpcB*, such that the gene order is *hpc(FE)CDB*.

### 3.2.6. APPARENT INDUCIBILITY OF *hpcB* AND *hpcD* IN pDR1930.

The apparent inducibility of *hpcB* was tested in 5K and *E.coli* C JJ221 using pDR1930 in order to independently test Fawcett's results with pTF302. These two constructs differ only in that pDR1930 contains an extra 0.5 Kbp *EcoR1-Sal1* fragment upstream of *hpcBand hpcD*. *E.coli* 5K and JJ221 cultures harbouring pDR1930 were grown in minimal glycerol media in order to test constitutive and regulated expression of *hpcB* and *hpcD*. The results of this experiment are shown in Table 3.2.6.
Table 3.2.6.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Protein Concentration (mg.ml⁻¹)</th>
<th>CHM isomerase Activity</th>
<th>Specific Activity</th>
<th>HPC dioxygenase Activity</th>
<th>Specific Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5K</td>
<td>3.4</td>
<td>24.0</td>
<td>7.1</td>
<td>22.6</td>
<td>6.6</td>
</tr>
<tr>
<td>JJ221</td>
<td>2.5</td>
<td>4.8</td>
<td>1.9</td>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Activity = μmol.min⁻¹.ml⁻¹
Specific Activity = μmol.min⁻¹.mg protein⁻¹
The results obtained here are analogous to those observed with pTF302 by Fawcett. CHM isomerase is apparently expressed only 3.7 fold higher in the unregulated strain, whereas the HPC dioxygenase is expressed at 33 fold higher levels in the unregulated background. These results would suggest therefore that the expression of the hpcB gene encoding HPC dioxygenase, is regulated in pDR1930 and thus in a different regulatory group to hpcD, which encodes CHM isomerase. However, as detailed in this and later chapters, this model for the gene order in the region encoding hpcB and hpcD is not correct. More rigorous analysis based on the presence or absence of enzyme activities in various subclones and upon the actual nucleotide sequence of this entire region confirms that hpcD lies downstream from hpcB. These results are discussed in detail in section 3.3.1. and in chapter 4.

3.2.7. DIRECTION OF TRANSCRIPTION OF THE HPC GENES ENCODED ON pJJ801.

The original model for the organisation and direction of transcription of the seven structural genes cloned by Jenkins was of two distinct regulatory groups, separated by a divergent operator/promoter region. Jenkins proposed that hpcGH were separated from hpcBCDEF by a divergent promoter region between hpcB and hpcG. However, several pieces of evidence contradict both the gene order as discussed previously and in later sections, and also the direction of transcription of hpcBCDEF encoded upon pJJ801. Based upon the expression characteristics of constructs under the control of the lac UV-5 promoter of pUC presented in this study and upon those reported by Fawcett (1989), the direction of transcription is from right to left with reference to pJJ801, ie hpc(EF)CBD. In addition the amino terminal region of hpcG encoding the HHED hydratase was found by Fawcett to lie 4bps downstream of the most internal EcoRV site of pJJ801, such that the direction of transcription
of this gene is in the same orientation as the intact genes encoded on pJJ801. This model for the direction of transcription of the gene on pJJ801 is based upon the effect of the strong lacUV-5 promoter of pUC on the expression of the hpc genes. In this system a gene whose transcription is in the same direction as the lac UV-5 promoter of the host plasmid, would be expected to have an increased number of transcripts which could therefore undergo translation, hence giving high specific activities for the enzymes encoded upon that transcript. When the direction of transcription of the cloned DNA is opposed to that of the lac UV-5 promoter of the host plasmid there would be a reduction in the number of transcripts, due to collision of RNA polymerases travelling in opposite directions (Brewer 1988) or the binding of sense and anti-sense mRNA's (Coleman et al, 1984).

In this study, active Hpc enzymes were only expressed in the constructs where the insert DNA corresponds to transcription in the right to left orientation with respect to pJJ801, i.e. pDR1930 and pDR1835. In the remaining two constructs where the orientation of the insert DNA is reversed with respect to the lac UV-5 promoter, no active enzymes are detected, i.e. in pDR1830 and pDR1935.

3.3. CONSTRUCTS REQUIRED TO DETERMINE THE GENE ORDER OF hpcD AND hpcB.

The exact order of the genes encode upon the 3.0 Kbp BamH1-SalI fragment of pJJ801 was determined in the first instance by the creation of a series of unidirectional deletion subclones in pDR1930. The hpcBD gene order was then derived on the basis of enzyme activities associated with certain subclones. The methodology required to create these subclones is dealt with in more detail in chapter 2 and more detailed descriptions of the individual subclones are given in chapter 5. In essence however, the technique relies on
the 5' to 3' exonuclease activity of Exonuclease III which under certain conditions can be used to create unidirectional deletions in a given DNA fragment. Deletions were made from the SalI end of pDR1930 such that the cloned DNA was deleted in part, but the pUC vector DNA including both the sequencing primer site and lac UV-5 promoter remained intact. This not only facilitates nucleotide sequencing of the inserted DNA, but effectively brings any open reading frame in the cloned DNA, that is in the correct orientation, nearer to the lac UV-5 promoter enabling possible high level expression.

A series of deletions from the SalI end of pDR1930 showed that the HPC dioxygenase enzyme activity was lost before that of the CHM isomerase enzyme. The deletion subclone pDR9305 which contains a 1.9 Kbp insert is the smallest clone of the series to still retain expression of the hpcB gene encoding HPC dioxygenase. Expression of hpcD encoding the CHM isomerase is maintained in deletion subclones which have inserts larger than approximately 1.0 Kbp, such that pDR9317 is the smallest deletion subclone to still retain full CHM isomerase activity. This is shown diagrammatically in Fig.3.3.1.1.

These results were augmented by the creation of a number of other subclones derived from pDR1930. The availability of restriction sites within the cloned fragment and also within the polylinker of pUC19 enabled these constructs to be quickly created. A 4.2 Kbp derivative of pDR1930 was created by digesting pDR1930 with NdeI and subsequent extraction of the 4.7 Kbp fragment from a 1% low melting point agarose gel. The intact fragment was ligated in the presence of T4 DNA ligase and single colonies analysed after transformation of competent cells with the ligation mixture. The resulting construct was called pDR947 and expresses HPC dioxygenase only, implying that the single NdeI site lies within or 5' to the CHM isomerase gene: hpcD.
Deletion subclones required to establish the *hpcBD* gene order.

A series of deletion subclones were created from the *Sal*I site of pDR1930 and the enzyme activities associated with individual subclones tested. The genes expressed by individual subclones are shown for four constructs. Only the insert DNA for the deletion subclones is shown, for clarity. The *Bam*HI and *Hind*III sites denote the boundaries of the insert DNA after *Exo*III digestion. The deletion subcloning strategy retains all of the pUC DNA except that spanning the *Sal*I, *Pst*I and *Sph*I restriction sites. Restriction sites are as follows: B=*Bam*H1, Ev=*Eco*RV, Nd=*Nde*I, Kp=*Kpn*I, E=*Eco*RI, Ss=*Sac*II, S=*Sal*I, Pv=*Pvu*I, H=*Hind*III, Sp=*Sph*I, Ps=*Pst*I, Sm=*Sma*I, Sc=*Sst*I (*Sac*I). The arrow denotes the direction of transcription with respect to the *lacUV-5* promoter of pUC.
Fig. 3.3.1.1.
Deletion subclones required to establish the hpcBD gene order.

EScKpSmB  I  I  I  I  Nd  SPsSpH  

pDR1930  pDR9305  hpcD+ hpcB+  
pDR9306  hpcD+ hpcB-  
pDR9317  hpcD+ hpcB-  
pDR9316  hpcD- hpcB-  

0  1 Kbp
Nucleotide sequencing work carried out by Fawcett (1989) showed that the open reading frame for the OHED hydratase enzyme of the pathway lies 4 bps downstream from the most internal EcoRV site with the direction of transcription toward the BamH1 site. The \textit{hpcD} gene must therefore lie upstream of this EcoRV site and downstream of the \textit{Kpn1} site, presumably with the \textit{Nde1} site occurring within the open reading frame for \textit{hpcD} since pDR947 does not express an active CHM isomerase. The constructs used to determine the \textit{hpcBD} gene order are shown in Fig.3.3.1.2.

These results suggest that the gene order is \textit{hpcBD} in this region of the cloned DNA and contradicts the gene order based from the apparent inducibility of the \textit{hpcB} gene and constitutive expression of \textit{hpcD}. These results were corroborated by nucleotide sequence analysis of the 3.0 Kbp \textit{Bam H1 - Sal1} region. The open reading frames identified in this region were confirmed by matching the deduced amino acid sequences of the open reading frames to the actual amino terminal amino acid sequences of the purified proteins. This work is detailed in chapter 4 and chapter 5.
Subclones created from pDR1930 to corroborate the gene order deduced from the deletion subcloning strategy.

pDR1930 expresses both \textit{hpcB} and \textit{hpcD} encoding HPC dioxygenase and CHM isomerase, respectively. When the 1.2 Kbp \textit{BamHI-NdeI} fragment is lost as in pDR947, only \textit{hpcB} is expressed. When a further 0.5 Kbp is lost from the same end to create pDR934, such that the 1.5 Kbp \textit{KpnI-SalI} remains, the expression of \textit{hpcB} is lost. Restriction sites are as follows: \textit{B=BamHI}, \textit{Ev=EcoRV}, \textit{Nd=NdeI}, \textit{Kp=KpnI}, \textit{E=EcoRI}, \textit{Ss=SstII (SacII)}, \textit{S=SalI}, \textit{Pv=PvuII}, \textit{H=HindIII}, \textit{Sp=SphI}, \textit{Ps=PstI}, \textit{Sm=SmaI}, \textit{Sc=SstI (SacI)}. The arrow denotes the direction of transcription with respect to the \textit{lacUV-5} promoter of pUC.
Fig.3.3.1.2.
Subclones created from pDR1930.
3.4. ORDER OF THE GENES ENCODING HHDD ISOMERASE AND COHED DECARBOXYLASE.

The order of the $hpcF$ and $hpcE$ genes encoding HHDD isomerase and COHED decarboxylase, respectively was originally postulated by Jenkins (1987) on the basis of enzyme activities expressed from subclones created by the isolation of restriction fragments from, and partial $Sau3A1$ digestions of, pJJ801. In the model according to Jenkins, the $hpcE$ gene lies upstream from $hpcF$, with the 3' end of the gene lying close to the single $PvuII$ site that occurs in pJJ801. Fawcett tested this model in the light of evidence that indicated the gene order and direction of transcription was different to that postulated by Jenkins. Fawcett produced a subclone of pJJ801 called pTF306 which contains the 5.3 Kbp $BamH1-PvuII$ fragment of pJJ801 in the vector pBR328. This construct is analogous to one created by Jenkins called pJJ006 except that the vector used by Jenkins was pBR322. Jenkins reported that pJJ006 expressed $hpcBCDE$, but not $hpcF$ and inferred that the $PvuII$ site lies within an open reading frame for $hpcF$. However, the subclone produced by Fawcett: pTF306 expressed $hpcBCDEF$, contradicting this result. Further experiments by Fawcett failed to distinguish the $hpc(EF)$ gene order.

Earlier trial experiments carried out in this study and in that conducted by Fawcett (1989) indicated that HHDD isomerase and COHED decarboxylase were proteins of approximately 50 KDa. As such, approximately 2.5 Kbp would be required to encode $hpcE$ and $hpcF$. It was thought that the creation of a unidirectional deletion series from the $BamH1$ end of pDR1835 would produce a construct where the 5' end of the gene cluster was brought within a few base pairs of the $lacUV-5$ promoter. Such a construct may have expressed HHDD isomerase and COHED decarboxylase at elevated levels, facilitating
purification of both enzymes. An early attempt to create a unidirectional deletion series of subclones from the BamH1 end of pDR1835 gave only a few useful subclones one of which was retained a 2.5 Kbp fragment. (This construct was later designated pDR83501) A 2.5Kbp EcoR1 - Pst1 fragment was isolated from pDR83501 and was cloned into the inducible high expression vector pKK223-3 which had been digested with the same restriction enzymes. Following ligation and subsequent transformation of E.coli JM105, single colonies were picked and screened for the desired recombinant subclone. Several colonies contained the same construct, which was called pDRK14. Inducible expression of hpcEF from pDRK14 was tested and compared to that produced by pDR1835 and pDR83504. The results are shown in Table 3.4. The construct pDRK14 is shown in Fig 3.4.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Protein concentration (mg.ml(^{-1}))</th>
<th>COHED decarboxylase (A)</th>
<th>COHED decarboxylase (SA)</th>
<th>HHDD isomerase (A)</th>
<th>HHDD isomerase (SA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDR1835</td>
<td>11.0</td>
<td>6.10</td>
<td>0.55</td>
<td>5.65</td>
<td>0.50</td>
</tr>
<tr>
<td>pDR83504</td>
<td>10.7</td>
<td>7.16</td>
<td>0.67</td>
<td>8.24</td>
<td>0.77</td>
</tr>
<tr>
<td>pDRK14</td>
<td>7.5</td>
<td>5.97</td>
<td>0.80</td>
<td>3.15</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Activity (A) = \(\mu\text{mol.min}^{-1}.\text{ml}^{-1}\)
Specific Activity (SA) = \(\mu\text{mol.min}^{-1}.\text{mg protein}^{-1}\)
pDRK14

pDRK14 was constructed by cloning the 2.5 Kbp EcoRI-SalI fragment of pDR83504 into the high expression vector pKK223-3. The cloned genes are expressed from the EcoRI end of the polylinker under the control of a plasmid encoded tac promoter. The direction of transcription with respect to the vector encoded tac promoter is shown by the arrow. Restriction sites are as follows: B=BamHI, Ev=EcoRV, Nd=NdeI, Kp=KpnI, E=EcoRI, Ss=SstII (SacII), S=SalI, PvI=PvuI, PvII=PvuII, H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI, Sc=SstI (SacI).
Fig. 3.4.

pDRK14
There would appear to be no increase in specific activity in the pDRK14 construct compared to pDR85304. It has been reported (D.B. Wigley. pers comm) that generally, no significant increase in specific activity is seen in pKK223-3 derived plasmids, until the vector encoded tac promoter is brought within 200bps of the ribosome binding site of the gene of interest. The reason for this lack of induction became clearer when a complete deletion series from the Sal1 end of pDR1835 was produced. This deletion series was being isolated and sequenced at the same time as the pDRK14 experiments were carried out. It was expected that deletions made from this Sal1 end would result in the loss of one enzyme activity rather than both, since all the region upstream of hpcEF including any Hpc promoter, was preserved in this deletion series. However, when enzyme activities from several of the pDR853 series subclones were tested on a number of occasions, no separation of HHDD isomerase and COHED decarboxylase enzyme activities could be made. The constructs expressed both enzymes or neither. Each of the constructs in this series was sequenced concurrently and showed the isolated subclones to be true members of the deletion series and not spontaneous recombinants, since the nucleotide sequence obtained from them overlapped. One possible explanation for these results would be that the HHDD isomerase and COHED decarboxylase activities were due to the same gene product.

This hypothesis was tested by the creation of a deletion series from the BamH1 end of pDR1835 as detailed in chapter 5. Both HHDD isomerase and COHED decarboxylase activities were detected in constructs which retained inserts of larger than 1.8 Kbp. However, in constructs smaller than this size both activities were lost. Several of the constructs created in this series, notably those which retained less that 200 bps between the lacUV-5 promoter of pUC and the genes in this 1.8 Kbp region, showed elevated levels of expression for the HHDD isomerase and COHED decarboxylase activities. One of these
constructs pDR83507, was used to purify a protein which displayed both enzyme activities and had a subunit molecular weight of 50KDa by SDS-PAGE. The purification and amino terminal amino acid sequence of this enzyme is detailed in chapter 4.

3.5.1. **INVESTIGATING THE POSITION OF THE GENES ENCODING HHED ALDOLASE AND THE Hpc REGULATORY PROTEIN.**

The 5' end of hpcG which encodes the OHED hydratase enzyme has been located 4 bps downstream of the most internal EcoRV site in pJJ801. Since pJJ801 does not express an active hydratase enzyme it is assumed that the vector encoded BamH1 site, lies within the open reading frame of the gene, approximately 600 bps downstream from the 5' end of hpcG. The OHED hydratase protein has been purified and the amino terminal sequence determined for the first 34 residues (Ferrer & Cooper 1988). The amino terminal amino acid sequence of the enzyme is in perfect agreement with the predicted amino acid sequence and has been used to positively identify the hpcG open reading frame (Fawcett 1989). The purified protein has a subunit molecular weight 32.5 KDa by SDS-polyacrylamide gel electrophoresis which would predict that the hpcG gene would encode an ORF of approximately 0.9 Kbp. The two genomic subclones that express all the enzymes for the complete degradation of HPC to pyruvate and succinic semialdehyde: pJJ200 and pJJ210 do not contain the left hand (as drawn) BamH1 site of pJJ801, since pJJ801 was created by a partial digestion of the genomic DNA using Sau3A1 followed by ligation into BamH1 restricted pBR328. As such, it is not possible to simply restrict pJJ200 or pJJ210 with BamH1 in order to derive clones to more conveniently study the complete genes which are not found in pJJ801. According to Jenkins (1987) the genomic subclones pJJ200 and pJJ210 contain an active regulatory gene since the expression of the pathway enzymes
in cells harbouring these plasmids, only occurs when the inducer 4-HPA is included in the growth medium. From a variety of different subcloning procedures Jenkins assigned the \textit{hpcR} gene encoding the regulator protein to lie adjacent to \textit{hpcH} encoding HHED aldolase. A variety of workers have tried to study the repressor gene product in more detail acting on this information but have met with little success. A more systematic approach was adopted in this study by \textit{EcoR} digest of pJJ200. From the plasmid maps produced by Jenkins in was known that an \textit{EcoR} digest of pJJ200 would yield two fragments of similar molecular weight which could only be separated by low percentage agarose gels. Approximately 5\(\mu\)g of pJJ200 plasmid DNA was digested overnight with \textit{EcoR} and then electrophoresed through a 0.5\% LMP agarose gel until two bands of approximately 7.0 Kbp and 6.5 Kbp were clearly separated. The two bands, designated \(F_1\) and \(F_2\) respectively, were extracted from the gel by the phenol method and a small sample of each run on an agarose gel to check the recovery. Approximately 100ng of each fragment were then ligated for 2 hours at room temperature before the ligation mixture was used to transform \textit{E.coli} 5K. The transformation mixes were then plated onto freshly made LB-ampicillin plates. After growth overnight at 37°C, only the \(F_1\) fragment gave rise to ampicillin resistant colonies.

This experiment established which fragment retained the majority of pBR322 and was therefore ampicillin resistant and viable. The ampicillin sensitive fragment, \(F_2\) was then cloned into \textit{EcoR} and CIP treated pUC18 to allow further study of the genes it encoded. This was achieved by using a 2:1 molar excess of \(F_2\) to pUC18 in the ligation reactions. The clones were initially screened using blue/white selection on LB-ampicillin plates supplemented with X-gal and IPTG. Three white showed the presence of the \(F_2\) fragment in pUC18 when analysed by mini-prep and subsequent digestion with \textit{EcoR}1. One such construct was designated pDRF\(_2\). Jenkins (1987) positioned the regulatory
gene: hpcR downstream of hpcG and hpcH. When this piece of information was combined with the gene order established in this study, it was though that the F2 fragment would express hpcBDGH in a regulated manner. This was tested by growth of pDRF2 in LB liquid cultures, in the presence and absence of the inducer 4-HPA, using E.coli 5K as the host strain. The expression of HPC dioxygenase was used as a marker in these experiments but no detectable activity was found. No activity was found for COHED decarboxylase in these extracts either, suggesting that the genes encoded on the F2 fragment are orientated in opposition to the lac UV-5 promoter of pUC18 in all the three constructs identified. This result would be corroborated by analysis of enzymes produced by the ampicillin resistant F1 fragment. However, when this plasmid was grown in the presence and absence of 4-HPA, COHED decarboxylase activity was only found associated with the induced cells. This result suggests that the regulatory gene hpcR is located at the opposite end of the gene cluster to that described by Jenkins (1987). A restriction map of pDRF1 is shown in Fig.3.5.1. The presence of hpcR in pDRF1 was independently confirmed by repression of hpc gene expression in cis. E.coli 5K (hpa+) is a derivative of E.coli 5K which has had the hpa genes introduced by P1 transduction. The hpa and hpc genes are constitutively expressed in this strain and it is able to grow on 4-HPA as the sole carbon source. When E.coli 5K (hpa+) was transformed with pDRF1, expression of the Hpc pathway enzymes became regulated as judged by measurement of the enzyme activities from cells grown in the presence and absence of 4-HPA (data not shown).
Fig. 3.5.1.

Restriction map of pDRF₁

pDRF₁ was created by an EcoR1 digest of pJJ200 (Fig. 3.1.1.) and retains the 4.0 Kbp BamH1-EcoR1 fragment of pBR322 and approximately 3.9 Kbp of the cloned E.coli C DNA. This construct was found to be inducible and expressed hpcEF. Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdeI, Kp=KpnI, E=EcoRI, Ss=SstII (SacII), S=SalI, PvI=PvuI, PvII=PvuII H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI, Sc=SstI (SacI).
Fig. 3.5.1.

Restriction map of pDRF₁

E S Pv' Pv'N B B S Pv'N d PsPv' E

0 1 Kbp

pDRF₁
3.5.2. OBTAINING EXPRESSION OF THE GENES ENCODED ON F\textsubscript{2} AND CONFIRMING THE \textit{hpcGH} GENE ORDER.

According to the available restriction maps of pJJ200 the 6.5 Kbp \textit{EcoR1-BamH1} fragment of pDRF\textsubscript{2} could not easily be orientated in pUC using the sites within the insert by simple subcloning procedure. Because pDRF\textsubscript{2} is derived from the pBR322 based plasmid pJJ200, it was possible to obtain a 6.8 Kbp \textit{EcoR1-HindIII} fragment, (where the \textit{HindIII} site is derived from pBR322) which could be orientated in pUC with respect to the vectors \textit{lac UV-5} promoter such that the cloned genes could be expressed. Approximately 2\mu g of pDRF\textsubscript{2} and 2\mu g of pUC18 were cut with \textit{EcoR1} and \textit{HindIII}, and the required fragments isolated from a 0.8% LMP gel. After ligation the mixture was used to transform competent \textit{E.coli DH5\alpha} and single colonies were analysed by mini prep and restriction digest. The clone isolated by this procedure was called pDR1865. Further restriction digests of pDR1865 showed the presence of a single \textit{BamH1} site whereas two were expected. It appears that a previously unmapped \textit{HindIII} site occurs just inside the most external \textit{BamH1} site in pDRF\textsubscript{2} such that pDR1865 carries no DNA in the insert derived from pBR322. According to Jenkins this fragment should carry the regulatory gene \textit{hpcR} so expression of the genes should be inducible contrary to the results reported in section 3.5.1. Expression of \textit{hpcBDGH} in the presence and absence of 4-HPA, was tested in pDR1865 by preparing crude extracts grown from \textit{E.coli DH5\alpha} transformed with pDR1865. The results of this analysis are detailed in Table 3.5.2. As can be seen from the specific activities recorded there seems to be no difference in the expression of these four genes in the presence or absence of 4-HPA. This finding corroborates the results detailed in the previous section where the repressor function was mapped to the other end of the cloned genes.
Table 3.5.2. 
Expression of Hpc enzymes from pDR1865 during induced and non-induced growth.

<table>
<thead>
<tr>
<th>Presence of 4-HPA</th>
<th>Protein concentration (mg.mL⁻¹)</th>
<th>HPC dioxygenase S.A.</th>
<th>CHM isomerase S.A.</th>
<th>OHED hydratase S.A.</th>
<th>HHED aldolase S.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>9.5</td>
<td>1.87</td>
<td>0.58</td>
<td>0.59</td>
<td>1.33</td>
</tr>
<tr>
<td>-</td>
<td>9.5</td>
<td>1.78</td>
<td>0.60</td>
<td>0.40</td>
<td>2.12</td>
</tr>
</tbody>
</table>

where S.A. = Specific activity measured in μmol.min⁻¹.mg⁻¹.
Once pDR1866 was isolated, two derivatives were quickly produced to confirm that \textit{hpcH} is located 3' to \textit{hpcG}. Two separate digests were set up. In one, pDR1865 was digested with \textit{BamH}I and \textit{Hind}III and in the other pDR1865 was digested with \textit{Sal}I and \textit{Hind}III. After the digest was allowed to go to completion the restricted DNA was precipitated and redissolved so that the fragments could be treated with mung bean nuclease to produce blunt ends. The blunted fragments were then run onto a 1.0% LMP agarose gel and the larger fragment from each digest was isolated, extracted from the gel and religated. Competent cells were then transformed with the ligation mixture after which single colonies were analysed by mini prep followed by restriction digest.

Two clones were identified. pDR832 contains the \textit{EcoRI} to \textit{SalI} region of pDR1865 in pUC18 whereas pDR836 contains the \textit{EcoRI} to \textit{BamH}I region of pDR1865 in pUC18. Competent \textit{E.coli} DH5\(\alpha\) were transformed with these plasmids and crude extracts prepared after overnight growth on LB-ampicillin. Measurement of the enzyme activities associated with these two plasmids showed the presence of HPC dioxygenase, CHM isomerase and OHED hydratase but not HHED aldolase. As such both the \textit{SalI} and \textit{BamH}I may lie within the \textit{hpcH} ORF and which lies 3' to \textit{hpcG}. Restriction maps of pDRF\(2\), pDR1835, pDR832 and pDR836 are shown in Fig.3.5.2.
Fig. 3.5.2.

Restriction maps of pDRF₂, pDR1865, pDR832 and pDR836

The arrow denotes the direction of transcription with respect to the lacZ promoter of pUC. The cloned DNA of pDRF₂ contains a short 300 bps region derived from pBR322 and is shown by the hatched area. The HindIII site within the cloned DNA of pDRF₂ was not identified until the construction of pDR1865.

pDRF₂ expresses none of the Hpc genes, since the direction of transcription from the lacUV-5 promoter is in the opposite orientation to which they are transcribed. pDR1865 expresses both HpcG, encoding OHED hydratase, and hpcH encoding HHED aldolase. pDR836 and pDR832 only express hpcG, indicating that the SalI and BamH1 sites lie within the hpcH open reading frame. Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdeI, Kp=KpnI, E=EcoRI, Ss=SstII (SacII), S=SalI, Pv=PvuII, H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI, Sc=SstI (SacI).
Fig. 3.5.2.

Restriction maps of pDRF<sub>2</sub>, pDR1865, pDR832 and pDR836.
3.5.3. PROBING THE REGION ENCODING THE Hpc REGULATOR GENE (hpcR).

The finding that hpcR is located upstream of hpcEF was unexpected. According to Jenkins (1987), when the 7.0 Kbp BamH1-BamH1 fragment of pJJ200 was isolated and cloned into pBR328, all the hpc genes with the exception of hpcH encoding HHED aldolase, were expressed in a constitutive manner. This was interpreted as meaning that the remaining 3.0 Kbp BamH1-BamH1 fragment encoded part of the aldolase and all of hpcR. But there is another possible explanation: there are two BamH1 sites upstream of the 3' end of hpcEF in pJJ200 which are very close to each other. If the most internal site were located within hpcR then the isolation of fragments with BamH1 ends from pJJ200 would result in the loss of regulation. This was tested simply by digesting pDRF1 with BamH1, followed by isolation of the single high molecular weight band seen in an LMP gel. The fragment was then religated and used to transform competent E.coli 5K. Three individual colonies were then grown in small liquid cultures and the plasmid DNA analysed by mini-prep and restriction digest. The plasmid isolated was designated pDRF1'. (Fig.3.5.3.1.). Two 100ml LB cultures were inoculated with 1ml of the small overnight cultures. One of the cultures was then supplemented with 4-HPA to 4mM to test for regulation of hpcEF. In each case the expression of hpcEF was unregulated, indicating that a small BamH1 fragment had been lost during the procedure which therefore destroyed the ability of hpcR to regulate hpcEF.
Fig. 3.5.3.1.

Restriction map of pDRF_1'

pDRF_1' was derived from pDRF_1 by BamH1 digestion followed by extraction of a 6.6 Kbp fragment from an LMP gel. This enabled pDRF_1' to be obtained which differs from pDRF_1 by the removal of a 600 bps BamH1-BamH1 fragment. pDRF_1' expresses \( hpcEF \) constitutively whereas pDRF_1 expresses \( hpcEF \) in a regulated manner. Restriction sites are as follows: B = BamH1, B' = Pseudo, nonfunctional BamH1 site of pBR322, Ev = EcoRV, Nde = NdeI, Kp = KpnI, E = EcoRI, Ss = SstII (SacII), S = SalI, Pvi = PvuI, Pvi" = PvuII, H = HindIII, Sp = SphI, Ps = PstI, Sm = SmaI, Sc = SstI (SacI).
Fig. 3.5.3.1.

Restriction map of pDRF'.
This was investigated further by nucleotide sequencing from the BamH1 end of pDRF₁ and is discussed in more detail in section 5.13. A synthetic oligonucleotide probe was produced which would anneal to the region of pBR322 immediately upstream of the vector encoded BamH1 site so that nucleotide sequence information at the far right hand end (as drawn) of the E.coli C DNA could be obtained. This procedure yielded several pieces of information. Firstly, the vector encoded BamH1 site has been lost during the original cloning strategy and two BamH1 sites occur within the cloned DNA. The inner of the two sites lies within 100 bps of the silenced pBR322 BamH1 site and has previously been mistakenly thought to be that of the vector. The region between the two BamH1 sites in the cloned DNA encodes the amino terminal region of the serine chemoreceptor protein (methyl accepting chemotaxis protein I). The ‹tsr› gene of E.coli K12 (alternative name: cheD), which encodes this protein, maps to minute 99 of the E.coli K12 linkage map (Backman 1987). The region between the most internal BamH1 site and the silenced BamH1 site of pBR322, does not encode the ‹tsr› gene and has not been identified during searches of available nucleic acid databases. The presence may have arisen during the original ligation reaction which created pJJ200. Alternatively this region may have been present in the pBR322 DNA originally used. The region between the two BamH1 sites in the cloned DNA may also contain part of the ‹hpcR› open reading frame, which would explain why a functional repressor is not expressed from pDRF₁' (Fig.3.5.3.2.)

As detailed in section 5.13, further oligonucleotide primers were constructed to obtain the nucleotide sequence of the region between the 5' end of the ‹tsr› gene to a point immediately upstream of the ‹hpc› structural gene promoter. This 900 bps region should therefore contain ‹hpcR›.
pDRF₁ is shown for reference, and the kilobase scale refers to it. Regions encoding genes are marked by hatched areas and are transcribed in the directions indicated by the arrows.

A region of approximately 1.25 Kbp is shown enlarged stretching from the 5' end of hpcEF to the non-functional (pseudo) BamH1 site of pBR322. Two functional BamH1 sites occur within the cloned DNA, the outer of which lies within 150 bps of the pseudo BamH1 site of pBR322. This outer functional BamH1 site has previously been mistaken for a vector encoded site due to its close proximity. The inner BamH1 site lies within the hpcR ORF since a construct derived from pDRF₁; pDRF₁', does not express a functional repressor protein. pDRF₁' was constructed from pDRF₁ by removal of the 0.7 kbp BamH1-BamH1 region. A previously unmapped SmaI and PvuII site were identified by nucleotide sequencing (see section 5.13) The 5' region of the tsr gene encoding the serine chemotaxis protein was also identified at the right hand end of pDRF₁ from the nucleotide sequence. A region of approximately 100 bps between the most external BamH1 site of the cloned DNA and the nonfunctional BamH1 site of pBR322 does not show any identity with tsr and was probably inserted during the original construction of pJJ200 from which pDRF₁ was derived.

Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdeI, E=EcoRI, Ss=SstII (SacII), S=SaiI, Pvu= PvuI, Pvu II= PvuII, Sp= SpI, Ps=PstI, Sm= SmaI, Sc= SstII (SacI). B' = Pseudo BamH1 site of pBR322.
Fig. 3.5.3.2.

Schematic diagram of right hand end of pJJ200.
3.6. **ORDER OF THE hpc GENES.**

From the evidence presented in this chapter, which is corroborated by nucleotide sequencing of the hpc gene cluster detailed in chapter 5, the order of the structural genes is \( hpc(EF)CBDGH \). The regulatory gene: \( hpcR \), occurs upstream of \( hpcEF \). (see Fig.3.6.1.)
Schematic diagram of the organisation, location and direction of transcription of the hpc genes with respect to the genomic clone pJJ200.
CHAPTER 4.
ENZYME PURIFICATION.
4.1.1. INTRODUCTION.

Several of the subclones created in this study express the Hpc pathway enzymes at levels in excess of 2% of total cell protein. This has greatly facilitated their purification. Purification of the individual enzymes has allowed two main areas to be studied. Firstly, the purified enzymes have been subjected to amino terminal amino acid sequencing which has enabled the positive identification of the open reading frames of the genes which encode them. This information corroborates the gene order deduced from measurement of enzyme activities from cells carrying the individual deletion subclones as detailed in chapter 3. Secondly, the availability of milligram quantities of the purified enzymes has allowed their physical characterisation and in two cases preliminary crystallisation experiments have begun.

4.1.2. PURIFICATION OF CHM ISOMERASE.

High level expression of CHM isomerase can be achieved from several of the pDR93 series deletion subclones, notably pDR9304 and pDR9317. CHM isomerase was purified from *E.coli* DH5α.(pDR9304). Cells grown overnight at 37°C in 800ml of culture medium were harvested by centrifugation at 10000 xg for 10 min at 0°C. Ultracentrifuged crude extracts were then prepared as detailed in methods and materials. The supernatant was treated with protamine sulphate (40mg.ml⁻¹ in 20mM Tris.Cl buffer pH 7.5) in the proportion of 1mg protamine sulphate to 20 mg bacterial protein, to remove nucleic acids. The extract was stirred at room temperature for 10 min and then centrifuged at 15,000xg for 15 min at 4°C to remove precipitated material. The supernatant was then passed through a 0.22μm acrodisc filter before application to an HR 10/10 Mono Q anion exchange column and chromatographed using a Pharmacia fast protein liquid chromatography (FPLC) system. The column was equilibrated in 20mM Tris.Cl buffer pH 7.5
before application of the sample. A 70ml gradient of 0.0-0.5M NaCl in the same buffer at a flow rate of 4ml.min⁻¹ was used to elute the enzyme. The CHM isomerase was eluted at approximately 0.4M NaCl and the fractions with the highest specific activities were pooled. The volume of the pooled fractions was reduced to 2ml by pressure dialysis, before applying to a Pharmacia HR 5/5 Phenylsuperose column after diluting with an equal volume of 2M ammonium sulphate. A 10ml gradient of 1.7-0.0 M ammonium sulphate in 0.1M sodium phosphate buffer pH 7.5 was applied to the column at a flow rate of 0.5ml.min⁻¹. The CHM isomerase was eluted at the end of the gradient in phosphate buffer only. The fractions with the highest specific activity were pooled and the volume reduced to 2ml by pressure dialysis. The concentrated fractions were then applied to two Pharmacia HR 10/30 Superose 12 gel filtration columns run in series, equilibrated in 50mM sodium phosphate buffer 0.15 M NaCl pH 7.5 at a flow rate of 0.4 ml.min⁻¹. The CHM isomerase was eluted as a single band at greater than 99% purity as estimated by SDS-PAGE (See Fig.4.1.2.) Details of the purification are given in Table 4.1.2. The gel filtration step was also used to estimate the native molecular weight of the purified protein. The gel filtration system was calibrated using proteins recommended by the column manufacturers: immunoglobulin G (160,000), bovine serum albumin (66,000), β-lactoglobulin (35,000) and cytochrome C (12,400). Under the condition stated CHM isomerise elutes with an apparent native molecular weight of approximately 30 KDa which when compared to the subunit molecular weight estimated by SDS-PAGE, suggests the enzyme may exist as a dimer.
Fig. 4.1.2.

SDS-PAGE analysis of fractions taken at each stage of the purification of CHM isomerase

Approximately 25μg of protein were loaded per well on a 7.5% to 20% SDS-polyacrylamide gel as shown.

Lane  A. SDS 7 (Sigma) marker proteins.
     B. Ultracentrifuged crude extract.
     C. Pooled Mono Q fractions.
     D. Pooled Phenylsuperose fractions.
     E. Pooled Superose 12 fractions.
     F. SDS 7 (Sigma) marker proteins.

Molecular weight marker protein sizes are as shown in KDa.
Fig. 4.1.2.
SDS-PAGE analysis of fractions taken at each stage of the purification of CHM isomerase.
Table 4.1.2.
A summary of the purification of CHM isomerase.

<table>
<thead>
<tr>
<th>PURIFICATION STEP</th>
<th>VOLUME (ml)</th>
<th>TOTAL PROTEIN (mg)</th>
<th>TOTAL UNITS (A)</th>
<th>SPECIFIC ACTIVITY (SA)</th>
<th>RECOVERY (%)</th>
<th>PURIFICATION (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultracentrifuged extract</td>
<td>17.5</td>
<td>446.0</td>
<td>2292</td>
<td>5.1</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>P.S. treated ultracentrifuged extract</td>
<td>17.0</td>
<td>395.0</td>
<td>2924</td>
<td>7.4</td>
<td>127.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Pooled Mono Q fractions</td>
<td>6.0</td>
<td>25.2</td>
<td>1656</td>
<td>65.9</td>
<td>72.2</td>
<td>12.9</td>
</tr>
<tr>
<td>Pooled Phenylsuperose fractions</td>
<td>6.0</td>
<td>11.4</td>
<td>1445</td>
<td>127.6</td>
<td>63.0</td>
<td>25.0</td>
</tr>
<tr>
<td>Pooled Superose 12 fractions</td>
<td>5.0</td>
<td>9.0</td>
<td>1335</td>
<td>148.3</td>
<td>58.2</td>
<td>29.1</td>
</tr>
</tbody>
</table>

P.S.= Protamine sulphate  
A=μmol.min⁻¹  
SA=μmol.min⁻¹.mg⁻¹
Approximately 100μg of purified CHM isomerase was divided between three wells of a 7.5%-20% gradient SDS-polyacrylamide gel. After electrophoresis, the gel was electro-blotted onto a PVDF membrane for 1.5h at 250mAmps. One of the blotted samples was then loaded into the protein sequencer. On the first occasion that the enzyme was sequenced, the presence of proline at position 1 was suggested. Because of the problems associated with complete Edman degradation of secondary amines, the repetitive yield of residues in successive cycles is reduced, such that accurate assignment of the amino acid sequence is not possible. As a consequence, a second sample of PVDF membrane blotted isomerase was submitted for sequencing, with an extended trifluoroacetic acid cleavage step during the first cycle, in order to ensure efficient cleavage of the alanylthiazolium (ATZ) intermediate in the first sequencing cycle. This procedure proved effective, and yielded unambiguous sequence for the first 19 residues with the exception of position 7 where no assignment was possible. To investigate this further, a sample of the blotted isomerase was submitted to the procedure developed by Amons (1987), to identify the presence of cysteine, which cannot be detected without side chain modification to yield a stable phenylthiohydantoin derivative. This technique proved successful and a cysteine was assigned at position 7. The first 19 residues of CHM isomerase are shown in Fig. 4.1.3.
### Fig. 4.1.3.

Amino terminal amino acid sequence for CHM isomerase.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Yield (pmol) run 2</th>
<th>Yield (pmol) run 3</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96.8</td>
<td>51.2</td>
<td>Pro</td>
</tr>
<tr>
<td>2</td>
<td>20.5</td>
<td>63.5</td>
<td>His</td>
</tr>
<tr>
<td>3</td>
<td>86.4</td>
<td>52.4</td>
<td>Phe</td>
</tr>
<tr>
<td>4</td>
<td>62.6</td>
<td>34.5</td>
<td>Ile</td>
</tr>
<tr>
<td>5</td>
<td>57.8</td>
<td>37.9</td>
<td>Val</td>
</tr>
<tr>
<td>6</td>
<td>47.2</td>
<td>21.2</td>
<td>Glu</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>8.8</td>
<td>Cys</td>
</tr>
<tr>
<td>8</td>
<td>13.5</td>
<td>4.3</td>
<td>Ser</td>
</tr>
<tr>
<td>9</td>
<td>27.1</td>
<td>9.6</td>
<td>Asp</td>
</tr>
<tr>
<td>10</td>
<td>24.8</td>
<td>6.6</td>
<td>Asn</td>
</tr>
<tr>
<td>11</td>
<td>14.0</td>
<td>8.6</td>
<td>Ile</td>
</tr>
<tr>
<td>12</td>
<td>0.9</td>
<td>5.5</td>
<td>Arg</td>
</tr>
<tr>
<td>13</td>
<td>17.3</td>
<td>6.3</td>
<td>Glu</td>
</tr>
<tr>
<td>14</td>
<td>18.8</td>
<td>8.0</td>
<td>Glu</td>
</tr>
<tr>
<td>15</td>
<td>16.4</td>
<td>5.3</td>
<td>Ala</td>
</tr>
<tr>
<td>16</td>
<td>13.0</td>
<td>3.8</td>
<td>Asp</td>
</tr>
<tr>
<td>17</td>
<td>12.1</td>
<td>7.7</td>
<td>Leu</td>
</tr>
<tr>
<td>18</td>
<td>12.8</td>
<td>2.5</td>
<td>Pro</td>
</tr>
<tr>
<td>19</td>
<td>9.0</td>
<td>2.4</td>
<td>Gly</td>
</tr>
</tbody>
</table>
4.1.4. **Km DETERMINATION FOR CHM ISOMERASE.**

Measurements of specific activity over a sixty fold change in substrate concentration were made for CHM isomerase against CHM using a Perkin-Elmer UV-5 spectrophotometer at 30°C. Reaction rates monitored at the two highest substrate concentrations were measured 'off peak' at 320nm, since the absorbance of such a concentration of CHM at 300nm would have been out of the range of the spectrophotometer used. The results were then analysed using a linear least squares regression analysis computer program (Walmsley & Lowe 1986). The results of this analysis indicated that saturating conditions had not been achieved during the kinetic measurements and that due to the high extinction coefficient for CHM saturating conditions were not possible in this assay system (See Fig.4.1.4.1.) However, a lower limit for the Km was estimated by the program at 536 μM +/- 24% with a Vmax of 1384 μmol.min^-1.mg^-1 +/- 18%.
Fig.4.1.3.1.

A plot of reaction velocity against substrate concentration for CHM isomerase against CHM.
At very low substrate concentrations, the velocity of the enzyme catalysed reaction is directly proportional to the substrate concentration governed by the equation:

\[ V = \frac{V_{\text{max}} \cdot [S]}{K_m} \]

Where \( V_{\text{max}}/K_m \) is the specificity constant (Fersht 1986). Using such an approximation, it is possible to derive the specificity constant for CHM isomerase against CHM by plotting specific activity against substrate concentration at low concentrations of substrate (See Fig.4.1.3.2.). Since the molecular weight of the enzyme is known, it is possible to express the specificity constant in terms of \( K_{\text{cat}}/K_m \). The specificity constant for CHM isomerase against CHM in such an analysis was found to be \( 6.01 \times 10^8 \text{M}^{-1}\text{s}^{-1} \). When a similar analysis was performed against the structurally related compound HHDD, (See Fig.4.1.4.2.) a specificity constant of \( 2.57 \times 10^2 \text{M}^{-1}\text{s}^{-1} \) was obtained, indicating that CHM isomerase has a 2400 fold preference for CHM over HHDD.

The activity of the enzyme in 0.1M sodium phosphate buffer, was tested over a pH range of 6.6 to 8.2. CHM isomerase displayed optimal activity at a pH of 7.6 under these conditions using a substrate concentration of 0.05mM CHM.
Fig. 4.1.3.2.

Plots of reaction velocity of CHM isomerase against CHM and HHDD at low substrate concentrations.

a. CHM

b. HHDD
4.2.1. **PURIFICATION OF HPC DIOXYGENASE.**

Several workers have previously attempted to purify the HPC dioxygenase enzyme of *E.coli* C, but have met with little success, since the enzyme appears to be extremely unstable. The addition of agents which are known to help stabilize other types of oxygenase and dioxygenase enzymes such as glycerol, dithiothreitol and iron salts, appear to have no effect with the *E.coli* enzyme (R.A. Cooper personal communication). This is in marked contrast to the corresponding enzyme of *Pseudomonas ovalis*, whose purification was reported as early as 1964 and was found to be very stable for over a year without any loss of activity, when stored under nitrogen and in the presence of ferrous iron. Moreover, this enzyme was crystallised from preparations of the enzyme at a concentration of 4mg.ml\(^{-1}\) at 5°C.

High level expression of the *E.coli* C enzyme from several of the pDR93 series deletion subclones, has partially overcome some of the difficulties in purification by providing a large amount of active enzyme in the initial crude extract. *E.coli* 5K(pDR9304) was grown overnight in 400ml of Luria broth containing ampicillin and harvested by centrifugation at 10,000 xg. Ultracentrifuged crude extracts were then prepared as detailed in Methods and Materials.

The ultracentrifuged extract was filtered through a 0.2μm acrodisc filter before chromatographic separation using an FPLC system. A volume of 0.5ml of the extract was applied to a Mono Q HR 5/5 anion exchange column equilibrated in the same buffer as the extract. A linear gradient of 0.0-1.0M NaCl applied over 20ml was used to determine the conditions for elution of the dioxygenase before scale up to a Mono Q HR 10/10 anion exchange column. Bulk purification of the ultracentrifuged extract then proceeded using this column with a 0.0M to 0.5M NaCl gradient over 80ml at a flow rate of
Approximately 4ml of the ultracentrifuged extract was loaded per run (112mg of protein). The dioxygenase enzyme was found to elute from the column at approximately 0.3M NaCl under these conditions. Fractions of 2ml corresponding to the highest specific activity from all the MonoQ HR 10/10 runs, were then reapplied to a Mono Q HR 5/5 column, after four fold dilution in 20mM Tris.Cl pH 7.5. SDS-PAGE analysis of the resultant fractions, containing active dioxygenase (See Fig 4.2.1.1.) revealed the presence of several contaminants. A further purification was achieved by hydrophobic interaction chromatography using a Phenylsuperose HR 5/5 column. This column was equilibrated in 0.1M sodium phosphate buffer pH 7.5 with 1.7M (NH₄)₂SO₄ and proteins were eluted with a linear decreasing gradient of the salt. Under these conditions the dioxygenase enzyme eluted at 0.2M (NH₄)₂SO₄. This last chromatographic step yielded a major fraction from which a 33 KDa band accounted for approximately 95% of the total protein. However, the major fractions containing a protein of 33KDa as estimated by SDS-PAGE (Fig 4.2.1.2.) showed barely detectable HPC dioxygenase activity when assayed immediately after elution from the Phenylsuperose column.
Fig. 4.2.1.1.

SDS-PAGE analysis of fractions from Mono Q column containing HPC dioxygenase

7.5% to 20% SDS-polyacrylamide gel analysis of fractions taken during the purification of the HPC dioxygenase. The molecular weight of the marker proteins in KDa is shown. The contents of each lane are as follows:

Lane   a. SDS 7 (Sigma) marker proteins.
        b. Ultracentrifuged extract.
        c. Pooled fractions from the Mono Q column.
        d.
        e. Fractions taken over peak from one run
        f.
        g.
        h. Fractions taken over peak from one run
        i.
        j. SDS 7 (Sigma) marker proteins.
Fig. 4.2.1.1.

SDS-PAGE analysis of fractions from Mono Q column containing HPC dioxygenase
Fig. 4.2.1.2.

SDS-PAGE analysis of fractions from Phenylsuperose purification step for HPC dioxygenase

7.5% to 20% SDS-polyacrylamide gel used to analyse fractions taken during the purification of the HPC dioxygenase. The molecular weight of the marker proteins in KDa is shown.

Lane  a. SDS 7 (Sigma) marker proteins.
     b.
     c. Fractions taken over peak from one run
     d.
     e.
     f. Fractions taken over peak from one run
     g.
     h. SDS 7 (Sigma) marker proteins.
Fig. 4.2.1.2.
SDS-PAGE analysis of fractions from Phenylsuperose purification step for HPC dioxygenase.

Lane: a b c d e f g h

M.P. (KDa)
-66
-45
-36
-29
-24
-20.1
-14.2
A different purification strategy was adopted by applying the semi-purified enzyme to a gel filtration column after the MonoQ step. This method also enabled an apparent native molecular weight for the HPC dioxygenase to be assigned. Two Superose 12 gel filtration columns were linked in series on the FPLC system, and were equilibrated in 50mM sodium phosphate buffer pH 7.5, 0.15M NaCl, at a flow rate of 0.4ml.min⁻¹. The column was calibrated using the proteins recommended by the column manufacturers: immunoglobulin G (160,000), bovine serum albumin (66,000), β-lactoglobulin (35,000) and cytochrome C (12,400). Since the sample applied to the Superose 12 column was known to be impure from SDS-PAGE analysis, fractions of 0.2ml were collected over the peak from the gel filtration separation, so that the HPC dioxygenase activity could be correlated with the protein profile. (Fig 4.2.1.3.) SDS-PAGE was also performed to assess the purity of the fractions collected from this separation, as shown in Fig 4.2.1.4. From this data the apparent molecular weight for the HPC dioxygenase is 80KDa ± 3.2 KDa.

The peak of the enzyme activity does not correspond exactly with the peak of the protein profile from the FPLC system. This is thought to be due to imprecise synchronization of the FPLC system fraction collector and UV₂₈₀ monitor.
Fig.4.2.1.3.a.

Protein profile from Superose 12 gel filtration column
Fig. 4.2.1.3.b.

Enlargement of Superose 12 column trace correlating the protein profile the activity of the fractions collected.

The protein profile is shown by the thick line, with the specific activity of the fractions collected over the peak, shown by the connected points. Fractions of 0.2ml were collected over the peak. The peak of the enzyme activity in fraction 6 does not correspond exactly to the peak of the $A_{280}$ peak due to imprecise synchronization of the FPLC system. Activities of the individual fractions are shown below.

<table>
<thead>
<tr>
<th>Fraction No</th>
<th>Activity (μmol.min$^{-1}$.ml$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>28.0</td>
</tr>
<tr>
<td>3</td>
<td>23.3</td>
</tr>
<tr>
<td>4</td>
<td>37.1</td>
</tr>
<tr>
<td>5</td>
<td>39.3</td>
</tr>
<tr>
<td>6</td>
<td>48.3</td>
</tr>
<tr>
<td>7</td>
<td>46.9</td>
</tr>
<tr>
<td>8</td>
<td>45.1</td>
</tr>
<tr>
<td>9</td>
<td>29.7</td>
</tr>
<tr>
<td>10</td>
<td>6.2</td>
</tr>
<tr>
<td>11</td>
<td>7.0</td>
</tr>
</tbody>
</table>
Fig. 4.2.1.3.b.
Enlargement of Superose 12 column trace detailing the protein profile to the HPC dioxygenase activity of the collected fractions
Fig. 4.2.1.4.

SDS-PAGE analysis of fractions collected over the main peak from the Superose 12 column

7.5% to 20% SDS-polyacrylamide gel showing fractions over the main peak from the superose 12 gel filtration column. Lanes are numbered according to the fractions collected over the peak. The majority of the HPC dioxygenase is found in fraction 6. The far right hand lane contains SDS 7 (Sigma) marker proteins ranging in size in KDa is shown.
Fig. 4.2.1.4.
SDS-PAGE analysis of fractions collected over the main peak from the Superose 12 column

Lane: 4 5 6 7 8 9 10

M.P. (KDa)
-66
-45
-36
-29
-24
-20.1
-14.2
4.2.2. **pH OPTIMUM FOR HPC DIOXYGENASE.**

The extinction coefficient ($\varepsilon_{380}$) of CHMS is known to be pH dependent, so before the pH optimum for the HPC dioxygenase could be assigned, the $\varepsilon_{380}$ for CHMS between pH 6.6 and pH 8.6 in 0.1M sodium phosphate buffer was determined. 0.1M Sodium phosphate solutions of different pH were initially produced by titration of aqueous disodium hydrogen phosphate with sodium dihydrogen phosphate. The absorbance of a standard amount of stock CHMS: 0.0228μmol (10μl of stock) was then measured in the sodium phosphate buffers of different pH. The absorbance values produced were then used to calculate the $\varepsilon_{380}$ at that particular pH. This is shown graphically in Fig 4.2.2.1. The pH dependence of the HPC dioxygenase enzyme was then calculated by the measurement of the enzyme catalysed reaction in the same sodium phosphate solutions at different pHs using the Mono Q pooled fractions. From such an analysis the pH optimum was found to be between pH 7.6 and pH 7.8. The results are shown graphically in Fig 4.2.2.2.

4.2.3. **Km FOR HPC DIOXYGENASE.**

The Km for HPC dioxygenase was determined over a 1μM to 1mM range with respect to HPC using semi-purified HPC dioxygenase obtained after fractionation of an ultracentrifuged extract by anion exchange. The results were analysed using Multifit, a flexible non-linear least squares regression analysis (Walmsley, A.R. & Lowe, A.G. 1985). A plot of HPC dioxygenase activity as a function of HPC concentration is presented in Fig. 4.2.3. The Km derived from this data was 16.1μM with a Vmax of 270μmol.min⁻¹.mg⁻¹.
Fig. 4.2.2.1.

pH dependent absorbance of CHMS in 0.1M sodium phosphate buffer at 380nm.
Fig. 4.2.2.2.

pH dependence on activity for HPC dioxygenase.
Fig. 4.2.3.

A plot of reaction velocity against substrate concentration for HPC dioxygenase.
Approximately 160μg of HPC dioxygenase obtained from fractions collected after the Phenylsuperose purification, were blotted onto a PVDF membrane after electrophoresis on a 7.5%-20% SDS-polyacrylamide gel. The sample was electroblotted onto the PVDF membrane at 250mA for 2.5h. A portion of the membrane-blotted sample was then loaded into the protein sequencer and a total of 21 residues were obtained. The yield for the first residue; glycine is much higher than any of the other amino acids sequenced. This is due to the presence of glycine in the buffer used for electroblotting, some of which is carried through to the sequencing step. No other amino acid was detected in the first sequencing cycle so glycine was thought to be the amino terminal residue. This was later confirmed when the nucleotide sequence of the gene encoding HPC dioxygenase \((hpcB)\) became available. From this information glycine is encoded as the second residue in the sequence, the first being methionine. No amino terminal methionine was detected during amino acid sequencing so it is believed that this residue is removed from the mature protein. The amino terminal sequence of the HPC dioxygenase is presented in Fig 4.2.4.
Amino terminal amino acid sequence for HPC dioxygenase.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Yield (pmol)</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>298.8</td>
<td>Gly</td>
</tr>
<tr>
<td>2</td>
<td>44.8</td>
<td>Lys</td>
</tr>
<tr>
<td>3</td>
<td>90.5</td>
<td>Leu</td>
</tr>
<tr>
<td>4</td>
<td>86.4</td>
<td>Ala</td>
</tr>
<tr>
<td>5</td>
<td>68.6</td>
<td>Leu</td>
</tr>
<tr>
<td>6</td>
<td>60.0</td>
<td>Ala</td>
</tr>
<tr>
<td>7</td>
<td>87.8</td>
<td>Ala</td>
</tr>
<tr>
<td>8</td>
<td>43.7</td>
<td>Lys</td>
</tr>
<tr>
<td>9</td>
<td>54.7</td>
<td>Ile</td>
</tr>
<tr>
<td>10</td>
<td>35.4</td>
<td>Thr</td>
</tr>
<tr>
<td>11</td>
<td>18.4</td>
<td>His</td>
</tr>
<tr>
<td>12</td>
<td>42.1</td>
<td>Val</td>
</tr>
<tr>
<td>13</td>
<td>26.4</td>
<td>Pro</td>
</tr>
<tr>
<td>14</td>
<td>8.2</td>
<td>Ser</td>
</tr>
<tr>
<td>15</td>
<td>18.3</td>
<td>Met</td>
</tr>
<tr>
<td>16</td>
<td>19.7</td>
<td>Tyr</td>
</tr>
<tr>
<td>17</td>
<td>14.2</td>
<td>Leu</td>
</tr>
<tr>
<td>18</td>
<td>3.0</td>
<td>Ser</td>
</tr>
<tr>
<td>19</td>
<td>9.1</td>
<td>Glu</td>
</tr>
<tr>
<td>20.</td>
<td>3.7</td>
<td>Leu</td>
</tr>
<tr>
<td>21.</td>
<td>4.2</td>
<td>Pro</td>
</tr>
</tbody>
</table>
4.3.1. **PURIFICATION OF HHDD ISOMERASE/COHED DECARBOXYLASE.**

It was previously thought that HHDD isomerase and COHED decarboxylase were encoded by separate genes: *hpcF* and *hpcE* respectively (Jenkins & Cooper 1989). However, analysis of the enzyme activities associated with deletion subclones created from both ends of the 3.8 Kbp *Sal1-BamH1* insert in pDR1835, indicated that the two enzyme activities could not be separated in the same way as that used to determine the *hpc B* and *hpcD* gene order (See section 3.4). The most likely explanation for these observations was that the two enzyme activities were associated with a single gene product of approximately 50 KDa in size. In order to test whether HHDD isomerase and COHED decarboxylase activities were due to the same gene product, both activities were purified simultaneously from the same cell extract.

Nucleotide sequencing of pDR83504 identified an open reading frame with associated ribosome binding site, occurring approximately 200 bps downstream of the *lacUV-5* promoter of the vector. This construct was the smallest to express HHDD isomerase and COHED decarboxylase activities, and was chosen for the purification procedure. A 500ml Luria broth culture, supplemented with ampicillin, was grown overnight at 37°C after inoculation with 10ml of an overnight starter culture of *E.coli* JM105 (pDR83504). An ultracentrifuged crude extract was then prepared as described in methods and materials. The protein concentration of the ultracentrifuged extract was estimated by the Biuret method and protamine sulphate added to the ultracentrifuged extract in the ratio of 1mg protamine sulphate to every 20mg of bacterial protein. The mixture was stirred at room temperature for 10 minutes before the precipitate was removed by centrifugation at 20,000 xg. The protamine sulphate treated ultracentrifuged extract, was then filtered through a 0.2µm filter before application to a Mono Q HR 10/10 anion exchange column.
Proteins were eluted with an 80ml gradient of NaCl to 0.5M. A major asymmetric peak eluting between 0.25M and 0.31M NaCl was found to contain the majority of the two activities. The major peak containing both enzyme activities, and the way in which the enzyme activities vary across the peak are shown in Fig.4.3.1.2. Fractions containing the majority of the two enzyme activities were pooled and concentrated by pressure dialysis before application to a Phenylsuperose HR 5/5 column. The column was equilibrated in 0.1M sodium phosphate buffer pH 7.5 containing 1.0M (NH$_4$)$_2$SO$_4$ at a flow rate of 0.7ml.min$^{-1}$. The pooled monoQ fractions were diluted 1:1 with the same sodium phosphate buffer containing 1.7M (NH$_4$)$_2$SO$_4$ before application to the column. Proteins were eluted with a linear decreasing gradient of (NH$_4$)$_2$SO$_4$ to 0.17M over 4mls. The gradient was then halved and the flow rate decreased to 0.25ml.min$^{-1}$ for the 4ml over which the major protein fraction elutes. The FPLC system was programmed to collect fractions over the peak of any protein eluted, with a maximum volume of 1.4mls. The protein profile from this purification step is shown in Fig.4.3.1.2. The purity of the fractions collected at this step was assessed by SDS-PAGE as shown in Fig.4.3.1.3. Analysis of the major peak fractions from the phenylsuperose column purification by SDS-PAGE showed that the major protein component was an apparent single subunit protein of molecular weight approximately 50 KDa.
The activity for HHDD isomerase and COHED decarboxylase were recorded over the main peak of activity from the MonoQ column separation as detailed in the text. Both activities were found to elute between 0.25M and 0.31M NaCl and correspond to the large A<sub>280</sub> peak shown by the thick line. COHED decarboxylase activity is shown by the circles joined by a continuous line (---), whereas the HHDD isomerase activity is shown by the squares joined by the dashed line (---).
Fig. 4.3.1.1.

Diagram representing the variation of HHDD isomerase and COHED decarboxylase enzyme activities over the protein profile from the Mono Q purification step.
Fig.4.3.1.2.

Protein profile from the Phenylsuperose HR 5/5 column purification step

The Phenylsuperose HR 5/5 column was equilibrated in 50mM sodium phosphate buffer pH 7.5 containing 1.0M ammonium sulphate and was ran initially at a flow rate of 0.7 ml.min\(^{-1}\). Once the gradient had reached 0.15M ammonium sulphate the flow rate was decreased to 0.25 ml.min\(^{-1}\). Fractions of up to 1.4ml were collected over the main peak, eluting at approximately 0.08M ammonium sulphate, which contained both HHDD isomerase and COHED decarboxylase activities.
Fig. 4.3.1.2.
Protein profile from the Phenylsuperose HR 5/5 column purification step.
Fig.4.3.1.3.

SDS-PAGE analysis of fractions taken from the Phenylsuperose column

A 12% SDS-polyacrylamide mini gel was used to assess the purity of fractions taken over the peak from the Phenylsuperose column as shown in Fig.4.3.1.2.. Molecular weight marker protein sizes are as shown in KDa.

Lane a. Fractions taken over peak at 0.08M
b. ammonium sulphate
c.
d. Fractions taken over the main peak
e.
f. 0.25x loading of peak fraction
g. 2x loading of peak fraction
h. 5x loading of peak fraction
i. SDS-6H (Sigma) molecular weight markers.
Fig. 4.3.1.3.
SDS-PAGE analysis of fractions taken from the Phenylsuperose column.
Fig. 4.3.1.4.
Protein profile from the Superose 12 gel filtration purification step detailing the variation in enzyme activity for both COHED decarboxylase and HHDD isomerase.

Fractions 1.4ml were collected over the main peak of the elution profile and the enzyme activity of COHED decarboxylase and HHDD isomerase was measured in each. The activity profile for COHED decarboxylase is shown with shaded circles and closed lines, the activity profile for HHDD isomerase with shaded boxes and dashed lines. The thick line denotes the protein profile from the FPLC trace.
It was found that after the phenyl superose column separation both the HHDD isomerase and COHED decarboxylase activities could be stimulated between 50 and 100 fold by preincubation with 5mM MgCl$_2$ at 0°C. The fractions containing the highest specific activity were pooled and applied to two Superose 12 gel filtration columns run in series. The columns were equilibrated in 50mM Tris.Cl pH 7.5, 0.15M NaCl and 5mM MgCl$_2$ at a flow rate of 0.4ml.min$^{-1}$. The protein profile from this column purification along with the associated enzyme activities are shown in Fig 4.3.1.4. A single protein which displays both HHDD isomerase and COHED decarboxylase activities was eluted as the major fraction in the protein profile and was judged to account for in excess of 99% of total protein as estimated by SDS-PAGE. (Fig. 4.3.1.5.). The gel filtration step was also used to make an estimation of the native molecular weight for the enzyme. The column was calibrated using the proteins recommended by the column manufacturers: immunoglobulin G (160,000), bovine serum albumin (66,000), β-lactoglobulin (35,000) and cytochrome C (12,400). Under the condition stated above, the enzyme eluted with an apparent native molecular weight of 30 KDa which is smaller than the subunit molecular weight of 50 KDa estimated by SDS-PAGE. It would appear then that this enzyme runs anomalously on the gel filtration system used. This may be due to the enzyme adopting a non-globular conformation which is assumed for accurate native molecular weight determinations by gel filtration. The native molecular weight of the enzyme was also investigated using sedimentation equilibrium centrifugation by Dr A.J. Rowe and was found to be 41200 KDa suggesting the enzyme exists as a monomer. Details of the purification are given in table 4.3.1.a and b.
Fig.4.3.1.4.
Protein profile from the Superose 12 gel filtration purification step detailing the variation in activity for both COHED decarboxylase and HHDD isomerase.
7.5% to 20% SDS-polyacrylamide gel of fractions taken at each stage of the purification. Approximately 25µg of protein were loaded per track. Molecular weight marker protein sizes are as shown in KDa.

Lane  a. SDS 7 (Sigma) marker proteins.
    b. Ultracentrifuged extract.
    c. Protamine sulphate treated ultracentrifuged extract
    d. Pooled MonoQ fractions.
    e. Pooled Phenylsuperose fractions.
    f. Pooled Superose 12 fractions.
    g. SDS 7 (Sigma) marker proteins.
Fig. 4.3.1.5.

SDS-PAGE analysis of fractions taken at each stage of the purification of COHED decarboxylase/HHD isomerase.
**Table 4.3.1.**

**a. Purification of the COHED decarboxylase activity.**

<table>
<thead>
<tr>
<th>PURIFICATION STEP</th>
<th>VOLUME (ml)</th>
<th>TOTAL PROTEIN (mg)</th>
<th>TOTAL UNITS (A)</th>
<th>SPECIFIC ACTIVITY (SA)</th>
<th>RECOVERY (%)</th>
<th>PURIFICATION (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultracentrifuged extract</td>
<td>10.0</td>
<td>228.0</td>
<td>1800.0</td>
<td>7.9</td>
<td>100</td>
<td>1.0</td>
</tr>
<tr>
<td>P.S. treated ultracentrifuged extract</td>
<td>10.0</td>
<td>190.0</td>
<td>1500.0</td>
<td>7.9</td>
<td>83</td>
<td>1.0</td>
</tr>
<tr>
<td>Pooled Mono Q fractions</td>
<td>15.0</td>
<td>54.0</td>
<td>1224.5</td>
<td>22.7</td>
<td>68</td>
<td>2.8</td>
</tr>
<tr>
<td>Pooled Phenylsuperose fractions</td>
<td>8.5</td>
<td>18.7</td>
<td>532.9</td>
<td>28.5</td>
<td>30</td>
<td>3.6</td>
</tr>
<tr>
<td>Pooled Superose 12 fractions</td>
<td>4.0</td>
<td>10.0</td>
<td>504.0</td>
<td>50.4</td>
<td>28</td>
<td>6.3</td>
</tr>
</tbody>
</table>

P.S. = Protamine sulphate  
A = μmol.min⁻¹  
SA = μmol.min⁻¹.mg⁻¹

**b. Purification of the HHDD isomerase activity.**

<table>
<thead>
<tr>
<th>PURIFICATION STEP</th>
<th>VOLUME (ml)</th>
<th>TOTAL PROTEIN (mg)</th>
<th>TOTAL UNITS (A)</th>
<th>SPECIFIC ACTIVITY (SA)</th>
<th>RECOVERY (%)</th>
<th>PURIFICATION (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultracentrifuged extract</td>
<td>10.0</td>
<td>228.0</td>
<td>1316</td>
<td>5.7</td>
<td>100</td>
<td>1.0</td>
</tr>
<tr>
<td>P.S. treated ultracentrifuged extract</td>
<td>10.0</td>
<td>190.0</td>
<td>1140</td>
<td>6.0</td>
<td>87.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Pooled Mono Q fractions</td>
<td>15.0</td>
<td>54.0</td>
<td>781.7</td>
<td>14.5</td>
<td>59.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Pooled Phenylsuperose fractions</td>
<td>8.5</td>
<td>18.7</td>
<td>840.7</td>
<td>44.9</td>
<td>64.0</td>
<td>7.9</td>
</tr>
<tr>
<td>Pooled Superose 12 fractions</td>
<td>4</td>
<td>11.2</td>
<td>458.8</td>
<td>45.8</td>
<td>35.0</td>
<td>8.0</td>
</tr>
</tbody>
</table>

P.S. = Protamine sulphate  
A = μmol.min⁻¹  
SA = μmol.min⁻¹.mg⁻¹
4.3.2. **AMINO TERMINAL AMINO ACID SEQUENCE FOR COHED DECARBOXYLASE/HHDD ISOMERASE.**

Approximately 100µg of the purified enzyme from the Superose 12 gel filtration step were blotted onto a PVDF membrane for 1.5h at 250mAmps after electrophoresis on a 7.5%-20% SDS-polyacrylamide gel. A portion of the blotted sample was then loaded into the protein sequencer which was programmed to sequence the first twenty residues of the protein. Unfortunately at this time the protein sequencer was not working to its maximum efficiency, especially with reference to positively charged amino acids such as arginine, histadine and lysine. However, a single amino acid sequence was obtained further supporting the evidence presented here, that HHDD isomerase and COHED decarboxylase activities are a function of the same protein. When the nucleotide sequence (see chapter 5) of the gene encoding HHDD isomerase/COHED decarboxylase became available, the identity of the unassigned arginine, histidine and lysine residues was confirmed from the deduced amino acid sequence. Nucleotide sequencing also confirmed that the amino acid sequence obtained, corresponds to a single open reading frame. Thus HHDD isomerase and COHED decarboxylase are not separate enzymes, and are a function of the same gene product.
**Fig. 4.3.2.**

**Amino terminal amino acid sequence of COHED decarboxylase/HHDD isomerase**

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Yield pmol</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>103</td>
<td>Met</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>Lys</td>
</tr>
<tr>
<td>3</td>
<td>120</td>
<td>Gly</td>
</tr>
<tr>
<td>4</td>
<td>93</td>
<td>Thr</td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>Ile</td>
</tr>
<tr>
<td>6</td>
<td>83</td>
<td>Phe</td>
</tr>
<tr>
<td>7</td>
<td>127</td>
<td>Ala</td>
</tr>
<tr>
<td>8</td>
<td>107</td>
<td>Val</td>
</tr>
<tr>
<td>9</td>
<td>111</td>
<td>Ala</td>
</tr>
<tr>
<td>10</td>
<td>91</td>
<td>Leu</td>
</tr>
<tr>
<td>11</td>
<td>39</td>
<td>Asp</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>ND</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>ND</td>
</tr>
<tr>
<td>14</td>
<td>24</td>
<td>Ser</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>ND</td>
</tr>
<tr>
<td>16</td>
<td>30</td>
<td>Leu</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>Asp</td>
</tr>
<tr>
<td>18</td>
<td>41</td>
<td>Ala</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>ND</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>Glu</td>
</tr>
</tbody>
</table>

ND = no data
4.3.4. **Kinetic Analysis of COHED Decarboxylase/HHDD Isomerase.**

Km determinations were carried out for the HHDD isomerase and COHED decarboxylase activities of the purified protein.

a. Km determination for COHED decarboxylase.

Measurements of specific activity over a twenty fold change in substrate concentration were made in the same manner as described in section 4.1.4. The assays were performed as detailed in methods and materials. An estimation of the equilibrium concentration COHED was made by measuring the change in absorption of CHM upon enzyme catalysed isomérisation. Approximately 33% of the CHM is converted to COHED at equilibrium. The results of the kinetic analysis were analysed using the linear least squares regression analysis as previously described. These results indicated that saturating conditions had not been achieved during the assays (Fig.4.3.4.) This is due to the high extinction coefficient for CHM used in the assays. However, a lower limit for the Km was estimated by the program at 32µM +/− 13% with a Vmax of 103.6 µmol.min⁻¹.mg⁻¹ +/− 6.6%.

b. Km determination for HHDD isomerase.

Measurements of specific activity over a 20 fold change in substrate concentration were made for HHDD isomerase as previously described. The results were then analysed using a least squares regression analysis and indicated that saturating conditions had not been achieved (Fig.4.3.4.2.) Assays at higher substrate concentrations than used could not be performed due to the high extinction coefficient of HHDD. However a lower limit for the Km was estimated at 40.8µM +/− 10% with a Vmax of 99.3 µmol.min⁻¹.mg⁻¹.
Using the approximation detailed in section 4.1.4., the specificity constant for HHDD isomerase against HHDD was calculated as $1.11 \times 10^6$ M$^{-1}$ s$^{-1}$. When measurements of HHDD isomerase activity were made where CHM was the substrate a specificity constant of $22.71$ M$^{-1}$ s$^{-1}$ was obtained indicting a 49000 fold preference for HHDD over CHM. Plots of reaction velocity against HHDD and CHM are presented in Fig.4.3.4.3.
Fig.4.3.4.1.

A plot of reaction velocity against substrate concentration for COHED decarboxylase.
Fig. 4.3.4.2.

A plot of reaction velocity against substrate concentration for HHDD isomerase.
Fig. 4.3.4.3.
Plots of reaction velocity of HHDD isomerase against HHDD and CHM at low substrate concentrations.

a. HHDD

b. CHM
CHAPTER 5.
NUCLEOTIDE SEQUENCING OF THE
\textit{hpc} GENES, TRANSCRIPT MAPPING
STUDIES AND SIMILARITY SEARCHES.
6.1.1. INTRODUCTION.

At the beginning of the project an order of priority was placed on which genes were to be sequenced (see section 1.8.). That order was:

1. The two isomerase genes
2. The decarboxylase gene.
3. Any other genes of the pathway that were accessible.

At that time the information concerning the gene order was still modelled on data reported by Jenkins (1987). As such it was thought that the first two objectives could be achieved by sequencing the 3.8 Kbp SalI-BamH1 region of pJJ801. However, as detailed in chapter 3, the gene order was found to be radically different to that proposed by Jenkins and an altered nucleotide sequencing strategy had to be adopted.

Once the position of the genes had been identified as detailed in chapter 3, it was then possible to begin detailed nucleotide sequencing studies.

5.2.1. DELETION SUBCLONING OF pDR1930.

A sequencing project of this size could be approached in a number of ways. Since it is only possible to read approximately 300 bps from a single set of sequencing reactions, an ordered sequencing strategy was required to sequence the hpc genes. In the past few years a variety of methods have become available from simply synthesizing oligonucleotides to known sequence to act as primers, to more advanced deletion strategies such as those of Dale (1985) and Henikoff (1984, 1987). For this study a unidirectional deletion subcloning system was adopted since this could generate the templates required for nucleotide sequencing and in addition would provide information on the gene order. Details of the deletion subcloning procedure are given in
Methods and Materials. In essence, however, the technique relies on the properties of Exonuclease (Exo) III. This enzyme will digest double stranded DNA from 5' overhangs or blunt ends but will not digest from 3' overhangs or blunt ends created by "filling in" with thionucleotides. By digestion of double stranded DNA with a combination of restriction enzymes which produce 3' and 5' overhangs, it is possible to direct the digestion of any subsequent attack by Exo III in one direction only. If aliquots are removed at various times in a controlled Exo III digest, it is possible to produce a series of clones which have a portion of the target DNA deleted from one end. A nested set of deletion subclones can be produced which differ, for example, by 200 bps, and are therefore ideal templates for sequencing.

Exo III digestions were made using a double stranded nested deletion kit purchased from Pharmacia LKB. When digestions were made under the conditions described by the manufacturer (as detailed in strategy 1 in section 2.2.8.) a deletion series of insufficient size was obtained. On increasing the temperature of the Exo III digestion from 25°C to 35°C, a series of deletion subclones which differed in size by approximately 200 bps were obtained. After ligation and subsequent transformation into competent E.coli, individual colonies were analysed by mini-prep and restriction enzyme digest. The series of subclones obtained by this method were designated pDR9301 to pDR9317 as detailed in Fig.5.2.1.
Fig.5.2.1.

Deletion subclones of the pDR93 series

The pDR93 deletion series was created by unidirectional digestion by ExoIII from the SalI site of pDR1930. Protection of the remaining portion of the pUC19 vector was achieved by digestion with SphI which produces a 3' overhang which is not susceptible to ExoIII. The insert region of the deletion subclones is shown only for clarity. Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdeI, Kp=KpnI, E=EcoRI, Ss=SstI (SacII), S=SalI, Pv=PvuII, H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI.
Fig. 5.2.1.
Deletion subclones of the pDR93 series
5.2.2. GENE ORDER DEDUCED FROM THE pDR93 DELETION SERIES.

Detailed descriptions of how the gene order of \textit{hpcD} and \textit{hpcB} was deduced from the pDR93 deletion series is presented in section 2.3.1. However, analysis of the enzyme activities associated with individual deletion subclones showed that \textit{hpcB} (encoding HPC dioxygenase) occurs 5' of \textit{hpcD} (encoding CHM isomerase). More exactly, pDR9305 is the smallest of the pDR93 deletion series to express active dioxygenase and isomerase enzymes. Deletion subclones smaller than pDR9305 but larger than pDR9317 express only CHM isomerase. Subclones smaller than pDR9317 do not express any of the Hpc enzymes.

5.3. NUCLEOTIDE SEQUENCING OF \textit{hpcD}.

Once the gene order had been deduced as detailed above, nucleotide sequencing of the individual genes could begin. Nucleotide sequencing of \textit{hpcD} was addressed first on the coding strand by sequencing of pDR9317, 9316, 9314 and pDR9313. Analysis of this single stranded nucleotide sequence using the University of Wisconsin (WIMP) package of programmes identified a single open reading frame (ORF) of 381 bps which is initiated 191 bps upstream of the \textit{NdeI} site and terminates 105 bps before of the most internal \textit{EcoRV} site in pJJ801. This ORF has a sequence six bps upstream of the ATG initiation codon, which bears similarity to the Shine-Dalgarno (Shine & Dalgarno 1975) ribosome binding site (RBS). Double stranded nucleotide sequence was obtained by using a combination of oligonucleotide primers and constructs derived from pDR1830 to sequence the non-coding strand. The 3' end of \textit{hpcD} was sequenced using the construct pDR821. This plasmid was created by digestion of pDR1830 with \textit{EcoRV} and \textit{BamH1} to remove the 0.6 Kbp \textit{BamH1-EcoRV} region followed by treatment of the linearised plasmid with Mung bean nuclease. This created a blunt ended, linear molecular which was self-ligated.
and used to transform competent *E.coli*. Single colonies were analysed by restriction digest after alkaline lysis mini-prep. One of the resulting constructs was named pDR821 and was sequenced from the *EcoR1* end of the polylinker. The double stranded sequence obtained from this procedure extended to a point mid-way through *hpcD*. Double stranded sequence extending past the amino terminal region of the gene was produced using a synthetic 17mer oligonucleotide probe: DR03. This primer was designed from the sequence information obtained from pDR821. The constructs required to sequence *hpcD* are shown in Fig.5.3.1. This ORF was confirmed as that encoding CHM isomerase when the predicted amino acid sequence was compared to the amino terminal sequence of the purified protein. The first 19 amino acid residues of the protein matched identically with residues 2 to 20 of the predicted amino acid sequence deduced from the nucleotide sequence. The amino terminal methionine specified by the nucleotide sequence was absent from the native protein when the amino terminal amino acid sequence was determined. The complete nucleotide sequence of *hpcD* and the translated ORF is presented in Fig.5.3.2.
Fig. 5.3.1.

Constructs required to sequence hpcD.
Fig. 5.3.1.

Constructs required to sequence hpcD.

Unidirectional deletion subclones were created from the SalI end of the insert DNA in pDR1930 by the ExoIII digestion procedure. The coding strand encompassing hpcD was sequenced with the deletion subclones pDR9317, pDR9316, pDR9314 and pDR9313 in the direction as shown. The arrows indicate the start, direction and extent of sequencing carried out using universal primer. The complimentary strand was sequenced from the polylinker end of pDR821 using reverse sequencing primer. This yielded nucleotide sequence information from a region outside the gene to approximately midway through it. The oligonucleotide probe:DR03 was then constructed on the basis of this information, to extend the nucleotide sequence beyond the 5' end of hpcD and is marked as the large arrow. Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdeI, Kp=KpnI, E=EcoRI, Ss=SstI (SacII), S=SalI, Pv=PsI, H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI
Fig. 5.3.2.

**Nucleotide sequence of hpcD.**

The nucleotide sequence is numbered in relation to the first nucleotide in the mRNA transcript that starts upstream of hpcEF. A region that bears homology to the Shine-Dalgarno site, or ribosome binding site (RBS) is located six bps upstream of the ATG initiation codon for hpcD as shown. The position of the Nde1 restriction site within the hpcD ORF is shown.
The nucleotide sequence and translated ORF of hpcD.
5.4. AMINO ACID COMPOSITION OF THE CHM ISOMERASE SUBUNIT AND SUBUNIT MOLECULAR WEIGHT.

Once the *hpcD* ORF had been identified, it was possible to calculate the subunit molecular weight of the CHM isomerase from the deduced amino acid sequence. The deduced ORF was analysed by the WIMP programmes EXTRACT and COMPOSITION to produce the required data. The subunit amino acid composition for CHM isomerase is detailed in Fig.5.4.1. This includes the amino terminal methionine which is apparently absent from the native protein. The *hpcD* gene encodes a polypeptide chain of 126 amino acids. From this data the subunit molecular weight, taking into account the missing amino terminal methionine was calculated as 14,266 Da which agrees well with a figure of 14 KDa deduced from SDS-PAGE calibrated with standard marker proteins. Analysis of the amino acid composition indicated that one tyrosine ($\varepsilon_{280}=1,100 \text{ M}^{-1}\cdot\text{L}^{-1}$), two tryptophans ($\varepsilon_{280}=5,050 \text{ M}^{-1}\cdot\text{L}^{-1}$) and nine phenylalanine residues occur per subunit. On this basis, the molar extinction coefficient for the purified enzyme at 280nm will be 11,200 per subunit which gives an absorbance of 0.8 $A_{280}\text{mg}^{-1}$. 
Fig. 5.4.1.

Subunit amino Acid composition for CHM isomerase

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>13</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>5</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>9</td>
</tr>
<tr>
<td>His (H)</td>
<td>7</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>5</td>
</tr>
<tr>
<td>Met (M)</td>
<td>5</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>5</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>6</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>6</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>2</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>1</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>10</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>7</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>7</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>15</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>5</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>5</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>6</td>
</tr>
<tr>
<td>Val (V)</td>
<td>6</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>1</td>
</tr>
</tbody>
</table>

Total Number of amino acids: 126
6.5. **NUCLEOTIDE SEQUENCING OF hpcB.**

Initially the nucleotide sequence of the coding strand for hpcB was determined by sequencing pDR9304, 05, 06, 07, 08, 09, 10, 11, 12 and pDR9315. This sequence information was used to design a series of oligonucleotide primers which were used to sequence the untranscribed strand. This strategy was augmented by sequencing pDR934 (see section 3.3) using universal sequencing primer. The strategy used to sequence hpcB is shown in Fig.5.5.1. An ORF was identified starting 387 bps upstream of the internal Kpn1 site in pDR1930 and was found to terminate 32 bps upstream of hpcD. A probable ribosome binding site was found six bps upstream of the ATG initiation codon of this ORF. The deduced amino acid sequence corresponded exactly between residues 2 and 22 to the amino terminal amino acid sequence of the purified HPC dioxygenase enzyme. As with CHM isomerase, amino terminal amino acid sequencing failed to detect the presence of methionine as the first residue in the sequence, even though this residue is predicted from the deduced ORF. The full nucleotide sequence of hpcB and translated ORF for HPC dioxygenase is presented in Fig.5.5.2. The single Kpn1 site that is present in the cloned DNA of pDR1930 lies within the ORF for HPC dioxygenase and corroborates the finding that pDR934 does not express an active HPC dioxygenase enzyme.
Fig.5.5.1.

Strategy require to sequence hpcB

The region encompassing hpcB is shown by the hatched box. The coding strand of hpcD was sequenced with the members of the 93 series deletion subclones as shown. The insert region of the deletion subclones is shown only for clarity. The small arrows denote the direction and extent of the sequence information obtained from these subclones. The nucleotide sequence of the untranscribed strand up to the Kpn1 site that occurs within the hpcB ORF, was obtained by sequencing pDR1930 using two oligonucleotides: DR12 and DR13 which were constructed on the basis of the nucleotide sequence obtained for the transcribed strand. Nucleotide sequence from the Kpn1 site was then obtained by sequencing from the EcoR1 end of the polylinker in pDR932 using universal sequencing primer. The double stranded nucleotide sequence was then extended to the 5' end of hpcB using a third oligonucleotide:DR14. The direction and extent of the nucleotide sequence information obtained with DR12, DR13 and DR14 is indicated by the the large arrows. Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdeI, Kp=Kpn1, E=EcoRI, Ss=SstI, S=SalI, Pv=PvuII, H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI, Ss'=SstII
Fig. 5.5.1.

Strategy required to sequence hpcB.
The sequence with homology to the Shine-Dalgarno, ribosome binding site was found seven base pairs upstream of the \textit{hpcB} initiation codon. The translated open reading frame is shown in the three letter code below the nucleotide sequence. The location of the recognition sites for \textit{PvuI} and \textit{KpnI} are labelled and overlined. The position of the nucleotides sequence is shown in relation to the first nucleotide of the mRNA transcript which starts upstream of \textit{hpcEF}. 
The nucleotide sequence and translated ORF of hpcB

---

Event: PvuI

---

Event: Kpn1

---

178
Fig. 5.5.2. The nucleotide sequence and translated ORF of *hpcB* continued:

```
CCGGAAGGATGAACAGCTACACCCCGCAGTTGCCACGCAGAGCTGGACGAGCGTGTG

3 3 5 9----- +------------------+------------------+------------------+------------------+----------------+----------------
3 4 1 8
CGCCTTCTTCCCTACTTGTCGATGTGGGCGCTCAAGCTGGCGGTCTACCTGCTCGCACAC
AlaGluGluMetAsnSerTyrThrArgGluPheAspArgGlnMetAspGluArgVal

3419
GTGAAGCTGTGGCGCGAAGGCCAGTTCAAAGAGTTCTGCAATATGCTGCCGGAGTACGCC

3 4 1 9 - + ----------------+------------------+------------------+------------------+----------------+--------------
3 4 7 8
CACTTCGACACCGCGCTTCCGGTCAAGTTTCTCAAGACGTTATACGACGGCCTCATGCGG
ValLysLeuTrpArgGluGlyGlnPheCysAsnMetLeuProGluTyrAla

3479
GACTACTGCTACGGCGAAGGCAATATGCACGACACGGTGATGCTGCTGGGGATGCTCGGC

3 4 7 9 - + ----------------+------------------+------------------+------------------+----------------+---------------
3 5 3 8
CTGATGACGATGCCGCTTCCGTTATACGTGCTGTGCCACTACGACGACCCCTACGAGCCG
AspTyrCysTyrGlyGluGlyAsnMetHisAspThrValMetLeuLeuGlyMetLeuGly

3539
TGGGATAAATACGCGGGAAGGATGGGAGTTTATCACCAGGCTATTCCCAAGCTCCTGG

3 5 9 9 - + ----------------+------------------+------------------+------------------+----------------+---------------
3 5 9 8
ACCCTATTATGCTGGCTGCTCAATCCCTAGTGCTGCTGCTAGGGGTCGAGGACC
TrpAspLysTyrAspGlyValTrpSerLeuSerProSerTyrSerGlnAlaSerTrp

3599
CACCCGTCAGGTTAAGCTGT

3620
GTGGCCAGTCATATTGCCACAAA
HisArgSerGly
```
5.6. **AMINO ACID COMPOSITION OF THE HPC DIOXYGENASE SUBUNIT AND SUBUNIT MOLECULAR WEIGHT.**

The translated ORF of *hpcB* was analysed as detailed in section 5.4.1. to produce the required data for amino acid analysis. The subunit amino acid composition for HPC dioxygenase is detailed in Fig. 5.6.1. which includes the amino terminal methionine which is apparently absent from the native protein. The *hpcB* gene encodes a single polypeptide chain of 277 amino acids. From this data the subunit molecular weight, taking into account the missing amino terminal methionine, was calculated as 31,332 Da which agrees well with a figure of 33 KDa deduced from SDS-PAGE calibrated with standard marker proteins.
Subunit amino acid composition for HPC dioxygenase

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>18</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>18</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>10</td>
</tr>
<tr>
<td>His (H)</td>
<td>12</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>16</td>
</tr>
<tr>
<td>Met (M)</td>
<td>12</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>8</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>16</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>9</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>5</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>6</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>20</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>22</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>13</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>21</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>11</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>7</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>20</td>
</tr>
<tr>
<td>Val (V)</td>
<td>18</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>14</td>
</tr>
</tbody>
</table>

Total Number of amino acids: 277
5.7. DELETION SUBCLONING OF pDR1835.

Initially a deletion series was created from the SalI site of pDR1835 in order to determine the gene order of hpcE and hpcF. The deletion series was created using procedure 2 as detailed in Methods and Materials, where the Exo III deletions were made at a temperature of 30°C. This procedure yielded a deletion series of 18 subclones which were designated pDR85301 through pDR85318 (See Fig.5.7.1). Measurement of enzyme activities associated with individual subclones in this series, showed that once 0.5 Kbp of DNA had been deleted both COHED decarboxylase and HHDD isomerase activities could no longer be detected. Nucleotide sequencing of each of the subclones in the pDR853 series showed them to be true members of the deletion series since the nucleotide sequence obtained overlapped and was continuous for 1.4 Kbp. A second deletion series was then created from the BamHI site of pDR1835 using procedure 3 as detailed in Methods and Materials where the ExoIII reactions were carried out at 35°C. This procedure yielded a series of 10 deletion subclones (See Fig.5.7.2.) Measurement of the enzyme activities associated with individual deletion subclones of the pDR835 series revealed that both COHED decarboxylase and HHDD isomerase activities were associated with all the subclones which retained inserts greater than 1.8 Kbp. Deletion subclones which were smaller than 1.8 Kbp did not appear to express either hpcE or hpcF. A series of other experiments as detailed in section 4.4.1. indicated that both COHED decarboxylase and HHDD isomerase were associated with the same gene product. Nucleotide sequencing of hpcEF encoding COHED decarboxylase/HHDD isomerase is described in section 5.8
Members of the pDR853 series of deletion subclones were created by the ExoIII digestion procedure from the \textit{SalI} end of pDR1835. The vector DNA was protected in the ExoIII digestion by cutting at the \textit{SphI} site in the polylinker which produces a 3' overhang which is not susceptible to digestion to ExoIII. Only the insert DNA for the deletion subclones is shown for clarity. The deletion subclones designated pDR85301 and pDR85302 differ in size by less than 20 bps so are effectively the same size. A similar situation is true for pDR85303 and pDR85304. Restriction sites are as follows: \textit{B}=\textit{BamHI}, \textit{Ev}=\textit{EcoRV}, \textit{Nd}=\textit{NdeI}, \textit{Kp}=\textit{KpnI}, \textit{E}=\textit{EcoRI}, \textit{Sc}=\textit{SstI}, \textit{S}=\textit{SalI}, \textit{PvI}=\textit{PvuI}, \textit{PvII}=\textit{PvuII}, \textit{H}=\textit{HindIII}, \textit{Sp}=\textit{SphI}, \textit{Ps}=\textit{PstI}, \textit{Sm}=\textit{SmaI}, \textit{Ss}=\textit{SstII}
Fig. 5.7.1.

pDR853 deletion subclone series.
Fig. 5.7.2.

pDR835 deletion subclone series

Members of the pDR835 series of deletion subclones were created by the ExoIII digestion procedure from the BamH1 end of pDR1835. The vector DNA was protected in the ExoIII digestion by cutting at the SstI site in the polylinker which produces a 3' overhang which is not susceptible to digestion by ExoIII. Only the insert DNA for the deletion subclones is shown for clarity. Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdeI, Kp=KpnI, E=EcoRI, Sc=SstI, S=SalI, Pvu= PvuII, H=HindIII, Sp= Sphi, Ps= PstI, Sm=SmaI, Ss=SstII
Fig. 5.7.2.

pDR835 deletion subclone series.
5.8. **NUCLEOTIDE SEQUENCING OF hpcEF.**

Double stranded nucleotide sequencing of the *hpcEF* ORF was achieved using subclones of the pDR835 and pDR853 deletion series and with several oligonucleotide primers. The coding strand of *hpcEF* was sequenced using pDR83507, 08, 09, 10, 11, 12, 13, 14 and pDR83516 using reverse sequencing primer. Three oligonucleotide primers: DR04, DR07 and DR08 were used to bridge gaps in the sequence generated from the subclones. DR04 was constructed from the sequence information obtained from pDR83507 and enabled the nucleotide sequence to be extended to that obtained from pDR83508. DR05 was constructed from the nucleotide sequence information obtained from pDR83509 and extended the sequence obtained from pDR83510. Finally DR08 was constructed from the sequence information obtained from pDR83514 and extended the sequence information obtained from pDR83515. The untranscribed strand of *hpcEF* was sequenced using members of the pDR853 deletion series. Nucleotide sequence from the 3' end of the gene was initially generated using pDR85301 and extended with successively smaller members of the deletion series up to and including pDR85313. The nucleotide sequence obtained from these subclones extended to a point approximately three quarters of the way along the non-coding strand. The remaining sequence information was then obtained with primers: DR09 and DR10. A diagram depicting the subclones and oligonucleotide primers required to sequence *hpcEF* is shown in Fig.5.8.1.
A large number of deletion subclones from both the pDR935 and pDR853 series were required to sequence hpcEF. The area encompassing hpcEF is shown by the hatched box. pDR1835 is shown for reference. The coding strand of the gene was sequenced with eleven members of the pDR835 series which are shown by the small arrows pointing from right to left. Three gaps in the nucleotide sequence were bridged using oligonucleotide primers: DR04, DR05 and DR08 which are shown by the large arrows. The non-coding strand of the gene was sequenced using eleven members of the pDR853 series deletion subclones which are shown by the small arrows pointing from left to right. Nucleotide sequence up to the 5' end of the gene was obtained by the construction of two oligonucleotide primers DR09 and DR10. Restriction sites are as follows: B=BamHI, Ev=EcoRV, Nd=NdI, Kp=KpnI, E=EcoRI, Sc=SstI, S=SalI, PvI=PvuI, PvII=PvuII, H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI, Ss=SstII
Fig. 5.8.1.

Sequencing strategy for *hpcEF*
The double stranded nucleotide sequence information was analysed using the WIMP package of programmes and a single ORF of 1215 bps was found to start 460 bps downstream of the $Pvu$II site of pJJ801 and terminates 323 bps upstream of the $Sst$II site that lies within $hpcC$ (see section 4.10.). This ORF was definitely assigned as that encoding COHED decarboxylase/HHDD isomerase when the amino terminal sequence of the purified protein was compared to that predicted from the ORF. As detailed in section 4.3.2., problems were experienced in assigning histidine, glutamine and arginine residues in the amino acid sequence due to technical problems with the protein sequencer at the time of analysis. Because of this problem, four residues at positions 12,13,15 and 19 in the amino acid sequence were not assigned. The residues predicted at those positions from the deduced ORF are histidine, arginine, glutamine and tryptophan respectively. The complete nucleotide sequence and the translated ORF of $hpcEF$ is presented in Fig.5.8.2.
A sequence with homology to the Shine-Dalgarno, ribosome binding site (RBS) was found six base pairs upstream of the $hpcEF$ initiation codon and is overlined. The translated open reading frame is shown in the three letter code below the nucleotide sequence. Transcript mapping studies (see section 5.16) showed that the first adenine in the mRNA transcript that encodes $hpcEF$, starts 40 bps upstream of the gene. This adenine has been designated +1 in the nucleotide sequence such that the $hpcEF$ ORF starts at +40.
The nucleotide sequence and translated ORF of hpcEF.

**Fig. 5.8.2.**

```
ACATATTAAAAATTGATGATCAGAATGTTATAGAAAGGCACACATATCTTGCGG
+-----------------------------------------------+-------------------+

TTGTTAAAAATTACTAGCTAGCTTGGATGATCGAATCCGAGGAGTGGTAATGAAAGGCACTATCTTCGCG
MetLysGlyThrThrValAla

GTGCCCTACGTGATGATCAGAATGTTATAGAAAGGCACACATATCTTGCGG
+---------------------------------------------------------------------+

CATCGCAACACTTGAGCCTGCGTTGAGACATGCTTGGATGATCGAATCCGAGGAGTGGTAATGAAAGGCACTATCTTCGCG
ValAlaLeuAsnHisArgSerGlnLeuAspAlaTrpGlnGluAlaPheGlnGlnSerPro

ATCAAAAGCCCAGCTAAATCTGGGTGTGGTTATTTAAACCCGCAATTACGGAATGTTGTT
IleLysAlaProLysThrAlaValTrpPheLysProArgAsnThrValAla

CTCGGCTAAGACGTACGGAGCCGCGGTACGTGCTGCGGTACGTGCTG
CysGlyPheProPheProGlnGluAsnLeuLeuSerGlyAlaAla

GGACATGCGTCCACCGTTGCTGACGTCCTGCCGGAAGAGAGCTTTTATCGCCGCGGTACGTGCTG
GlyTyrAlaLeuAsnLeuAspValSerLeuProGluSerPheProAlaPhe

AAAGCAAAATTCGCTGATTCTGTTCTGCAGAATGTTGTTAAGGAGCCGCGGTACGTGCTG
LysAlaLysCysArgAspGlyPheCysProIleGlyGluThrValAla

GATAATCTGACCATCTACCAAGATCAAGCGGCGCTGCTGAGCGCCCTGAGCAGAATTTGCCACACTGAAT
AspAsnLeuThrIleTyrThrGluAsnLeuAspHisTrpAsnThrSer

GATTACGCAAATTCGCTGATATCCGCGCTGCTGAGCGCCCTGAGCAGAATTTGCCACACTGAAT
AspAsnLeuThrIleTyrThrGluAsnLeuAspHisTrpAsnThrSer

CCAGGGATGTCATCTGTCCTGGAACGCCACCAAGGGCGGCGGTGAAATACACAGCCGAGGT
ArgValArgValLeuLeuGlyPheProProLeuGluAsnProValValAspGluArg

188
```
Fig. 5.8.2. The nucleotide sequence and translated ORF of hpcEF continued:

661  GAAGTGGACCAGGCAGAGGCCTCCACAGCTGCCACACCCGACGGTACGTTGGTTGCC
     ------------------------------+
     GluValThrThrArgLysSerPheProThrProHisProHisGlyThrLeuPheAla

720

781  CTGGGCTTGTCAGAGGCGCAGACCGCACTGGGCGAATTTATTCAACCGACCATACCGTGCGCCAGCG
     +-------------------+-------------------+-------------------+-------------------+
     LeuValPheLeuAlaAspHisAlaSerGluLeuGluPheLysProPheGluGluPro

840

901  CGTAACGTCAGCGAAGCCGATGCCATGGATTATGTCGCGGGCTACACCGTGTAACGAC
     ------------------------------+
     ArgAsnValSerGluAlaAspAlaMetAspTyrValAlaGlyTyrThrValCysAsnAsp

1020

1201

189
5.9 AMINO ACID COMPOSITION OF COHED DECARBOXYLASE/HHDD ISOMERASE AND SUBUNIT MOLECULAR WEIGHT.

The predicted ORF from the nucleotide sequence for $hpcEF$ was analysed as detailed in section 5.4.1. to produce the amino acid composition. The subunit amino acid composition of COHED decarboxylase/HHDD isomerase is detailed in Fig.5.9. The $hpcEF$ gene encodes a polypeptide of 406 amino acids with a molecular weight of 44,514 Da which agrees well with a figure of 50 KDa deduced from SDS-PAGE calibrated with standard marker proteins. The nucleotide sequence indicates that the isomerase/decarboxylase subunit contains three tryptophan, fourteen tyrosine and sixteen phenylalanine residues per subunit. On that basis the molar extinction coefficient for the enzyme at 280nm was calculated at 30,550 M$^{-1}$.L$^{-1}$ which gives an absorbance of 0.68 A$_{280}$.mg$^{-1}$
Fig. 5.9.  
Subunit amino acid composition of COHED

decarboxylase/HHDD isomerase

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>40</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>21</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>16</td>
</tr>
<tr>
<td>His (H)</td>
<td>7</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>16</td>
</tr>
<tr>
<td>Met (M)</td>
<td>7</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>33</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>20</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>29</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>3</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>4</td>
</tr>
<tr>
<td>Glu (G)</td>
<td>28</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>25</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>21</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>39</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>23</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>13</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>18</td>
</tr>
<tr>
<td>Val (V)</td>
<td>26</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>14</td>
</tr>
</tbody>
</table>

Total Number of amino acids: 405
5.10. **NUCLEOTIDE SEQUENCING OF \textit{hpcC}.**

The position of \textit{hpcC} with respect to the cloned \textit{E.coli} C DNA has been known for some time. The internal \textit{EcoR1} and \textit{Sal1} sites of pJJ801 and the corresponding sites in pJJ200 and pJJ210 were known to lie within the \textit{hpcC} ORF on the basis of subcloning experiments carried out by Jenkins (1987) and Fawcett (1989). Southern blotting experiments carried out by Fawcett (1989) positioned the 5' end of the gene close to the \textit{Sal1} site. Further experiments by Fawcett (1989) identified the direction of transcription of \textit{hpcC} such that the \textit{EcoR1} site lies downstream of the \textit{Sal1} site in the \textit{hpcC} ORF. Double stranded nucleotide sequencing of \textit{hpcC} was achieved with a variety of deletion subclones, synthetic oligonucleotides and other subclones.

The amino terminal region of the gene was sequenced on the coding strand using the deletion subclones pDR83515 and pDR83516. This sequence was checked by sequencing the untranslated strand from the \textit{Sal1} site of pDR1835 using universal sequencing primer. A second plasmid was constructed to extend this sequence past the 5' end of the gene toward the 3' end of \textit{hpcEF}. A single \textit{SstII} site occurs approximately 150 bps upstream of the \textit{Sal1} site which allowed the intervening region to be removed after appropriate manipulation. Approximately 2\mu g of pDR1835 was cut with \textit{PstI} and \textit{SstII} and then treated with mung bean nuclease to produce blunt ends. The linearised plasmid was ligated and used to transform competent \textit{E.coli}. The resultant transformants were then prepared by mini-prep and the plasmid DNA analysed by restriction digest. A construct which had lost the 150 bp \textit{Sal1-SstII} was identified. This construct was designated pDR1837.

The central region of \textit{hpcC} between the \textit{Sal1} and \textit{EcoR1} sites was sequenced on both strands after isolation of the 0.5 Kbp region in the construct pDR1905. This plasmid was created by digestion of pDR1930 with \textit{EcoR1}
followed by isolation of a 3.2 kbp fragment from a 1% low melting point gel. After extraction, the fragment was ligated and used to transform competent cells. Several single colonies were grown in small scale LB cultures from which plasmid DNA was prepared. The DNA was analysed by restriction digest and a construct which retained the 0.5 Kbp EcoR1-SalI identified. This construct was designated pDR1905. The 0.5 Kbp SalI-EcoR1 region was then sequenced on both strands using universal and reverse sequencing primers. The nucleotide sequence obtained was extended using two synthetic oligonucleotides. The oligonucleotide DR06 extended the nucleotide sequence of the translated strand toward the EcoR1 site on the translated strand whereas DR07 was used to extend the nucleotide sequence toward the SalI site on the untranslated strand. The nucleotide sequence obtained from sequencing pDR1930 using DR06 confirmed that single EcoR1 site exists in the hpcC ORF. Sequencing of pJJ801 with DR07 confirmed that a single SalI site occurs in the hpcC ORF.

The construct pTF802, (Fawcett 1989) which contains the 2.5 Kbp EcoR1-BamH1 fragment of pJJ801 in pUC18 was used to extend the nucleotide sequence from the EcoR1 site toward the 3' region of hpcB. Two members of the pDR93 deletion series: pDR9301 and pDR9302 were then used to complete sequencing of the translated strand as far as the 5' end of hpcB which lies downstream of hpcC. The untranslated strand of this region was then sequenced as far as the EcoR1 site using two oligonucleotides primers: DRT02 and DR11. Complete restriction maps of pDR1987, pDR1905 and pTF802 are given in Fig.5.10.1. The constructs and sequencing strategy required to sequence hpcC are shown in Fig.5.10.2. Analysis of the nucleotide sequence obtained identified an ORF encoding a polypeptide of 449 amino acids which started 250 bps upstream of the single SalI site. This ORF had an associated ribosome binding site 10 bps upstream of the ATG initiation codon. The ORF
was positively identified as that encoding *hpcC* when the predicted amino terminal sequence was compared to the first thirty four amino acids determined by amino terminal amino acid sequencing of the purified enzyme (Fawcett *et al* 1989). The complete nucleotide sequence of *hpcC* and the translated ORF is presented in Fig.5.10.3

5.11 **AMINO ACID COMPOSITION OF THE CHMS DEHYDROGENASE SUBUNIT AND SUBUNIT MOLECULAR WEIGHT.**

The predicted ORF for CHMS dehydrogenase which encodes a polypeptide of 449 amino acids was analysed as detailed in section 5.4.1. to provide the data for amino acid analysis. The amino acid composition of the enzyme is detailed in Fig.5.11.1. From this data the subunit molecular weight of the enzyme is 48,775 Da which is consistent with a figure of 52,000 deduced from SDS-PAGE analysis (Fawcett *et al* 1989).
Fig. 5.10.1.

Restriction maps of pDR1937, pDR1905 and pTF802.

The cloned DNA of the genomic clone pJJ801 is shown for reference. The subclone pDR1837 contains the 3.7 Kbp SstII-BamH1 region of pJJ801 and was constructed from pDR1835 such that the SstII site was destroyed during construction. pDR1905 contains the 0.5 Kbp EcoR1-SalI region in pUC19 and was constructed from pDR1930. pTF802 contains the 2.5 Kbp BamH1-EcoR1 region in pUC19 and was constructed by Fawcett (1989). Restriction sites are as follows: B=BamH1, Ev=EcoRV, Nd=NdI, Kp=KpnI, E=EcoRI, Sc=SstI, S=SalI, Pvi=PvuI, PvII=PvuII, H=HindIII, Sp=Sphi, Ps=PstI, Sm=SmaI, Ss=SstII.
Fig. 5.10.1.

Restriction maps of pDR1937, pDR1905 and pTF802.
A variety of clones and oligonucleotides were required to sequence hpcC. The cloned DNA of the genomic clone pJJ801 is shown for reference with the position and extent of hpcC shown by the hatched box. The 5' region of the gene was sequenced on the coding strand using the deletion subclones pDR83515 and pDR83516. The non-coding strand was then sequenced using pDR1835 and pDR1837. The mid region of the gene between the SalI and EcoR1 sites, encompassing approximately 500 bps was then sequenced using pDR1905. Double strand sequence was then obtained using two oligonucleotide primers DR06 and DR07. The remaining part of the gene from the EcoR1 site to the 3' terminus was then sequenced on the coding strand using pTF302, pDR9301 and pDR9302. pDR9301 and pDR9302 were derived from the pDR93 series of deletion subclones (see section 5.2.1.). The non-coding strand was then sequenced after the construction of two oligonucleotide probes: DRT02 and DR11. The small arrows indicate the start, direction and extent of nucleotide sequencing. carried out using universal or reverse primer as appropriate. The large arrows refer to the extent and direction of sequencing carried using oligonucleotide primers constructed to the sequence obtained from the subclones. Restriction sites are as follows: B=BamHI, Ev=EcoRV, Nd=NdI, Kp=KpnI, E=EcoRI, S=SalI, Pvl=PvuI, PvII=PvuII, H=HindIII, Sp=SphI, Ps=PstI, Sm=SmaI, Ss=SstII
Fig. 5.10.2.

Constructs required to sequence *hpcC*
Fig.5.10.3.

**Nucleotide sequence of hpcC.**

The nucleotide sequence is numbered in relation to the first nucleotide in the mRNA transcript that starts upstream of *hpcEF*. A region that bears homology to the Shine-Dalgarno site, or ribosome binding site (RBS) is located nine bps upstream of the ATG initiation codon for *hpcC* as shown. The positions of the SstII, SalI, and EcoRI restriction sites within the *hpcC* ORF are shown.
**Fig. 5.10.3.**

The nucleotide sequence and translated ORF of *hpcC*.

<table>
<thead>
<tr>
<th>RBS</th>
<th><em>hpcC</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>GTTGAGTGGAAACACGCGAAATGAAAAAGTAAATCATTTGATACACCGGCAAAAAATGTT</td>
<td>1358</td>
</tr>
<tr>
<td>CAACCTCACTCGTCTTCGTTTTATGTTACGATGCAGGTGGAAGCCGTTGCATTACAA</td>
<td>1358</td>
</tr>
<tr>
<td>MetLysValAsnHisTrpIleAsnGlyLysAsnVal</td>
<td></td>
</tr>
<tr>
<td>GCAGGTAACGACTACTTCCTGCCACCAATCCGGCAACGGGTGAAGTGCTGGCGGATGTG</td>
<td>1538</td>
</tr>
<tr>
<td>TAGCTGGTCTTGCACGGTCTCTAGCGGCCTACCTTTGGCGCCTGTGCCCGGACGGCTAG</td>
<td>1598</td>
</tr>
<tr>
<td>HisGlnThrLysAsnValLeuIleAsnMetGlulThrAlaAspGlyGluAspAla</td>
<td></td>
</tr>
<tr>
<td>GCCTCTGGCGGTAAAGCGGAGATCAATCACGCGGTATGGCGCTTGCGGATGTG</td>
<td>1418</td>
</tr>
<tr>
<td>TAGCTGGTCTTGCACGGTCTCTAGCGGCCTACCTTTGGCGCCTGTGCCCGGACGGCTAG</td>
<td>1598</td>
</tr>
<tr>
<td>HisGlnThrLysAsnValLeuIleAsnMetGlulThrAlaAspGlyGluAspAla</td>
<td></td>
</tr>
<tr>
<td>GCCAGGCGGCCACCTCGCTCAGTACCGGCCATCGGTGCTCCCTGCATACAGC</td>
<td>1478</td>
</tr>
<tr>
<td>AlaSerGlyGluAlaGluIleAsnGlnAlaValThrAlaAspVal</td>
<td></td>
</tr>
</tbody>
</table>

**Sst2**

| GTTGGTTCAGAAGACGCTGTTGAGGTGTTGAGGGTTTATGACC | 1718 |
| GACCAGCGCGCCATCGCCAAACGACTGACACAGCGCATCGAGCGACGTGGCAGCAATACTGG | 1778 |
| LeuValGluValGlyValGlyValAlaValSerProThrAspAspGlyEurAsnTrpThr |
| GCACCTGGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1838 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1838 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1838 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1898 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1598 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1598 |

**Sal1**

| CTGGTTCAGAAGACGCTGTTGAGGTGTTGAGGGTTTATGACC | 1778 |
| GACCAGCGCGCCATCGCCAAACGACTGACACAGCGCATCGAGCGACGTGGCAGCAATACTGG | 1778 |
| LeuValGluValGlyValGlyValAlaValSerProThrAspAspGlyEurAsnTrpThr |
| GCCACCTGGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1838 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1838 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1838 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1898 |
| GCCACCTCGGCAAGATCGCCGCTGTTGAGGTGTTGAGGGTTTATGACC | 1598 |

Sst2
Fig. 5.10.3. The nucleotide sequence and translated ORF of hpcC continued:

1959 CATCATGACGCCGCGGTGTGTCTACACCGGCGTTACCGGCGTCTCGGCT 2018
GTAGATCGCAACCCACCGTCGCGGCTGCGGTGGGAAGATATCTATCTATCGTT
HisHisAspValArgAlaValSerPheThrGlyThrAlaThrGlyArgAsnIleMet

1990 AAAAAGCCGGGGCTGGAAAAATACTCCATGGAACTGGGCGGTAAATCGCCGGTGCTGATT
2078 GGTAGTACTGCACGCACGGCACAGCAAGTGGCCGCCATGCCGCTGGCCCGCGTTATAGTAC
HisHisAspValArgAlaValSerPheThrGlyThrAlaThrGlyArgAsnIleMet

2039 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2138 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2079 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2138 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2119 AACGCCGACGCCGCTCAGCCGGCTGGGGCAATCTTTATTTCAACAAAGCATCTACCGGAA
2198 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
HisHisAspValArgAlaValSerPheThrGlyThrAlaThrGlyArgAsnIleMet

2159 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2238 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2199 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2258 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2239 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2318 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2279 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2378 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2319 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2438 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2359 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2498 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2399 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2558 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2439 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2618 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2479 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2678 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle

2519 TTTTGGCCGCCACTTTTTTTATTAGGACGATCTGACCCCGGCAATTGCGGCAACGACTAA
2678 LysAsnAlaGlyLeuLysLysTyrSerMetGluLeuGlyGlyLysSerProValLeuIle
**Fig. 5.11.1.**

Subunit amino Acid composition for CHMS dehydrogenase

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>53</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>23</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>19</td>
</tr>
<tr>
<td>His (H)</td>
<td>8</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>21</td>
</tr>
<tr>
<td>Met (M)</td>
<td>12</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>23</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>24</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>25</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>6</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>5</td>
</tr>
<tr>
<td>Glu (D)</td>
<td>23</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>38</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>22</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>42</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>23</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>15</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>16</td>
</tr>
<tr>
<td>Val (V)</td>
<td>38</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>9</td>
</tr>
</tbody>
</table>

Total Number of amino acids: 449
The location of \textit{hpcG} which encodes OHED hydratase was assigned by Fawcett (1989) by a combination of Southern blotting and nucleotide sequencing. Using the clones available, Fawcett located the 5' region of \textit{hpcG} to lie 4 bps downstream from the most internal of the two EcoRV sites in pJJ801. The amino terminal sequence of the purified enzyme is known (Ferrer and Cooper 1988) and corresponded exactly to the ORF identified.

The nucleotide sequence obtained by Fawcett was extended and checked on the untranscribed strand using two constructs during this study. The 5' region of the gene was sequenced on both strands using pTF302 which contains the 279 bp EcoRV-EcoRV region of pJJ801 in pUC18. The remaining region of \textit{hpcG} encoded on pJJ801 was then sequenced using the construct pDR1903. The 263 bps region between the most external EcoRV site and the left hand (as drawn) \textit{BamH}1 site of pJJ801 was isolated by digestion of pDR1930 with EcoRV and \textit{SalI} followed by treatment with mung bean nuclease. The linearised fragment was then recovered from a low melting point gel, ligated and used to transform competent cells. The resultant 3.0 Kbp plasmid isolated was designated pDR1903 and is shown with pTF830 in Fig.5.12.1.
The subclone pDR1903 was created from pDR1930 as detailed in the text to isolate the 0.3 Kbp BamHI-EcoRV region. The 0.3 Kbp EcoRV-EcoRV region was isolated and cloned into pUC18 by Fawcett (1989) to produce pTF830. The subclone pDR1930 which contains the 3.0 Kbp BamHI-SalI region cloned into pUC19 is shown for reference. Restriction sites are as follows: B=BamHI, Ev=EcoRV, Nd=NdeI, Kp=KpnI, E=EcoRI, Sc=SstI, S=SalI, Pv= PvII, H=HindIII, Sp= SphI, Ps= PstI, Sm= SmaI, Ss= SstII.
Fig. 5.12.1.

Restriction maps of pDR1903 and pTF830.
The cloned DNA was then sequenced using universal and reverse sequencing primers. To ensure that the sequence obtained was continuous through the most external *EcoRV* site, pDR1930 was sequenced using universal sequencing primer. Similarly, the nucleotide sequence adjacent to the most external *EcoRV* site was checked using the smallest of the pDR93 deletion subclones: pDR9313. An open reading frame of 547 bps encoding the first 184 amino acids of OHED hydratase was identified. The first 34 residues of the predicted ORF corresponded exactly to the amino terminal amino acid sequence which has been identified from the purified enzyme (Ferrer & Cooper 1988). The available nucleotide sequence and translated ORF of *hpcG* are shown in Fig.5.12.2.

The amino acid composition of the first 184 residues of OHED hydratase make up some 20,901 Da of the total subunit molecular weight of the enzyme. The purified enzyme has a subunit molecular weight of 32,500 as assessed by SDS-PAGE so the region identified in this study makes up in excess of 60% of the total primary sequence. This represents a significantly large portion of the amino acid sequence with which to carry out a similarity search against the database. Data base searches are detailed in section 5.18.
Using the clones available it was possible to sequence the first 541 bps of \textit{hpcG}. The nucleotide sequence is numbered in relation to the first nucleotide in the mRNA transcript that starts 5' to \textit{hpcEF}. A region that bears homology to the Shine-Dalgarno site, or ribosome binding site (RBS) is located seven bps upstream of the ATG initiation codon for \textit{hpcG} as shown. The position of the \textit{EcoRV} and \textit{BamH1} restriction sites within the \textit{hpcG} ORF are marked as shown. The \textit{EcoRV} restriction site that occurs outside the \textit{hpcG} ORF overlaps with the RBS. The \textit{BamH1} site shown within the \textit{hpcG} ORF is not a true \textit{BamH1} site and forms part of the polylinker of pUC18 in the clones used for sequencing and is also the \textit{BamH1} site of pBR328 which was originally used to clone this part of the \textit{hpc} gene cluster. Other genomic clones isolated by Jenkins (1987) which encode all of the \textit{hpc} genes do not contain a \textit{BamH1} site in the same position.
Available nucleotide sequence and translated ORF of hpcG.

**EcoRV**

**hpcG**

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGCATCGGCAATGCCCACTGCTATGTTGTAATAAACACACCACACCCTCTGATC</td>
<td>MetPheAspLysHisThrHisThrLeuIle</td>
</tr>
<tr>
<td>GCCCGTGCTCTGGGATCAGGCAGAAAAACAGCGCGAACAGATCCGCGCGATCTCGCTGGAT</td>
<td></td>
</tr>
<tr>
<td>CGGGTGCTCCTAGCTTGCTCTCCCCTTTCCTCAGGCGCTGCTAGAGCGACCTA</td>
<td>AlaGlnArgLeuAspGlnAlaGluLysGlnArgGluGlnIleArgAlaIleSerLeuAsp</td>
</tr>
<tr>
<td>ATGGCGGAGATCCATCGGAGGCCTTTAGCTGCTGCACTTACCCAAGCGGACTTT</td>
<td>TyrProGluIleThrIleLysAlaThrAlaValGlnArgGluTrpValArgLeuLys</td>
</tr>
<tr>
<td>ATCGCCGAAGTGCCAGCTGAAGCCCAAAAATCGGCGCTGACCTCGAAAGCGATGCA</td>
<td></td>
</tr>
<tr>
<td>GCAAGCTCCGATACCCGAGCCGATACCCGCTCGCTGATGCAAATGCTGCTGCA</td>
<td>AlaSerSerGlnIleSerGluProAspTyrGlyAlaLeuLeuHisAspAsnValLeuPro</td>
</tr>
<tr>
<td>GCAAAACCGCTGCGGCAATGCCCACTGCTATGTTGTAATAAACCCACACCCTCTGATC</td>
<td></td>
</tr>
<tr>
<td>CGCTACCTTGAATGCTTGCTGCTGCACTTACCCAAGCGGACTTT</td>
<td>TyrArgAlaArgThrGluThrLeuLeuHisGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>GCGGGCGCTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>AlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>TACCTGGGCTTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>IleAlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>ATCGCCGAAGTGCCAGCTGAAGCCCAAAAATCGGCGCTGACCTCGAAAGCGATGCA</td>
<td></td>
</tr>
<tr>
<td>GCAAGCTCCGATACCCGAGCCGATACCCGCTCGCTGATGCAAATGCTGCTGCA</td>
<td>AlaSerSerGlnIleSerGluProAspTyrGlyAlaLeuLeuHisAspAsnValLeuPro</td>
</tr>
<tr>
<td>GCAAAACCGCTGCGGCAATGCCCACTGCTATGTTGTAATAAACCCACACCCTCTGATC</td>
<td></td>
</tr>
<tr>
<td>CGCTACCTTGAATGCTTGCTGCTGCACTTACCCAAGCGGACTTT</td>
<td>TyrArgAlaArgThrGluThrLeuLeuHisGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>GCGGGCGCTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>AlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>TACCTGGGCTTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>IleAlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>ATCGCCGAAGTGCCAGCTGAAGCCCAAAAATCGGCGCTGACCTCGAAAGCGATGCA</td>
<td></td>
</tr>
<tr>
<td>GCAAGCTCCGATACCCGAGCCGATACCCGCTCGCTGATGCAAATGCTGCTGCA</td>
<td>AlaSerSerGlnIleSerGluProAspTyrGlyAlaLeuLeuHisAspAsnValLeuPro</td>
</tr>
<tr>
<td>GCAAAACCGCTGCGGCAATGCCCACTGCTATGTTGTAATAAACCCACACCCTCTGATC</td>
<td></td>
</tr>
<tr>
<td>CGCTACCTTGAATGCTTGCTGCTGCACTTACCCAAGCGGACTTT</td>
<td>TyrArgAlaArgThrGluThrLeuLeuHisGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>GCGGGCGCTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>AlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>TACCTGGGCTTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>IleAlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>ATCGCCGAAGTGCCAGCTGAAGCCCAAAAATCGGCGCTGACCTCGAAAGCGATGCA</td>
<td></td>
</tr>
<tr>
<td>GCAAGCTCCGATACCCGAGCCGATACCCGCTCGCTGATGCAAATGCTGCTGCA</td>
<td>AlaSerSerGlnIleSerGluProAspTyrGlyAlaLeuLeuHisAspAsnValLeuPro</td>
</tr>
<tr>
<td>GCAAAACCGCTGCGGCAATGCCCACTGCTATGTTGTAATAAACCCACACCCTCTGATC</td>
<td></td>
</tr>
<tr>
<td>CGCTACCTTGAATGCTTGCTGCTGCACTTACCCAAGCGGACTTT</td>
<td>TyrArgAlaArgThrGluThrLeuLeuHisGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>GCGGGCGCTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>AlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>TACCTGGGCTTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>IleAlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>ATCGCCGAAGTGCCAGCTGAAGCCCAAAAATCGGCGCTGACCTCGAAAGCGATGCA</td>
<td></td>
</tr>
<tr>
<td>GCAAGCTCCGATACCCGAGCCGATACCCGCTCGCTGATGCAAATGCTGCTGCA</td>
<td>AlaSerSerGlnIleSerGluProAspTyrGlyAlaLeuLeuHisAspAsnValLeuPro</td>
</tr>
<tr>
<td>GCAAAACCGCTGCGGCAATGCCCACTGCTATGTTGTAATAAACCCACACCCTCTGATC</td>
<td></td>
</tr>
<tr>
<td>CGCTACCTTGAATGCTTGCTGCTGCACTTACCCAAGCGGACTTT</td>
<td>TyrArgAlaArgThrGluThrLeuLeuHisGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>GCGGGCGCTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>AlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
<tr>
<td>TACCTGGGCTTCCAGCGTGCGACTTTCCGGTGTTTTAGCCGGACTGGAGCTTTCGCTACGTC</td>
<td>IleAlaGluGluGlyArgThrLeuLysGlyHisLysIleGlyLeuThrSerLysAlaMetGln</td>
</tr>
</tbody>
</table>

**BamHI**

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGAGGTGCAGCACTAGCATGAAGCCCAACTCTCGGGGTGACCTC</td>
<td>GlyGlyArgProIleLysProAspGluLeuAsp</td>
</tr>
<tr>
<td>GCTCCACAGGTAGCTGCTGCTGACTAGGCGCCG</td>
<td></td>
</tr>
</tbody>
</table>

203
From an analysis of restriction digests of pDRF and pDRF', it was estimated that only a few hundred base pairs would lie between the BamH1 sites that define the ends of the cloned DNA in the two constructs. This was investigated by nucleotide sequencing. This strategy required the construction of an oligonucleotide primer: P1219, which anneals to a region of pBR322 within 3 bps of the vector encoded BamH1 site, and would allow nucleotide sequencing toward the HindIII site of pBR322. Both pDRF and pDRF' were sequenced with this primer which showed an identical region of approximately 100 bps, to occur in both constructs in the region closest to the P1219 annealing site. This identical region was terminated with a BamH1 site followed by a further region of non-identical DNA.

When this non-identical region of pDRF was analysed with the WIMP package of programmes, an ORF was identified which bears near perfect identity with the amino terminal region of the serine chemoreceptor protein of E.coli K12 (Boyd et al 1983). The gene tsr, which encodes this protein, maps to minute 99 on the genetic linkage map of the K12 genome (Bachman 1987).

The occurrence of a shared region of DNA between pDRF and pDRF' was unexpected. It seems likely that this region was obtained by accident in the original construction of pJJ200 from which pDRF and pDRF' were subsequently derived. The cloned DNA was originally obtained by Sau3A1 digestion of chromosomal E.coli C DNA followed by ligation into BamH1 digested pBR322 to create pJJ200. Thus it is possible that a small piece of DNA was obtained during this process, which have no relationship to the major piece of cloned DNA which encodes the Hpc genes and part of tsr. The other explanation: that the pBR322 vector originally used had an insertion between what should have been the vector encoded, and now non-functional BamH1 site
and the BamH1 site found approximately 100 bps away, seems more unlikely. The pBR322 BamH1 site lies within the tetracycline resistance gene (tet). pJJ200 was originally selected on the basis of insertion inactivation and the ability to complement an hpc mutation in the strain E.coli JJ200 (Jenkins 1987). Such a large insert in the tet gene of pBR322 may have resulted in tetracycline sensitivity of the uncut plasmid, and should have been noticed in some of the control experiments performed.

The original purpose of sequencing pDRF1 with P1219, was to locate an extra BamH1 site within the cloned DNA which would lie within the hpcR ORF. This would explain why pDRF1' expresses hpcEF constitutively, whereas pDRF1 does not. Deletion of the region between the two BamH1 sites might produce a truncated form of the repressor, or severely cripple any protein produced from the remainder of the ORF. A second BamH1 site was not found upon inspection of the nucleotide sequence obtained from pDRF1, sequenced with P1219. A further oligonucleotide primer: DR15 was then constructed to the sequence obtained from pDRF1 which enabled a further 280 bps to be obtained, but again no BamH1 site was identified in this new sequence. A third oligonucleotide primer: DR16 was constructed to extent the nucleotide sequence. Using this primer, an internal BamH1 site was identified which lies 700 bps from the non-functional BamH1 site of pBR322. The nucleotide sequence upstream of this internal BamH1 site corresponded exactly to that obtained from pDRF1' when sequenced with the primer P1219.

One further question was then addressed to the nucleotide sequence at this end of the cloned DNA: where does the BamH1 site of pJJ801 occur in the hpcR gene? Sequencing of pDR1835 (see section 3.2.1.) from the BamH1 end of the cloned DNA did not show any similarity with that found from pDRF1 and pDRF1' which would have been expected if the BamH1 end of pJJ801 was
identical to the \textit{BamH1} end of pDRF$_1$. A fourth oligonucleotide primer; DR17 was constructed to extended the sequence obtained from pDRF$_1$ sequenced with DR16, to see if any match could be made to pDR1835. When the nucleotide sequence obtained from DR17 was analysed, it became apparent that an AT rich region occurs approximately 900 bps from the non-functional \textit{BamH1} site of pDRF$_1$. This AT rich area was exactly the same as that obtained from the deletion subclones derived from pJJ801. Furthermore, the \textit{hpcEF} ORF was found to occur 3' to this AT rich region of pDRF$_1$.

It is apparent that not all of the genomic DNA that Jenkins (1987) cloned into pBR328, to create pJJ801, was from the \textit{hpc} gene cluster. Approximately 1.2 Kbp if DNA which lies upstream of \textit{hpcEF} and the \textit{hpc} promoter region does not encode \textit{hpcR} and is not contiguous with the rest of the \textit{hpc} structural genes. This would explain why pJJ801 constitutively expresses \textit{hpc(EF)CBD}. The nucleotide sequence obtained from pDRF$_1$ in the region between the \textit{hpcEF} ORF and the most external of the two \textit{BamH1} sites, represents DNA of the \textit{hpc} gene cluster since \textit{hpcR} is encoded in this region.

5.14. DOUBLE STRAND NUCLEOTIDE SEQUENCING OF \textit{hpcR}.

Double stranded nucleotide sequence of the region encoding \textit{hpcR} was obtained using oligonucleotide primers, complementary to the sequence obtained as detailed above. Four primers were produced: DR18, 19, 20 and 21, which were used to sequence a region of approximately 1000 bps. This encompassed the region immediately upstream of the \textit{hpc} structural gene promoter and extended to the most external of the two \textit{BamH1} sites of pDRF$_1$. The sequence obtained was analysed using the WIMP package of programmes and several possible ORFs were identified. Only one of these ORFs extended through the internal \textit{BamH1} site and is in the opposite orientation to the rest of the \textit{hpc} genes. A region which bears similarity to the Shine-Dalgarno,
ribosome binding site, was found approximately six bps upstream of the initiation codon for this ORF. None of the other possible ORFs identified by the computer analysis had a ribosome binding site upstream of an initiation codon or extended through the most internal BamHI site. The \textit{hpcR} ORF is separated from \textit{hpcEF} by 191 bps and terminates 118 bps before the \textit{tsr} gene. The sequencing strategy used in the region encoding \textit{hpcR} is shown in Fig.5.14.1. The nucleotide sequence of this region is shown in Fig.5.14.2.

5.14.3. **AMINO ACID COMPOSITION OF THE Hpc REPRESSOR AND SUBUNIT MOLECULAR WEIGHT.**

The translated ORF for the Hpc repressor was analysed as described in section 4.5.1. to determine the amino acid composition and subunit molecular weight. The \textit{hpcR} ORF encodes a polypeptide of 148 amino acids which has a molecular weight of 17,183 Da. The amino acid composition of the Hpc repressor from the translated ORF is presented in Fig.5.14.3.
Fig. 5.14.1.

Sequencing strategy used for the region encoding *hpcR*

The nucleotide sequence on one strand was obtained by sequencing pDRF1 and pDRF1' with an oligonucleotide probe (P1219) which anneals to a site within 4 bps of b' in the vector DNA. The sequence between pDRF1 and pDRF1' was obtained using oligonucleotides DR15 and DR16 which were constructed from the sequence obtained from pDRF1. The sequence obtained from pDRF1' was then extended to the *hpc* structural gene promoter region using a third oligonucleotide DR17. The nucleotide sequence of the other strand was then obtained using five oligonucleotides DRT01, DR18, DR19, DR20 and DR21 as shown. The position of an ORF thought to be that of *hpcR* which is transcribed from left to right as drawn is shown by the shaded box. Restriction sites are as follows: B=*BamH1*, Nd=NdEl, E=E*coRI*, S=SalI, PvuI=PvuI, PvuII=PvuII, Ps=PstI, Sm=SmaI, B'=silence *BamH1* site of pBR322
Fig. 5.14.1.
Sequencing strategy used for the region encoding hpcR
Fig. 5.14.2.

Nucleotide sequence of the region encoding hpcR.

The nucleotides are numbered from the first base upstream from the first nucleotide of the transcript encoding the pathway enzymes. Analysis of the nucleotide sequence identified an ORF which extends for 444 bps and is orientated in the opposite direction to the rest of the hpc genes. This ORF extends through the most internal BamH1 site of pDRF1 and terminates 119 bps from the initiation codon for tsr (see section 5.13). Restriction sites within this region are marked.
Fig. 5.14.2.

Nucleotide sequence of the region encoding hpcR.

RBS

hpcR

AGGGAAAGATAATGCACGACTCTCTAACCATTGCGTTGCTCCAGGCGCGAGAACGGCG

218  — + ---------------+-------------------+------------------+------------------+------------------+------------ 277

TCCCCTTTCTATACGTGAGAGGAGTTGGAACGCAACGAGGTCCGCGCTCTTCCCCG

MetHisAspSerLeuThrIleAlaLeuLeuGlnAlaArgGluAlaAla

ATGAGTTATTTTGCCTCAAATGCGATATTTGACTGAGCAACAGTGGCGCATC

278  — + ---------------+-------------------+------------------+------------------+------------------+------------ 337

TACTCAATAAAGCCTGTAGCTGAGAGATTGGTAACGCAACGAGGTCCGCGCTCTTCCCCG

MetSerTyrPheArgProIleValLysArgAsnLeuThrGluGlnGluThrArgIle

GTGCGTATTTCTGCGGAAAGCCATATGCGATATTTGACTGAGCAACAGTGGCGCATC

338  — + ---------------+-------------------+------------------+------------------+------------------+------------ 397

CACGCATAGGCCCTTTCGCGTAGTATCCATAAAGTCTAGACCCCATAGCCGAGACG

ValArgIleLeuAlaGluSerProSerMetAspPheHisAspLeuAlaTyrArgAlaCys

BamH1

ATTTTCGCCTCAAGCGAGATCGTCCACGCGTTAGGCAATTTGACTGAGCAACAGTGGCGCATC

398  + --------+------------------+------------------+ ------------------+-------------- 457

TAAAAGGCCGTTTCCCTGAGAACATGCTACCTTTGCGCTGCCAAATACACAAC

IleLeuArgProSerLeuThrGlyIleLeuThrArgMetGluArgAspGlyLeuValLeu

CGATAAAGCCGATCATGATACCCGAGAGGCTGACTTATATCCGTAGCCAAAGAGGGGCAG

458  — + -----------------+-------------------+-----------------+------------------+------------------+------------ 517

GCTAATTTCGCTGATAGATCTAGTCCCTAGATAGACCCCATAGCCGACTTCTCTCTCCCCGTC

ArgLeuUysProIleAsnAspGlnArgLysLeuTyrIleSerLeuThrLysGluGlyGln

GGCCTGTATTACCCGCTGTTGGCGACTGCTACTTCTGGAATGCTATTGCTATTGCTGTT

518  — + -----------------+-------------------+-----------------+------------------+------------------+------------- 577

GCCGACATATGGCGAGGATCTTGTCCTACTTTCTCTGAGATGCTTAATTCTGCTGATC

AlaLeuTyrAsnArgAlaGlnThrGlnIleGluAlaTyrArgGlnIleGluAlaGln

TTTACTGCGCGAAATAAGCTAACCCTCTGTTGATAGATTTATTCTCTGGTCT

578  + -----------------+-------------------+-----------------+------------------+------------------+------------- 637

AAAAATGCCGTTTTTGTACATCTTGAGGAAACATCTCTGACAAATGAAACTGACCCA

PheThrAlaGluLysMetGlnGlnLeuHisLeuLeuGluPheIleAlaLeuGly

Sma1

AAATTCCGCAAGAAGACATGCGGCGCAATTAAATGAAATTTTTTAGTATCTG

638  + -----------------+-------------------+-----------------+------------------+------------------+------------- 690

TTAACGCGCAGTTCTCTCTCTGTAGGCAGGCGCTATATTACATTTAAAATCATAGAC

AsnSerArgGlnGluAspIleProGlyAspAsnGlu
Fig. 5.14.3.
Subunit amino Acid composition of the Hpc repressor.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A):</td>
<td>14</td>
</tr>
<tr>
<td>Asp (D):</td>
<td>7</td>
</tr>
<tr>
<td>Phe (F):</td>
<td>4</td>
</tr>
<tr>
<td>His (H):</td>
<td>4</td>
</tr>
<tr>
<td>Lys (K):</td>
<td>5</td>
</tr>
<tr>
<td>Met (M):</td>
<td>4</td>
</tr>
<tr>
<td>Pro (P):</td>
<td>5</td>
</tr>
<tr>
<td>Arg (R):</td>
<td>14</td>
</tr>
<tr>
<td>Thr (T):</td>
<td>8</td>
</tr>
<tr>
<td>Trp (W):</td>
<td>1</td>
</tr>
<tr>
<td>Cys (C):</td>
<td>1</td>
</tr>
<tr>
<td>Glu (D):</td>
<td>13</td>
</tr>
<tr>
<td>Gly (G):</td>
<td>5</td>
</tr>
<tr>
<td>Ile (I):</td>
<td>12</td>
</tr>
<tr>
<td>Leu (L):</td>
<td>19</td>
</tr>
<tr>
<td>Asn (N):</td>
<td>5</td>
</tr>
<tr>
<td>Gln (Q):</td>
<td>12</td>
</tr>
<tr>
<td>Ser (S):</td>
<td>7</td>
</tr>
<tr>
<td>Val (V):</td>
<td>3</td>
</tr>
<tr>
<td>Tyr (Y):</td>
<td>5</td>
</tr>
</tbody>
</table>

Total Number of amino acids: 148
5.15 **NUCLEOTIDE SEQUENCING FROM THE HindIII END OF pDR1865.**

The identification of the *tsr* gene at one boundary of the *hpc* gene cluster as described in section 5.13, prompted a similar investigation at the other end of the cloned DNA. The construct pDR1856 (see section 3.5.2.) was chosen for this purpose. Nucleotide sequencing from the *HindIII* of pDR1865 was carried out from the universal sequencing primer site of the vector. The first 295 bps were obtained on one strand only. When this data was compared to the EMBL nucleotide sequence database, a region with 97% similarity was identified. This match occurs within the region encoding the P15B and P35B genes of *E.coli* K12 (Sands *et al.* 1988). These two genes are found in an area between 68 and 69 minutes on the *E.coli* K12 genome in a region which is thought to encode proteins involved with transcription and translation. The proteins involved with these particular functions in *E.coli* K12 are transcribed clockwise in two operons: the *metY* operon, containing *metY*, P15A, *nusA*, *infB*: and about a 1000 bps further downstream, the *rpsO* and *pnp* operon. The P15B and P35B genes have been shown to lie within the intervening region and are cotranscribed with *rpsO*.

5.16 **TRANSCRIPT MAPPING STUDIES.**

With the nucleotide sequence and gene order established as described above, some more detailed aspects of the expression of the *hpc* genes could be addressed. Fawcett (1989) carried out Northern blotting experiments to try and determine the number and size of transcripts produced by pJJ200 grown in induced and uninduced conditions. In total three transcripts were identified one of which was assigned to the repressor gene transcript and two others encoding the structural genes. These two transcripts were of approximately 2.7 kbp and 4.5 kbp approximately. The existence of a promoter structure upstream of *hpcEF* seemed certain since no structural genes for the Hpc
pathway enzymes were found upstream of \textit{hpcEF}. A single transcript which would encompass both \textit{hpcEF} and \textit{hpcC} would have a minimum length of 2630 bps and might therefore correspond to the 2.7 kbp transcript identified by Fawcett. Assuming this to be the case, a second promoter would be located in front of \textit{hpcB} and direct the transcription of the 4.5 Kbp transcript identified by Fawcett. A third promoter, directing the transcription of \textit{hpcR}, would be expected to lie 5' to \textit{hpcEF}. The existence of a promoter upstream of \textit{hpcEF} was investigated using transcript mapping.
Various lines of evidence, as detailed in previous sections, indicated that a promoter structure would be located upstream of \textit{hpcEF}. It was known that the direction of transcription of the \textit{hpc} genes was such that \textit{hpcEF} would be the first ORF to be transcribed in any polycistronic mRNA molecule, since no other gene encoding an enzyme of the pathway was located upstream of \textit{hpcEF}. This hypothesis seemed more likely when the nucleotide sequence upstream of \textit{hpcEF} became available. Simple inspection of this sequence identified several areas in which a promoter may have been located. Since nucleotide sequence information was available, transcript mapping by primer extension, rather than S1 mapping was performed. An oligonucleotide primer: DRT01 was synthesized which would anneal to the first 21 residues of the \textit{hpcEF} ORF in any mRNA transcript. Messenger RNA was prepared from cells harbouring pDR1835 and from cells harbouring pUC18 as a control. The primer DRT01 was hybridized to mRNA prepared from these cells as described in Methods and Materials which was the extended to the 5' end of the mRNA molecule using M-MuLV reverse transcriptase, incorporating \textsuperscript{35}S-dATP into the extension product. Three control extension reactions were performed: DRT01 hybridized to mRNA from pUC18, mRNA from pDR1835 without any primer and an extension reaction containing DTR01 only. The first control reaction was performed to check that the mRNA from any transcript derived from pDR1835, was not forming secondary structures which could act as templates for the reverse transcriptase enzyme. The reactions were performed and electrophoresed through 6% urea/acrylamide sequencing gel, along side nucleotide sequence generated from pDR1835 using DRT01 as the sequencing primer. This allowed a direct comparison of any bands indicating the start of a transcript, to the nucleotide sequence from which it came. As such any
promoter would lie close upstream. The result of this experiment is shown in Fig.5.16.1.

Bands indicating the start of mRNA transcripts are seen only in the track from the extension reaction containing mRNA from pDR1835 and DRT01. There are a variety of bands indicated, but one is by far the most intense and occurs 39 bps upstream of the first nucleotide in the hpcEF ORF. The most intense band maps to a T in the sequencing track. Since the nucleotide sequence is obtained from the non-transcribed strand, the mRNA transcript therefore starts with an A. Approximately 10 bps upstream of this A is a sequence TAAAAG which bears resemblance to the -10 consensus sequence of E.coli promoters (TTGACA\textsubscript{T}N_{15-19} TANANT: Hawley & McClure 1983). A -35 region is also found with the sequence TTGCCA which bears similarity to the consensus sequences found for E.coli promoters. This is shown in Fig.5.16.2.
Reverse transcriptase mapping upstream of *hpcEF*.

The Nucleotide sequence of the region immediately upstream of *hpcEF* is shown by the sequencing tracks A,C,G and T. This sequence was obtained using the oligonucleotide DRT01 which was also hybridized to total cellular RNA for reverse transcriptase mapping.

Lane 1: pDR1835 mRNA + DRT01
Lane 2: pUC18 mRNA + DRT01
Lane 3: pDR1835 mRNA only
Lane 4: DRT01 only

The arrow indicates the most intense reverse transcriptase product. The sequence in the region of transcriptional initiation is shown at the left and the initiating adenine is indicated by an asterisk.
Fig.5.16.2.

Promoter structure upstream of \textit{hpcEF}

Analysis of the nucleotide sequence immediately upstream of the initiating adenine of the mRNA transcript, identified two regions which bear homology to the consensus sequence for \textit{E.coli} promoters according to Hawley and McClure (1983). The nucleotide sequence of the \textit{hpc promoter} (hpc seq) is shown above the consensus sequence with exact match shown between.
Fig. 5.16.2.

Promoter region upstream of hpcEF.

-35

hpc seq 5' CTTTGCCAAAGTACGCAATCGGAATAAAAGTAATCATTA 3'

match :: ::

Consensus tcTTGACat

-10

TAtAaT

+1
5.17.1. **TRANSCRIPTIONAL TERMINATION 5' TO hpcR.**

The nucleotide sequence from a region 5' to hpcR to the limit of the sequence information obtained from pDRF₁ was analysed using the WIMP programme TERMINATOR which searches for rho-independent RNA polymerase terminators according to the method of Brendel & Trifonov (1984). This programme searches for GC-rich regions which are immediately followed by an AT-rich sequence which are characteristic of rho-independent transcriptional terminator. The programme also evaluates whether hair-pin structures would be formed in the GC-rich region. One region was identified which would have a high tendency to form a GC-rich hair-pin loop and was immediately followed by a run of T residues. This sequence was located 3' to hpcR, just inside the tsr gene ORF. This sequence is shown in Fig.5.17.a.

5.17.2. **CATABOLITE ACTIVATOR PROTEIN (CAP) BINDING SITE.**

It was shown by Jenkins (1987) that the expression of the Hpc enzymes of both *E.coli* C and *E.coli* 5K (pJJ210) were subject to glucose repression. This could be partially relieved by the addition of 5mM cAMP to the growth medium, suggesting a role for CAP in control of the hpc genes. The DNA sequence in the vicinity of the promoter upstream of hpcEF was inspected for inverted repeats resembling the consensus sequence for a CAP binding site (de Crombrugghe *et al* 1984) and a good candidate was found 13 bps upstream of the -35 region (Fig.5.17.b.).
a. Possible rho-independent transcriptional terminator.

Nucleotides which could form a hairpin-loop structure are shown by the arrows:

```
5' TTGTGACCGCTTACTGCTGGTTTTGGCCGT 3'
```

b. CAP binding site.

Consensus sequence: \(-AA\text{-}TGTGA\text{-}T\text{-}\text{-}\text{-}\text{-}\text{-}\text{-}\text{-}\text{-}TCAXATT-\)

Match: \:

\textit{Hpc sequence:} TGTAAATAGTTGTTAATTAGATCACATTTACATC
5.18. **OPERATOR SITE IN THE REGION BETWEEN hpcR AND hpcEF.**

It seems likely that an operator site would occur in the region between hpcR and hpcEF since this would allow coordinated control of expression of the regulator gene and the hpc pathway structural genes. This area contains a high proportion of AT residues, which may cause the topology of the DNA in this region to be altered providing a site for the repressor to bind. A large number of operator sites are marked by regions of diad symmetry, the trp operator site for example (Klig et al 1988). Computer analysis of the region between hpcR and hpcEF identified an inverted repeat sequence of 12 nucleotides, separated by 56 nucleotides, which may function as a site for the repressor protein to bind (Fig.5.18.).
The nucleotides are numbered from the first base upstream from the first nucleotide of the transcript encoding the pathway enzymes. The -10 and -35 regions of the structural gene promoter on the lower strand are marked by the hatched boxes. The catabolite activator protein (CAP) binding site is marked by the shaded area on the lower strand. Inverted repeat sequences (IR) which may act as an operator site are marked by the solid arrows. An ORF which has a good ribosome binding site six bases upstream, and which may encode hpcR, is initiated at nucleotide 230.
Fig. 5.18.

Nucleotide sequence of the region encoding hpcR and tsr, indicating the CAP binding site, the operator/promoter site and a possible rho-independent terminator.

```
1  AATGATTACTTTATGGGGATTGCGTAGTTTGGCAAAAGTAAGTGATGTAAATGTGATCT
10
60

61  TTACTAATGAAATAACCCTAAGCGATCAAACCGTTTTCTACACTACATTTCATGAG
120

121  TATGCTCAATTTGATTAATTGAAGGTAATTGGGAAAGATATGCCAGCTACT
180

181  AATTAACAACTATTTACATTTAAAGATAATAAAATATTAAATCATCAAGTTAACAAAA
240

241  TCTAACCATTGCCTGCTCAAGCAGGAGGAGATGTATTATTCTTGGGCAATCTG
300

301  TAAACGGAATATTGACTGACCAACAGTGGGATCTGTAGTATCTCGGAAAGGCC
360

361  ATGTAATTGTTGATAAATGTAAATTTCTATTATTTTATAATTTATAGTAGTTCAATTGTTTT
420

421  TTAGGATATTTCCGATATCGTATGCTACCTTCTGTCGCTGCCAATTCAATACCAGCTTATATTGCAT
480

481  GCGGAAGCTCTATATCTCGCTGACCAAAGAGGGGCAGGCGCTGTATAACCGCGCTCAGAC
540

541  GCGCCGCACTATATCTCGCTGAGCTCTCTTCTCGGCCCCAGCTTCTGACATTTTCTTCTGAC
600

601  GTTAAACCATCCTGATTGAGAATTATTCTGCTGGATATCTCCCGACAAAGAGCAGATC
660
```
A consensus sequence for the *tsr* gene encoding the serine chemotaxis protein occurs down stream of the termination codon for the *hpcR* ORF. The 5' region of this gene contains a sequence which is predicted to form a *rho*-independent terminator of transcription.
5.19. **SEQUENCE COMPARISON STUDIES.**

Sequence comparisons were addressed in two main areas. Firstly, database searches were carried out to find any similarities between the Hpc pathway enzymes and sequences already deposited in the databases. Comparative studies were then carried out between the enzymes of the Hpc pathway. Two protein sequence databases were used: the National Biomedical Research Foundation (NBRF) database, release 20.0 March 31st 1989 and the SWISSPROT database, release 10 March 1989. In addition two nucleotide sequence databases were used: the European Molecular Biology Laboratory nucleotide sequence data library release 19, May 1989 and the Genetic Sequence Data Bank (GENBANK) sequence data bank, release 46 November 1986.

5.19.1 **DATABASE SEARCHES.**

a. **HPC DIOXYGENASE.**

The NBRF and SWISSPROT protein sequence databases were searched using the FASTP protein comparison programme from the Lipman-Pearson suite of programmes available at Leicester University. When both protein sequence data bases were searched for proteins with similar amino acid sequences to that of HPC dioxygenase, no obvious matches were found. However, the nucleotide sequence of a number of different enzymes involved in fission of the aromatic ring in ortho- and meta-positions are now available. As such some specific comparisons could be made, in the knowledge that these enzymes catalyse similar ring fission reactions to that of the HPC dioxygenase on structurally related compounds. It has been reported that the bacterial aromatic ring-cleavage enzymes are classified into two different gene families (Harayama & Rekik 1989). These families are divided upon the basis of the type of ring-fission reaction they catalyse i.e. either intradiol (ortho) or extradiol
(meta) fission of the aromatic nucleus. Comparisons of the HPC dioxygenase (a meta-ring fission dioxygenase) were addressed to both major groups.

Catechol is an intermediate in both meta and ortho-ring fission pathways. The catechol 2,3-dioxygenase of the *Pseudomonas putida* mt-2 TOL plasmid, catalyses the meta-ring fission of catechol to give 2-hydroxymuconate semialdehyde. The *xylE* gene encoding this enzyme, has been sequenced (Nakai *et al* 1983) and encodes a protein of 35 KDa. The catechol 2,3-dioxygenase of the NAH7 plasmid of *Pseudomonas putida* (Harayama *et al* 1987) which participates in the degradation of naphthalene, share 84% sequence similarity with the isofunctional enzyme of the TOL pathway. The amino acid sequence of both isofunctional catechol 2,3-dioxygenases were obtained from the NBRF database and compared with that of HPC dioxygenase using the WIMP programme GAP. The results of this analysis are detailed in Fig.5.19.1.

As can be seen the level of similarity between either two catechol 2,3 dioxygenases and HPC dioxygenase is much lower than the similarity between the two catechol enzymes. This is perhaps unsurprising given that they catalyse the same reaction and the fact that the catechol 2,3 dioxygenases are plasmid encoded, whereas HPC dioxygenase is chromosomally encoded. The considerable similarity between a variety of catechol 2,3-dioxygenases has been noted previously (Ghosal *et al*, 1987, Harayama *et al* 1987, Bartilson & Shingler, 1989). In addition similarities in gene order and restriction maps have been found between the meta-pathway genes of a number of catabolic plasmids. This fact, in addition to the observed similarities, has led to the speculation that the meta-pathway genes of several different plasmids may have been inherited as metabolic modules (Cane & Williams 1986).
Alignment of the amino acid sequences of two catechol 2,3-dioxygenases and HPC dioxygenase to maximize similarities

The alignment of dioxygenase nucleotide sequences was carried out using the WIMP program GAP available in Leicester. The amino acid sequence of the *E.coli* C HPC dioxygenase (*hpcB*) is shown on the top line. The amino acid sequences of the TOL plasmid catechol 2,3-dioxygenase (*xylE*) of *P.putida* is shown on the middle line. The amino acid sequence of the corresponding enzyme from the NAH plasmid (*nahH*) is shown on the bottom line. Amino acid residues are numbered with respect to the HPC dioxygenase sequences. The dots below the sequence indicate residues shared by all three sequences. The overall level of similarity is not higher than 14% as assessed by this method.
Fig. 5.19.1.

Alignment of the amino acid sequences of two catechol 2,3-dioxygenases and HPC dioxygenase to maximize similarities.
This might explain then the striking similarity of the catechol 2,3-dioxygenases. The similarity between the HPC dioxygenase and the two catechol 2,3-dioxygenases is at best 14% as determined by the GAP programme, so it is unclear whether there is any evolutionary relationship between these enzymes. The low level of similarity depicted in Fig.5.19.1. shows that any evolutionary relationship is extremely distant.

The other major class of dioxygenase enzyme of which amino acid sequences are known, are the ortho-fission dioxygenase enzymes of the β-keto adipate pathway. Protocatechuate 3,4-dioxygenase is formed by the association of α (Kohlmiller & Howard 1979) and β (Iwaki et al 1979) subunits which bind ferric iron. The genes encoding the protocatechuate 3,4-dioxygenase subunits are chromosomally encoded (Doten et al 1987, Hartnett et al 1990) and catalyse the ortho-ring-fission of protocatechuate as the committal step of the β-keto adipate pathway. The catechol 1,2-dioxygenases are homopolymers which catalyse the ortho-ring cleavage of catechol in the catechol branch of the β-keto adipate pathway. The catA gene encoding the catechol 1,2-dioxygenase of A. calcoaceticus has been sequenced (Neidle et al 1988). Comparison of the amino acid sequences of the protocatechuate 3,4-dioxygenase subunits and the catechol 1,2-dioxygenase showed them to be evolutionary related and indicated that there are conserved residues that may be associated with the ligation of ferric ion and may make significant contributions to secondary or tertiary structure (Neidle et al 1988). More recently the crystal structure of the Pseudomonas protocatechuate 3,4-dioxygenase has been solved (Ohlendorf et al 1988) and indicates that an iron atom is bound by two histidyl and two tyrosyl residues. The genes encoding the A. calcoaceticus α and β subunits (pcaG and pcaH respectively) have been cloned and sequenced (Hartnett et al 1990). As such, interspecies comparisons of the enzyme have been made in addition, to the comparisons with catechol
1,2-dioxygenase. Previous evidence based on the marked difference in transcriptional control of the structural genes of the β-ketoadipate pathway, indicated a great evolutionary distance between the genera *Acinetobacter* and *Pseudomonas*. However, comparison of the amino acid sequences of the protocatechuate 3,4-dioxygenases indicates sequence identity exceeding 50% when the sequences of the corresponding subunits from the different species are aligned. Identical amino acids occupy 18% of the aligned amino acid sequences for the α and β subunits of the *A.calcoaceticus* and *P.putida* enzymes which reinforces the idea that the α and β subunits also share a common ancestry. Similarities with the HPC dioxygenase were tested by comparing the aligned amino acid sequences of the α and β subunits of both *A.calcoaceticus* and *P.putida* enzymes using the WIMP programme GAP. In each case the percentage similarity did not exceed 11%. The alignments made for one subunit did not match the alignments made for the corresponding subunit from the other species. Moreover, the similarities were generally found in discrete areas. As such is was not possible to make a coherent comparison with the protocatechuate 3,4-dioxygenase subunits, to identify conserved sequences between all the polypeptides. Unless a more rigourous comparison algorithm becomes available it must be concluded that there is no similarity or a very low level of sequence similarity between the HPC dioxygenase and any of the known protocatechuate 3,4-dioxygenase subunits.

Recently, the amino acid sequence of a different type of protocatechuate dioxygenase has been reported (Noda *et al.* 1990). It has been reported by Harayama and Rekik (1988) that the various dioxygenase enzymes involved in aromatic degradation fall into two gene families. The protocatechuate 4,5-dioxygenase reported by Noda *et al* (1990) presents an interesting test of this hypothesis, because unlike all the other protocatechuate dioxygenases reported so far, this is a meta-fission (extradiol) enzyme. Database searches carried out
by Noda et al failed to find similarities with either the intradiol or extradiol dioxygenases and concluded that protocatechuate 4,5-dioxygenase has a separate evolutionary origin to any of the other dioxygenase enzymes reported so far. There are significant similarities between the reaction carried out by protocatechuate 4,5-dioxygenase and HPC dioxygenase. Both enzymes catalyse meta-ring fission reactions upon substrates which are very similar, HPC being the higher chemical homologue. However, in common with the other protocatechuate dioxygenases, protocatechuate 4,5-dioxygenase is comprised of two types of subunit whereas HPC dioxygenase is comprised of a single type of subunit. Comparison of the α and β subunits of the protocatechute enzyme to HPC dioxygenase using the WIMP programme GAP, showed similarities of 12% and 17% for the α and β subunits respectively.

b. CHMS DEHYDROGENASE.

The complete amino acid sequence of CHMS dehydrogenase was compared to entries in the SWISSPROT and NBRF databases. Three isoforms of human aldehyde dehydrogenase (EC 1.2.1.3.) were initially identified as having a high level of identity with CHMS dehydrogenase. The human mitochondrial isoform isolated from fetal muscle, had the highest degree of identity so a more accurate comparison with CHMS dehydrogenase was made at the amino acid level using the WIMP programme GAP. Alignment of the two amino acid sequences with the introduction 4 gaps shows the two dehydrogenases to share 40% similarity. With the exception of the first 35 residues of the aldehyde dehydrogenase, the homologous residues occur throughout the polypeptide chain and are not confined to any one area.

Aldehyde dehydrogenases have a broad substrate specificity and participate in ethanol metabolism through the clearance of acetaldehyde. As such there is some interest in the use of inhibitors which could be used
clinically, for alcohol aversion therapy. The amino acid sequence of the cytoplasmic isoenzyme is known, and some characterisation of this enzyme has taken place (Hempel et al 1984). The recently identified human mitochondrial enzyme (Braun et al 1987) and the cytoplasmic isoenzyme show a similarity of approximately 70%. Studies have identified a cysteine residue at position 302 of the cytoplasmic isoenzyme, which is sensitive to the clinically used inhibitor disulfiram. A second cysteine residue that is reactive with iodoacetamide was located to either cysteine-455 or cysteine-463. The cysteine-302 residue appears to be conserved in the human mitochondrial enzyme and CHMS dehydrogenase. The mitochondrial aldehyde dehydrogenase has a cysteine at position 477 which corresponds to cysteine-455 of the cytoplasmic isoenzyme but no cysteine is found in the aligned sequence of CHMS dehydrogenase. Cysteine-463 of the cytoplasmic aldehyde dehydrogenase is replaced with proline in the mitochondrial isoenzyme and alanine in CHMS dehydrogenase. Thus the only conserved residue implicated in catalysis, that is conserved in the three aligned sequences is cysteine-302 of the cytoplasmic aldehyde dehydrogenase. The alignments are shown in Fig.5.19.2.
Alignement of the amino acid sequences of CHMS dehydrogenase and two isoforms of human aldehyde dehydrogenase to maximize similarities

The alignment of the *E.coli* C CHMS dehydrogenase amino acid sequence with those of the mitochondrial and cytoplasmic isoforms of human aldehyde dehydrogenase was carried out in the same way as described in Fig.5.19.1. The top line shows the sequence of the mitochondrial aldehyde dehydrogenase (*adm*), the second line indicating the cytoplasmic isoform of the enzyme (*adc*). The sequence of CHMS dehydrogenase (*hpcC*)is shown on the bottom line. The numbering is with respect to the mitochondrial isoform of aldehyde dehydrogenase. Conserved residues are shown in boxes with the cysteine implicated in catalysis marked by a dot.
Fig. 5.19.2.
Alignment of the amino acid sequences of CHMS dehydrogenase and human aldehyde dehydrogenase to maximise similarities.
c. CHM ISOMERASE, COHED DECARBOXYLASE/HHDD ISOMERASE, OHED HYDRATASE AND HPCR REPRESSOR PROTEIN.

When the databases were searched against the amino acid sequences of CHM isomerase, COHED decarboxylase/HHDD isomerase, Hpc repressor and the partial sequence obtained for OHED hydratase (see section 5.12). However, no obvious matches were found.

5.19.2. SEQUENCE COMPARISONS BETWEEN THE Hpc ENZYMES.

Similarities between the Hpc enzymes were addressed using the WIMP programme GAP available at Leicester. A total of fifteen pairwise comparisons were made and the results are shown in table 5.14.2. There are no obvious large regions of similarity between any of the Hpc enzymes at the amino acid sequence level.
Table 5.19.2.
Summary of the percentage similarities between the Hpc enzymes as assessed by the programme GAP.

<table>
<thead>
<tr>
<th></th>
<th>hpcR</th>
<th>hpcEF</th>
<th>hpcC</th>
<th>hpcB</th>
<th>hpcD</th>
<th>hpcG'</th>
</tr>
</thead>
<tbody>
<tr>
<td>hpcR</td>
<td>100</td>
<td>10.8</td>
<td>13.5</td>
<td>11.5</td>
<td>15.4</td>
<td>10.8</td>
</tr>
<tr>
<td>hpcEF</td>
<td>100</td>
<td>11.4</td>
<td>9.8</td>
<td>12.7</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>hpcC</td>
<td>100</td>
<td>13.1</td>
<td>16.9</td>
<td>12.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hpcB</td>
<td>100</td>
<td>11.9</td>
<td>10.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hpcD</td>
<td></td>
<td>100</td>
<td>11.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hpcG'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

where:
- hpcR = Hpc repressor protein
- hpcEF = COHED decarboxylase/HHDD isomerase
- hpcC = CHMS dehydrogenase
- hpcB = HPC dioxygenase
- hpcD = CHM isomerase
- hpcG' = Partial sequence available for OHED hydratase
CHAPTER 6
CRYSTALLISATION OF CHM ISOMERASE
AND HHDD ISOMERASE/COHED
DECARBOXYLASE.
6.1. **INTRODUCTION.**

During the course of this study clones have been created which express several of the Hpc enzymes at levels in excess of 1% of total cell protein. This has been advantageous in several ways. High level expression of the individual enzymes has facilitated their purification. The purified enzymes have been N-terminal amino acid sequenced such that the open reading frames encoding them can be definitely assigned. In addition the availability of large quantities of the individual enzymes has made investigation of some of their properties much easier. Experiments of this type have described in chapter 4.

The ability to grow protein crystals with a view toward determining X-ray structures, generally depends on the availability of such large quantities of pure enzyme. Since milligram quantities of both CHM isomerase and HHDD isomerase/COHED decarboxylase were available, some effort has been spent in trying to grow crystals.

6.2. **CRYSTALLISATION OF CHM ISOMERASE.**

There are a large number of different parameters which govern the growth of protein crystals. Consequently an initial screen was carried out to find conditions under which crystals would form. The crystallisation was achieved by vapour diffusion in hanging drops according to the method of McPherson (1982). Tissue culture trays containing 24 wells were used so that a range of conditions could be tried. Hanging drops used either polyethylene glycol (PEG) or ammonium sulphate as the precipitant at a range of concentrations. PEGs of molecular weights 600, 1000, 4000 and 6000 were used in doubling concentrations between 5% and 32%. Trials using ammonium sulphate as the precipitant, were set up using doubling concentrations from 1% to 30%. These initial experiments used CHM isomerase at a concentration
of 5mg.ml\(^{-1}\) in 20mM Tris.Cl pH 7.5. Drops of 20\(\mu\)l comprising 10\(\mu\)l of enzyme solution 10\(\mu\)l of solution from the well, were hung over 2ml wells. The wells containing PEG were buffered with 20mM Tris.Cl at pH 6.5, pH 7.0 or pH 7.5. The wells containing ammonium sulphate were buffered with 50mM Tris.Cl at pH 7.5. Drops of 25\(\mu\)l were placed on plastic coverslips which were inverted and sealed above the well with high vacuum grease. The trays were then left for approximately 4 weeks at room temperature and were periodically checked for the appearance of crystals in the hanging drops. A precipitate formed in the drops containing PEG buffered to pH 6.5 at concentrations greater than 10% and also in the drops containing PEG at pH 7.0 at concentrations greater than 20%. Precipitates were also found in the drops containing ammonium sulphate between 31% and 34%. After approximately four weeks the ambient temperature increased and two crystals of a bipyrimidal shape with a rectangular cross section were seen in the well containing 25% ammonium sulphate pH 7.5. By the following day, when the ambient temperature had decreased, these crystals had started to redissolve. It was thought from this observation that one of the main factors affecting the growth of CHM isomerase crystals was temperature. The trays were then placed in a 37°C incubator overnight and examined the following day. The bipyramidal crystals had disappeared completely by this stage and in their place a large number of long rectangular rod shaped crystals had formed with a cross section of up to 100\(\mu\)m X 200\(\mu\)m and up to 4mm in length. These rod shaped crystals (Fig.6.2.1.) were designated type I crystals and appeared to be stable at 20°C when harvested into 35% ammonium sulphate, 20mM Tris.Cl pH 7.5.
Fig. 6.2.1.

Type I CHM isomerase crystals.

Type I needle shaped CHM isomerase crystals are the most common crystalline form of to be isolated which grow up to 4mm in length. with a maximum cross section of 100μm x 200μm. The bar represents 500μm.
Fig. 6.2.1.
Type I CHM isomerase crystals.
This form was analysed using synchrotron radiation of 1.488 Å wavelength at the SERC synchrotron source at Daresbury, U.K. The crystals were stable in the X-ray beam and diffracted to beyond 3.0 Å resolution (Fig. 6.2.2.). The space group was determined to be orthorhombic with unit cell dimensions a=88 Å, b=89 Å, c=121 Å (Wigley et al. 1989). Systematic absences indicated a body centered lattice and showed that the space group was either $I222$ or $I2_{1}2_{1}2_{1}$.

Since that time a number of other crystal forms have been grown from ammonium sulphate. Large numbers of tetragonal type II crystals have been grown from 27% ammonium sulphate at 30°C and a typical example is shown in Fig. 6.2.3. These crystals diffract to beyond 2.0 Å are stable in the X-ray beam and have a space group of $P4_{3}2_{1}2$. Precession photographs along the four-fold and two-fold axes are shown in Fig. 6.2.4. Type II crystals have been used to collect a native data set to 2.1 Å (D.B. Wigley pers. comm.) and data sets for platinum and gold derivatives have been collected at the EMBL synchrotron source in Hamburg, FRG. This data has been used to calculate a 3.7 Å electron density map of the enzyme. Higher resolution data has now been collected from both derivatives to aid interpretation of the electron density maps.
Fig.6.2.2.

Diffraction pattern obtained from type I crystal form showing diffraction to 2.8Å at the edge of the photograph.
Fig. 6.2.2.

Diffraction pattern for type I crystals.
Fig.6.2.3.

Photograph of type II CHM isomerase crystal form. Crystals of up to 700\(\mu\)m along one edge have been obtained to date.
Fig. 6.2.3.

Type II crystals.
Fig.6.2.4.

12° precession photographs of a type II CHM isomerase crystal soaked in 1mM K$_2$PtCl$_4$. 
Fig. 6.2.4.

Precision photographs along the four-fold and two-fold axes.

a. Four-fold axis
b. Two-fold axis
It seems that the formation of rod shaped, type I crystals is favoured over that of the other crystal forms. The appearance of rod shaped crystals in a drop seems to precede the disappearance of any other crystal form already in the drop. The majority of the type II crystals have been grown from preparations of the enzyme which have not been purified by gel filtration chromatography after the hydrophobic interaction step (see section 4.1.2.) The presence of a small amount of contaminating protein in these preparations seems to help prevent the formation of the type I crystals.

Several other crystal forms have been produced as shown in Fig.6.2.5. These crystal types have not all been characterised however. The different crystal forms of CHM isomerase are summarised in Table 6.2.
Fig. 6.2.5.
Photographs of type III, IV and V crystal forms of CHM isomerase.

a. Type III crystal form which has a cubic space group (see table 6.2). Crystals of up to 1.2 mm from point to point have been obtained at pH 7.5. The bar indicates 100 μm.

b. Type IV crystal form of unknown space group. Several large crystals of this type have been obtained. The crystal pictured is over 1 mm across. This crystal was grown from ammonium sulphate at pH 6.5.
Fig. 6.2.5.
Type III, IV & V crystals.
Fig.6.2.5. continued

c. Type V crystals of unknown space group. These type of crystals flat plates. Crystals of this type over 1mm in length have been produced. Bar represents 50μm.
Fig. 6.2.5. continued
Table 6.2.  
Summary of the different CHM isomerase crystal forms.

<table>
<thead>
<tr>
<th>Crystal Type</th>
<th>Description</th>
<th>Space Group</th>
<th>Diffraction Limit</th>
<th>Derivatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Rods</td>
<td>$I222/2_12_1$</td>
<td>$&gt;3.0\text{Å}$</td>
<td>-</td>
</tr>
<tr>
<td>II</td>
<td>Rectangular</td>
<td>$P4_32_12$</td>
<td>$&gt;2.0\text{Å}$</td>
<td>Pt, Au</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>bipyrimids</td>
</tr>
<tr>
<td>III</td>
<td>Cubic space</td>
<td>$P4_332/1$</td>
<td>28.0Å</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>group</td>
<td>$P4_132$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Cubic?</td>
<td>N.C.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>V</td>
<td>Flat plates</td>
<td>N.C.</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

N.C. = Not characterised  
Pt = $K_2PtCl_4$  
Au = $NaAuCl_4$.  

243
COHED decarboxylase/HHDD isomerase was purified as detailed in section 4.3 and then concentrated by pressure dialysis to approximately 10mg.ml\(^{-1}\). Hanging drops of total volume 20\(\mu\)l were set up using a range of different precipitants. Trials with PEG 600, 1000, 4000, 8000 and 20,000 were set up between 5\% and 30\% with a 5\% change in concentration between each well. A series of hanging drop trials containing ammonium sulphate as the precipitant were also tried, with 5\% changes in a concentration range between 5\% to 60\%. All the wells in these trials were buffered to pH 7.5 with 20mM Tris.Cl.

The trays were then left at ambient temperature at periodically checked for several weeks. Precipitates formed in nearly all the ammonium sulphate drops after two weeks and precipitates had also formed in the higher molecular weight PEG drops. After nearly one month, two large clumps of crystals had formed in the well containing 30\% PEG 1000 (Fig.6.3.1.). These crystals were quite obviously twinned, but had grown to a size suitable for X-ray analysis. The large crystal shown in Fig.6.3.1. was mounted and analysed at the Daresbury X-ray synchrotron source and was found to diffract to beyond 2.5\(\AA\). (data not shown) Because the crystal was twinned, no regular array of spots could be identified in the diffraction pattern so the space group could not be assigned. Subsequent attempts to produce more crystals of this enzyme have been unsuccessful to date, and it therefore seems that more work is needed to identify the conditions under which these crystals will grow.
Fig. 6.3.1.

Crystal of COHED decarboxylase/HHDD isomerase. The twined nature of this crystal form can be easily seen. The bar represents 450μm.
Fig. 6.3.1.
CHAPTER 7

DISCUSSION AND FUTURE DIRECTIONS
7.1. **SYNOPSIS**

The work detailed in this thesis represents the current state of knowledge regarding the organisation, order and sequence of the hpc genes of *E.coli* C. It is hoped that this may form the basis for more advanced studies to probe how this aromatic degradative operon is controlled and perhaps investigate the origins of these particular genes.

7.2. **SUMMARY OF THE RESULTS**

Molecular analysis of the hpc genes, through nucleotide sequencing and in some cases purification and study of their protein products, has given a new perspective of the Hpc pathway. As a consequence of the deletion subcloning work, which was carried out to provide the templates for nucleotide sequencing, we now know that the gene order and direction of transcription of the hpc operon is very different to that originally proposed by Jenkins (1987). The genes encoding the enzymes of the Hpc pathway are transcribed in the order hpc(EF)CBDGH. The presence of a promoter in front of hpcEF which could direct the transcription of these genes has been demonstrated. The possibility of a second promoter in the vicinity of hpcB which would corroborate the Northern blotting data of Fawcett (1989), is also still a possibility. No experiments to directly prove or disprove this hypothesis have been conducted during this study although inspection of the nucleotide sequence between the 3' terminus of hpcC and the 5' end of hpcB does show the presence of a -24(GG) -12(GC) sequence which may function as a promoter. These type of promoters are known to occur in other aromatic degradative pathways including the TOL pathway. Several features have been identified from the nucleotide sequence including the presence of a CAP binding site which may participate in the regulation of the Hpc pathway genes. The presence of an operator site for the Hpc repressor has not been positively confirmed although two inverted repeat
sequences are found near the promoter for the Hpc pathway genes. An ORF which is transcribed in the opposite direction to the rest of the hpc genes would seem likely to encode the Hpc repressor protein. Direct confirmation of the hpcR ORF through matching the amino terminal sequence of the purified protein to that predicted, is lacking. The nucleotide sequence of hpcH encoding HHED aldolase and the 3' portion of hpcG encoding OHED hydratase remains to be completed. If this were completed, the hpc operon would be one of the longest aromatic degradative pathways whose entire nucleotide sequence were known. Comparable information for the TOL pathway or β-ketoadipate pathway is not available even though molecular analysis of their genes has been under way for a longer period of time than for the Hpc pathway. Such nucleotide sequence information would then form the basis for investigating the inter-relatedness of the various aromatic degradative pathways and hence give clues to their origins.

The original hypothesis: that the two isomerase enzymes may have evolved from the same ancestral precursor, given their similar substrate structures and close physical location of their genes, seems to be untrue. Kinetic evidence determined for the purified enzymes indicates a very high degree of substrate specificity. However, from a very early stage in the project, it became apparent that their respective subunit molecular weights were very different. From this evidence alone it might be reasonable to suggest that, if the genes are related they must have undergone large changes with time to give rise to such differently sized protein products. This appears to be borne out by comparison of their amino acid sequences which overall show very low levels of similarity. In addition, there seem to be no areas in the amino acid sequence comparison, which indicate that certain regions have been conserved over the rest of the primary structure indicating the location of the substrate
binding site for example. The situation is further complicated by the finding that the COHED decarboxylase and the HHDD isomerase activities are due to the same protein. Comparisons of the Hpc enzyme amino acid sequences to each other have not highlighted any degree of similarity which would disfavour the proposal of Horowitz (1945) regarding evolution of pathways and favour a process whereby the required activities are recruited from other parts the genome on the basis of their existing catalytic properties. With the exception of CHMS dehydrogenase, comparisons of the Hpc enzymes to the databases have not identified any homologues. This does not necessarily mean that the Hpc enzymes are unique, merely that their homologues may have not yet been identified and deposited in the databases so that searches of this type can be made. The identification of two isoforms of human aldehyde dehydrogenase which have approximately 40% similarity to the CHMS dehydrogenase was unexpected. The similarity extends all through the primary sequence of the dehydrogenase and is not confined to any particular area indicating a nucleotide binding domain or a similar substrate binding site for example. The basis for the overall similarity can be speculated upon, and indeed other dehydrogenases with similarity to CHMS dehydrogenase may be found in the future.

Nucleotide sequencing of the cloned DNA at both ends of the genomic clone pJJ200, revealed regions which are highly homologous to regions of the E.coli K12 genome. The nucleotide sequence 3' to hpcR contains a region which is very similar to the gene tsr. This gene maps to minute 99 within the K12 genetic map. The nucleotide sequence obtained from the other end of pJJ200 bears a high degree of similarity to a region encoding the P15B and P35B genes of E.coli K12, which are located at approximately minute 69 in the genetic map. These two loci are separated by approximately 1400 Kbp on the
E. coli K12 chromosome and yet are separated only by approximately 9 Kbp on the E. coli C genomic clone pJJ200. There are several possible explanations for this result. It is possible that the P15B and P35B are located in a similar region to the tsr gene in E. coli C. This is difficult to check since the E. coli C genetic map is poorly documented compared to that of K12 but there are some similarities (Wiman et al 1970). However, a more plausible explanation might be that a fragment of DNA from the minute 69 locus was inadvertently cloned into pBR322 along with the hpc gene cluster during the original ligation reaction that created pJJ200 which is therefore not composed of DNA that is contiguous in the organism. This is supported by the fact that the tsr gene is located less than 50 bps from the 3' terminus of hpcR and there are no Sau3A1 sites in the intervening region which would have resulted if the tsr gene had been introduced as a separate fragment in the original ligation reaction. In addition, mapping experiments carried out by Skinner (1981) located the hpc genes to a locus around minute 0 in the E. coli C genome.

7.3 FUTURE WORK.

The work detailed in this thesis could form the basis for a large number of more advanced studies on the control and regulation of the hpc genes. Given the nucleotide sequence information available one of the first tasks would be to determine whether a second promoter structure occurs within the hpc gene cluster thus establishing how many transcriptional units are involved in regulation. Probing the regulation of the pathway would then entail the positive identification the hpcR ORF. The hpcR gene could then be cloned into a high expression vector by a PCR methodology enabling large quantities of the Hpc repressor to be purified. Hopefully such a strategy would then allow a large number of the regulatory aspects of the pathway to be
addressed such as footprinting of the operator site(s) and investigating the binding of HPC to the repressor protein.

Comparison of the Hpc enzymes at a structural level is under way. This project resembles that being undertaken for the β-ketoadipate pathway enzymes. Crystallographic analysis of CHM isomerase, and COHED decarboxylase/HHDD isomerase is under way at the University of York under the direction of Dr D.B. Wigley. A 3.7Å electron density map of CHM isomerase has already been calculated which indicates that a large proportion of the enzymes structure consists of β sheets. Better crystals of COHED decarboxylase/HHDD isomerase are being sought in order to collect high resolution data. A project concerning crystallographic analysis of CHMS dehydrogenase is also planned. X-ray crystallography structures of the these proteins may yield further information on the relatedness of the Hpc enzymes.

As more sequence information of the TOL and β-ketoadipate pathways becomes available a more global view of the evolution of aromatic pathways may become possible.1 pollution.
REFERENCES
hydroxyphenylacetic acid in \textit{Pseudomonas ovalis}.

Amons, R. (1987) Vapor-phase modification of sulfhydryl groups in proteins
\textit{FEBs Lett} 212: 68-71

\textit{TIBs} 14: 473-475

operon on NAH plasmid pWW60-22 and TOL plasmid pWW53-4 and its
evolutionary significance.
\textit{J. Gen. Microbiol.} 134: 2769-2778

and molecular biology. Vol II. ed. F.C. Neidhardt. American society of
microbiology, Washington D.C.

hydroxyphenylacetic acid by \textit{Pseudomonas putida}. Biochem. Biophys.
Res. Commun. 76: 565-571.


Belfaiza, J., Parsot, C., Martel, A., de la Tour, C. B., Margarita, D., Cohen, G.
enzymes catalysing consecutive steps in methionine biosynthesis
originate from a common ancestor.
\textit{P.N.A.S. USA} 83: 867-871.

*Nature* 301: 623-626

*Nucleic Acid Res* 15: 3179

*Proc Natl Acad Sci USA.* 81: 6929


*J. Bacteriol* 155: 113-121

*P.N.A.S. USA* 70: 1841-1845.

Cane, P.A. & Williams, P.A. (1986) A restriction map of Naphthalene catabolic plasmid pWW60-1 and the location of some of its catabolic genes.
*J. Gen. Microbiol* 132: 2919-2929


*J. bacteriol* 114: 974-979


Fawcett, T., Garrido-Pertierra, A. & 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.  


*P.N.A.S. USA* 78: 7458-7462.

In: *The bacteria* 10: 295-323

Ghosal, D. You, I-S., & Gunsalus (1987) Nucleotide sequence and expression of
gene nahH of plasmid HAN7 and homology with gene xylE of TOL
pWW0.
Gene 55: 19-28

degradation of halogenated compounds.
Science 288: 135-228

society of microbiology, Washington D.C.

lactonizing enzyme at 3.0A resolution.
J. Mol. Biol 194: 143-153

proteins by means of the biuret reaction.
J. Biol. Chem. 177: 751-756

Gralla, J.D. (1989a) Bacterial gene regulation from distant DNA sites.

Gralla, J.D. (1989b) Specific repression of the lac operon - the 1988 version. In
DNA-Protein interactions in transcription:3-10. Alan R. Liss (Pub) New
York USA.

promoter DNA sequences.
Nucleic Acid Res 11: 2237-2249


*J. Bacteriol* 172: 956-966


Henikoff, S. 1984 Unidirectional digestion with Exonuclease III creates break points for DNA sequencing


Henikoff, S. 1987 Unidirectional digestion with Exonuclease III in DNA sequence analysis.

*Meth in Enzymol* 155: 156-165


*P.N.A.S. USA* 82: 7525-7529
*P.N.A.S. USA* 78: 7069-7072

*P.N.A.S. USA* 31: 153-157

Holloway, B.W. & Morgan, A.F. (1986) Genome organisation in *Pseudomonas*

*J. Gen. Microbiol* 134: 2877-2887


*J. Bacteriol* 150: 188-194

Iwaki, M. Kagamiyama, H. & Nozaki, M. (1979) The complete amino acid sequence of the β subunit of protocatechuate 3,4-dioxygenase  
*J. Biochem* 86: 1159-1162

*J. bacteriol.* 171: 3181-3186

*J. Mol. Biol.* 205: 557-571

*J. Bacteriol* 164: 887-895

Kita, H., & Senoh, S. 1979. 3,4-dihydroxyphenylacetate-2,3-oxygenase (*Pseudomonas ovalis*)  

Kita, H. (1965) Crystallization and some properties of 3,4-dihydroxyphenylacetate 2,3-oxygenase from Pseudomonas ovalis.
   J. Biochem 58: 116-122.

   Cell 50: 495-508

   J.Bacteriol 171: 4326-4333

   P.N.A.S. USA 80: 687-691.

Kohlmiller, N.A. & Howard, J.B. (1979) The structure of the α subunit of protocatechuate 3,4-dioxygenase
   J. Biol. Chem. 254: 7309-7315

   Biotechniques 6: 544-547

Kunz, D.A. & Chapman, P.J. (1981) Isolation and characterisation of spontaneously occurring TOL plasmid mutants of Pseudomonas putida HS1
   J. Bacteriol 146: 952-964


*P.N.A.S. USA* 86: 476-480


*Ann Rev Genet.* 16: 135-168


*J. Bacteriol.* **168:** 815-820.

*J. Bacteriol.* **169:** 5496-5503.

*J. Bacteriol* **170:** 2412-2413

*J. Bacteriol.* **172:** 2704-2709.


*Nature* **306:** 203-206


*American society for Microbiology Microbiology* 1981: 140-143


*Escherichia coli* K12

*J. Gen. Microbiol.* 133: 347-351


*EMBO J.* 5: 3013-3019


*J.Mol. Biol.* 204: 435-445


*Proc. Natl. Acad. Sci. USA* 72. 784-788


*Mol. Microbiol.* 1: 293-300


Translational coupling at the intracistronic boundary of the *Escherichia coli* galactose operon. *Cell* 30: 865-871.


Shine, J., & Dalgarno, L. (1975). The 3'-terminator sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to non-sense triplets and ribosome binding sites. *P.N.A.S.* 71: 1342-1346


*Methods and programs in Biomedicine* 21: 113-118.


*Genetics* 66: 245-266


*J. Bacteriol.* 120: 416-423

Wiman, M., Bertani, G. Kelly, B. & Sasaki, I. (1970) Genetic map of *Escherichia coli* strain C

*Mol. Gen, Genet* 107: 1-31


*J. Bacteriol* 124: 7-13


*FEMS Microbiol Lett* 51: 181-186


Naphthalene/salicylate oxidation.
*P.N.A.S. USA* 79: 874-878

mediated dissimilation of naphthalene and salicylate in *Pseudomonas
putida* PMD-1
*J.Bacteriol* 147: 836-843

regulation of the *Pseudomonas cepacia* protocatechuic acid oxygenase
genes.
*J. Bacteriol* 171: 5907-5914
Appendix 1
Oligonucleotides used in this study.
<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>length</th>
<th>Sequence 5'...3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal Primer</td>
<td>24mer</td>
<td>CGCCAGGTTTTCCCATCGAC</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>20mer</td>
<td>CACTATCGACTACGCAGTCA</td>
</tr>
<tr>
<td>P1219</td>
<td>21mer</td>
<td>CCACGATGCCTCGCGGTAG</td>
</tr>
<tr>
<td>DRT01</td>
<td>21mer</td>
<td>CGCGAAGAATAGTGGCTTTTCAT</td>
</tr>
<tr>
<td>DRT02</td>
<td>21mer</td>
<td>TGCTAATGCTAACTTACCAT</td>
</tr>
<tr>
<td>DR03</td>
<td>21mer</td>
<td>GAAGTGCGTTTATAATCGTTC</td>
</tr>
<tr>
<td>DR04</td>
<td>21mer</td>
<td>CTACGCAATCGGGAATAAAAAG</td>
</tr>
<tr>
<td>DR05</td>
<td>21mer</td>
<td>GTGCGTGTTTCTCGCAGAAGGT</td>
</tr>
<tr>
<td>DR06</td>
<td>21mer</td>
<td>GAGCTGGCGCTGGAAGGCGGT</td>
</tr>
<tr>
<td>DR07</td>
<td>21mer</td>
<td>ACGACACGGGACCGCACGTCAT</td>
</tr>
<tr>
<td>DR08</td>
<td>21mer</td>
<td>TATGACCCTGAATCCGGAAG</td>
</tr>
<tr>
<td>DR09</td>
<td>21mer</td>
<td>GCCGAGCAGAATGGCATCGCC</td>
</tr>
<tr>
<td>DR10</td>
<td>21mer</td>
<td>TCGTTAGCCAGCGCATATCCG</td>
</tr>
<tr>
<td>DR11</td>
<td>21mer</td>
<td>CAGATGTACGACGCAGGCCA</td>
</tr>
<tr>
<td>DR12</td>
<td>21mer</td>
<td>GATAAAGTGCGGCATAAAAACG</td>
</tr>
<tr>
<td>DR13</td>
<td>21mer</td>
<td>TCGCGGGTGTAGCCTTACAT</td>
</tr>
<tr>
<td>DR14</td>
<td>21mer</td>
<td>CATTTCCCCGGCAACGCTTGC</td>
</tr>
<tr>
<td>DR15</td>
<td>20mer</td>
<td>TGCTGCTGGCGAATCCTTG</td>
</tr>
<tr>
<td>DR16</td>
<td>20mer</td>
<td>CATTATGCGCCGGGATGCT</td>
</tr>
<tr>
<td>DR17</td>
<td>20mer</td>
<td>TGGGCTTTCCGCAATACGC</td>
</tr>
<tr>
<td>DR18</td>
<td>20mer</td>
<td>GCGCAGTCACGGTAACGTGA</td>
</tr>
<tr>
<td>DR19</td>
<td>20mer</td>
<td>ATTTTCACGATCTTTCCGAT</td>
</tr>
<tr>
<td>DR20</td>
<td>20mer</td>
<td>GGAATTTATGCTCTGGTGA</td>
</tr>
<tr>
<td>DR21</td>
<td>20mer</td>
<td>AGGCGGTCTGTCTTTAATG</td>
</tr>
</tbody>
</table>
Appendix 2.
Nucleotide sequence of the hpc gene cluster obtained in this study.
Restriction enzyme map with open reading frames of the E. coli C gene cluster for homoprotocatechuate degradation.
<table>
<thead>
<tr>
<th>Start Position</th>
<th>End Position</th>
<th>Length</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2701</td>
<td>2760</td>
<td>60</td>
<td>DNA Sequences</td>
</tr>
<tr>
<td>2761</td>
<td>2820</td>
<td>60</td>
<td>Protein Sequences</td>
</tr>
<tr>
<td>2821</td>
<td>2880</td>
<td>60</td>
<td>DNA Sequences</td>
</tr>
<tr>
<td>2881</td>
<td>2940</td>
<td>60</td>
<td>Protein Sequences</td>
</tr>
<tr>
<td>2941</td>
<td>3000</td>
<td>60</td>
<td>DNA Sequences</td>
</tr>
</tbody>
</table>

**DNA Sequences**

- **2701-2760**
  - DNA sequence: GCCCGATGGAAACCGCGGACACGGGCCTGCCGATCCATCAGACCAAAAATGTGTTGATC
  - Length: 60 nucleotides

- **2761-2820**
  - DNA sequence: CGGCGCTACCTTTGGCGCCTGTGCCCGGACGGCTAGGTAGTCTGGTTTTTACACAACTAG
  - Length: 60 nucleotides

- **2821-2880**
  - DNA sequence: ACTTATCGCCTGACCGACAAGATGTGGTACGTGACGCGGTAAGCCCGTGGTTTGT
  - Length: 60 nucleotides

- **2881-2940**
  - DNA sequence: GCACGTGGGCTTCAGCTCGGGTGAAGTGGTACGGTGACGCGGTTG
  - Length: 60 nucleotides

**Protein Sequences**

- **2701-2760**
  - Protein sequence: AMETADTGLPIHQTKNVLI
  - Length: 12 amino acids

- **2761-2820**
  - Protein sequence: TRFSQ
  - Length: 4 amino acids

- **2821-2880**
  - Protein sequence: VSTCGTPQA
  - Length: 6 amino acids

- **2881-2940**
  - Protein sequence: RPVHRKGGGLPDLRRT
  - Length: 15 amino acids

**Other Details**

- **2701-2760**
  - Gene name: FBN
  - Start position: 1
  - End position: 16
  - Number of amino acids: 12

- **2761-2820**
  - Gene name: SC
  - Start position: 1
  - End position: 16
  - Number of amino acids: 4

- **2821-2880**
  - Gene name: S
  - Start position: 1
  - End position: 16
  - Number of amino acids: 5

- **2881-2940**
  - Gene name: C
  - Start position: 1
  - End position: 16
  - Number of amino acids: 6
<table>
<thead>
<tr>
<th>S</th>
<th>SE</th>
<th>EST</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH</td>
<td>f</td>
<td>HNcfa</td>
<td>s</td>
</tr>
<tr>
<td>h</td>
<td>a</td>
<td>pcroep</td>
<td>p</td>
</tr>
<tr>
<td>a</td>
<td>N</td>
<td>aIFNE</td>
<td>E</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>211111</td>
<td>1</td>
</tr>
</tbody>
</table>

```
TCGCGCATCTTTATTCAAAACAGATCTACCGGGAATTCGTGAAATTTGCCGAAACGGCC
AGCGCTGAGAAATAATGGTGCATTGCAGGACCTTAAACAGCTCTTGCCGCTGGCG
3301

RASLFNKRASANS*
3360

RM * CLM
3420

B C
B s H v
s m P i
r A a J
1 1 2 1

AACCGTGTGCGCGTGGGCGATCCGACCGATCCGAATACCCAGGTTGGGGCGCTTATCAGC
3480

CAGCAACACTGGGAAAAAGTCTCCGGCTATATCCGTCTGGGCATTGAAGAAGGCGGCAAC
3540

CTCCTGGCGGGCGGCCGGAATACCGTATGCGCGTTGCCCAGGAAGAGAAGGACGCGGGTTGCCACGACCGCCTGCATCTATTGGCATACGCGCAACGGGTCCTTCTC
3600
```
Enzymes that do cut:

Aat2  Acc1  Acy1  Afl1  Afl2  Alu1  AlwNI  ApaLI  Asn1  Asul  Asu2
Ava1  Ava2  BamHI  Ban1  BbrI  Bbv1  Bbv2  Bcl1  BglI  BlnI  BssAI
BsaAI  BsaI  BsmFI  BamAI  BamI  Bsp12  BspHI  BspM2  BsrI  BssH2
BstNI  BstXI  Cfr1  Cfr10  Clal  CviJ1  Dde1  Dra1  Dra3  DsaI  EcoB
EcoR  EcoRl  EcoRV  EcoSI  Eco47  Eco57  Fnu4H  FokI  Gdi2  GsuI  HaeI
Hae2  Hae3  Hgl1  HgiE2  HhaI  Hinc2  Hinf1  HpaI  Hpa2  HphI
KpnI  Ksp63  Mae2  Mae3  Mbo1  Mbo2  Mlu1  Mse1  MnlI  MseI  MaeI
Mci1  Mco1  Mda1  Mla3  Mla4  Mru1  Msi1  Nsp1  NapB2  FflMI  PflI
PvuI  Pvu2  Rsal  SalI  Sau3AI  Sau96  ScrF1  ScaI  SfaNI  SmaI
SspI  SstI  Sty1  TaqI  ThaI  TspE1  Tsp45  Tth32  VspI  XcmI  XhoI
XmnI

Enzymes that do not cut:

ApaI  Avr2  BclI  Ban2  BglII  BstEII  Dra2  DrdI  Earl  EcoNI  EspI
FinI  FsiI  HindIII  Mael  MstI  NheI  NheII  NotI  PmaCI  PvuII  PstI
Rsr2  Scal  SfiI  SnaB1  SspI  Spel  SphI  SpI  StuI  Stul  Tth31
XbaI  XhoI  Xma3
Appendix 3.
Publications.
Subcloning and nucleotide sequence of the 3,4-dihydroxyphenylacetate (homoprotocatechuate) 2,3-dioxygenase gene from *Escherichia coli* C.

David I. Roper and Ronald A. Cooper

Department of Biochemistry,
University of Leicester,
Leicester LE1 7RH,
UK

Received:

Correspondence: Dr R. A. Cooper
Department of Biochemistry
University of Leicester
Leicester LE1 7RH
UK

Telephone: 0533 - 523455
Fax: 0533 - 523369

Summary
A cloned gene encoding the *Escherichia coli* C homoprotocatechuate (HPC) dioxygenase, an aromatic ring cleavage enzyme, was used to produce large amounts of the protein. Preparations of *E. coli* C HPC dioxygenase, whether expressed from the cloned gene or produced by the bacterium, lost activity very rapidly. The pure protein showed one type of subunit of Mr 33000. The first 21 N-terminal amino acids were sequenced and the data used to confirm that the open reading frame of 831bp, identified from the nucleotide sequence, encoded HPC dioxygenase. Comparison of the derived amino acid sequence with those of other extradiol and intradiol dioxygenases showed no obvious similarity to any of them.
Keywords: Homoprotocatechuate 2,3 dioxygenase;

Protein purification; N-terminal sequence;

Gene sequence; Derived primary structure;

*Escherichia coli* C
INTRODUCTION

The aromatic nucleus is one of the most abundant units of chemical structure in the biosphere. Because it is strongly stabilised by resonance, the ring must be hydroxylated under aerobic conditions to make it sufficiently reactive to undergo degradation. Normally, two hydroxyl groups need to be present for ring opening. The hydroxyls are usually required on adjacent carbon atoms to give a cis-hydroxylated aromatic ring. The activated ring is then opened by dioxygenase action. In ortho-fission (intradiol cleavage) the carbon-carbon bond between the cis-hydroxyls is cleaved. In meta-fission (extradiol cleavage) a carbon-carbon bond immediately adjacent to the cis-hydroxyls is the one broken [1].

Sequence information is now available for a number of these intradiol and extradiol dioxygenases. The sequences of three intradiol enzymes have been compared and it was concluded that two distinct catechol 1,2-dioxygenases have a common ancestry which is shared, but more distantly, with a protocatechuate 3,4-dioxygenase [2]. Consideration of the sequences of five extradiol dioxygenases led to the conclusion that they were members of the same gene super family, which was distinct from the family of intradiol dioxygenases [3]. However, a recent report of the sequence of a sixth extradiol dioxygenase, protocatechuate 4,5-dioxygenase, concluded that it was unrelated to any of the other extradiol or intradiol dioxygenases [4].

We have determined the sequence of a seventh extradiol dioxygenase whose substrate, homoprotocatechuate (HPC; 3,4-dihydroxyphenylacetate), is the next higher chemical homolog of protocatechuate. As reported here the sequence of *Escherichia coli* HPC 2,3-dioxygenase has no marked similarity to any other extradiol or intradiol dioxygenase yet described.

2. MATERIALS AND METHODS

2.1. *Bacteria, plasmids and growth conditions*

Cells were grown aerobically at 37°C in Luria broth [5] containing 100 µg·ml⁻¹ ampicillin. The *E. coli* K12 strain 5K [6] was used as host for the plasmids used in this study. The construction of pDR1830, pDR1930 and pDR934 is
described in the text. pDR1930 was used to construct a nested set of deletions (pDR9304 - 9312, pDR9315; Fig. 2) using a Pharmacia exonuclease III deletion kit.

2.2. **Enzyme assays**

HPC dioxygenase and 5-carboxymethyl-2-hydroxymuconate (CHM) isomerase were assayed as described previously [7,8]. For measurement of the dioxygenase pH optimum allowance was made for the pH dependence of the extinction coefficient of the substance monitored, 5-carboxymethyl-2-hydroxymuconic semialdehyde [9]. Km values were obtained by non-linear regression analysis of v against s [10].

2.3. **Purification of HPC dioxygenase**

The enzyme was purified from *E.coli* 5K (pDR9304). An ultracentrifuged crude extract was prepared from 400 ml Luria broth-ampicillin grown cells as described previously [11]. The ultracentrifuged extract (4ml per run) was applied to a Pharmacia HR 10/10 Mono Q anion exchange column and chromatographed using a Pharmacia fast protein liquid chromatography (FPLC) system. An 80 ml gradient of 0.0—>0.5 M NaCl in 20 mM Tris.HCl buffer pH 7.5, with a flow rate of 4 ml.min⁻¹ was used and the fractions with the highest specific activities, eluted at approx. 0.3 M NaCl, were pooled. The enzyme at this stage appeared to be about 80% pure as judged by SDS-PAGE so further purification was attempted. A sample was applied to a Pharmacia Phenyl Superose HR 5/5 hydrophobic interaction column and eluted with a 10ml decreasing gradient of 1.7—>0.0M (NH₄)₂SO₄ in 0.1M sodium phosphate buffer pH 7.5 at a flow rate of 0.5ml.min⁻¹. The HPC dioxygenase eluted at 0.2M (NH₄)₂SO₄ but virtually all the enzyme activity had been lost. A further sample from the Mono Q step was subjected to gel filtration using two Pharmacia HR10/30 Superose 12 columns connected in series. The columns were equilibrated with 0.05M sodium phosphate/0.15M NaCl buffer pH 7.5 and run at a flow rate of 0.4 ml.min⁻¹. The recovery of HPC dioxygenase activity in this case was much better but some activity was lost and the specific activity decreased. Details of the purification are given in Table 1. Protein in the various fractions was measured as described previously [11].
2.4. Molecular mass estimation

Subunit molecular masses were estimated by SDS-polyacrylamide gel electrophoresis (PAGE) [12] using gradient gels of 7.5% to 20% acrylamide. The proteins used to calibrate the gels were bovine serum albumin (66000); ovalbumin (45000); glyceraldehyde 3-phosphate dehydrogenase (36000); carbonic anhydrase (29000); trypsinogen (24000); trypsin inhibitor (20100) and bovine lactalbumin (14200).

2.5. Amino acid sequencing

The peak fraction after Phenyl Superose chromatography was run on a 7.5% - 20% SDS-PAGE gradient gel and electroblotted onto a polyvinylidene difluoride (PVDF) membrane [13] using 50 mM glycine -50 mM Tris, pH 10, transfer buffer and stained with Coomassie blue R-250. The stained Mr 33000 protein band was excised and loaded into an Applied Biosystems 470A gas-phase sequencer without polybrene.

2.6. DNA manipulations

Small-and large-scale plasmid preparations were carried out by standard procedures [14]. Restriction endonucleases, T4 DNA ligase and exonuclease III deletion kit were used according to the manufacturer's instructions. Isolation of DNA fragments from low melting point agarose gels was as described [15]. Transformations were carried out using the morpholinepropane sulphonic acid (MOPS) - RbCl method [16].

2.7. Nucleotide sequencing and oligonucleotide synthesis

Sequencing was carried out using T7 DNA polymerase according to Tabor and Richardson [17]. The nucleotide analog 7-methyldeazaguanosine was included in the nucleotide mixes during sequencing to overcome base compressions. Plasmid DNA for sequencing was prepared as described [18]. Oligonucleotides were synthesized with an Applied Biosystems 380 B DNA synthesizer using cyanoethylphosphoramidite chemistry.

2.8. Chemicals

HPC was purchased from Sigma and CHM was prepared as described [6]. Restriction endonucleases were from Gibco-BRL and Pharmacia, T7 DNA polymerase and the exonuclease III deletion kit were from Pharmacia.
Nucleotides were from United States Biochemicals and the $^{35}\text{S-dATP}$ (1000 mCi/mmol) was from Amersham. All other chemicals were of the highest grade commercially available.

3. RESULTS AND DISCUSSION

3.1. Subcloning of the HPC dioxygenase gene (hpc B)

*E. coli* C genes involved in the catabolism of HPC have been cloned [6]. A 3.0kbp *BamHI-SalI* fragment from one of these clones (pJJ801) was introduced into the polylinker region of pUC18 and pUC19 to give pDR1830 and pDR1930, respectively. When pDR1830 and pDR1930 were propagated in the *E. coli* K12 strain 5K (which itself is devoid of the *hpc* genes) two enzymes of the HPC pathway, HPC dioxygenase and CHM isomerase were present at high activity for 5K(pDR1930) but were barely detectable for 5K(pDR1830) (data not shown). When the two cell-free extracts were analysed by SDS-PAGE, that from 5K (pDR1930) showed two very strong bands (Mr 14000 and 33000) that were not seen in the 5K (pDR1830) extract. The Mr14000 band is known to be CHM isomerase [11] so it seemed likely that the Mr 33000 band was HPC dioxygenase.

When the insert DNA of pDR1930 was shortened unidirectionally from the *SalI* end by exonuclease III digestion a series of deletion subclones was obtained. The shortest subclone that still expressed HPC dioxygenase activity was pDR9304. The shorter subclones (see Section 3.4) enabled the position of the HPC dioxygenase gene to be defined. Such shorter subclones like pDR9315 still expressed CHM isomerase activity. When the *KpnI* site in the insert DNA was used along with the *Kpn I* site in the polylinker to delete DNA from the opposite (*BamHI*) end of the insert DNA the resulting plasmid (pDR934; Section 3.4) expressed neither HPC dioxygenase nor CHM isomerase activity. Since the *KpnI* site in the insert DNA had been deleted in pDR9315 which still produced CHM isomerase it seemed that this *KpnI* site was within the HPC dioxygenase gene.

3.2. Enzyme purification

The HPC dioxygenase produced by *E. coli* C cells grown on 4-hydroxyphenylacetate [7] was very unstable and the crude extract lost all its activity after storage for 24 h at 0°C. The addition of reagents such as glycerol, acetone, dithiothreitol or iron salts that are known to stabilize some dioxygenases had no effect on the *E. coli* enzyme. Because pDR9304 strongly
expressed HPC dioxygenase, extracts prepared from 5K(pDR9304) have 100 x higher initial specific activities than those from *E. coli* C and therefore facilitated purification of the enzyme. However, these preparations also lose activity quite rapidly so the emphasis in the purification was to monitor both the catalytic activity and the Mr 33000 protein.

The major protein peak obtained after Phenyl Superose chromatography of the HPC dioxygenase from the Mono Q step had only slight catalytic activity but the Mr 33000 protein accounted for more than 95% of the total protein in the fraction. No other fractions had HPC dioxygenase activity. The recovery of active HPC dioxygenase after Superose 12 gel filtration was higher than after Phenyl Superose chromatography and exactly matched the distribution of the Mr 33000 protein. However, in the best fractions, the Mr 33000 protein accounted for about 85% of the total protein and so was slightly less pure than after Phenyl Superose chromatography. The Mr 33000 protein from the Phenyl Superose step was the one used for N-terminal sequence estimation. Using the enzyme from the Mono Q column, the effect of pH on the rate of reaction was measured in 0.1M sodium phosphate buffer over the pH range 6.6-8.6. The optimum region was from 7.2 to 7.8. The Km for HPC measured at pH 7.5 using solutions saturated with O₂ was 16 ±3 μM.

3.3 *N-terminal amino acid sequence*

Although the PVDF blot was well washed to remove contaminating buffer, very small amounts of glycine are always retained. In the first sequencing cycle the amount of glycine detected was much greater than that of any amino acid in subsequent cycles. But no other amino acid was seen in the first cycle so glycine appeared to be the amino terminal residue. Unambiguous sequence information was obtained for the first 21 residues and this is given in Fig. 1.

3.4 *Nucleotide sequence of the HPC dioxygenase gene*

The sequencing strategy used is described in the legend to Fig. 2. Because the deletion subclones had an average size difference of 100 bp and the sequence of around 250 residues was obtained for each construct there was very significant repetition of sequence measurement. The single-strand sequence was thus obtained with a high degree of accuracy. This sequence information was used to design two synthetic oligonucleotides (DR12 and 13) that were used in
sequencing part of the opposite strand. The remaining part of the opposite strand was sequenced using the deletion subclone pDR934. Sequencing from the KpnI site of pDR934 utilising the universal primer site of pUC19 gave information for about 250 bases and this sequence was extended past the start of the dioxygenase gene by use of a further synthetic oligonucleotide, DR14, Fig. 2.

Computer analysis of the nucleotide sequence revealed an open reading frame (ORF) of 831 bp with associated ribosome binding site [19] 10 bp before the ATG start codon. This ORF was confirmed as that encoding HPC dioxygenase by its predicted amino acid sequence for residues 2-22 which corresponded exactly to the amino terminal amino acid sequence obtained from the purified protein. The methionine at position one of the ORF was absent from the purified protein. There are 34 bp between the end of the HPC dioxygenase gene and the start of the CHM isomerase gene whose sequence has already been described [11]. The nucleotide sequence which shows a single KpnI site and the predicted amino acid sequence for HPC dioxygenase are shown in Fig. 3. A KpnI site was known from the deletion subcloning [see Section 3.1] to be within the HPC dioxygenase gene. The molecular mass calculated from the predicted amino acid sequence was 31332 which agrees well with that of 33000 for the purified protein measured by SDS-PAGE.

The predicted amino acid sequence of HPC dioxygenase was compared to those of the various extradiol and intradiol dioxygenases already reported [2,3,4] using the GAP programme of the University of Wisconsin Molecular Biology package. The extent of similarity varied between 8% and 14% indicating that there was no striking relationship between HPC dioxygenase and any of the other aromatic ring-fission dioxygenases. No obvious similarity to any other protein was apparent when the HPC dioxygenase sequence was compared to the NBRF (release 20) and the SWISSPROT (release 10) protein data bases. There was no apparently significant sequence similarity between HPC dioxygenase and CHM isomerase [11], two constituent enzymes of the HPC catabolic pathway. So the relationship of HPC dioxygenase to other ring-fission dioxygenases or to other HPC pathway enzymes is still an open question.
Acknowledgements: We are indebted to Dr. K. S. Lilley for the amino acid sequencing, Mr. John Keyte for oligonucleotide synthesis, and Michelle Kelly for typing the manuscript. D. I. R. thanks the SERC for a research studentship.
REFERENCES


Legends to Figures

Fig. 1. N-terminal amino acid sequence of *E.coli* C HPC dioxygenase.

Fig. 2. Restriction map of subclones and sequencing strategy for the *E.coli* C HPC dioxygenase gene. The 3.0 kbp *BamHI*-*SalI* fragment (heavy line) cloned into the polylinker site of pUC19 to give pDR 1930 is shown with major restriction sites. The restriction sites of the polylinker are not drawn to the same scale. Deletion of the 1.5 kbp *BamHI*-*KpnI* fragment of the insert DNA gives pDR934. The physical location of the HPC dioxygenase gene (*hpcB*) is shown by the hatched box. Unidirectional deletion subclones were created from the *SalI* end of the insert DNA of pDR 1930 by the exonuclease III digestion procedure. The coding strand encompassing HPC dioxygenase was sequenced using the deletion subclones pDR9304-9312 and pDR9315 as shown. The small arrows indicate the start, direction and extent of dideoxynucleotide sequencing carried out using the M13 reverse sequencing primer. The complementary strand was sequenced using the M13 universal primer. The broad arrows indicate the start, direction and extent of sequencing carried out using the synthetic oligonucleotides DR 12-14 with the small arrow indicating, similarly, sequencing from the *KpnI* end of the insert DNA of pDR934. Restriction sites are: *Nd*, *NdI*; *E*, *EcoRI*; *Ss*, *SstI*; *Kp*, *KpnI*; *Sm*, *SmaI*; *B*, *BamHI*; *Ev*, *EcoRV*; *Ss"SstII*; *S*, *SalI*; *Ps*, *PstI*; *Sp*, *SphI*; *H*, *HindIII*; *Pv*, *PvuII*.

Fig. 3. Nucleotide sequence and deduced amino acid sequence of the *E.coli* C HPC dioxygenase gene (*hpcB*). The sequence is written in the 5'→3' direction of the coding strand with the deduced amino acid sequence below. The amino terminus of the gene was identified by matching the predicted sequence to the sequence obtained from the purified protein. The *KpnI* site within the gene and the ribosome binding site (RBS) are indicated. The start of the CHM isomerase gene (*hpcD*) is 32 bases after the end of *hpcB* as shown. The numbers refer to the nucleotide positions with the A of the ATG start codon numbered 1.
TABLE 1

A summary of the purification of E. coli C HP dehydrogenase.
Gly Lys Leu Ala Leu Ala Ala Lys Ile Thr

His Val Pro Ser Met Tyr Leu Ser Glu Leu Pro
1. Preliminary crystallographic analysis of 5-carboxymethyl-2-hydroxymuconate isomerase from *Escherichia coli*.
*J. Mol. Biol.* 210, 883-884.

2. Purification, some properties and nucleotide sequence of 5-carboxymethyl-2-hydroxymuconate isomerase of *Escherichia coli C.*
*FEBS Lett.* 266: 63-66
Preliminary Crystallographic Analysis of 5-Carboxymethyl-2-hydroxymuconate Isomerase from *Escherichia coli*

Dale B. Wigley, David I. Roper and Ronald A. Cooper
Preliminary Crystallographic Analysis of 5-Carboxymethyl-2-hydroxymuconate Isomerase from Escherichia coli

Escherichia coli 5-carboxymethyl-2-hydroxymuconate (CHM) isomerase was purified from an overexpressing cell line. The enzyme has been crystallized from ammonium sulphate in two different crystal forms. One of these has been analysed and found to be orthorhombic \( I222 \) or \( I2_{1}2_{1}2_{1} \) with cell dimensions \( a = 88 \) Å, \( b = 89 \) Å, \( c = 121 \) Å. The asymmetric unit contains two dimers (\( V_{n} = 211 \) Å\(^3\)/dalton). The crystals diffract to beyond 3.0 Å resolution and are stable to irradiation with X-rays. Data have been collected to 3.0 Å resolution and a search for potential heavy-metal derivatives is in progress.

Bacterial aerobic degradation of aromatic compounds involves fission of the benzene nucleus by dioxygenases with the subsequent conversion of the ring-fission products to intermediates of the citric acid cycle (Dagley, 1975). In the meta-cleavage degradation of 3,4-dihydroxyphenylacetic acid (homoprotocatechuic acid, HPC; Fig. 1) there are two distinct isomerization steps in the reactions leading eventually to pyruvate and succinic semialdehyde (Jenkins & Cooper, 1988). The first is the isomerization of 5-carboxymethyl-2-hydroxymuconate (CHM) to 5-carboxymethyl-2-oxo-hex-3-ene-1,6-dioate catalysed by CHM isomerase. The second is the analogous isomerization of 2-hydroxyhepta-2,4-diene-1,7-dioate (HHDD) to 2-oxohept-3-ene-1,7-dioate catalysed by HHDD isomerase. Chromosomal genes encoding these proteins have been cloned from Escherichia coli C and shown to be closely linked, being separated from one another only by other genes of the same catabolic pathway (Jenkins & Cooper, 1988). Each enzyme is apparently specific for its own substrate, despite the structural similarity of CHM and HHDD. We are seeking structural information to improve our understanding of the interactions between these enzymes and their ligands and hence to determine the basis of their specificity.

CHM isomerase was purified from an E. coli K12 strain harbouring pUC19 carrying the coding sequence and promoter region of E. coli C CHM isomerase. The enzyme was assayed as described (Garrido-Pertierra & Cooper, 1981) and comprised 2% of the soluble cell protein, (D.I.R. and R.A.C., unpublished results). Cells were grown overnight at 37°C in 400 ml of Luria broth containing 100 μg ampicillin/ml and harvested by centrifugation. After washing in 80 ml of 20 mM-Tris·HCl buffer (pH 7.5), the cells were resuspended in 16 ml of the same buffer and disrupted by ultrasonication at 0°C. This crude extract was centrifuged at 180,000 g for 90 minutes at 4°C, and then 1 mg of protamine sulphate added per 20 mg of protein. After gentle stirring at 0°C for 15 minutes, the sample was centrifuged at 15,000 g for a further 15 minutes at 4°C. The supernatant was passed through a 0.22 μm pore size filter, then applied to an HR 10/10 Mono Q anion-exchange column and chromatographed using a Pharmacia fast protein liquid chromatography system. The enzyme was eluted using a 70-ml gradient of 0.0-0.5 M NaCl in 20 mM-Tris·HCl buffer (pH 7.5). The peak fractions were pooled, an equal volume of 2 M-ammonium sulphate added, and then applied to an HR 5/5 Phenyl-Superose column equilibrated with 50 mM-sodium phosphate buffer (pH 7.5) containing 1.7 M-ammonium sulphate. A 10-ml gradient of 1.7 to 0.0 M-ammonium sulphate in 50 mM-sodium phosphate buffer (pH 7.5) was used to elute the enzyme. The peak enzyme fractions were then subjected to gel filtration on an HR 10/30 Superase 12 column. The peak fractions from this column were of greater than 99% purity as judged by SDS/polyacrylamide gel electrophoresis. For crystallization trials, the protein was concentrated to around 10 mg/ml by pressure dialysis and the buffer exchanged for 20 mM-Tris·HCl (pH 7.5). The specific activity of this final product was 300 μmol per minute per milligram of protein.

Crystals were obtained by vapour diffusion on a small scale in hanging drops (20 μl), and on a larger scale (200 μl) in sitting drops. A saturated solution of ammonium sulphate in 20 mM-Tris·HCl (pH 7.5) was added to the protein sample (at 10 mg/ml) to a final concentration of 20%. Vapour diffusion was carried out against 25 to 30% saturated ammonium sulphate solution in 20 mM-Tris·HCl (pH 7.5). The temperature at which crystallization was achieved was very important. At 25°C only a precipitate was seen, but at higher temperatures two different crystal forms were obtained. The first were rectangular bipyramids that were around 50 μm x 75 μm across and 100 μm in length. These

† Abbreviations used: HPC, homoprotocatechuic acid; CHM, 5-carboxymethyl-2-hydroxymuconate; HHDD, 2-hydroxyhepta-2,4-diene-1,7-dioate.
crystals grew at temperatures between 25°C and 30°C but were very sensitive to temperature and tended to redissolve if disturbed. The second form were long rectangular rods with a cross-section of 100 μm x 200 μm and up to 4 mm in length. These grew overnight at 30°C or 37°C and could be stabilized by harvesting into 35% saturated ammonium sulphate, 20 mM-Tris-HCl (pH 7.5). This second form was analysed using synchrotron radiation of 1.488 Å (1 Å = 0.1 nm) wavelength at Daresbury, U.K. The space group was determined to be orthorhombic with unit cell dimensions a = 88 Å, b = 89 Å, c = 121 Å. Systematic absences indicate a body-centred lattice and show the space group to be either I222 or I2_2_2. The crystals are stable in the beam and diffract to beyond 3.0 Å resolution. The size of the unit cell suggests two dimers (subunit molecular weight = 14,000) per asymmetric unit with V = 21 Å³/dalton, within the usual range for protein crystals (Matthews, 1968). The position of the non-crystallographic 2-fold axes should easily be determined by the rotation function (Rossmann & Blow, 1962) and will thus aid analysis of any potential heavy-atom derivatives. The protein contains a single cysteine residue per subunit (D.I.R. & R.A.C., unpublished results), which may be useful as a heavy-atom site. The search for suitable heavy-atom derivatives is in progress.

We would like to thank Dr J. P. Derrick for assistance with data collection.

Dale B. Wigley  
David I. Roper  
Ronald A. Cooper  
Department of Biochemistry  
University of Leicester  
Leicester LE1 7RH  
U.K.  
Received 10 August 1989

References

Edited by A. Klug
Purification, some properties and nucleotide sequence of 5-carboxymethyl-2-hydroxymuconate isomerase of *Escherichia coli* C

David I. Roper and Ronald A. Cooper

*Department of Biochemistry, University of Leicester, Leicester LEI 7RH, UK*

Received 2 March 1990; revised version received 17 April 1990

As part of an investigation into the evolution of catabolic pathway enzymes a cloned gene encoding the *Escherichia coli* C 5-carboxymethyl-2-hydroxymuconate (CHM) isomerase, an enzyme of the homoprotocatechuate catabolic pathway, was used to produce large amounts of the protein. The isomerase was purified to homogeneity and some of its properties determined. The reaction occurred optimally at pH 7.6 and the specificity constant was 5.8 x 10^6 M^-1 s^-1 with CHM and 6.0 x 10^6 M^-1 s^-1 with 2-hydroxyhepta-2,4-diene-1,7-dioate, the substrate of a second isomerase in the pathway. The pure protein showed one type of subunit of Mr 14000 whilst the molecular mass of the native enzyme was 30000, suggesting that it was a dimer of identical subunits. The first 19 N-terminal amino acids were sequenced and the data used to confirm that the open reading frame of 378 bp, identified from the nucleotide sequence, encoded the CHM isomerase.

**INTRODUCTION**

How the enzymes that catalyse successive reactions in catabolic pathways have arisen is unclear. They may have evolved from a single ancestral gene, by a stepwise process of tandem gene duplication followed by mutation of the extra gene copy, and would thus be structurally related. Alternatively, they may have arisen by recruitment, from a variety of sources, of proteins with the appropriate catalytic properties and would therefore be structurally unrelated. It is also possible that, within a single pathway, both of these processes are occurred.

The homoprotocatechuate (HPC) pathway is the bacterial route whereby 4-hydroxyphenylacetate, a product of phenylalanine and tyrosine breakdown and one of several compounds released during the degradation of lignin, is catabolized. HPC is converted to succinic aldehyde and pyruvate by a sequence of reactions initiated by *meta*-cleavage with dioxygen to give 5-carboxymethyl-2-hydroxymuconate semialdehyde (HMS) which in turn is oxidized in an NAD-dependent reaction to give 5-carboxymethyl-2-hydroxymuconate (CHM). The subsequent reactions of HMS involve, sequentially, isomerization, decarboxylation, isomerization, hydration and aldol cleavage (Fig. 1) [1-3], all of which depend on the presence of an enolizable oxo-group in the molecule [4]. So once a protein capable of isomerizing CHM was available, mutations of duplicated genes might have given rise fairly easily to altered proteins that could bring about the similar electronic shifts needed for decarboxylation, isomerization, hydration and aldol cleavage [4].

Analysis of the cloned HPC pathway genes and proteins therefore offers an opportunity to see whether enzymes catalyzing chemically related reactions have arisen by modification of a single ancestral gene. This paper describes the nucleotide sequence, the derived amino acid sequence and some properties of CHM isomerase to serve as a basis for comparison when information on the other pathway enzymes becomes available.

**2. MATERIALS AND METHODS**

2.1. Bacteria, plasmids and growth conditions

Cells were grown aerobically at 37°C in Luria broth [5] containing 100 μg · ml⁻¹ ampicillin. Liquid media were solidified as required by the incorporation of 1.6% (w/v) Bacto Agar (Difco Laboratories). The *E. coli* strain DH5α (Gibco-BRL) was used as host for the pUC18- and pUC19-based plasmids used in this study. The 3.0 kbp BamHI-SalI fragment of pJ801 [3] was incorporated into pUC18 and pUC19 to give pDR1830 and pDR1930, respectively. pDR1930 was used to prepare a nested set of deletions that included pDR9304 and pDR9313-9317 (Fig. 3) using a Pharmacia exonuclease III deletion kit. pDR1830 was digested with *Bam*HI and *Eco*RV to delete 0.75 kbp of cloned DNA, the cohesive ends removed by digestion with S1 nuclease and the blunt ends ligated to form pDR921 (Fig. 3).

2.2. Assay of CHM isomerase

The enzyme was assayed as described previously [2].

2.3. Purification of CHM isomerase

CHM isomerase was purified from *E. coli* DH5α(pDR9304). Cells grown overnight at 37°C in 800 ml medium were harvested by cen-
trifugation at 10000 × g for 10 min at 4°C. The pellet was washed in 160 ml 20 mM Tris-HCl buffer, pH 7.5, and finally resuspended in 16 ml of the same buffer. Cells (4 ml portions) were ultrasonicated as described previously [6]. The crude extract was ultracentrifuged at 180000 × g for 90 min at 4°C. The supernatant was treated with protamine sulphate (40 mg ml⁻¹ in 20 mM Tris-HCl buffer, pH 7.5) in the proportion 1 mg protamine sulphate to 20 mg bacterial protein. The mixture was stirred gently at room temperature for 10 min then centrifuged at 15000 × g for 15 min at 4°C. The supernatant was applied to an HR 10/10 Mono-Q anion exchange column and chromatographed using a Pharmacia fast protein liquid chromatography (FPLC) system. A 70 ml gradient of 0.0–0.5 M NaCl in 20 mM Tris-HCl buffer, pH 7.5, with a flow rate of 4 ml min⁻¹ was used and the fractions with the highest specific activities, eluted at approx. 0.25 M NaCl, were pooled. The volume of the pooled fractions was reduced to 2 ml by ultrafiltration, 2 ml of 1 M ammonium sulphate was added and the solution applied to an HR 5/5 Phenyl Superose column. A 10 ml gradient of 1.7–0.0 M ammonium sulphate in 0.1 M sodium phosphate buffer, pH 7.5, was applied at a flow rate of 0.5 ml min⁻¹. The CHM isomerase was eluted at the end of the gradient. Fractions with the highest specific activities were pooled and the volume reduced to 2 ml by ultrafiltration. Two Pharmacia HR 10/30 Superose 12 gel filtration columns connected in series were equilibrated with 50 mM sodium phosphate/0.15 M NaCl buffer, pH 7.5. The concentrated fraction was applied to this system and eluted at a flow rate of 0.4 ml min⁻¹. The CHM isomerase obtained was greater than 99% pure as judged by SDS-PAGE. Details of the purification are given in Table I. Proteins used to calibrate the gels were bovine serum albumin (66000), ovalbumin (45000), glyceraldehyde 3-phosphate dehydrogenase (36000), carbonic anhydrase (29000), trypsinogen (24000), trypsin inhibitor (20100) and bovine lactalbumin (14200).

2.4. Molecular mass estimation

Native molecular masses were estimated by gel filtration as described under purification of CHM isomerase. The system was calibrated using IgG (160000); bovine serum albumin (66000); β-lactoglobulin (35000) and cytochrome c (12400) as standards. Subunit molecular masses were estimated by SDS-polyacrylamide gel electrophoresis (PAGE) [9] using gradient gels of 7.5% to 20% acrylamide. The proteins used to calibrate the gels were bovine serum albumin (66000), ovalbumin (45000); glyceraldehyde 3-phosphate dehydrogenase (36000); carbonic anhydrase (29000); trypsinogen (24000); trypsin inhibitor (20100) and bovine lactalbumin (14200).

2.5. Amino acid sequencing

The pure enzyme was run on a 7.5%–20% SDS-polyacrylamide gradient gel and electroblotted onto a polyvinylidene difluoride (PVDF) membrane [10] using 50 mM glycine–50 mM Tris, pH 8, transfer buffer and stained with Coomassie blue R-250. The stained protein band was excised and loaded into an Applied Biosystems 470A gas-phase sequencer without polybrene.

2.6. DNA manipulations

Small and large scale plasmid preparations were carried out by standard procedures [11]. Restriction endonucleases, T4 DNA ligase and exonuclease III deletion kit were used according to the manufacturer’s instructions. Isolation of DNA fragments from low melting point agarose gels was as described [12]. Transformations were carried out using the morpholinepropane sulphonic acid (MOPS)-Richter method [13].

Table I

A summary of the purification of E. coli C CHM isomerase

<table>
<thead>
<tr>
<th>Purification step</th>
<th>Volume (ml)</th>
<th>Total protein (mg)</th>
<th>Total units (µmol·min⁻¹)</th>
<th>Specific activity (µmol·min⁻¹·mg⁻¹ protein⁻¹)</th>
<th>Recovery (%)</th>
<th>Purification (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultracentrifuged extract</td>
<td>17.5</td>
<td>446.0</td>
<td>2292</td>
<td>5.1</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Protamine sulphate treated</td>
<td>17.0</td>
<td>395.0</td>
<td>2924</td>
<td>7.4</td>
<td>127.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Ultracentrifuged extract</td>
<td>6.0</td>
<td>25.2</td>
<td>1566</td>
<td>65.9</td>
<td>72.2</td>
<td>12.9</td>
</tr>
<tr>
<td>Pooled Mono-Q fractions</td>
<td>6.0</td>
<td>11.4</td>
<td>1445</td>
<td>127.6</td>
<td>63.0</td>
<td>25.0</td>
</tr>
<tr>
<td>Pooled Superose fractions</td>
<td>5.0</td>
<td>9.0</td>
<td>1335</td>
<td>148.3</td>
<td>58.2</td>
<td>29.1</td>
</tr>
</tbody>
</table>
Nucleotide sequencing was carried out with T7 DNA polymerase according to Tabor and Richardson [14]. Plasmid DNA for sequencing was prepared by the method of Kraft et al. [15].

Chemicals

CHM was prepared as described [3]. Restriction endonucleases were from Gibco-BRL and Pharmacia, kits for DNA sequencing and preparing deletions were exonuclease III were from Pharmacia. All other chemicals were of the highest grade commercially available.

RESULTS AND DISCUSSION

Purification of the enzyme

Contrary to a previous report [3] the 3.0 kbp BamHI-SalI fragment of pJJ801 carries the gene for CHM isomerase (Roper and Cooper, unpublished results) as well as the HPC dioxygenase gene. When pDR1930 that carries this BamHI-SalI fragment was shortened by exonuclease III digestion one of the resulting plasmids (pDR9304) (Fig. 3), expressed much increased amounts of CHM isomerase and HPC dioxygenase and was therefore used as a source of CHM isomerase.

Some properties of the purified enzyme

The purified enzyme was stable for at least 4 weeks at 4°C. The effect of pH on the rate of reaction was measured in 0.1 M sodium phosphate buffer over the pH range of 6.6-8.6 and the optimum was at pH 7.6. At this pH the specificity constant (kcat/Km) [16] was 1.8 x 10^7 M^-1 s^-1 with CHM and 6.0 x 10^2 M^-1 s^-1 with 2-hydroxyhepta-2,4-diene-1,7-dioate, the substrate of the second isomerase in the pathway, indicating a strong preference for CHM. On SDS-PAGE the enzyme showed a single subunit of Mr 14000. The Vd of the native enzyme estimated by gel filtration was 10000 ± 2000 suggesting that the enzyme is a dimer of identical subunits.

N-terminal amino acid sequence

The first analysis indicated the presence of proline at position 1 so a second sample was subjected to an extended trifluoroacetic acid cleavage step during the first cycle. An unambiguous sequence for the first 19 residues was obtained with the exception of residue 7 where no assignment was possible. A sample of the dotted protein was analysed by the procedure of Imans [17] and this resulted in the identification of cysteine at position 7. The first 19 residues of CHM isomerase are shown in Fig. 2.

<table>
<thead>
<tr>
<th>1</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>His</td>
</tr>
<tr>
<td>3</td>
<td>Phe</td>
</tr>
<tr>
<td>4</td>
<td>Ile</td>
</tr>
<tr>
<td>5</td>
<td>Val</td>
</tr>
<tr>
<td>6</td>
<td>Gli</td>
</tr>
<tr>
<td>7</td>
<td>Gli</td>
</tr>
<tr>
<td>8</td>
<td>Cys</td>
</tr>
<tr>
<td>9</td>
<td>Ser</td>
</tr>
<tr>
<td>10</td>
<td>Asp</td>
</tr>
<tr>
<td>11</td>
<td>Asp</td>
</tr>
<tr>
<td>12</td>
<td>Ser</td>
</tr>
<tr>
<td>13</td>
<td>Asn</td>
</tr>
<tr>
<td>14</td>
<td>Arg</td>
</tr>
<tr>
<td>15</td>
<td>Glu</td>
</tr>
<tr>
<td>16</td>
<td>Ala</td>
</tr>
<tr>
<td>17</td>
<td>Asp</td>
</tr>
<tr>
<td>18</td>
<td>Leu</td>
</tr>
<tr>
<td>19</td>
<td>Pro</td>
</tr>
<tr>
<td>20</td>
<td>Gly</td>
</tr>
</tbody>
</table>

3.4. Nucleotide sequence of the CHM isomerase gene

Unidirectional deletions were made from the SalI site in pDR1930 to produce a series of subclones. The enzyme activity expressed by individual subclones was tested using crude extracts and CHM isomerase activity was found for subclones that retained greater than 1.1 kbp of DNA from the BamHI site of the construct. The smallest deletion subclone to retain CHM isomerase activity was pDR9317. Restriction maps and the nucleotide sequencing strategy employed are shown in Fig. 3.
DNA sequencing was begun from the HindIII site of pDR9317 towards the BamHI site utilizing the reverse primer site of pUC19. Because the deletion subclones had an average size difference of 100 bp and sequence for at least 300 residues was obtained from each construct there was very significant repetition of sequence measurement. The single-strand sequence was thus obtained with a high degree of accuracy. Sequence for the opposite strand was obtained by analysis of pDR821. The reverse primer site of pUC18 was used to obtain the sequence of the first 280 bases. A synthetic oligonucleotide corresponding to the sequence of bases 256–276 was then constructed and used as a primer to obtain the sequence for a further 260 bases. Computer analysis of the nucleotide sequence revealed an open reading frame (ORF) of 378 bp with associated ribosome binding site [18] 186 bp upstream of the Ndel site. This ORF was confirmed as that encoding CHM isomerase since the predicted amino acid sequence for residues 2–20 corresponded exactly with the amino termi-

minal amino acid sequence obtained from the pure protein. The methionine at position 1 of the ORF is absent from the purified protein. The molecular mass calculated from the predicted amino acid sequence was 14044 which agrees very well with that of the purified enzyme measured by SDS-PAGE. The nucleotide sequence and predicted amino acid sequence for CHM isomerase are shown in Fig. 4.

When various data bases were searched for other proteins with amino acid sequences similar to that of CHM isomerase no obvious matches were found. However, it will be possible to see whether CHM isomerase has any obvious sequence similarities to the other enzymes of the HPC catabolic pathway once the nucleotide and predicted amino acid sequences for those enzymes becomes available. In turn this might shed some light on how catabolic pathway enzymes have evolved.

Acknowledgements: We are indebted to Dr K.S. Lilley for the amino acid sequencing, Mr John Keyte for oligonucleotide synthesis, Mr A. Taylor and Ms Z. Master for typing the manuscript. D.I.R. thanks the SERC for a research studentship.

REFERENCES