# STUDIES ON THE aroA GENE OF ESCHERICHIA COLI

by

Kenneth Duncan

Thesis submitted for the degree of Ph.D. in the Faculty of Medicine, University of Glasgow.

December 1984



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

CONTENTS

CONTENTS

2.18 Preparation of crude extracts of *E.coli*

2.19 Enzyme assays

2.20 Determination of protein concentration

2.21 Polyacrylamide gel electrophoresis in the presence of SDS

2.22 DNA sequencing

2.23 M13 vector DNA

2.24 Digestion of RF DNA and storage

2.25 Ligation of RF DNA to insert DNA

2.26 Transformation of *E.coli* JM101

2.27 Preparation of single-stranded template DNA

2.28 Annealing of primer and template

2.29 The sequencing reactions

2.30 Reaction mixes

2.31 Polyacrylamide gel electrophoresis

2.32 Autoradiography

2.33 T-track analysis

2.34 Preparation of RF DNA *in vitro*

2.35 Clone turn-around

2.36 Computer programs for manipulation of DNA sequences

2.37 Determination of $^{32}P$ in DNA samples

2.38 5'-end labelling of DNA fragments

2.39 Preparation of RNA from *E.coli*

2.40 Nuclease S1 mapping

2.41 Blotting of DNA and RNA to introcellulose

2.42 Hybridisation solution for Northern blot

List of Figures

List of Tables

6.4 Growth of *E.coli* KL282

6.5 Specific activity of PSAT in crude *E.coli* extracts

6.6 Amino acid composition of PSAT

7.1 Active site lysine of 3-dehydroquinases

7.2 Active site lysine of aminotransferases

7.3 Quarternary structure of the shikimate pathway enzymes

Abbreviations

The abbreviations used are those recommended in "Policy of the Journal and Instructions to Authors", Biochemical Journal (1983), 209, 1-27, with the exception of:

| | |
|---|---|
| bp | base pairs |
| BSA | bovine serum albumin |
| DAHP | deoxy-D-arabino heptulosonate-7-phosphate |
| DHQ | 3-dehydroquinate |
| DHS | 3-dehydroshikimate |
| EPSP | 5-enolpyruvylshikimate 3-phosphate |
| E-4-P | erythrose-4-phosphate |
| FPLC | fast protein liquid chromatography |
| HPLC | high pressure liquid chromatography |
| kb | kilo base pairs |
| LMT | low melting temperature |
| PEP | phosphoenolpyruvate |
| PSAT | phosphoserine aminotransferase |
| SDS-PAGE | polyacrylamide gel electrophoresis in the presence of SDS |
| TCA | trichloroacetic acid |
| TEMED | NNN'N'-tetramethylethylenediamine |

Acknowledgements

I am indebted to my supervisor, Dr John Coggins, who has guided and advised me and whose enthusiasm and encouragement kept me going in difficult times. I wish to thank my colleagues on D-floor for their help and for many useful discussions, especially Ann, Ian, Gary and Martin.

Within the Biochemistry Department, I am grateful to Dr Ian Hamilton, who, as my 'auditor' took a keen interest in my progress; and to Dr Bob Eason for help with the computing facilities.

I am also grateful to Dr Andrew Coulson, Department of Molecular Biology, Edinburgh University, for introducing me to the Wisgen package; and to Dr Mark Edwards, G.D. Searle & Co., High Wycombe, for the *tyrB* sequence prior to publication.

I wish to thank the Medical Faculty, University of Glasgow, for the award to a Postgraduate Studentship; Professor R.M.S. Smellie for providing the facilities in the Biochemistry Department; and the SERC for the running expenses for the DNA sequencing experiments.

I am extremely grateful to Mrs L. Peedle, who undertook the typing of this thesis at short notice, and who completed it with great skill and patience.

Finally, special thanks go to my wife, Elaine, for her patience, support and understanding, and for reminding me that there's more to life than the shikimate pathway.

SUMMARY

1. The E.coli aroB gene, encoding 3-dehydroquinate synthase has been sub-cloned from pLC29-47, a member of the Clarke-Carbon E.coli gene bank. Cells harbouring an aroB recombinant plasmid pKD101, overexpress the enzyme 20-25 fold. Further sub-cloning of a 3.5 kb EcoRI-BglII fragment, and comparison with an independently isolated aroB clone, pJB14, has located the coding sequence of the gene.

2. The E.coli structural gene encoding 3-dehydroquinase (aroD) has been subcloned on a 1.8 kb ClaI fragment from the plasmid pJKK12. The resulting recombinant, pKD201, overexpresses the enzyme approximately 100-fold. The entire DNA sequence of the ClaI fragment has been determined and the coding sequence of the aroD gene identified. The aroD gene encodes a 240 amino acid polypeptide of calculated $M_r$ 28117.

3. The aroA gene of E.coli, encoding 5-enolpyruvylshikimate 3-phosphate synthase was sub-cloned from the tranducing bacteriophage λpserC. Further sub-cloning located the gene on a 1.9 kb ClaI-PvuII fragment. Cells harbouring a number of aroA recombinant plasmids overexpress EPSP synthase approximately 100-fold, including pKD501 which is now used as a source of the purified enzyme; these cells are also tolerant to high levels of the herbicide glyphosate, the

target of which is EPSP synthase. It was found, however, that cloning from a ClaI site in the phage reduced the level of expression to that found in wild type cells. Polyacrylamide gel electrophoresis in the presence of SDS of crude extracts of plasmid carrying cells showed that in those clones which overexpress EPSP synthase at a high level, a heavily stained 40 000 $M_r$ band is present.

4. The DNA sequence of the aroA gene has been determined. The gene encodes a 427 amino acid polypeptide with a calculated $M_r$ of 46112. The location of the coding sequence of aroA in the DNA sequence revealed that it is positioned 700 bp away from the ClaI site and examination of the sequence revealed that there was no obvious promoter for the gene. Further cloning experiments showed that the aroA promoter was located at least 1 200 bp upstream of the initiation codon.

5. DNA sequencing and analysis of the sequences upstream of aroA revealed the presence of another gene. This gene encodes a polypeptide of 362 amino acids and calculated $M_r$ 39834, corresponding to the 40 000 $M_r$ band seen on SDS-PAGE of the crude extracts of cells carrying EPSP synthase overexpressing plasmids. The gene was identified as serC, which encodes the enzyme phosphoserine aminotransferase, the second enzyme on the three-step serine biosynthetic pathway.

6. Northern blot analysis and S1 nuclease mapping have been used to show that the two genes are linked to form the serC-aroA operon. The operon is expressed from a single promoter, located 55 bp upstream of the initiation codon for serC. Two messages are produced: the first is monocistronic, encoding only the serC gene; the second is polycistronic, and encodes both genes.

# CHAPTER 1 INTRODUCTION

## 1.1 Overview of the shikimate pathway, and its organisation

### 1.1.1 The shikimate pathway

The shikimate, or common aromatic amino acid, pathway is responsible for the biosynthesis of chorismic acid from the carbohydrate precursors erythrose-4-phosphate (E-4-P) and phosphoenol pyruvate (PEP). It takes its name from the central intermediate, shikimic acid. The conversion of E-4-P and PEP to chorismate is a seven step process (Figure 1.1). Bacteria, fungi and plants are capable of synthesising chorismate using the same chemical steps; the pathway is not found in other eukaryotic organisms.

Chorismate is the branch point intermediate for pathways leading to synthesis of the aromatic amino acids phenylalanine, tryptophan and tyrosine, and for a number of aromatic ring containing compounds (Figure 1.2).

The steps in the pathway were elucidated by Davis and his co-workers in the 1950's. Auxotrophic mutants of Escherichia coli were isolated which identified blocks in the pathway (Davis, 1951), and it was shown that shikimate was the common precursor of several aromatic metabolites. The field has been extensively reviewed (Haslam, 1974; Weiss & Edwards, 1980; Herrmann, 1983) and is discussed briefly in the sections which follow.

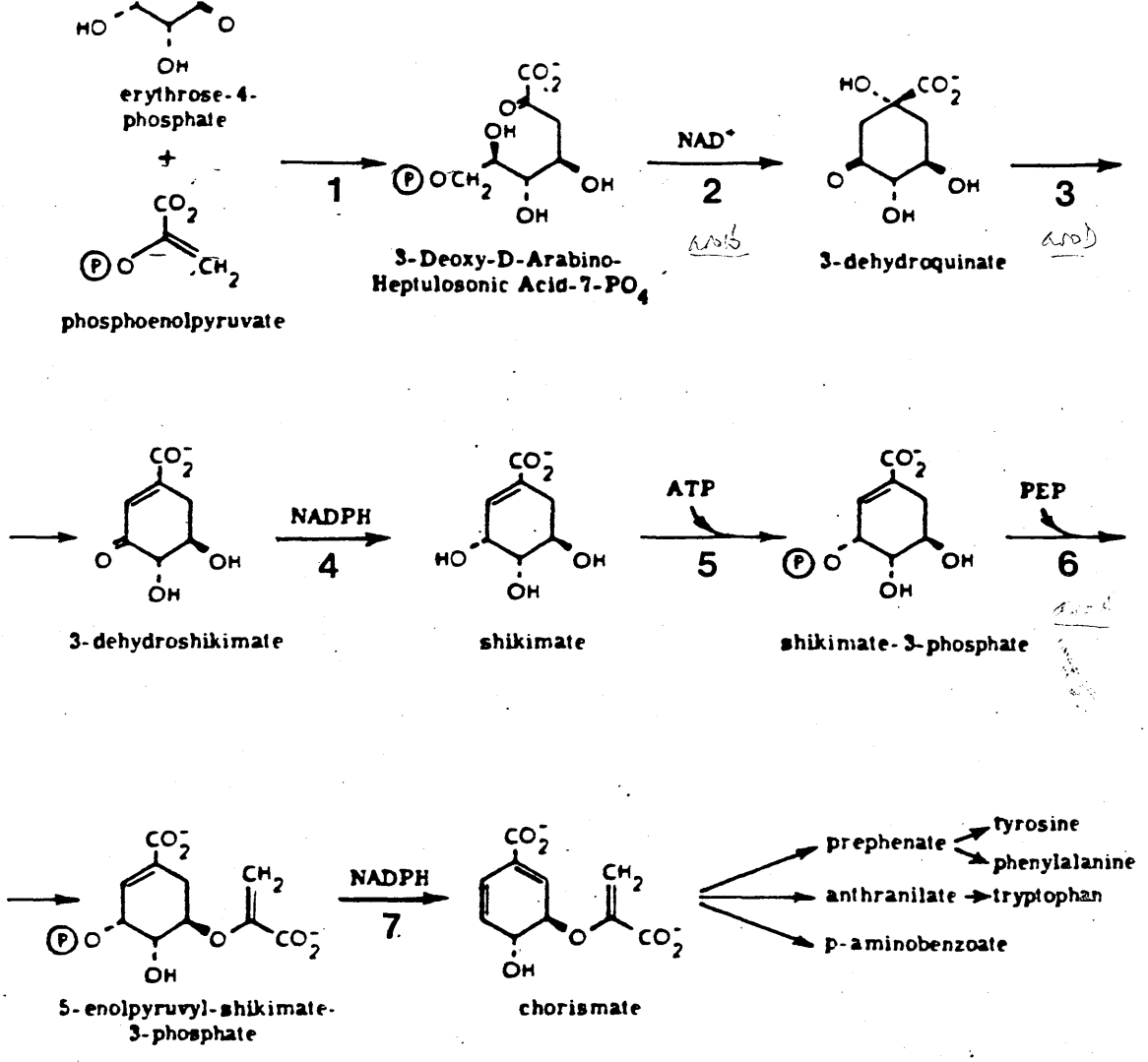



Figure 1.1: The shikimate pathway.

Enzymes:
1 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase
2 3-dehydroquinate synthase
3 3-dehydroquinase
4 shikimate dehydrogenase
5 shikimate kinase
6 5-enolpyruvylshikimate 3-phosphate synthase
7 chorismate synthase

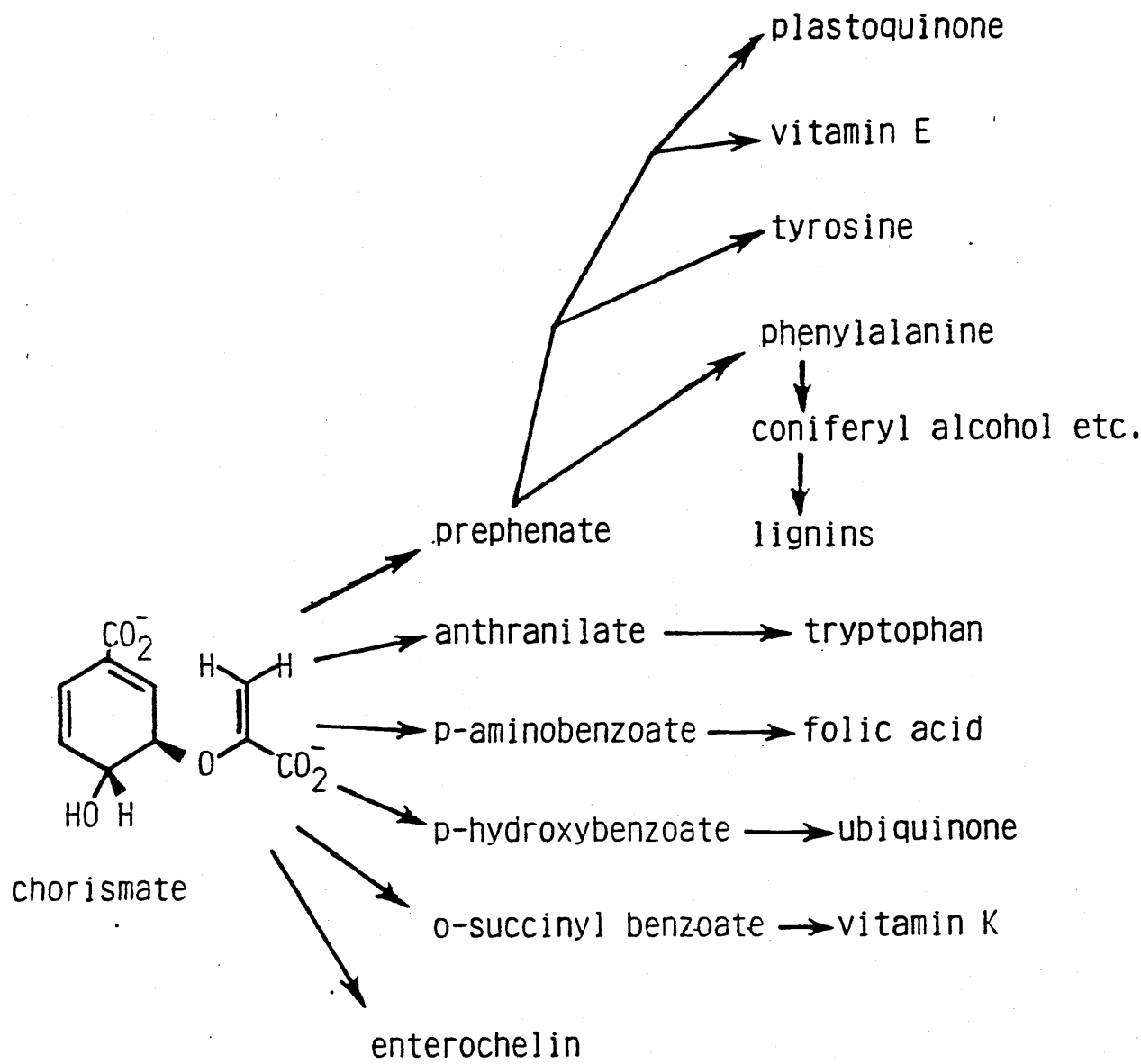



Figure 1.2: The aromatic biosynthetic pathways beyond chorismate.

### 1.1.2 Organisation of the enzymes and genes in the pathway

In *E. coli* and in most bacteria which have been studied, each pathway enzyme occurs on a separate polypeptide chain, which can be purified free of the activity of any other pathway enzyme (Coggins *et al.*, 1985). The structural genes for these enzymes are widely scattered about the genome (Bachmann, 1983; Sanderson & Roth, 1983; Henner & Hoch, 1980).

In contrast, most fungi contain a pentafunctional enzyme, the *arom* complex in which the five central pathway activities (steps 2-6, Figure 1.1) occur on a single polypeptide chain, the product of a single gene (Lumsden & Coggins, 1977; Gaertner & Cole, 1977; Larimer *et al.*, 1983).

In plants, two of the enzyme activities, 3-dehydroquinase and shikimate dehydrogenase are always found associated even after various procedures are employed to separate them (Boudet & Lécusson, 1974; M.S. Campbell & D.M. Mousdale, unpublished results). Each of the other pathway enzymes is clearly separable (Mousdale & Coggins, 1984b).

### 1.1.3 The shikimate pathway as a model system

The shikimate pathway is therefore a good model system for studying the organisation of a biosynthetic pathway. Several questions can be asked of the system. For example:
Why is there a multifunctional polypeptide in some organisms but not in others?
Why are there multifunctional enzymes catalysing the

reactions of some biosynthetic pathways and not all biosynthetic pathways within the same organism? How might the gene encoding a multifunctional enzyme have evolved?

The answers to these and a number of related questions may be obtained by detailed studies of the enzymology of the pathway and by cloning and DNA sequence analysis of the genes encoding the enzymes of the pathway, from a spectrum of organisms.

### 1.1.4 Definition of a multifunctional enzyme

The arom complex represents an example of a special class of enzymes, the multifunctional enzymes (Kirschner & Bisswanger, 1976; Bisswanger & Schimcke-Ott, 1980). Multifunctional proteins/enzymes may be defined by two characteristic properties. Firstly, they have multiple catalytic or binding functions and secondly, these functions are located on the same polypeptide chain. The active centres and/or binding sites may be generated by the folding of contiguous stretches of the polypeptide chain to yield autonomous globular domains. The active sites may lie within a domain, between domains, or at intersubunit interfaces (Kirschner & Bisswanger, 1976).

Examples of multifunctional enzymes will be discussed later in this chapter, along with the catalytic properties of these enzymes and their possible modes of evolution.

## 1.2 Genetic studies on the shikimate pathway

### 1.2.1 Bacteria

The structural genes for the shikimate pathways enzymes have been most intensively studied in E.coli. Mutants completely lacking each of the enzymatic activities of the pathway, except shikimate kinase, have been identified. The distribution of five of the genes (aroA - aroE) was deduced by Pittard & Wallace (1966), and the auxotrophic mutations were correlated with enzymatic deficiencies (Figure 1.3). There are three isoenzymes for DAHP synthase; mutations in each gene (aroF, aroG, aroH) were isolated and mapped by Wallace & Pittard (1967). The regions of the chromosome flanking the three genes described in detail in this work (aroA, aroB and aroD) are shown in Figure 1.4.

The aro genes of Salmonella typhimurium have been studied (Gollub et al., 1967; Sanderson & Roth, 1983); their distribution is scattered and very similar to that of E.coli. In Bacillus subtilis, there is some clustering of the aro genes (Henner & Hoch, 1980), but it is not known if the gene clustering represents a functional unit of transcription.

### 1.2.2 Control of aro gene expression in E.coli

It has been shown (Tribe et al., 1976) that the shikimate pathway genes are expressed constitutively, with the exceptions of the DAHP synthases and shikimate kinase. Synthesis of the enzymes is unaffected by mutations in tyrR and trpR, the regulatory genes for tyrosine and

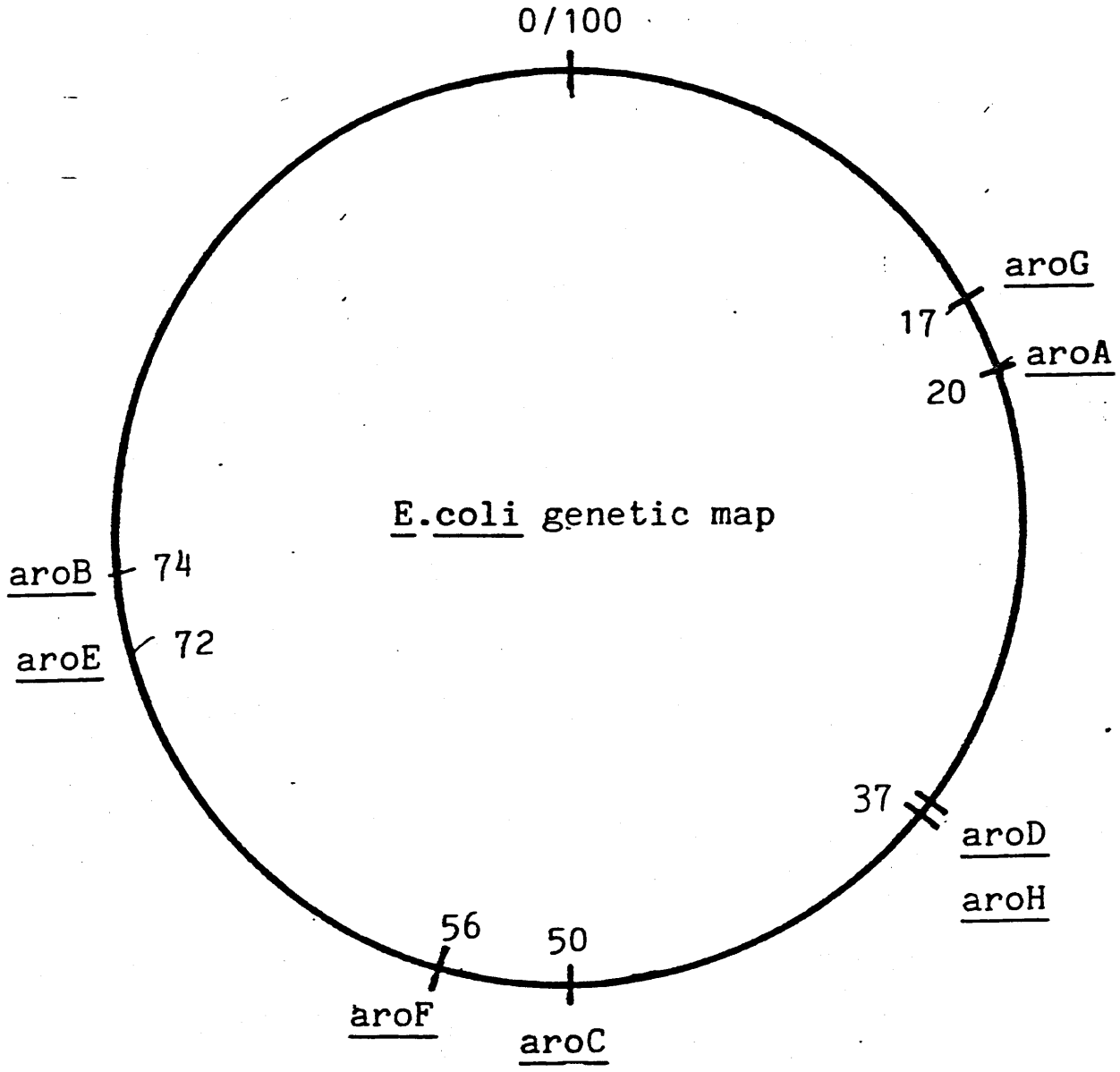



Figure 1.3: Location of the aro genes on the 100 minute E.coli genetic map (Bachmann, 1983).

| Pathway step | gene | enzyme |
|---|---|---|
| 1 | aroF | DAHP synthase (tyr-repressible) |
| 1 | aroG | DAHP synthase (phe-repressible) |
| 1 | aroH | DAHP synthase (trp-repressible) |
| 2 | aroB | 3-dehydroquinate synthase |
| 3 | aroD | 3-dehydroquinase |
| 4 | aroE | shikimate dehydrogenase |
| 6 | aroA | EPSP synthase |
| 7 | aroC | chorismate synthase |

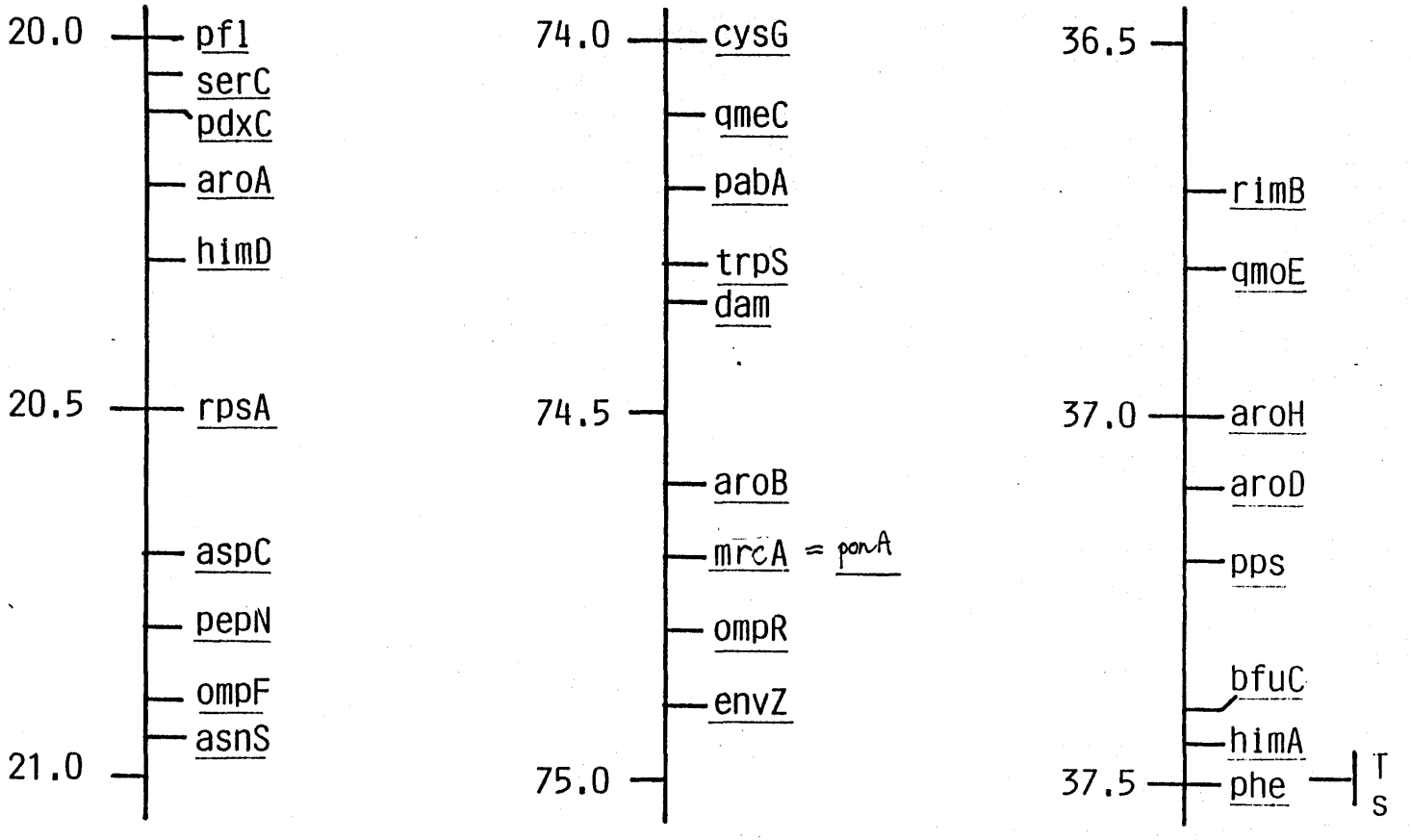



Figure 1.4: Detailed genetic map around the aroA, aroB and aroD genes of E.coli (from Bachmann, 1983).

tryptophan synthesis, respectively. With the exception of the three DAHP synthases, there is no specific regulation by either the aromatic amino acids or by DAHP or chorismate. There is, however, an unexplained change in the rate of enzyme synthesis co-ordinated with the growth rate.

There are two shikimate kinase isoenzymes in E.coli. Shikimate kinase I is expressed constitutively, but shikimate kinase II is under the control of the tyrR regulatory protein (Ely & Pittard, 1979).

The aroH (tryptophan sensitive DAHP synthase) gene is regulated by trpR (trp repressor). Zurawski et al. (1981) showed that the operator sequence of aroH has a 14/18 bp similarity in the DNA sequence to the trp operator, suggesting that this gene is regulated by repression.

Expression of aroG (phenylalanine sensitive DAHP synthase) is controlled by phenylalanine and tryptophan through the action of the tyrR repressor (Brown & Somerville, 1971).

The aroF gene (tyrosine repressible DAHP synthase) is a member of the tyr operon, along with tyrA, the structural gene for chorismate mutase/prephenate dehydrogenase. Control of the tyr operon depends on the concentration of tyrosine within the cell and is mediated by tyrR (Brown & Somerville, 1971).

The DAHP synthase genes are also derepressed by growth in medium deficient in iron (McCray & Herrmann, 1976). This is presumably to allow the replenishment of the intracellular pool of chorismate, which is the precursor of the iron chelator enterochelin.

1.2.3 Fungi

There are marked differences in the organisation of the enzymes and genes of the shikimate pathway in fungi. In stark contrast to the separable enzymes/scattered genes found in E.coli for the five central enzymes of the pathway, the enzymes in most fungi that have been studied are encoded by the arom gene cluster. The arom gene cluster was first identified in Neurospora crassa (Giles et al., 1976a), and has subsequently been found in a number of organisms, including other fungi (Ahmed & Giles, 1969), Euglena (Berlyn et al., 1970), Saccharomyces cerevisiae (de Leeuw, 1967) and Schizosaccharomyces pombe (Strauss, 1979). Figure 1.5 shows the structure of the arom gene cluster from N. Crassa. Giles et al. (1967a) obtained evidence for a gene cluster from the study of a number of polyaromatic mutants. These mutants displayed biochemical pleiotropic effects, polarity effects and the genetic localisation of completely non-complementing mutants was asymmetric, suggesting that the cluster was normally transcribed in a polarised fashion as a polycistronic mRNA. All five enzyme activities were identified after sucrose density gradient centrifugation of Neurospora crude extracts as part of an aggregate of $M_r$ 200 000.

The 'arom aggregate' was first purified by Burgoyne et al. (1969), who showed that the five activities were inseparable during purification and provided preliminary evidence for smaller components - the five polypeptide chains from which, it was supposed, the aggregate was

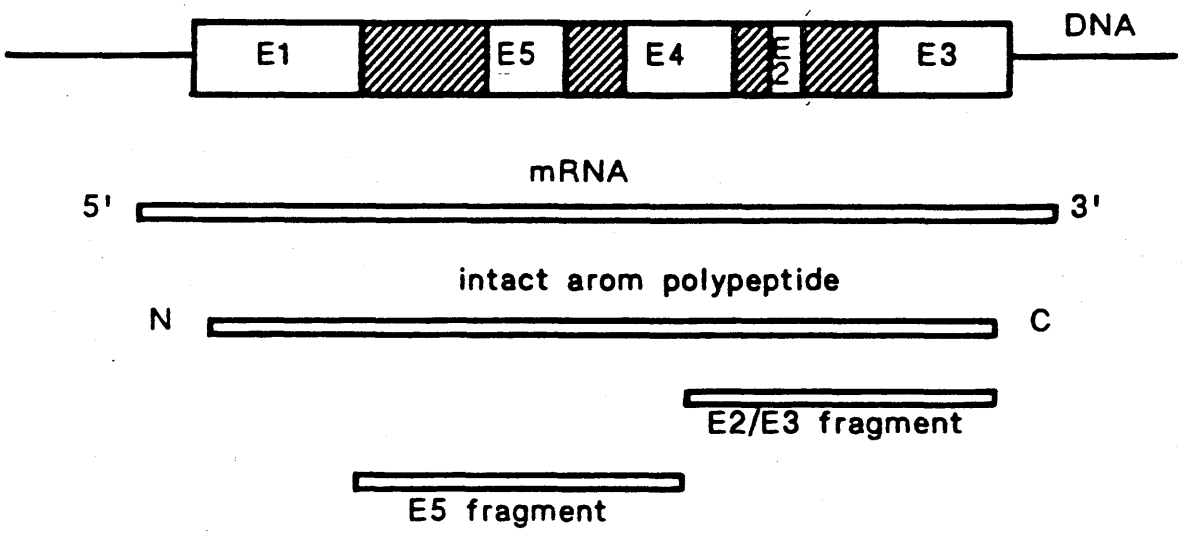



Figure 1.5: Structure of the arom gene of N.crassa (Rines et al., 1969). Mutations in the five enzyme activities of the arom complex, E1-E5 (corresponding to steps 2-6 in the pathway, respectively) were mapped in the DNA as shown above. Limited proteolysis has been used to isolate fragments of arom, carrying the activities shown (Smith & Coggins, 1983; M.R. Boocock, Ph.D. Thesis, University of Glasgow, 1983).

composed. However, when the arom complex was purified rapidly in the presence of proteinase inhibitors, it was found to be composed of two subunits of $M_r$ 165 000 (Lumsden & Coggins, 1977; Gaertner & Cole, 1977), which are identical (Lumsden & Coggins, 1978).

### 1.2.4 Cloning of fungal arom genes

The arom gene cluster has recently been cloned from three organisms. Recombinants carrying the arom gene were identified by complementation of various E.coli auxotrophic mutants:

| Organism: | Complements E.coli: | |
|---|---|---|
| Aspergillus nidulans | aroD | (Kinghorn & Hawkins, 1982) |
| S.cerevisiae | aroA, aroB, aroD, aroE | (Larimer et al., 1983) |
| S.pombe | aroD, aroE | (Nakanishi & Yamamoto, 1984) |

The observation that the S.cerevisiae gene complements a number of E.coli mutants suggests that the protein coding region is not interrupted by introns, as these would not be processed in the E. coli host.

## 1.3 The shikimate pathway in plants

The organisation of the shikimate pathway enzymes was investigated in a number of photosynthetic organisms for evidence of enzyme aggregation (Berlyn et al., 1970). Sucrose density gradient centrifugation showed that in Anabaena, Chlamydamonas, Nicotiana and Physcomitrella, all the pathway enzymes were separable, with the exception of

3-dehydroquinase and shikimate dehydrogenase. It was also shown that Euglena has an arom-like enzyme complex. Boudet & Lécusson (1974) used several techniques to separate 3-dehydroquinase and shikimate dehydrogenase from higher plants but found that they were always associated, suggesting the occurrence of either a multienzyme complex or a bifunctional enzyme. It has recently been shown that the shikimate pathway enzyme activities are located in the chloroplast (Mousdale & Coggins, 1984b), and that in pea seedlings 3-dehydroquinase and shikimate dehydrogenase occur on a bifunctional polypeptide of $M_r$ 60 000 (M.S. Campbell & D.M. Mousdale, unpublished results).

## 1.4 The shikimate pathway: enzymology

### 1.4.1 Step 1: 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (E.C. 4.2.1.15)

The condensation of E-4-P and PEP to form the seven carbon compound DAHP and inorganic phosphate is the first committed step in the shikimate pathway. The enzyme which catalyses this step, DAHP synthase, is very important in that it is at the gateway to the synthesis of a number of compounds and it therefore regulates the amount of carbon flowing through the pathway. Genetic and biochemical analysis has shown that there are three DAHP isoenzymes in E. coli, each of which is repressible by a different aromatic amino acid (Doy & Brown, 1965; Wallace & Pittard, 1967).

All three DAHP synthases have been purified to homogeneity and shown to have a similar structure. DAHP synthase (phe) is a tetrameric enzyme with a subunit $M_r$ of 35 000 (McCandliss *et al.*, 1978). Both DAHP synthase (tyr) and DAHP synthase (trp) are dimeric enzymes with a subunit $M_r$ of 39 000 (Schoner & Herrmann, 1976; Poling *et al.*, 1981). The DNA sequences of *aroF* (Shultz *et al.*, 1984), *aroG* (Davis & Davidson, 1982) and *aroH* (Zurawski *et al.*, 1981) show that there are regions of strong homology in the genes, suggesting a common ancestry.

Each DAHP synthase isoenzyme is inhibited specifically by a single aromatic end product. Phenylalanine and tyrosine inhibit DAHP synthase (phe) and DAHP synthase (tyr), respectively, by more than 95%. Tryptophan does not inhibit DAHP synthase (trp) by more than 40%, ensuring a continuous supply of chorismate for the biosynthesis of aromatic compounds, even when the aromatic amino acids are in excess (Herrman, 1983). Feedback regulation of DAHP synthase is the major regulator of carbon flow through the shikimate pathway.

In common with *E.coli*, *N.crassa* contains three DAHP synthases, each of which is inhibited by a different aromatic amino acid end product. The tryptophan sensitive isoenzyme has been purified to homogeneity; it is a tetramer with a subunit $M_r$ of 52 000 (Nimmo & Coggins, 1981).

1.4.2 Step 2: 3-dehydroquinate synthase (E.C. 4.6.1.3)

The second enzyme of the shikimate pathway, 3-dehydroquinate synthase, is responsible for the cyclisation of DAHP to form 3-dehydroquinate (DHQ). During the course of this reaction, the enzyme catalyses an oxidation, a β-elimination, an intramolecular aldol condensation and a reduction. The E.coli enzyme has been purified to homogeneity from wild type E.coli (Maitra & Sprinson, 1978) and from both wild type and an overproducing strain of E.coli (Frost et al., 1984). The subunit $M_r$ of this monomeric enzyme has been reported as 56 000 (Sucrose density gradient centrifugation; Berlyn & Giles, 1969); 57 000 (gel filtration and SDS-PAGE; Maitra & Sprinson, 1978); and 40 000 (SDS-PAGE; Frost et al., 1984). The sub-cloning of the aroB gene of E.coli, encoding 3-dehydroquinate synthase from pLC29-47, has been reported (Duncan & Coggins, 1983; K. Duncan, this work; Frost et al., 1984).

The E.coli enzyme requires catalytic amounts of $NAD^+$ and $Co^{2+}$ (Maitra & Sprinson, 1978). The 3-dehydroquinate synthase activity of the arom complex also requires $NAD^+$, but requires $Zn^{2+}$ rather than $Co^{2+}$ (Lambert, Boocock & Coggins, 1984).

The purified B.subtilis enzyme has an $M_r$ of 17 000 and is found associated with chorismate synthase and an NADPH-dependent flavin reductase activity (Hasan & Wester, 1978a).

### 1.4.3 Step 3: 3-dehydroquinase (E.C. 4.2.1.10)

The first double bond of the aromatic ring system is introduced by the third enzyme in the pathway, 3-dehydroquinase, which converts dehydroquinate to dehydroshikimate (DHS). The enzyme has been purified to homogeneity from *E.coli* K12 (Chaudhuri & Coggins, 1984b), and from an overproducing strain of *E.coli* (Dr S. Chaudhuri, unpublished results); it is a monomeric enzyme with a subunit $M_r$, estimated by SDS-PAGE, of 29 000.

Boudet & Lécusson (1974) showed that in a number of plant species, 3-dehydroquinase activity is inseparable from shikimate dehydrogenase activity. Bifunctional 3-dehydroquinase/shikimate dehydrogenase has been purified to homogeneity from the moss *Physcomitrella patens* (Polley, 1978), from *Phaseolus mungo* (Koshiba, 1979), and from *Pisum sativum* (M.S. Campbell & Dr D.M. Mousdale, unpublished results). The *P.patens* enzyme has an $M_r$ of 49 000 while the *Pisum sativum* enzyme has an $M_r$ of 60 000; both of these enzymes are monomeric.

The amino acid sequence around the active site lysine of the 3-dehydroquinase of the *arom* multifunctional protein has recently been determined (Dr S. Chaudhuri, unpublished results).

### 1.4.4 Step 4: shikimate dehydrogenase (E.C. 1.1.1.25)

The dehydrogenation of dehydroshikimate to form shikimate is catalysed by shikimate dehydrogenase. The $NADP^+$ specific enzyme has been isolated from *E.coli* K12;

it has a subunit of $M_r$ 31 000 (by SDS-PAGE) and is unusual in that it is a monomeric dehydrogenase (Chaudhuri & Coggins, 1984a). The *E.coli* structural gene for shikimate dehydrogenase, *aroE*, has been cloned and the DNA sequence determined (Anton & Coggins, 1983; I. Anton, Ph.D. Thesis, University of Glasgow, 1984).

### 1.4.5 Step 5: shikimate kinase (E.C. 2.7.1.71)

Shikimate is phosphorylated to shikimate 3-phosphate by shikimate kinase. Mutants which totally lack shikimate kinase have never been isolated from *E.coli* (Pittard & Wallace, 1966) and at least two peaks of shikimate kinase activity were observed on sucrose density gradient centrifugation of an *E.coli* crude extract (Berlyn & Giles, 1969), indicating that there are two isoenzymatic forms, shikimate kinase I and shikimate kinase II. Shikimate kinase I is synthesised constitutively, but the shikimate kinase II gene (*aroL*) is under the control of the *tyrR* repressor (Ely & Pittard, 1979). The $M_r$ of the enzyme has been estimated to be approximately 20 000 by gel filtration (Ely & Pittard, 1979).

There is a single shikimate kinase in *B.subtilis*. It has been purified to homogeneity and is a polypeptide of $M_r$ 10 000 which is complexed with a bifunctional DAHP synthase-chorismate mutase (Huang *et al.*, 1975). Shikimate kinase may represent the key allosteric control point of the pathway in this organism.

### 1.4.6 Step 6: 5-enolpyruvylshikimate 3-phosphate synthase (E.C.2.5.1.19)

The reaction catalysed by EPSP synthase (alternative name: 3-phosphoshikimate 1-carboxyvinyltransferase) is the transfer of the enolpyruvyl moiety of PEP to shikimate 3-phosphate, yielding 5-enolpyruvylshikimate 3-phosphate and inorganic phosphate (Levin & Sprinson, 1964). EPSP synthase has been purified to homogeneity from *E.coli* K12 (Lewendon & Coggins, 1983) and from pea seedlings (Mousdale & Coggins, 1984a). Both enzymes are monomeric, with a subunit $M_r$ of 49 000.

A great deal of interest has been focussed on this enzyme recently, following the discovery that glyphosate (N-phosphonomethylglycine, 'Roundup'), a successful broad spectrum, post emergence herbicide acts by blocking aromatic amino acid biosynthesis at the level of EPSP synthase (Steinrucken & Amrhein, 1980). Until recently, detailed mechanistic studies of herbicidal action on EPSP synthase have been hindered by the lack of sufficient quantities of the purified enzyme. At first the only pure preparations of EPSP synthase available were those of the *arom* enzyme complex of *N.crassa* and a detailed kinetic study of the effect of glyphosate on this enzyme has appeared (Boocock & Coggins, 1983). The *E.coli* enzyme is now available in milligram quantities (Duncan, Lewendon & Coggins, 1984a) and is being studied as a model for the action of the herbicide on the plant enzyme.

### 1.4.7 Step 7: Chorismate synthase (E.C. 4.6.1.4)

Chorismate synthase introduces the second double bond in the aromatic ring system by trans-1,4-elimination of orthophosphate from EPSP to yield chorismate. The enzyme has been partially purified from *E.coli* (Morell *et al.*, 1967) and purified to homogeneity from *B.subtilis* (Hasan & Nester, 1978b). The *B.subtilis* enzyme is part of a trimeric complex with 3-dehydroquinate synthase and NADH-dependent flavin reductase. In *Neurospora*, the chorismate synthase and flavin reductase reside on the same polypeptide chain (Welch *et al.*, 1974). Recently, a much improved purification of the *N.crassa* enzyme has been reported (M.R. Boocock, Ph.D. Thesis, University of Glasgow, 1983) but there is very little mechanistic information about the enzyme.

## 1.5 The qa gene cluster of N.crassa

In *N.crassa*, there is a second 3-dehydroquinase activity, which has a function in the catabolism of quinic acid (Giles *et al.*, 1967b). The catabolic dehydroquinase gene, *qa*-2, is a member of the *qa* cluster which consists of five structural genes and two regulatory genes, all of which are tightly linked (Giles, 1978; Patel *et al.*, 1981; Huiet, 1984). Transcription of the structural genes is induced 300-1 000 fold by quinic acid, which is plentiful in the rotting vegetation upon which *Neurospora* thrives (Patel *et al.*, 1981). Transcription is controlled by the

two regulatory genes, qa-1F and qa-1S (Huiet, 1984). The gene products of qa-2 (catabolic 3-dehydroquinase), qa-3 (quinate/shikimate dehydrogenase) and qa-4 (dehydroshikimate dehydratase) are responsible for the first three steps of quinate and shikimate catabolism (Figure 1.6).

qa-2 was the first Neurospora structural gene to be cloned, by complementation of an E.coli aroD mutant (Vapnek et al., 1977). Subsequently, the entire qa region was cloned (Schweizer et al., 1981) and its genetic organisation and transcriptional regulation investigated (Patel et al., 1981; Alton et al., 1982; Huiet, 1984; Tyler et al., 1984).

Catabolic dehydroquinase has been purified to homogeneity (Hautala et al., 1975; Chaudhuri & Coggins, 1984c). It exists as a dodecamer with a subunit $M_r$ of 20 000 (Chaudhuri & Coggins, 1984c). The DNA sequence of the qa-2 gene has been determined (Alton et al., 1982; Dr M.E. Case, unpublished results).

The availability of the DNA sequence of the E.coli aroD gene and the Neurospora arom gene will allow the evolutionary relationships of the three dehydroquinases to be deduced.

## 1.6 The aromatic pathways beyond chorismate

Chorismate is an important substrate for a range of biosynthetic pathways, too numerous and too complex to describe in detail here. The end products of these pathways are shown in Figure 1.2. It is interesting, however, to

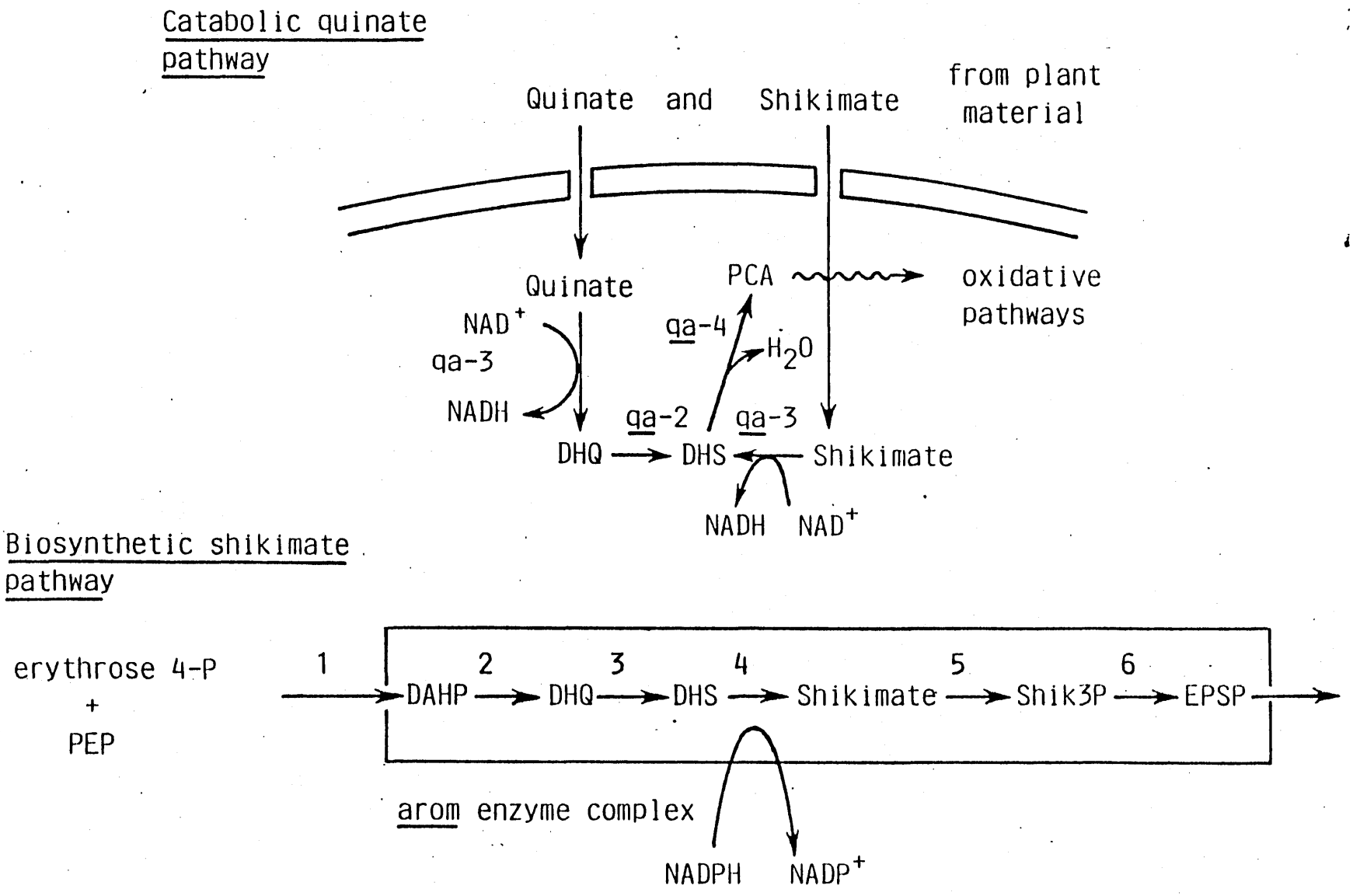



Figure 1.6: Relationship between the catabolic quinate pathway and the biosynthetic shikimate pathway. PCA is protocatechuic acid.

note the diversity in the organisation of the enzymes and genes of the aromatic amino acid biosynthetic pathways of different organisms.

### 1.6.1 Phenylalanine

The phenylalanine biosynthetic pathway has been recently reviewed and a full account is given in Garner & Hermann (1983). Chorismate is converted to phenylalanine in three steps by one of two routes (Figure 1.7). The phenylpyruvate route is found in E.coli, Salmonella, Bacillus, Neurospora and Saccharomyces. The arogenate pathway occurs in blue-green bacteria and Pseudomonas aeruginasa.

In E.coli, Salmonella and Klebsiella there is a bi-functional chorismate mutase/prephenate dehydratase, but in Bacillus, Brevibacterium, Neurospora and Saccharomyces, the two activities reside on different polypeptides.

### 1.6.2 Tyrosine

An account of tyrosine biosynthesis is given in Camakaris & Pittard (1983). The conversion of chorismate to tyrosine is a three step process (Figure 1.7), via 4-hydroxypyruvate (E.coli, Salmonella, Klebsiella, Bacillus, Pseudomonas, and Neurospora) or arogenate (blue green algae, Pseudomonas and Neurospora).

In E.coli, S.typhimurium and K.aerogenes a bifunctional chorismate mutase/prephenate dehydrogenase is found. In B.subtilis the single chorismate mutase provides prephenate for the monofunctional prephenate dehydrogenase and prephenate dehydratase enzymes.

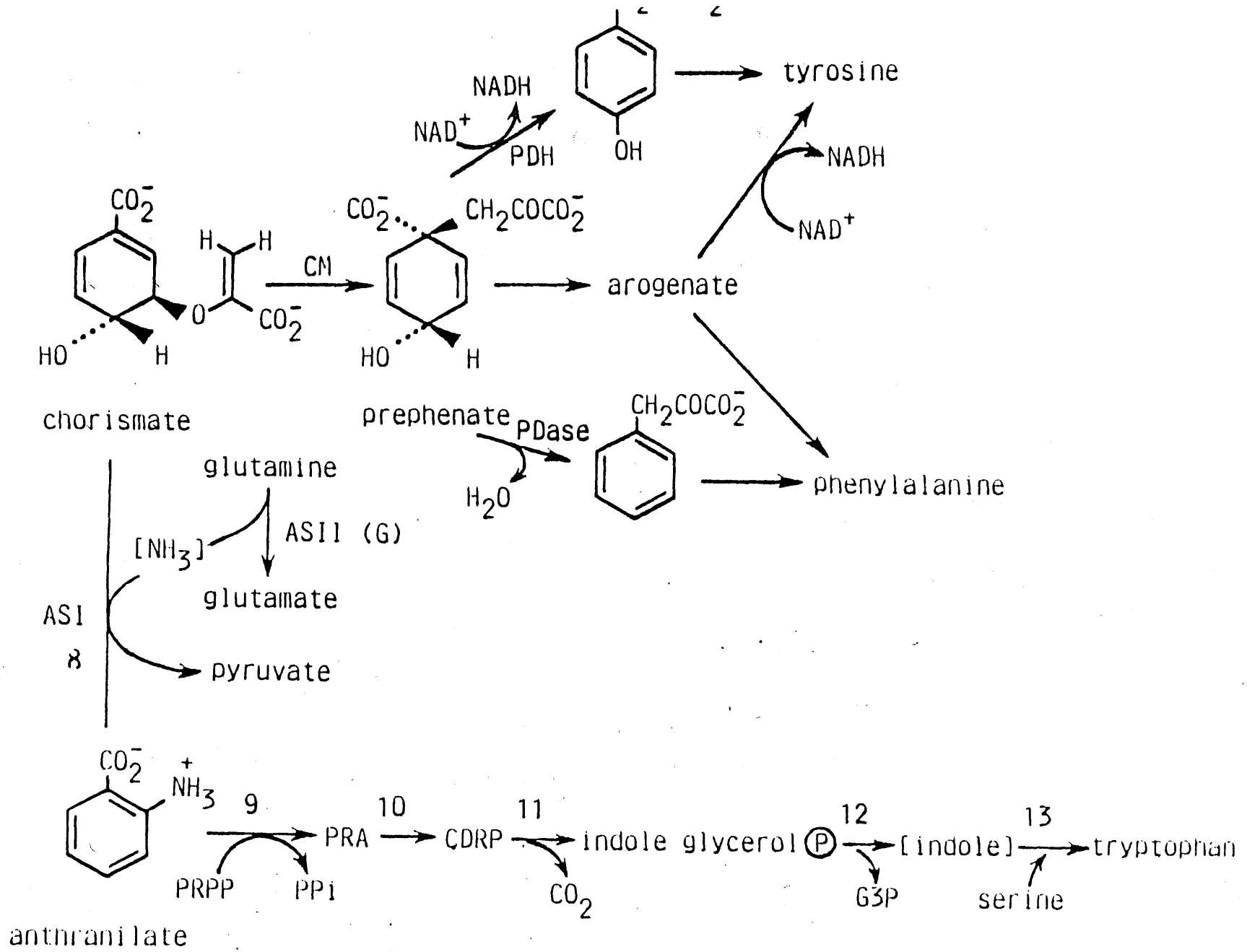
Figure 1.7: The biosynthesis of tyrosine, phenylalanine and tryptophan from chorismate.

Enzymes:

| | |
|---|---|
| ASI (8) | anthranilate synthase catalytic subunit |
| ASII (G) | anthranilate synthase glutaminase subunit |
| 9 | anthranilate phosphoribosyl transferase |
| 10 | phosphoribosyl anthranilate isomerase |
| 11 | indoleglycerol phosphate synthase |
| 12 | tryptophan synthase step A |
| 13 | tryptophan synthase step B |
| CM | chorismate mutase |
| PDH | prephenate dehydrogenase |
| PDase | prephenate dehydratase |
| D | diaphorase (NADP-dependent flavin reductase) |

Intermediates:

| | |
|---|---|
| PRA | phosphoribosyl anthranilate |
| CDRP | (O-carboxyphenylamino)-1-deoxyribulose 5-phosphate |
| PRPP | phosphoribosyl pyrophosphate |
| G3P | glyceraldehyde 3-phosphate |


tyrosine
NADH
NAD+
PDH
OH
NADH
NAD+
CO2-
CH2COCO2-
H
H
CM
O
CO2-
HO
H
HO
H
arogenate
chorismate
prephenate
PDase
CH2COCO2-
H2O
phenylalanine
glutamine
[NH3]
ASII (G)
glutamate
ASI
pyruvate
CO2-
+
NH3
9
10
11
12
13
PRA
CDRP
indole glycerol P
[indole]
tryptophan
PRPP
PPi
CO2
G3P
serine
anthranilate

1.6.3 Tryptophan

The tryptophan biosynthetic pathway provides the most striking diversity of organisation to be found. The enzymes which catalyse the reactions in this five step pathway (Figure 1.7) are organised very differently in different organisms. For example, some organisms have multifunctional enzymes for certain steps in the pathway, whereas others have monofunctional polypeptides. The combinations of enzymatic steps found associated with one another is also different in different organisms.

The organisation of the shikimate and tryptophan biosynthetic pathway enzymes is shown in Figure 1.8. The tryptophan pathway enzymes are organised in the following way.

The first reaction in tryptophan synthesis, the conversion of chorismate and ammonia to anthranilate and pyruvate, is catalysed by anthranilate synthase, which is subject to feedback regulation by tryptophan. E.coli anthranilate synthase is made up of two components: ASI, which catalyses the $NH_3$-dependent synthesis of anthranilate from chorismate and is encoded by the trpE gene, and ASII, which forms a covalent glutaminyl intermediate to deliver the amide of glutamine to ASI for anthranilate synthesis. The ASII component is fused to the second enzyme in the pathway, anthranilate phosphoribosyltransferase; this bifunctional protein chain is the product of the trpD gene. In Neurospora, the ASII component is fused, not to the second enzyme of the pathway, but to the third and fourth

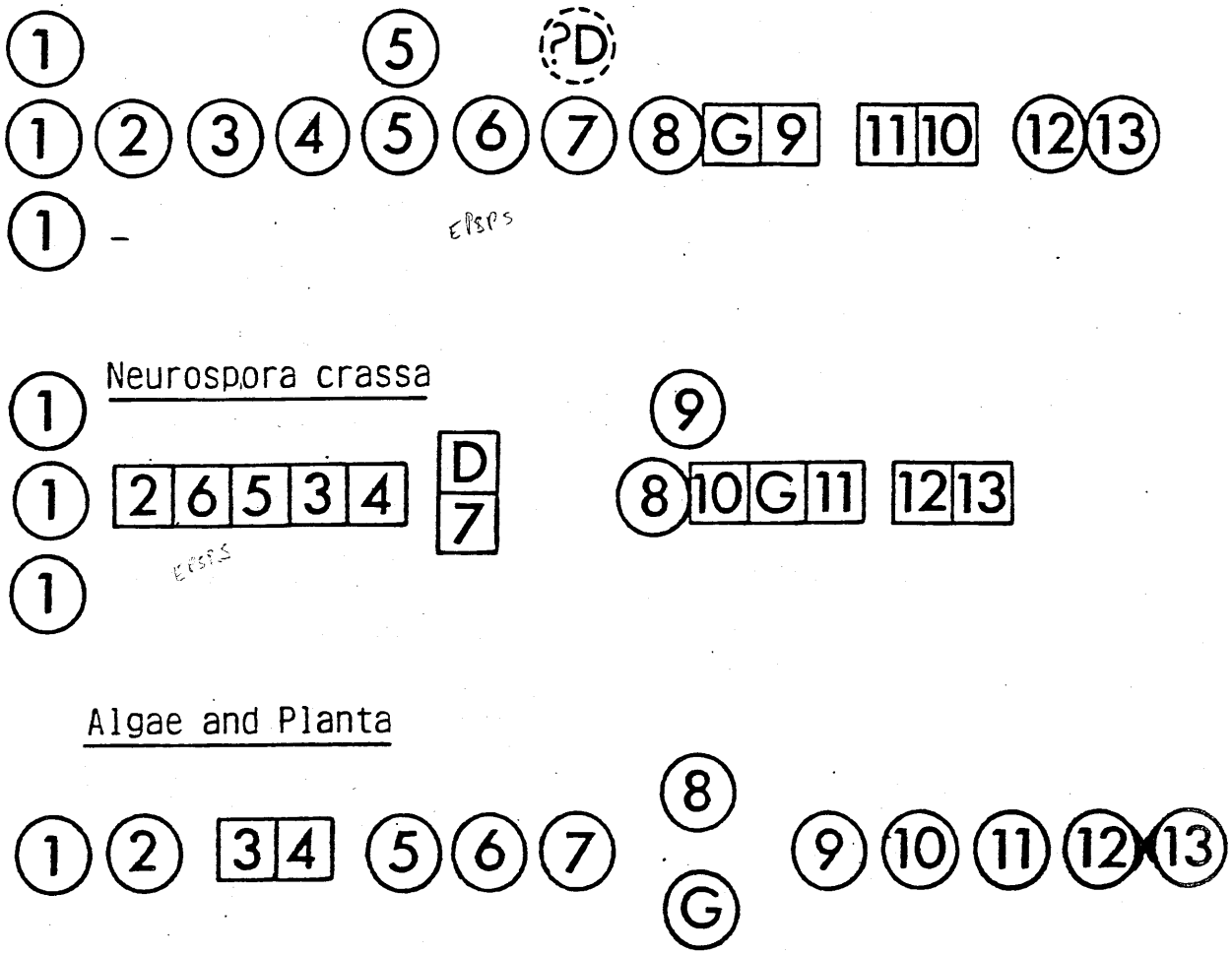



Figure 1.8: Structural organisation of enzymes of the shikimate and tryptophan pathways of a range of organisms.

The enzymes are numbered as in Figures 1.1 and 1.7.

Key: Rectangles - multifunctional polypeptides

Circles - monofunctional polypeptides

Joined circles - multienzyme complexes

enzymes of the pathway, phosphoribosyl anthranilate isomerase and indoleglycerol phosphate synthase, respectively. The yeast ASII component is fused only to indoleglycerol phosphate synthase (Zalkin, 1980). Phosphoribosyl anthranilate isomerase and indoleglycerol phosphate synthase form a bifunctional enzyme in E.coli, encoded by the trpC gene.

Tryptophan synthase catalyses the conversion of indoleglycerol phosphate to tryptophan via the enzyme bound intermediate indole. In prokaryotes, tryptophan synthase is a tetramer of the $\alpha_2\beta_2$ type, but in eukaryotes such as Neurospora or Saccharomyces, the enzyme occurs as a dimer of identical subunits, although it catalyses the same overall and partial reactions (Crawford, 1980).

In E.coli, the structural genes for the enzymes of this pathway are organised into the trp operon. A 7 kb region of DNA encodes five trp genes, which are transcribed as a polycistronic message, trp EDCBA. The primary promoter for this operon is located 200 bp upstream of trpE. Transcription of the operon is regulated by repressor-operator interactions. Trp repressor binding to the trp operator is controlled by tryptophan (Platt, 1978).

A second control process, attenuation, was first described for the trp operon by Yanofsky (1981). During transcription, RNA polymerase must traverse a 162 bp leader region before encountering the trpE gene. The leader contains a transcription terminator, and whether or not termination occurs depends on the availability of aminoacylated $tRNA^{trp}$. If $tRNA^{trp}$ is uncharged, the full length

transcripts are produced, and tryptophan synthesis can go ahead.

## 1.7 Non-aromatic amino acid biosynthetic pathways

The structural genes for the enzymes which catalyse the reactions involved in the biosynthesis of the non-aromatic amino acids show a wide diversity in organisation and regulation. Like the aromatic amino acid biosynthetic pathways, differences are found both between the organisation of the same pathway in different species and between the types of organisation of different pathways in the same organism. Examples can be found in *E.coli* of such diverse modes of organisation as scattered, constitutively expressed genes; scattered, regulated genes; operons; and many multifunctional proteins. A few examples of different modes of gene organisation are given below.

### 1.7.1 Serine biosynthesis in E.coli

The three step conversion of 3-phosphoglycerate to serine is described in detail in Chapter 6. McKitrick & Pizer (1980) have shown that the *serA*, *serB* and *serC* structural genes are constitutively expressed. The genes are scattered about the genome (Bachmann, 1983) and so this organisation is similar to that found for the shikimate pathway structural genes (other than DAHP synthase and shikimate kinase genes).

### 1.7.2 Arginine biosynthesis in E.coli

Arginine is synthesised in eight steps from glutamine. Four of the genes of arginine biosynthesis are clustered (*arg* ECBH) and the others are scattered about the genome (Bachmann, 1983). The synthesis of the *arg* biosynthetic enzymes is under arginine control; the genes are repressed by the *arg* repressor, the product of the *argR* gene (Gorini *et al.*, 1961; Mass, 1961). The genes are thus organised as a 'regulon' where the product of one regulatory gene governs the expression of unlinked but functionally related structural genes.

### 1.7.3 Histidine biosynthesis in E.coli

The most extreme case of coordinate regulation of expression of the structural genes for a biosynthetic pathway is in the *his* operon. The *his* operon contains the structural genes encoding the nine enzymes which catalyse the synthesis of histidine from phosphoribosyl pyrophosphate and ATP. Histidine biosynthesis requires the expenditure of a large amount of metabolic resources and is unnecessary under conditions where histidine is available in the growth medium. Biosynthesis is regulated by a number of effectors, including operon repression mediated by attenuation; a transcriptional control responsive to the general availability of amino acids; and feedback inhibition of the first enzyme of the pathway (Artz & Holzschu, 1983).

### 1.7.4 Coordinate regulation of gene expression in biosynthetic pathways

The modes of regulation of the structural genes on the pathways described above suggest that coordinate synthesis of the enzymes is important to the economy of the cell. Under conditions where the end product of a pathway is freely available in the growth medium there is no need to synthesise that compound and thus no requirement for the catalytic enzymes of the pathway. The substrate for the enzyme may be an intermediate which is not normally present in the cell and so synthesis of that enzyme on its own would be extremely wasteful as there would be no substrate, and/or no use for the product of the reaction. For this reason, the structural genes for most biosynthetic enzymes are regulated at the transcription level. It is likely that only the genes of those pathways, the products of which are required in large amounts, or are constantly depleted by the cell, will be constitutively expressed. But what of multifunctional enzymes? It is possible that they have evolved as a solution to this same problem. Having two or more activities on a single polypeptide chain ensures that the activities are co-ordinately expressed. It is also possible that bacterial operons represent the half-way stage between scattered genes and a multifunctional enzyme.

It is unlikely, however, that co-ordinate expression represents the only reason for the existence of multifunctional enzymes. The evolution of multifunctional enzymes is discussed in the sections which follow.

## 1.8 Evolution of multifunctional enzymes

A large number of multifunctional enzymes have been described (Kirschner & Bisswanger, 1976); they are found throughout nature, in prokaryotes, lower eukaryotes and in higher eukaryotes. How might such enzymes have evolved and is there any evidence to support the hypothesis that they confer some advantage to an organism?

Three reasons for the existence of multifunctional enzymes are:

(i) Co-ordinate expression of enzyme activities: having more than one enzyme activity on a polypeptide chain ensures that the enzyme activities are synthesised in equimolar amounts.

(ii) Co-ordinate regulation of enzyme activity, the presence of a number of enzyme activities on a single polypeptide chain, possibly along with a regulatory domain, may allow the co-ordinate allosteric control of a number of enzyme activities by a single effector molecule.

(iii) Metabolic channeling: the segregation, within the cell, of intermediates may be important in preventing an intermediate being used as a substrate on a competing biosynthetic pathway. Multifunctional enzymes might be able to achieve this by the slow dissociation of a product from the complex, providing a high steady state concentration for the next enzyme.

It has been suggested (Bonner *et al.*, 1965) that multifunctional enzymes arose by the fusion of ancestral monofunctional polypeptide chains, perhaps via a stage where the independent components were associated. This would

depend on the genes specifying the individual components being located adjacent to one another on the genome and fusion taking place at the DNA level. There are known to be two mechanisms by which the individual genes might be brought together in the genome. The first is transposition. Transposable elements are very common and have been shown to be responsible for a number of rearrangement events. Scattered ancestral genes might have been transposed into close proximity leading either to operons (common in bacteria) or, after fusion, to multifunctional enzymes (common in eukaryotes). The second mechanism is gene duplication, followed by divergence and fusion. A mutation in a duplicate copy of a gene will not have a detrimental effect on the cell, and if the mutation provides a new function then it might confer a selective advantage.

Giles (1978) has suggested that there are two routes by which multifunctional enzymes may have arisen. The first is, as suggested above, the transposition of scattered ancestral genes to a single site. A second possibility is that the ancestral organism possessed a true gene cluster - a cluster of non-contiguous, independent genes which have either been scattered as in the case of the *E.coli aro* genes or have been fused to form a multifunctional enzyme (as in *arom*), or an operon (*E.coli trp* genes).

A number of examples are given below of detailed investigations of the evolution of multifunctional enzymes which provide evidence in support of the above hypotheses.

1.8.1 Tryptophan Synthase

In E.coli and other prokaryotes, tryptophan synthase is a tetramer of the form $\alpha_2\beta_2$. Each subunit catalyses a partial reaction. The sequence of the trpA and trpB genes, encoding the α and β subunits respectively, have been determined both from E.coli and S.typhimurium (Nichols & Yanofsky, 1979; Crawford et al., 1980). The Neurospora and Yeast enzymes are homodimers which catalyse the same overall and partial reactions, but are slightly larger than E.coli α + β. The DNA sequence of the S.cerevisiae tryptophan synthase gene (TRP5) has recently been determined (Zalkin & Yanofsky, 1982).

The E.coli and yeast native enzymes are therefore similar in size and catalytic properties implying that the bifunctional enzyme has arisen by fusion of E.coli-like α and β subunits. The homology between the E.coli α and β sequences and the α and β domains of the yeast enzyme is striking. TRP5 amino acids 1-269 are equivalent to the E.coli α subunit (trpA) and amino acids 298-708 are equivalent to the E.coli β subunit (trpB). This implies a fusion in the order α followed by β, but in all prokaryotes which have been studied, the order of the structural genes in the trp operon is β followed by α. In the yeast gene there is a 28 amino acid connecting region which may reflect the DNA sequence associated with the genetic events leading to the fusion event i.e. preceding trpB or following trpA in the ancestral organism. The sequence was found to be homologous to the 3' end of E.coli trpC and the trpC-trpB intercistronic region.

In E.coli, the trpB-trpA intercistronic region has the sequence:

trpB stop

UGAUG

start trpA

A single base addition (or deletion) mutation could therefore fuse the genes and yet a bifunctional enzyme with a β-α fusion has not been found. This suggests two things. Firstly, the yeast enzyme has an advantage over a β-α fusion enzyme; the connecting region may be necessary. For example, to permit correct folding and interaction of the domains. Secondly, the fusion in yeast ensures coordinate expression of the subunits. In E.coli, it is possible that the close association of the stop and start codons allows the cell to bypass some of the normal steps in termination and initiation in order to translate the polycistronic trpB-trpA message more efficiently (Platt, 1978).

The sequence of the yeast TRP3 gene (ASII/indoleglycerol phosphate synthase) has been compared to the corresponding Neurospora gene (trp-1; ASII/phosphoribosyl anthranilate isomerase/indoleglycerol phosphate synthase). It has been shown that in each case there is a connecting region between the ancestral E.coli-like trpG-trpC sequences, but the sequence of the connector is not conserved. The connector is therefore probably essential for the structural integrity of the multifunctional enzyme, but its sequence is not important (Zalkin et al., 1984).

It is also interesting to note that in the trifunctional yeast HIS4 enzyme, the component activities are fused in the order:

HIS4A (Step 3) - HIS4B (Step 2) - HIS4C (Step 10)

In the S.typhimurium his operon the genes are linked in the order:

hisD (Step 10) ....(5 genes) ....hisI (step 3) - hisE (Step 2)

During evolution of HIS4, a rearrangement of the ancestral components must have taken place before fusion, possibly allowing the most favourable interactions between domains (Donahue et al., 1982).

### 1.8.2 Aspartokinase/Homoserine dehydrogenase from E.coli

In E.coli, the first and third reactions of the common pathway leading to threonine and methionine are catalysed by two bifunctional enzymes, Aspartokinase/Homoserine dehydrogenase, or AKI/HDHI and AKII/HDHII, encoded by the thrA and metL genes respectively. The native enzyme is a tetramer with each chain carrying on discrete domains the AK and HDH activities. The sequence of the genes has been determined (Katinka et al., 1980; Zakin et al., 1983) and comparison of the two amino acid sequences established that they are derived from a common ancestor.

Analysis of the thrA gene (Katinka et al., 1980) has revealed how the bifunctional enzyme has arisen. Met-249 in the sequence is the start of the dehydrogenase part of

the chain and is preceded by the sequence GAGGU which is homologous to the consensus E.coli ribosome binding site (Shine & Dalgarno, 1975). Also in the preceding sequence, two separate base changes could have abolished two stop codons, yielding a fused polypeptide. The formation of the fused polypeptide may have given bacteria an advantage in co-ordinately controlling both activities at the transcription and translation levels, in addition to allosteric control of both activities by threonine.

A computer study has shown that there are internal homologies within the gene (Ferrara et al., 1984). It has been postulated that there were two ancestral genes $AK_o$ and $HDH_o$ which duplicated and fused to give $AK_o$' $AK_o$ and $HDH_o$'$HDH_o$. These then fused to give $AK_o$'$AK_oHDH_o$'HDH, which is the common ancestor of thrA and metL. This is an example of both a gene duplication and fusion event followed by divergence of the ancestral function and a fusion of two different enzyme activities.

The thrA gene is a member of the thr operon, along with thrB and thrC. The thrA-thrB and thrB-thrC junctions have been studied and it has been shown that only one nucleotide separates thrA and thrB, and that the thrB and thrC genes are adjacent (Cossart et al., 1981). Only two nucleotide changes would be necessary to make the thr operon genes into a single tetrafunctional gene/enzyme. That this has not happened may reflect the necessity for connecting regions between protein domains, or it may suggest that the proximity of the genes to one another allows for enhanced

translational efficiency, as is the case with the trpB-trpA junction, and that this is the primary function of the operon, i.e. co-ordinate regulation of transcription and translation.

### 1.8.3 Fatty acid synthase

The fatty acid synthase (FAS) of mammals represents the most extreme case known of a multifunctional enzyme - seven enzyme activities on a single polypeptide chain. The mechanism of FAS is essentially the same in all organisms which have been studied and yet there is a complex variety of structural forms. FAS is ubiquitous; it is essential for the growth of membranes and the conversion of excess carbohydrate into fat. In bacteria and plants the FAS enzyme activities occur as discrete polypeptide chains but in most other species there are various combinations of multifunctional enzymes. Fungal FAS has two subunits, $\alpha$ and $\beta$, whereas most other organisms have a single subunit FAS (McCarthy & Hardie, 1984).

The simplest hypothesis to explain the evolution of FAS is a gene fusion model. The independent polypeptides fused to form the vertebrate seven-function polypeptide seen today, and the fungal two subunits represents a more primitive between-stage. This, however is not the case. There are a number of differences between the vertebrate and fungal FAS enzymes; in particular the active sites of both enzymes have been mapped, and their order along the polypeptide chain is different (McCarthy & Hardie, 1984).

The gene fusion events leading to the vertebrate and fungal FAS must therefore have occurred independently. This is an example of convergent evolution; the different FAS enzymes must confer some advantage to the organisms which possess them over the monofunctional enzymes.

### 1.8.4 Arom

Evidence has been obtained from limited proteolysis and chemical modification studies that the active sites of the arom complex are spatially distinct and that at least four of the activities can be isolated on stable proteolytic fragments of the pentafunctional polypeptide (Coggins et al, 1981; Coggins, 1982; Smith & Coggins, 1983; M.R. Boocock, Ph.D. Thesis, University of Glasgow, 1983). This suggests that the complex may have evolved by fusion of E.coli-like gene sequences. What are the advantages to fungi of the arom polypeptide?

The first, most obvious advantage is in co-ordinate expression of the activities of a number of enzymes on the same biosynthetic pathway. The second advantage might be in co-ordinate regulation of the enzyme activities. It has been suggested that the catalytic constants of the arom enzymes are altered by DAHP, the first substrate of the complex. An activation is seen in the presence of DAHP (Welch & Gaertner, 1976). The final advantage might be in the channeling of intermediates through the pathway. Giles et al. (1967a, 1967b) suggested that the function of the arom complex was to prevent the build up of DHQ in

the cell by channeling the intermediates from DAHP to EPSP down the shikimate pathway, preventing the induction of the catabolic dehydroquinase. Kinetic evidence for compartmentation of the intermediates has been obtained (Welsh & Gaertner, 1975), although independent experiments on more highly purified and more active *arom* complex do not support these results (Dr J.M. Lambert and Dr M.R. Boocock, unpublished results).

## 1.9 Aims of this project

Structural and mechanistic studies in our laboratory on the shikimate pathway enzymes from both *Neurospora* and *E.coli* have been hindered by a lack of adequate quantities of the enzymes in purified form. This is especially true for EPSP synthase, where interest in the mechanism of action of the herbicide glyphosate has stimulated the need for a three-dimensional model of the enzyme.

The contrast between the separable *E.coli* enzymes and the pentafunctional *arom* complex of *Neurospora* provides a model system for the structural and functional organisation of a biosynthetic pathway. A detailed knowledge of gene sequences might allow us to delineate possible modes of evolution leading to these very different levels of gene/enzyme organisation.

Both of these problems can be approached by cloning and DNA sequence analysis of the genes from both *E.coli* and *Neurospora*. The aim of this project was to isolate and

study a number of the E.coli genes for the shikimate pathway enzymes. This thesis describes the sub-cloning of the aroB gene (Chapter 3); the sub-cloning and DNA sequence analysis of the aroD gene (Chapter 4); the sub-cloning and DNA sequence analysis of the aroA gene (Chapter 5); and the investigation of the expression of the aroA gene (Chapter 6). Cloning of the E.coli aro genes has assisted the purification of the corresponding enzymes from overproducing strains; the study of EPSP synthase (Ann Lewendon, Ph.D. Thesis, University of Glasgow, 1984) has progressed in parallel with this work. No DNA sequence for an arom gene is yet available, and so an investigation of the evolution of this multifunctional enzyme has not been carried out. However, a limited and preliminary comparative study of the genes sequenced in this work is presented in Chapter 7.

## 1.10 Strategy employed in cloning the aro genes

The successful completion of the work described in this thesis has been due mainly to three factors, namely, the identification of the aro genes by relief of auxotrophy of suitable E.coli mutants; the close association of gene and enzyme studies; and the choice of the M13/dideoxy method of DNA sequencing. These techniques are outlined below.

### 1.10.1 Identification of cloned genes by relief of auxotrophy

At the outset, the strategy envisaged for obtaining clones of these genes was to produce a plasmid gene bank of

E.coli DNA, by random 'shotgun' cloning of genomic DNA into one of the unique restriction enzyme cleavage sites of the plasmid vector pAT153 (Twigg & Sherratt, 1980). Recombinants carrying the gene(s) of interest were to be identified by transforming suitable E.coli aro strains, and selecting for relief of auxotrophy by the ability to grow on unsupplemented medium.

It was realised, however, that for an E.coli gene, it is not always necessary to look in a genomic library of DNA fragments. A large proportion of the genome has now been characterised on either:

(a) the Clarke & Carbon (1976) ColE1 gene bank and a gene/protein index is also available for these plasmids (Neidhart et al., 1983)

(b) a number of λ-transducing phages.

If the location of a gene of interest is known, a plasmid or phage carrying the region of the chromosome near to the gene can often be found after careful searching of the literature. The aroB and aroA genes were sub-cloned from a plasmid and a phage, respectively, which were used initially in studies of neighbouring genes.

### 1.10.2 Correlation with enzymatic studies

As stated earlier, work on the E.coli genes has proceeded in parallel with and complementary to studies on the corresponding enzymes. For example, the ease with which the enzymes can be assayed spectrophotometrically in our laboratory has helped to answer a number of questions about

the expression of the genes. The amino terminal sequences of the enzymes and amino acid compositions have proved beyond doubt the location of the gene coding sequences on sequenced DNA fragments.

### 1.10.3 DNA sequencing studies

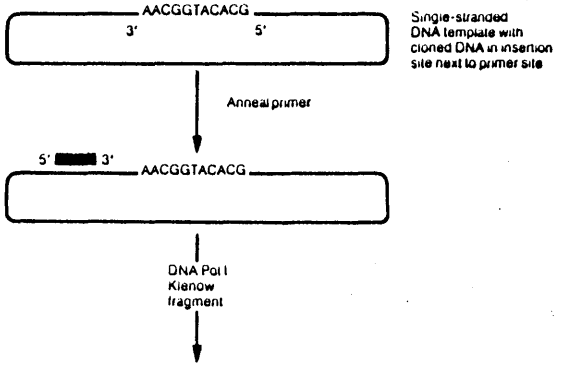
The DNA sequencing technique employed in this work was the M13/dideoxy method. The Sanger method of chain termination sequencing (Sanger, 1981) is based upon the use of deoxynucleotide analogues which are randomly incorporated into a growing DNA strand to give specific chain termination. The 'Klenow' fragment of *E.coli* DNA polymerase I is used to synthesise a complementary strand from a single-stranded template. During sequence analysis, four different reactions are carried out; each contains all four deoxynucleotides (one of which is radioactively labelled), but a different dideoxynucleotide, at a low concentration. Since dideoxynucleotides lack a 3'-hydroxyl group, the chain can no longer be extended beyond the point of incorporation and the end result of each reaction is a family of DNA molecules, each starting at the same point (the primer) and ending at a different dideoxynucleotide. The fragments can then be separated by electrophoresis on polyacrylamide gels and detected by autoradiography. This is summarised in Figure 1.9a which is taken from the Amersham "M13 cloning and sequencing handbook".

Figure 1.9: The M13/dideoxy method of DNA sequencing.

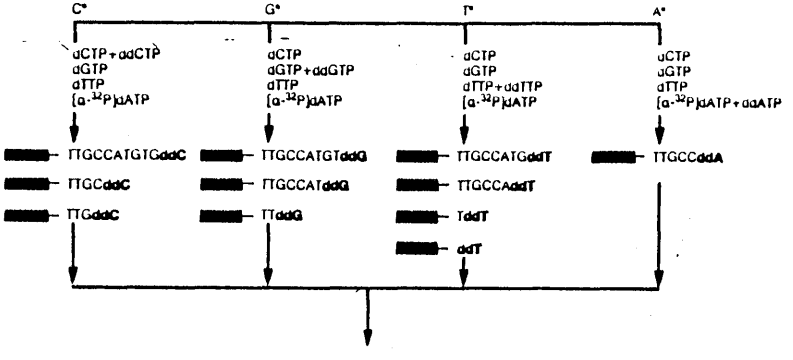
(a) The dideoxy sequencing reaction, beginning with single-stranded template DNA

(b) The life cycle of phage M13

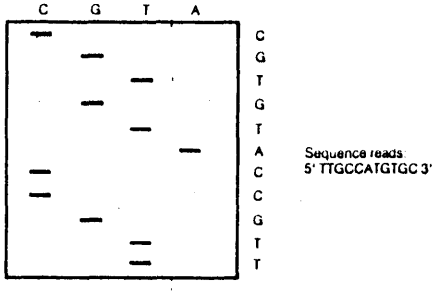
## The annealing reaction

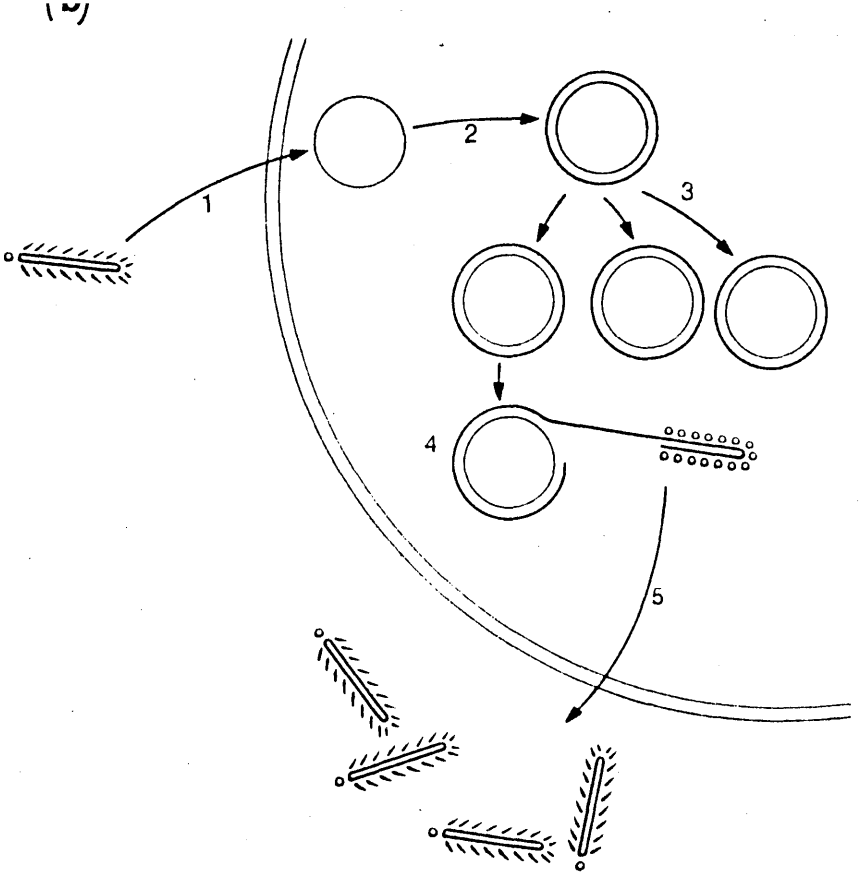



## The sequencing reaction




## Gel electrophoresis and autoradiography




**1)** Phage enters bacterial host cell via pilus
**2)** Single-stranded DNA converted to double- stranded RF (replicative form)
**3)** Replication of DNA to give progeny RF (100+ per cell)
**4)** Single strand synthesis by rolling ball method
**5)** Single-stranded DNA is processed and packaged into mature virus and extruded from cell (200+ per cell generation)

M13 sequencing and cloning takes advantage of this system. M13 RF DNA is used as a cloning vector. On introducing this into a cell single-stranded copies are made, and hence of any DNA cloned into it. This is a simple way of producing large amounts of the single-stranded template necessary for 'dideoxy sequencing'

This sequencing method depends on a supply of pure single-stranded template DNA. The life cycle of the single-stranded filamentous phage M13 can be exploited to prepare template DNA (Figure 1.9b). DNA for sequencing can be cloned into the double stranded replicative form of the phage in a series of vectors developed by Messing (1983), prior to preparation of template for the sequencing reaction.

# CHAPTER 2 MATERIALS AND METHODS

## 2.1 Materials

All reagents, with the exception of the speciality reagents mentioned in the text, were purchased from BDH Chemicals, Poole, Dorset.

Suppliers of speciality reagents:

Amersham - Amersham International, Amersham, Buckinghamshire

Boehringer - BCL, Lewes, East Sussex

BRL - Bethesda Research Laboratories, Paisley, Scotland

Difco - Difco Laboratories, Detroit, USA

Sigma - Sigma Chemical Company, Poole, Dorset

## 2.2 Bacterial strains

The bacterial strains used in this work are all derivatives of *E.coli* K12. They are listed in Table 2.1, along with the markers carried on the strains, and the source of the strain.

Stocks of each strain, and of plasmid carrying strains were maintained as follows: A 10 ml LB culture was inoculated with a single colony isolate of the strain and grown overnight at 37° with shaking. A 1 ml sample was taken and tested for either markers carried by the strain or for a plasmid present within the strain. 9 ml of sterile 80% glycerol was added to the remainder of the culture and after thorough mixing was placed at -20° for long term storage.

| Strain | Genotype | Origin/Reference |
|---|---|---|
| E.coli K12 | wild type ATCC 14948<br>$F^-$, λ lysogenic derivative of K12; Lederberg strain W3100. | American Type Culture Collection (Rockville, Maryland USA) |
| E.coli AB2826 | aroB, supE42, $\lambda^-$ | CGSC (E.coli Genetic Stock Centre Dept. of Human Genetics Yale University New Haven, USA) |
| E.coli AB2848 | aroD352, tsx356, supE42, $\lambda^-$ | CGSC |
| E.coli AB2829 | aroA354, supE42, $\lambda^-$ | CGSC |
| E.coli KL282 | serC13, (serS14, serS16), phoA4, relA1, tonA22, $T^R_2$, supD32 | CGSC |
| E.coli JM101 | Δ(lac pro), thi, supE, F'traD36, proAB, $lacI^q$, lacZΔM15 | Messing et al. (1981) |
| E.coli HB101 | $F^-$, pro, leu, thi, lacY, hsdR, endA, recA, rpsL20, ara14, galK2, xyl5, mtl1, supE44 | Bolivar & Backman (1979) |

Table 2.1: E.coli strains described in this study.

## 2.3 Plasmids and phage described in this study

The plasmids and phage described in this study are shown in Tables 2.2 and 2.3. Table 2.2 lists those plasmids and phage which have been used as vectors, or source material; Table 2.3 describes the plasmids constructed during the course of this work.

## 2.4 Growth media

The growth media used in the course of this work are listed in Table 2.4. All media were sterilised by autoclaving; (+ ' x ') indicates that this item was added after autoclaving.

In making MM plates, the M9Salts and agar were autoclaved separately, as 2 x M9 salts and 30% agar. Then equal volumes were mixed before plating.

Any supplements to plates were added as concentrated stock solutions after the medium had cooled to 55°, and immediately before pouring.

Throughout the text, the notation LB/amp is used to indicate L broth supplemented with ampicillin at 50 µg/ml.

## 2.5 Supplements to growth media

Supplements to growth media are listed in Table 2.5 or in the text of the results chapters. They were sterilised by autoclaving, where possible, or by passage through a 0.22 µM filter (Millipore).

| Plasmid | Markers | Origin/Reference |
|---|---|---|
| pAT153 | *amp*$^R$, *tet*$^R$ | Twigg & Sherratt (1980) |
| pBR322 | *amp*$^R$, *tet*$^R$ | Sutcliffe (1977) |
| pKK223-3 | *amp*$^R$ | J. Brosius, unpublished (Dr J. Knowles) |
| pLC29-47 | *aroB*$^+$, *ponA*$^+$ | Takeda *et al.* (1981) (Dr S.R. Kushner) |
| *Phage* | | |
| $\lambda$*pserC* | *serC*$^+$, *aroA*$^+$, *rpsA*$^+$ | Kitakawa *et al.* (1980) (Dr K. Isono) |

*Table 2.2*: Plasmid vectors; plasmid and phage source material.

| Plasmid | Markers | DNA inserted at | Vector |
|---|---|---|---|
| pKD101 | *aroB*$^+$ *tet*$^R$ | Pst I | pAT153 |
| pKD102 | *aroB*$^+$ *tet*$^R$ | Pst I | pAT153 |
| pKD103 | *aroB*$^+$ *tet*$^R$ | Pst I | pAT153 |
| pKD104 | *aroB*$^+$ *tet*$^R$ | Pst I | pAT153 |
| pKD105 | *amp*$^R$ | BamHI | pAT153 |
| pKD106 | *aroB*$^+$ *amp*$^R$ | EcoRI - BamHI | pAT153 |
| pKD201 | *aroD*$^+$ *amp*$^R$ | Cla I | pAT153 |
| pKD202 | *amp*$^R$ | Cla I | pAT153 |
| pKD501 | *aroA*$^+$ *serC*$^+$ *tet*$^R$ | Pst I | pAT153 |
| pKD502 | *aroA*$^+$ *serC*$^+$ *tet*$^R$ | Pst I | pAT153 |
| pKD503 | *aroA*$^+$ *serC*$^+$ *amp*$^R$ | BamHI - Hind III | pAT153 |
| pKD504 | *aroA*$^+$ *serC*$^+$ *amp*$^R$ | BamHI - AvaI | pAT153 |
| pKD505 | *aroA*$^+$ *amp*$^R$ | ClaI - AvaI | pAT153 |
| pKD506 | *aroA*$^+$ *amp*$^R$ | ClaI - PvuII | pBR322 |
| pKD508 | *aroA*$^+$ *serC*$^+$ *tet*$^R$ | PstI | pAT153 |

Table 2.3: Plasmids constructed during the course of this work.

| Medium | Composition, per litre | (final concentration) |
|---|---|---|
| L-Broth (LB) | 10 g bactotryptone (Difco)<br>5 g yeast extract (Difco)<br>10 g NaCl<br>(+ 5 ml 20% glucose) | <br><br><br>(0.1%) |
| L-Agar (LA) | as LB, + 15 g agar (Difco) | |
| Minimal medium | 1 g $NH_4Cl$ | (0.02 M) |
| (M9 Salts, MM) | 0.13 g $MgSO_4$ 7 $H_2O$ | (0.5 mM) |
| | 3 g $KH_2PO_4$ | (0.02 M) |
| | 6 g $Na_2HPO_4$ | (0.04 M) |
| | (+ 1 ml 100 mM $CaCl_2$) | (100 $\mu$M) |
| | (+ 2 ml thiamine (1 mg/ml)(Sigma) | (2 $\mu$g/ml) |
| | (+ 10 ml 20% glucose) | (0.2%) |
| MM plates | as MM, + 15 g agar | |
| 2TY | 16 g bactotryptone<br>10 g yeast extract<br>5 g NaCl | |
| H plates | 10 g bactotryptone<br>8 g NaCl<br>12 g agar | |
| H top agar | 10 g bactotryptone<br>8 g NaCl<br>8 g agar | |

Table 2.4: Growth media.

| Supplement | Concentration (μg/ml) | Stock solution |
|---|---|---|
| Ampicillin | 50 μg/ml | 25 mg/ml |
| Tetracycline | 20 μg/ml | 10 mg/ml in ethanol |
| Chloramphenicol | 170 μg/ml | 34 mg/ml in ethanol |
| Aromatics: | | |
| (i) phenylalanine | 300 μg/ml | |
| (ii) tryptophan | 150 μg/ml | |
| (iii) tyrosine | 300 μg/ml | |
| (iv) p-aminobenzoate | 1.2 μg/ml | 300 μg/ml |
| (v) p-hydroxybenzoate | 1.2 μg/ml | 300 μg/ml |
| Serine | 20 μg/ml | |
| Pyridoxine | 0.075 μg/ml | 75 μg/ml |

Table 2.5: Supplements to growth media.

## 2.6 Preparation of plasmids (small scale)

For small scale ('mini-prep') plasmid preparations, the method of Holmes & Quigley (1981) was followed. Lysis buffer is: 10mM-tris (pH 8.0), 50mM -EDTA, 8% sucrose, 0.5% Triton X-100. Lysozyme was prepared as a stock solution (10 mg/ml in TE) and stored in small aliquots at -20°. TE buffer is 10mM-tris (pH 8.0), 2mM-EDTA.

### 2.6.1(a) 1 ml Preparations

1 ml of plasmid carrying cells in medium was placed in a 1.5 ml microfuge tube and the cells harvested by centrifugation (1 min). The pellet was resuspended in 350 µl lysis solution and 25 µl lysozyme (10 mg/ml) was added. The tube was placed in a boiling water bath for 45 s and then spun for 5 min. The supernatant was removed to a fresh tube and the plasmid DNA recovered by the addition of an equal volume of isopropanol, followed by chilling at -70° (dry ice/methanol) for 10 min and centrifugation (10 min). The plasmid DNA pellet was resuspended in 30 µl TE. 5 µl was sufficient for a single restriction enzyme digest (approximately 1 µg of DNA).

### 2.6.2(b) 10 ml Preparations

As above, except that: cells harvested by centrifugation in an MSE18 centrifuge (9 000 x g, 2 min); resuspension volume was 700 µl lysis solution + 50 µl 10 mg/ml lysozyme; plasmid DNA was resuspended in 50 µl TE.

A large amount of RNA is present in these preparations which obscures the lower part of an agarose gel. If bands with lengths 500 bp were to be visualised, the preparation was treated with RNAse A (2.5 µl of 1 mg/ml stock; 37°, 15 min) before analysis or storage.

## 2.7 Preparation of plasmids (large scale)

The method used is described in Maniatis et al. (1982), and is a modification of the alkali lysis technique of Birnboim & Doly (1979).

### 2.7.1(a) Growth of plasmid carrying cells

A single plasmid carrying colony was inoculated into 10 ml LB/antibiotic and shaken overnight. 0.1 ml of this culture was used to inoculate 25 ml LB/antibiotic and the flask shaken till late log phase was reached ($A_{650} = \sim 0.8$). The whole 25 ml culture was then poured into 500 ml LB/antibiotic (prewarmed to 37°) and shaken exactly 2½ h. Then 2.5 ml of chloramphenicol (34 mg/ml in ethanol) was added (170 µg/ml final concentration) and incubation continued a further 16-20 hours.

The cells were then harvested and washed in 100 ml ice-cold STE (0.1M - NaCl, 10mM-tris (pH 7.8), 1mM-EDTA).

### 2.7.2(b) Alkali lysis

The bacterial pellet was resuspended in 10 ml 50mM-glucose, 25mM-tris (pH 8.0), 10mM-EDTA and 1 ml lysozyme (50 mg/ml) added. The solution was transferred to a 100 ml

polycarbonate tube and 20 ml 0.2N - NaOH, 1% SDS was added and mixed. After standing the tube on ice for 10 min, 15 ml of 5M-potassium acetate (pH 4.8) was added, mixed, and the tube placed on ice a further 10 min. The cell DNA and debris was pelleted by centrifugation (MSE Prepspin 50, 30 000 x g, 30 min). The supernatant was divided between two 50 ml polycarbonate tubes and 15 ml isopropanol added to each. After standing at room temperature for 15 min, the DNA was recovered by centrifugation (MSE 18, 12 000 x g, 30 min). The pellet was washed with 70% ethanol then dried briefly under vacuum. The DNA was resuspended in 8 ml TE (10mM-tris pH 8.0, 2mM-EDTA), before treatment with RNAse. RNAse A (pre-treated at 100° for 10 min) and RNAse $T_1$ (pre-treated at 80° for 10 min) were added to a final concentration of 100 µg/ml and 100 U/ml respectively and incubated for 1 h at 37°.

2.7.3(c) Equilibrium centrifugation in CsCl - ethidium bromide gradients

The volume of the DNA solution was carefully measured. 1 g of solid CsCl was added for every ml of solution and the contents mixed to dissolve the salt. The volume was again measured exactly. Ethidium bromide (10 mg/ml) was added (0.8 ml per 10 ml CsCl solution). The final density was 1.55 g/ml and the ethidium bromide concentration 600 µg/ml.

The solution was transferred into a 10 ml polycarbonate tube, suitable for a Beckman Ti50 rotor and centrifuged in

a Beckman L8-M, 45 000 rpm, 20°, for 48 h.

The band containing the plasmid DNA was viewed under long-wave UV light and removed by pumping through a needle inserted through the top of the tube to just below the band. Ethidium bromide was removed by extraction with 1-butanol saturated in TE buffer, till all the pink colour had been extracted. The aqueous DNA solution was dialysed against three changes, each of 1 000 x sample volume, of TE buffer. The DNA concentration was determined by reading the $A_{260}$ ($A_{260}$ = 1.0 ≡ 50 µg/ml DNA) and the concentration of the solution adjusted to 1 mg/ml by the addition of TE buffer.

## 2.8 Preparation of E.coli K12 genomic DNA

DNA was prepared from E.coli K12 as follows:
A single colony of E.coli K12 was inoculated into 100 ml LB and shaken at 37° till the $A_{650}$ reached 0.7. The cells were harvested, washed with 50 ml TE, then resuspended in 2 ml 0.15M-NaCl, 0.1M-EDTA. 0.2 ml lysozyme (10 mg/ml) was added and the tube incubated at room temperature for 15 min. It was then frozen to -70° (dry ice/methanol). 12.5 ml of 0.1M NaCl, 0.1M-tris (pH 8.0), 1% SDS was added and the tube placed in 65° water bath till the frozen material thawed. The tube was again frozen to -70° and thawed to 65°. The sample was kept at 65° for several days to dissolve the DNA; very gentle inversion of the tube at intervals assisted this process.

The volume was estimated and 1.282 g CsCl/ml DNA solution added. Using a refractometer, the concentration

of the solution was adjusted till the refractive index was 1.39970 (1.70 g/ml). The solution was centrifuged to equilibrium (Prepspin 50, 10 x 10 ml Rotor, 190 000 x g, 64 h) and the gradient fractionated by taking samples down the tube with a wide-bore pipette. DNA containing fractions were pooled and dialysed against TE buffer.

## 2.9 Preparation of λpserC

The method used was modified from that described in detail in Maniatis *et al*. (1982).

### 2.9.1(a) Preparation of phage

A single colony of *E.coli* KL282, lysogenic for λpserC was inoculated into 100 ml LB and shaken overnight at 30°. Two 500 ml batches of LB, prewarmed to 30°, were then inoculated with sufficient of the overnight culture to give a starting $A_{650}$ of 0.05. The flasks were shaken at 30° till the $A_{650}$ reached 0.5. The culture was induced by incubation at 45° in a shaking water bath. The induced culture was shaken a further 4 h at 37°. 10 ml of chloroform was added to each flask and shaking continued a further 30 min.

The lysed cultures were cooled to room temperature and pancreatic DNAse and RNAse (Sigma) were added, both to a final concentration of 1 μg/ml. After a 30 min incubation, solid NaCl was added to 1 M final concentration and the flasks allowed to stand for 1 h. Debris was removed by centrifugation (11 000 x g, 10 min). Solid polyethylene

glycol (PEG 6 000) was added to 10% w/v and dissolved by slowly stirring on a magnetic stirrer. Phage particles were precipitated by standing the flask on ice for 1 h. The precipitate was recovered by centrifugation (11 000 x g, 10 min). The phage pellet was then resuspended in 16 ml SM(50mM-tris (pH 7.5) 0.1M-NaCl, 7.5mM-$MgSO_4$, 0.01% gelatin); to this was added 16 ml chloroform and the two solutions vortex mixed. The phases were separated by centrifugation (1 600 x g, 15 min) and the aqueous phase retained. 0.5 g/ml of solid CsCl was added and the phage suspension layered on top of a CsCl step gradient, prepared with CsCl solutions of 1.45, 1.50 and 1.70 g/ml densities. Gradients were centrifuged in a Prepspin 50 6 x 14 ml swing-out rotor at 33 000 rpm for 2 h.

The phage particles were seen as a band at the inter-phase of the 1.45 and 1.50 g/ml layers; the band was removed with a pasteur pipette. CsCl solution (1.5 g/ml) was added to the phage suspension to make up to the volume necessary for a polycarbonate tube for a Prepspin 50 10 x 10 ml rotor. The suspension was then centrifuged at 40 000 rpm for 36 h and the phage band recovered. Phage were then stored at $4^{o}$ in the CsCl solution.

2.9.2(b) Extraction of phage DNA

1 ml of purified phage suspension was dialysed against two changes of 1 l of 10mM-NaCl, 50mM-tris (pH 8.0), 10mM-$MgCl_2$. The following were added (final concentration in brackets):

| | |
|---|---|
| EDTA | (20 mM) |
| proteinase K (Boehringer) | (50 µg/ml) |
| SDS | (0.5%) |

Incubation was for 1 h at 65°.

The DNA was extracted by phenol and chloroform, transferring the aqueous phase with a wide-bore pipette. It was then dialysed against three 1 000-fold volumes of TE.

## 2.10 Restriction enzyme digestions

Restriction enzymes will tolerate a wide variation in the conditions under which they will cleave DNA, although with a possible reduction in activity under non-ideal conditions. For this reason, five buffers were used, according to the conditions recommended by the manufacturer. The composition of four of the buffers is shown below:

| | | high salt | medium salt | low salt | SmaI |
|---|---|---|---|---|---|
| NaCl | (mM) | 100 | 60 | 0 | - |
| tris (pH 7.5) | (mM) | 50 | 10 | 10 | 10 (pH 8.0) |
| $MgCl_2$ | (mM) | 10 | 10 | 10 | 10 |
| DTT | (mM) | 1 | 1 | 1 | 1 |
| KCl | (mM) | - | - | - | 20 |

SmaI buffer was used for SmaI only. The fifth buffer was 'Core Buffer' (BRL); a number of enzymes use this standard buffer.

Restriction enzyme digests were routinely carried out in a final volume of 20 µl, but larger volumes were also used, where appropriate. A typical digest contained:

| | |
|---|---|
| DNA (0.5 - 1.0 µg) | 1 - 5 µl |
| 10 x restriction enzyme buffer | 2 µl |
| 10 x BSA | 2 µl |
| water | to 20 µl |

The BSA stock concentration (10 x) was 1 mg/ml (BRL, nuclease-free).

In multiple digests where a different buffer was required, the lowest salt buffer digestion was carried out, then the higher salt buffer + second enzyme was added, and digestion carried out for a further period.

Digests were routinely incubated with the manufacturer's recommended quantity of restriction enzyme at 37° for 1 h, except for SmaI (30°) and TaqI (55° oven).

## 2.11 Phenol/Chloroform extraction of DNA and ethanol precipitation

DNA was routinely purified free of enzymes by phenol/chloroform extraction and ethanol precipitation. Phenol was redistilled before use, saturated with TE buffer and stored at -20°.

Phenol/chloroform extractions were performed as follows: volume of sample to be extracted was adjusted to 50 µl with TE, if necessary; 50 µl phenol was added and vortex mixed; centrifuged 1 min; upper aqueous layer transferred to fresh microfuge tube and phenol extraction repeated;

500 µl chloroform added and vortex mixed; centrifuged 5 s; aqueous layer transferred to fresh tube and chloroform extraction repeated.

Ethanol precipitation: 1/10 volume of 2.5M-Na acetate pH 5.2 and 2½ volumes of ethanol added and vortex mixed; placed -20° overnight, or -70° (dry ice/methanol) for 15 min; centrifuged 10 min.

## 2.12 Gel electrophoresis of DNA

DNA and DNA fragments were separated by gel electrophoresis on agarose or polyacrylamide, as follows:

| size range: | gel concentration: |
|---|---|
| 10 - 100 kb | 0.5% agarose |
| 0.5 - 10 kb | 1.0% agarose |
| 0.5 - 5 kb | 1.2% agarose |
| 0.1 - 2 kb | 2.0% agarose or 5% polyacrylamide |

### 2.12.1(a) Agarose gel electrophoresis

Agarose (BRL) gels were prepared by heating to boiling point in a reflux apparatus the desired quantity of tris-borate buffer (89mM-tris, pH 8.0, 89mM-borate, 2mM-EDTA) containing the appropriate concentration of agarose. The molten agarose was allowed to cool to 55°, ethidium bromide (10 mg/ml) was added to give a final concentration of 0.5 µg/ml, and the gel poured. Gels were run immersed in the same buffer containing ethidium bromide (0.5 µg/ml) at 100 - 150 mA.

DNA samples were prepared for electrophoresis by the addition of 1/10 volume of 'dye loading buffer' (30% sucrose, 0.1% bromophenol blue, in 1 x tris borate).

2.12.2(b) LMT gels (low melting temperature)

LMT gels were prepared exactly as described above, except that they were poured and run in the cold room. LMT agarose was purchased from BRL.

2.12.3(c) Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis was carried out as described in Maniatis et al. (1982).

2.13 Photography of gels

DNA was visualised by ethidium bromide fluorescence on a transilluminator (UV Products, Inc.). Gels were photographed with a Polaroid CU-5 camera and type 665 positive/negative film.

2.14 Recovery of DNA fragments from LMT gels

DNA fragments were recovered from LMT gels for sub-cloning free of other fragments in a restriction digest, or for further sub-digestion of a specific region of a plasmid. These two procedures differ in the degree of purity of DNA required.

2.14(a) Recovery of fragments for sub-cloning

Efficient sub-cloning requires purified DNA fragment for ligation to vector DNA. The fragment of interest was excised from a gel and placed in a 1.5 ml microfuge tube. 5 volumes of LMT extraction buffer (20mM-tris (pH 8.0), 1mM-EDTA) were added (100-200 µl) and the tube heated at 65° for 15 min. An equal volume of phenol was added, vortex mixed, and the phases separated by centrifugation (2 min). The aqueous layer was removed to a fresh tube and the phenol layer was back-extracted with 50 µl TE. The aqueous phases were combined and the phenol extraction repeated two times. This was followed by three 1 ml chloroform extractions and the DNA was recovered by ethanol precipitation.

2.14.2(b) Preparation of gel slices for further digestion

In this case, the DNA was prepared for digestion, without further purification. Gel slices, containing the DNA fragment of interest were placed in a 1.5 ml microfuge tube and 1 ml ice-cold water added. The tube was kept on ice for 1 h, with one change of water, then changed to 1 x restriction enzyme buffer for a further 1 h. The buffer was removed and the tube heated at 65° for 10 min, before being quickly transferred to a 37° bath. 10 x restriction enzyme buffer, 10 x BSA and water were added to adjust the components to the correct concentrations. Restriction enzyme (approximately three times more than usual) was added and digestion carried out for 1-2 h.

## 2.15 Phosphatase treatment of DNA

Phosphatase treatment of DNA removes the 5'-phosphate groups, preventing self-religation, and thereby increasing the yield of recombinants.

Plasmids were treated with phosphatase as follows. The plasmid DNA was cleaved with restriction enzyme and purified by phenol/chloroform extraction and ethanol precipitation. DNA was resuspended in:

| | |
|---|---|
| 10 x CIP buffer | 5 µl |
| water | to 50 µl |

(CIP buffer is 0.5mM-tris (pH 9.0), 10mM-$MgCl_2$, 1mM-$ZnCl_2$)
To this was added calf-intestinal-phosphatase (Boehringer) and samples were incubated as follows:

5' - protruding ends - 0.05 U, 37°, 30 min then a further 0.05 U, 37°, 30 min.

blunt ends ) 0.05 U, 37°, 15 min then 56°, 15 min
) followed by a further
3' - protruding ends ) 0.05 U, 37°, 15 min then 56°, 15 min

40 µl of water, 10 µl of 10 x STE and 5 µl of 10% SDS were added, and the tube was then heated to 65° for 10 min to inactivate the phosphatase and the DNA recovered by phenol/chloroform extraction and ethanol precipitation.

## 2.16 Ligation of DNA fragments

Ligation reactions were carried out in mixes containing the following:

| | |
|---|---|
| insert DNA | 2 - 5 µl (20ng) |
| vector DNA | 2 - 5 µl (100ng) |
| 10 x ligase reaction buffer | 2 µl |

ATP (5mM) 2 µl

dithiothreitol (100mM) 2 µl

water to 20 µl

The DNA ligase (BRL) was added, and incubated for:

sticky ends 0.1 U, $14^{o}$, 4 - 6 h

blunt ends 0.5 U, $0^{o}$, 16 - 24 h

The amount of DNA in the ligation reaction was adjusted to meet a target molar ratio of 5:1 insert ends:vector ends.

## 2.17 Transformation of E.coli

### 2.17.1(a) Preparation of competent cells

The transformation protocol was modified from the method of Dagert & Ehrlich (1979). This method involves the treatment of logarithmically growing cells with $CaCl_2$ and storage of the transformation competent cells on ice for 24 h before use. The authors claimed that this treatment increases the proportion of competent cells approximately 20-30 fold, despite a reduction in cell viability. Experiments with the strains most frequently used in this work (E.coli AB2826, AB2829, and AB2848) showed that the increased proportion of competent cells gained on overnight storage was matched by the decrease in viability of the cells. For this reason, competent cells were routinely made on the day they were needed. The transformation efficiency of the strains was optimised, yielding $5 \times 10^5$ - $1 \times 10^6$ transformation/µg covalently closed circular pAT153 DNA.

Competent cells were prepared from a fresh overnight culture of E.coli grown in 10 ml LB. 0.5 ml of this culture was inoculated into 50 ml LB and shaken vigorously at 37°. When the $A_{650}$ reached 0.3 - 0.5, the cells were placed on ice (10 min) then harvested (MSE18, 7 000 x g, 5 min). All subsequent steps were performed on ice and the cells handled carefully (i.e. no vortexing) to avoid damage of the fragile $CaCl_2$-treated cells. The cells were resuspended in 20 ml ice-cold 50mM-$CaCl_2$ and kept on ice 20 min. They were then harvested and resuspended in 0.5 ml 50mM-$CaCl_2$.

2.17.2(b) Transformation

100 µl of competent cells was added to DNA in a sterile tube. Ligation mixes were used without further treatment. The DNA/cell mixture was incubated on ice (15 min) then placed in a 37° water bath. After 5 min, 2 ml LB (pre-warmed to 37°) was added and incubation continued for 1½ h, before plating.

2.18 Preparation of crude extracts of E.coli

E.coli crude extracts were prepared from 100 ml cultures, grown in 2 x 50 ml batches (for aeration in 250 ml flasks). Minimal medium, minimal medium supplemented with aromatics, or LB was used. 50 µl of -20° glycerol stock cells was inoculated into 10 ml LB (+ antibiotic for plasmid selection, if present) and shaken overnight at 37°. A 1 ml sample was taken for confirmation of plasmid, if present. Each 50 ml batch of medium was inoculated with 0.1 - 1.0 ml of

overnight culture, depending on the strain and/or amount of growth required and the culture shaken at 37°. Cultures were harvested during late-log phase ($A_{650}$ = 0.8 - 0.9) or stationary phase ($A_{650}$ = 1.3 - 1.5).

The cells were washed with sonication buffer and resuspended in 5 ml of the same buffer. The two buffers used were:

sonication buffer 'A' = 200mM-KCl, 200mM-$KH_2PO_4$, 2mM-$MgCl_2$, 1mM-β-mercaptoethanol (pH 7.0)

sonication buffer 'B' = 200mM-tris, 200mM-KCl, 1mM-EDTA, 1mM-dithiothreitol (pH 7.5).

The cells were then disrupted by 3 x 30s bursts of sonication, with a 30s cooling period between (Dawe soniprobe Type 7532A; output 2, 80W).

The extracts were then centrifuged (MSE Prepspin 50; 200 000 x g, 2 h) to remove debris and membrane bound enzymes.

## 2.19 Enzyme assays

The instrument used for all spectrophotometric assays was a Gilford-Unicam model 252 spectrophotometer (chart full scale 0.1A or 0.05A). Assays (total volume 1 ml in 1 cm path length cells) were all conducted at 25°. One unit of enzyme activity is defined as the amount of enzyme that catalyses the conversion of 1 µmol substrate/min. Specific activity is defined as the units of activity / mg protein.

### 2.19.1(a) 3-dehydroquinase

3-dehydroquinase was assayed in 0.1 M-potassium phosphate buffer, pH 7.0. Extract was added to the cuvette and a

baseline $A_{234}$ obtained. The reaction was initiated by the addition of 3-dehydroquinate to a final concentration of 0.2 mM. The ammonium salt of 3-dehydroquinate was a gift from Dr S. Chaudhuri.

2.19.2(b) 3-dehydroquinate synthase

3-Dehydroquinate synthase was assayed by coupling to the 3-dehydroquinase reaction and measuring the rate of change of $A_{234}$. The final concentrations of the components in the assay mix were:

| | |
|---|---|
| glycine/KOH pH 8.4 | 100 mM |
| $CoCl_2$ | 200 μM |
| $NAD^+$ | 50 μM |
| DAHP | 400 μM |
| partially purified *E.coli* 3-dehydroquinase | 0.02 U |

The measured change in reaction rate was initiated with DAHP and the baseline rate at 234 nm was subtracted to give the true rate. DAHP and partially purified *E.coli* 3-dehydroquinase were gifts from Ms A. Coia.

2.19.3(c) EPSP synthase

EPSP synthase activity was measured in the reverse direction by coupling the formation of PEP to the pyruvate kinase and lactate dehydrogenase reactions. The assay mix contained (final concentration):

| | |
|---|---|
| potassium phosphate pH 7.0 | 100 mM |
| ADP | 2.5 mM |
| $MgCl_2$ | 2.5 mM |
| NADH | 0.1 mM |
| EPSP | 50 µM |
| PK/LDH | 2.5 µl (Boehringer) |

The rate of change of $A_{340}$ was measured and a blank rate (due to NADH oxidase) obtained. EPSP was added to the cuvette to initiate the reaction and the change in the rate of decrease in $A_{340}$ measured. EPSP was a gift from Ms A. Lewendon.

2.19.4(d) PSAT

The assay for PSAT activity is based on that described in Hirsch-Kolb & Greenberg (1971), and modified by Ann Lewendon. Assay mixes (1 ml) contained:

| | |
|---|---|
| tris (pH 8.2) | 50 mM |
| $NH_4$ acetate | 32 mM |
| glutamate | 2 mM |
| NADH | 0.2 mM |
| β-hydroxypyruvate | 2.5 mM (or phosphohydroxypyruvate) |
| Glutamate dehydrogenase | 0.02 mg |

The reaction was initiated by the addition of β-hydroxypyruvate and the change in the rate of absorbance change at 340 nm measured.

## 2.20 Determination of protein concentration

Protein concentrations were determined by the method of Bradford (1976). A standard curve was set up with varying concentrations of bovine serum albumin in the same buffer in which the protein determinations were to be carried out.

## 2.21 Polyacrylamide gel electrophoresis in the presence of SDS

Polyacrylamide gel electrophoresis in the presence of SDS was performed by the method of Laemmli (1970). The percentage acrylamide used is described in the text; a 3% stacking gel was used. Samples were boiled for 3 min in 3% SDS, 50 mM-tris (pH 6.8), 10% glycerol, 2% bromophenol blue before loading. Gels were stained for protein in 50% methanol, 10% acetic acid, 0.1% Coomassie blue (Serva G-250) for 45 min at $40^{\circ}$ and destained overnight in 10% methanol, 10% acetic acid ($40^{\circ}$).

Molecular weight markers were:

| | |
|---|---|
| Bovine serum albumin | (68 000) |
| Catalase | (60 000) |
| Glutamate dehydrogenase | (53 000) |
| Aldolase | (40 000) |
| Glyceraldehyde phosphate dehydrogenase | (38 000) |
| Carbonic anhydrase | (29 000) |

Approximately 2 µg of each was loaded.

## 2.22 DNA sequencing

All DNA sequencing was carried out using the M13/dideoxy method (Sanger, 1981; Messing, 1983). Initially, the components for M13 cloning and DNA sequencing were purchased in the form of kits; latterly, individual components were purchased and substituted. All the protocols for cloning and sequencing were supplied with the kits in the form of the "M13 cloning and sequencing handbook" (Amersham International plc), and these were strictly adhered to. In the following sections, page numbers in the format (handbook p.1) refer to the Amersham handbook.

## 2.23 M13 vector DNA

The paired vectors M13mp8 and M13 mp9 (Messing & Vieira, 1982) were used. The double stranded replicative form (RF) of the DNA was supplied with the cloning kit, or was a gift from Mr G. Millar.

## 2.24 Digestion of RF DNA and storage (handbook p.22)

2 µg of RF DNA was digested with the appropriate enzyme(s) and the extent of digestion determined by gel electrophoresis. Completely digested vector was purified by phenol extraction and ethanol precipitation. It was resuspended in TE buffer to give a final concentration of 10 ng/µl and stored at -20°. Each preparation was checked by transformation of competent *E.coli* JM101 with 10 ng cut vector and with 10 ng cut and religated vector.

## 2.25 Ligation of RF DNA to insert DNA (handbook p.24)

Ligation mixes for ligation of RF DNA to insert DNA were as follows:

| | |
|---|---|
| RF DNA (20 ng) | 2 µl |
| insert DNA (100 ng) | 5 µl |
| 10 x ligase reaction buffer | 1 µl |
| ATP 10 mM | 1 µl |
| Dithiothreitol | 1 µl |
| T4 DNA ligase | 1 µl |

T4 DNA ligase: for sticky ends 0.1 unit/µl
for blunt ends 0.4 units/µl

Incubation was for 4-6 h at 14° (sticky ends) or 0° for 24 h (blunt ends).

Final volume was varied from 10-20 µl, depending on the volume of buffer the insert DNA was resuspended in.

## 2.26 Transformation of E.coli JM101 (handbook p.25)

E.coli JM101 was streaked out on MM and after overnight growth the plate was stored at 4° (for approximately 1 month before re-streaking). A single colony was inoculated into 10 ml 2TY and shaken overnight at 37°. 40 ml 2TY was inoculated with 2 ml fresh overnight culture and shaken (37°) for 3 h. 10 ml 2TY was also inoculated with a drop of overnight culture to provide fresh cells. Cells from the 40 ml 2TY culture were harvested (5 000 x g, 5 min), resuspended in 20 ml 50mM-$CaCl_2$ and left on ice for 20 min. Cells were then harvested (5 000 x g, 5 min) and resuspended

in 4 ml 50mM-$CaCl_2$. 0.3 ml of competent cells was added to 5 µl of DNA ligation mix and kept on ice 40 min. Cells were then heat shocked (42°, 3 min) and returned to ice. The following were added to each tube:

| | |
|---|---|
| 100mM-IPTG (Isopropyl-β-D-thio-galactopyranoside) | 40 µl |
| 2% X-gal (5-bromo-4-chloro-3-indolyl-β-galactoside) (in dimethylformamide) | 40 µl |
| fresh cells | 200 µl |
| molten H top agar (kept at 42°) | 3 ml |

and the tube poured onto a pre-warmed (37°) H plate.

2.27 Preparation of single-stranded template DNA (handbook p.27)

A single plaque was inoculated using a sterile Eppendorf pipette tip into 1.5 ml 2TY containing 1 ml/100 ml medium of an overnight E.coli JM101 culture. This culture was shaken (37°) for 5 h then centrifuged 5 min (microfuge). At this stage, supernatants were sometimes stored at -20° for later analysis. The pellet of cells produced at this stage was also broken open and RF DNA extracted using the mini-prep technique. The RF DNA produced could then be investigated with restriction enzymes.

The supernatant was transferred to another tube and recentrifuged (5 min). The supernatant was added to 200 µl PEG/NaCl (20% polyethylene glycol 6 000/2.5M-NaCl), shaken and left for 15 min. This was then centrifuged 5 min and the supernatant discarded. Following a further 2 min spin, all remaining traces of PEG were removed using a drawn out pasteur pipette. 100 µl TE and 50 µl phenol were added to

the viral pellet and mixed by vortexing (15 s). After standing the tubes for 15 min, the contents were re-mixed (vortex, 15 s) and centrifuged 3 min. The upper layer was transferred to a fresh centrifuge tube and 1 ml chloroform added. The chloroform was mixed (vortex, 15 s) and separated (microfuge, 1 s). 10 µl 3M-Na acetate (pH 6) and 250 µl ethanol were added and the DNA precipitated by placing the tube at -70° (dry ice/methanol) for 15 min. The DNA was recovered by centrifugation (10 min) and washed with 1 ml (-20°) ethanol. The ethanol was poured off and the tube left to dry. The DNA pellet was redissolved in 50 µl TE and stored at -20°.

2.28 Annealing of primer and template (handbook p.33)

For each clone, the following were placed in a 1.5 ml microfuge tube:

| | |
|---|---|
| single stranded template DNA | 5 µl |
| primer | 1 µl |
| *10 x Klenow reaction buffer | 1.5 µl |
| distilled water | 2.5 µl |

(*100mM-Tris pH 8.5; 100mM-$MgCl_2$)

The primer used is a 17mer with the sequence:

5'-GTAAAACGACGGCCAGT-3'

The mixture was incubated in an oven at 55° for 2 h.

2.29 The sequencing reactions (handbook p.34)

To the annealed primer/template mix was added 1.5 µl (15 µCi) of ($\alpha$ - $^{35}$S)dATP $\alpha$ S at >600 Ci/mmol (Amersham

SJ304) and 1 µl (1 U/µl) of Klenow fragment of E.coli DNA polymerase I. 2.5 µl of the mixture was placed on the rim of each of four tubes, marked A, C, G or T in a microfuge rotor. 2 µl of the relevant dNTP/ddNTP mix (see below) was placed inside the rim of each tube and a brief spin mixed the contents. After 20 min, 2 µl of chase mixture (an 0.5 mM uniform mixture of dATP, dCTP, dGTP and dTTP) was placed on the rim of the tube and mixed in with a brief spin. After a further 15 min chase reaction, 4 µl formamide dye mix was added. Formamide dye mix was prepared as follows: 100 ml of formamide was deionised by stirring with 5 g Amberlite (BDH) for 30 min, then filtered to remove the resin; to this was added 0.03 g xylene cyanol FF, 0.03 g bromophenol blue and 4 ml 0.5M-EDTA.

2.30 Reaction mixes (handbook p.32)

The final concentration of each reagent in a sequencing reaction mix is listed in Table 2.6. Three sets of reaction mixtures were used:

A* - or 'normal' reaction mixes; for use with labelled dATP used in most experiments

A*/low dideoxy - in these mixtures the dideoxy concentration was 0.75 times that in the A* mixes; for long runs on linear gels, allowing reading of up to 400 bp

C* - for labelling with radioactive dCTP.

(a) A*

| | A reaction mix | C reaction mix | G reaction mix | T reaction mix |
|---|---|---|---|---|
| dATP | 16.6 | 16.6 | 16.6 | 16.6 |
| dCTP | 28 | 0.3 | 28 | 28 |
| dGTP | 28 | 28 | 0.3 | 28 |
| dTTP | 28 | 28 | 28 | 0.3 |
| ddATP | 22.2 (16.6) | - | - | - |
| ddCTP | - | 22.2 (16.6) | - | - |
| ddGTP | - | - | 66.6 (49.8) | - |
| ddTTP | - | - | - | 111 (83) |

(b) C*

| | A reaction mix | C reaction mix | G reaction mix | T reaction mix |
|---|---|---|---|---|
| dATP | 0.3 | 28 | 28 | 28 |
| dCTP | 16.6 | 16.6 | 16.6 | 16.6 |
| dGTP | 28 | 28 | 0.3 | 28 |
| dTTP | 28 | 28 | 28 | 0.3 |
| ddATP | 22.2 | - | - | - |
| ddCTP | - | 22.2 | - | - |
| ddGTP | - | - | 66.6 | - |
| ddTTP | - | - | - | 111 |

Table 2.6: Final concentration of reagents in sequencing reaction mixes. All concentrations are $\mu$M.

Numbers in brackets refer to the final concentrations of dideoxy nucleotide in the 'low dideoxy' mixes.

## 2.31 Polyacrylamide gel electrophoresis (handbook pp.36 and 45)

The DNA fragments in the sequencing reaction mixes were separated by electrophoresis on polyacrylamide gels (20 x 40 x 0.4 cm). The notched plate was siliconised before use with Repelcote (BDH).

Two types of gel were employed, linear and buffer gradient. The quantities of reagents necessary to pour a 6% acrylamide, 8M-urea tris-borate gel are shown in Table 2.7. Before loading, the wells were flushed out with running buffer (1 x tris-borate) to remove unpolymerised acrylamide and urea, which leaches into the gel slots, preventing the samples from forming a precise layer.

The samples were heated to 95$^{o}$ for 3 min then loaded immediately onto the gel. Gels were run at 28 mA for varying lengths of time (see Chapters 4, 5 and 6).

Gels were fixed after removing the notched plate by soaking in a 2:1 bath of 10% acetic acid, 10% methanol for 30 min to remove the urea. The gel was removed from the bath and a sheet of Whatman 3MM paper laid on top. The gel was peeled off with the paper backing, placed on a Bio-rad Model 1125 gel drier and covered by Saranwrap. Thirty minutes under vacuum was sufficient to dry the gel completely.

## 2.32 Autoradiography

The dried gel was placed in a light-proof cassette and covered by a sheet of Fuji RX film. The film was

| | | linear | buffer gradient | |
|---|---|---|---|---|
| | | | top | bottom |
| urea | (g) | 21 | 19.2 | 3.8 |
| acrylamide stock | (ml) | 7.5 | 6 | 1.125 |
| 10 x tris-borate | (ml) | 5.0 | 2 | 1.875 |
| sucrose | (g) | - | - | 0.75 |
| bromophenol blue (0.01 g/ml) | (ml) | - | - | 0.075 |
| water, to | (ml) | 50 | 40 | 7.5 |
| TEMED | (µl) | 50 | 80 | 15 |
| AMPS (%age) | (µl) | 300 (10) | 80 (25) | 15 (25) |

Table 2.7: Quantities of reagents necessary to pour a single polyacrylamide gel for DNA sequence analysis.

Abbreviations: TEMED - NNN'N'-tetramethylethylenediamine

AMPS - ammonium persulphate

Acrylamide stock - 40%; 38:2 acrylamide: NN' methylenebisacrylamide, deionised by stirring with 5 g Amberlite MB1 and filtered through a sintered glass funnel then passed through a Millipore filter (0.45 µm).

Urea - ultrapure (BRL).

exposed overnight at room temperature before developing.

2.33 T-track analysis (handbook p.39)

This technique was used to classify clones into groups containing the same insert (see Chapter 5). For ten clones the following mix was made up:

4 µl primer, 6 µl 10 x Klenow reaction buffer, 12 µl water. 2 µl priming mix was added to 2 µl template DNA and the tube incubated at 55° for 2 h.

A sequencing mix was made up:

16 µl dTTP/ddTTP, 3 µl ($\alpha$ - $^{35}$S)dATP $\alpha$S, 2 µl (2U) Klenow fragment. 2 µl sequencing mix was added to the annealed primer/template and the reaction started by a brief spin (microfuge). After 20 min, 1 µl chase solution was added and the reaction stopped after a further 15 min by the addition of 1 µl formamide dye mix.

2.34 Preparation of RF DNA in vitro (handbook p.29)

Primer and single stranded template DNA were annealed as for a sequencing reaction. 1 U of Klenow fragment and 1 µl chase solution were added and the mixture incubated at room temperature for 30 min. The reaction was stopped by heating at 70° for 10 min.

2.35 Clone turn-around

Clone turn-around was used to obtain the opposite strand of a small fragment, or to sequence the other end of a long fragment.

RF DNA was prepared in vitro, then the following were added: 2 µl 10 x restriction enzyme buffer, 5 U EcoRI, 5 U HindIII. The sample was incubated for 1 h at 37° and the fragments separated by electrophoresis on a LMT agarose gel. Insert DNA was recovered from the gel and ligated to the complementary vector, prepared by cleavage with EcoRI and HindIII.

## 2.36 Computer programs for manipulation of DNA sequences

The following programs were utilised in the compilation, manipulation and analysis of the DNA sequences. Programs (a) - (d) were run on a Digital PDP 11-34 computer, with a multi-user facility, in the Biochemistry Department, University of Glasgow.

2.36.1(a) BATIN: This program is used to store new gel readings onto a disk. Any number of separate gel readings may be entered and stored in individual files.

2.36.2(b) TTEM: A version of DBCOMP (Staden, 1980), modified by Dr R. Eason (Department of Biochemistry, University of Glasgow) so that it runs more efficiently. This program searches for overlaps between sequences, comparing each sequence in a set of sequences with each other sequence and other complementary sequence.

2.36.3(c) DBUTIL: Described in detail in Staden (1980). This program enables the user to establish and change the relationship between gel reading data, and build up a consensus sequence.

2.36.4(d) TRNTRP: This program translates DNA sequences into peptide sequences (Staden, 1978).

2.36.5(e) WISGEN: A number of the programs of the WISGEN (University of Wisconsin Genetics Computer Group) package (Devereux et al., 1984) were run on the ERCC (Edinburgh Regional Computing Centre) VAX 11/750 computer. This package contains programs for the analysis and investigation of DNA sequences and for comparison of sequences with those in the EMBL database (European Molecular Biology Laboratory, Heidelberg, West Germany) and the NIH-Genbank database (Bolt, Baranek & Newman, Inc., Cambridge MA, USA).

## 2.37 Determination of $^{32}P$ in DNA samples

### Precipitation with trichloroacetic acid (TCA)

A known volume of the sample to be assayed was spotted onto a Whatman GF/C glass fibre disk (2.4 cm diameter). An equal volume was added to a tube containing 100 µl of salmon sperm DNA (500 µg/ml in 20 mM-EDTA). 5 ml of ice-cold 10% TCA was added and the tube kept on ice 15 min. The precipitate was collected by filtering through another GF/C disc and washing the disk six times with 5 ml of ice-cold 10% TCA, followed by 5 ml of 95% ethanol. Both

filters were dried under a heat lamp then placed into scintillation vials. 8 ml of scintillation fluid was added and the samples counted in an Intertechnique SL4 000 scintillation counter. The scintillation fluid used was 4g/l PPO (2,5-diphenyloxazole) in 2:1 toluene; triton X-100. The first filter measures total radioactivity; the second measures the radioactivity incorporated into nucleic acids.

## 2.38 5'- end labelling of DNA fragments

25 pmol of 5'-termini, purified free of salts and proteins by phenol/chloroform extraction and ethanol precipitation were resuspended in 80 µl 50mM-tris (pH 8.2) and 0.5U calf-intestinal phosphatase (Boehringer) was added. After incubation at 37° for 30 min, the reaction was stopped by the addition of 20 µl 50 mM-EDTA and further incubation at 65° for 15 min. The DNA was purified by phenol/chloroform extraction and ethanol precipitation. It was then resuspended in 20 µl 50 mM-tris (pH 8.2), 10 mM-$MgCl_2$, 1 mM-EDTA, 50 mM-dithiothreitol, 1 mM-spermidine. To this was added: 10 µl ($\gamma^{32}$P)ATP (5 000 Ci/mmol, Amersham PB10218) and a 1 µl sample taken to determine total radioactivity. 8U T4 polynucleotide kinase (Boehringer) was added and incubated for 30 min at 37°. A 1 µl sample was taken for determination of TCA insoluble radioactivity. The DNA was then purified by phenol/chloroform extraction and ethanol precipitation.

## 2.39 Preparation of RNA from E.coli

RNA was prepared from *E.coli* AB2829/pKD501, as follows: 100 ml LB/tet, inoculated with *E.coli* AB2829/pKD501 was grown to $A_{650}$ = 0.3. The cells were harvested and re-suspended in 6 ml buffer R (0.02M-Na acetate (pH 5.5), 0.5% SDS, 1mM-EDTA). 12 ml phenol (equilibrated in buffer R) was added and the tube incubated at 60° for 10 min with gentle shaking. The sample was centrifuged to separate the phases and the aqueous phase removed to a fresh tube. This was re-extracted with phenol at room temperature. The aqueous phase was again separated and removed to a fresh tube. 2.5 volumes of ethanol were added and the tube chilled to -70° for 30 min. The precipitate was collected by centrifugation (12 000 x g, 30 min) and resuspended in 3 ml buffer R. Ethanol precipitation was repeated twice more and the final precipitate dissolved in 500 µl TE buffer.

## 2.40 Nuclease S1 mapping

Nuclease S1 was purchased from BRL. S1 buffer was:

80% formamide (deionised before use); 20mM-PIPES (Sigma); 0.4M-NaCl.

## 2.41 Blotting of DNA and RNA to nitrocellulose

Both DNA and RNA agarose gels were blotted onto nitrocellulose (Schleicher & Schuell BA85) by the method of Southern (1975), as described in detail in Maniatis *et al.* (1982).

## 2.42 Hybridisation solution for Northern blot

Both pre-hybridisation and hybridisation of nitrocellulose filters for Northern blot analysis was carried out with the same buffer, except that during hybridisation, the radioactive probe was added:

6 x SSC

5 x Denharts

100 µg/ml salmon sperm DNA (sheared) (Sigma).

0.5% SDS

SSC: 0.15M-NaCl, 0.015M-sodium citrate

50 x Denharts: 1% Ficoll, 1% polyvinylpyrrolidone (Sigma), 1% BSA.

# CHAPTER 3 STUDIES ON THE aroB GENE of E.coli

## 3.1 Introduction

### 3.1.1 The aroB gene

A purification scheme for 3-dehydroquinate synthase, the product of the E.coli aroB gene has been reported (Maitra & Sprinson, 1978). 560 µg of homogeneous enzyme was obtained from 380 g of E.coli in 6% overall yield after a 5-step purification. The purification factor was nearly 3 000 fold. The apparent molecular weight estimated both by gel filtration under native conditions and by SDS-PAGE was 57 000, suggesting that the enzyme is monomeric.

Since the enzyme is present in such low concentrations in wild type E.coli cells and the purification is difficult, mechanistic work on this enzyme has been hindered. The enzyme purification would be greatly simplified if an overproducing strain could be constructed. This requires the cloning of the aroB gene and its insertion into a multicopy plasmid vector. Reasonable quantities of the pure enzyme should then be available for detailed structural and mechanistic studies.

The aroB gene has been identified on pLC29-47, a member of the Clarke-Carbon E.coli genomic library (Clarke & Carbon, 1976; Takeda et al., 1981). pLC29-47 was the source of the aroB gene for the sub-cloning described in this study.

### 3.1.2 Previous work on pLC29-47

Takeda *et al.* (1981) searched the Clarke-Carbon plasmid collection for markers closely linked to penicillin-binding protein genes. One such marker is *aroB*; it is closely linked to the *ponA* (*mrcA*) gene, at minute 74 on the *E.coli* genetic map (Bachmann, 1983; Figure 1.4).

The vector used to construct the Clarke-Carbon library was ColE1. This vector has two disadvantages for over-expression of enzyme genes - it exists in low copy number within the cell and it has no readily selectable phenotype. We therefore wanted to subclone the genomic portion of pLC29-47 onto a more readily manipulated vector. pAT153 (Twigg & Sherratt, 1980) was used throughout.

Comparison of the limited restriction enzyme analysis of pLC29-47 carried out by Takeda *et al.* (1981) with the restriction map of ColE1 (Dougan *et al.*, 1978) showed that a 7.2 kb PstI fragment of pLC29-47 carried most of the genomic material (Figure 3.1a). The first objective was to sub-clone this fragment and confirm that it carried the *aroB* gene.

## 3.2 Sub-cloning and analysis of the aroB gene

### 3.2.1 Construction of pKD101 and pKD102

pLC29-47 was digested with PstI and the digestion products separated by electrophoresis on 1% LMT agarose. A band at approximately 7.2 kb was excised and the DNA purified. The fragment was ligated into the PstI site of phosphatase-treated pAT153 (Sections 2.15 and 2.16).

Figure 3.1: Restriction mapping of various aroB subclones.

(a) pLC29 -47; data from Takeda et al. (1981).

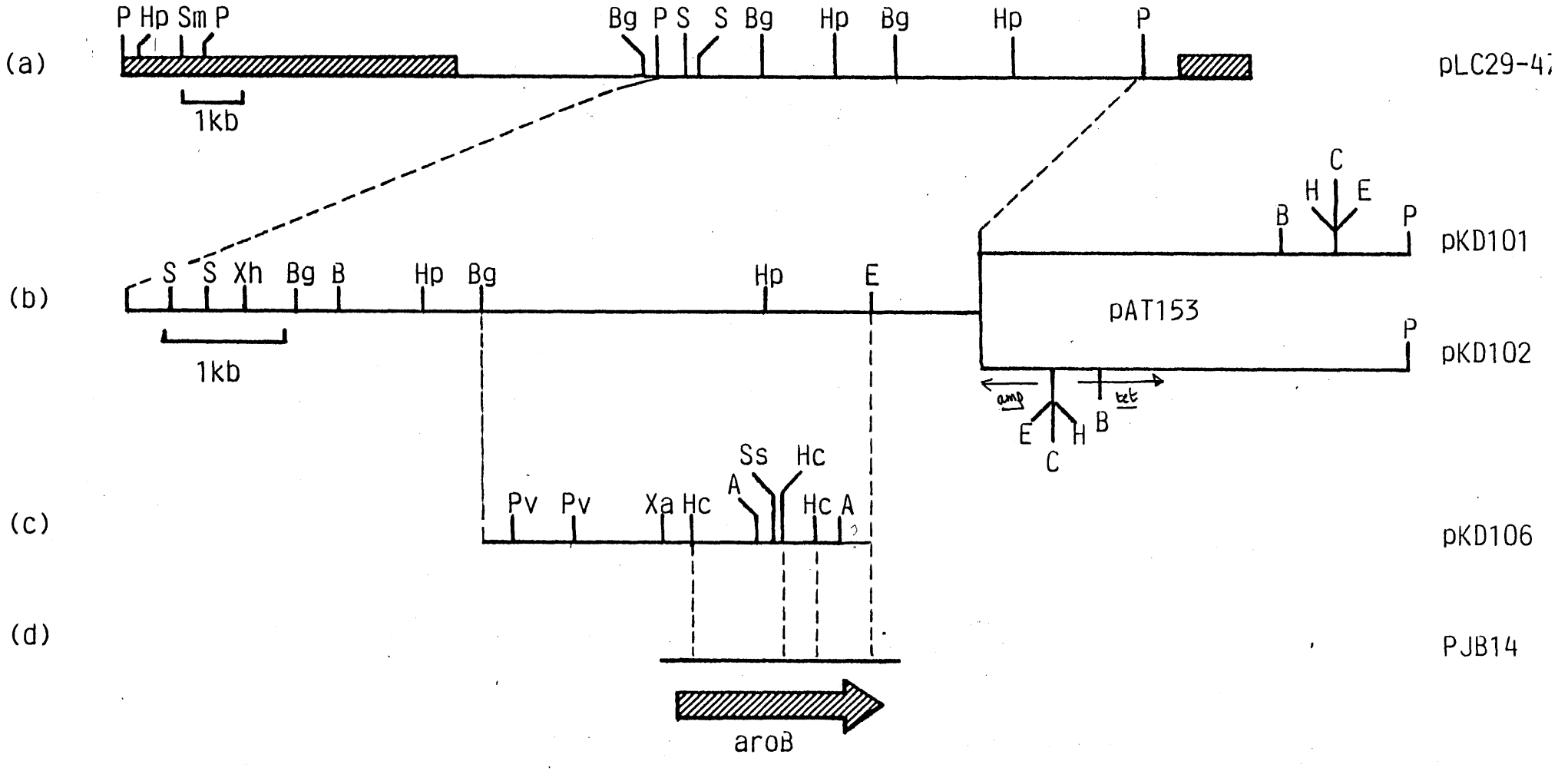
(b) pKD101/pKD102; mapping and orientation with respect to pAT153.

(c) pKD106; detailed restriction map.

(d) pJB14; region in common with pKD106, and orientation of the aroB gene.

Key to restriction enzymes:

| | | | |
|---|---|---|---|
| A | - AvaI | Hp | - HpaI |
| B | - BamHI | P | - PstI |
| Bg | - BglII | S | - SalI |
| C | - ClaI | Sm | - SmaI |
| E | - EcoRI | Ss | - SstII |
| H | - HindIII | Xa | - XorI |
| Hc | - HincII | Xh | - XhoI |
| | | Pv | - PvuII |


(a)
P Hp Sm P
Bg P S S Bg Hp Bg Hp P
1kb
pLC29-47
(b)
S S Xh Bg B Hp Bg Hp E
1kb
C H E B P
pKD101
pAT153
P
pKD102
amp
tet
E H B C
(c)
Ss Hc A
Pv Pv Xa Hc Hc A
pKD106
(d)
PJB14
aroB

The ligation mix was used to transform E.coli AB2826 and plated on LA/tet (Section 2.17). After overnight growth at 37°, colonies were transferred by replica plating to minimal medium and grown a further 24 h at 37°. 100 of the colonies which grew on minimal medium were picked off and gridded onto LA/tet, LA/amp and MM to check their growth properties. Colonies containing plasmids carrying the aroB gene at the PstI site within the ampicillin resistance gene of pAT153 should be $tet^R$, $amp^S$, $aroB^+$. 10 such colonies were picked and grown in 10 ml LA/tet. Plasmid DNA was extracted and analysed after digestion with various restriction enzymes. The asymmetric location of a pair of SalI sites within the PstI fragment was used to determine the orientation of the PstI fragment relative to the vector. Two clones were retained for further analysis. Each had the 7.2 kb PstI insertion, but in opposite orientations; they were designated pKD101 and pKD102 (Figures 3.1b and 3.2).

Two controls were performed in this experiment, and indeed in all experiments involving identification of clones by relief of auxotrophy. A crude preparation of recombinant plasmid was used to retransform the auxotrophic mutant employed to identify the clone (in this case E.coli AB2826) and all resulting transformants were $tet^R$, $amp^S$, $aroB^+$. The same crude preparation was used to transform a second auxotrophic aro mutant (E.coli AB2829, aroA). In this case, all transformants were $tet^R$, $amp^S$ but remained unable to grow on unsupplemented medium.

### 3.2.2 Restriction mapping of pKD101/pKD102

To design a strategy for further subcloning of the *aroB* gene, a more detailed restriction map of pKD101 than that published by Takeda *et al.* (1981) was required. A series of single and double enzyme digests were carried out on crude preparations of pKD101 and pKD102 using enzymes that would be useful in further cloning. The resulting map is shown in Figure 3.1b. The two BglII sites and the single EcoRI site are the most useful for further subcloning.

### 3.2.3 Construction of pKD103/pKD104

pKD101 and pKD102 were each cleaved with BglII and the resulting DNA fragments separated by electrophoresis on 1% LMT agarose. The slowly migrating fragment (9.3 kb) was excised and the DNA purified. This DNA was then self-religated and transformed into *E.coli* AB2826. After overnight growth on LA/tet, colonies were checked for the presence of recircularised plasmid of the correct size and for growth on minimal medium. Two plasmids, designated pKD103 and pKD104 were constructed by recircularisation of pKD101 and pKD102 respectively. Both are capable of complementing the *aroB* mutation in *E.coli* AB2826.

### 3.2.4 Construction of pKD105

From the same BglII digest of pKD101 described in the previous section, the smaller fragment (1.45 kb) was excised

and the DNA purified. This was then ligated to BamHI cleaved pAT153 and transformed into E.coli AB2826. After overnight growth on LA/amp, colonies were screened for inserts into the tetracycline resistance gene and 6 representatives were checked for an insert at the BamHI site by isolating samples of DNA and checking their size. These clones were found not to grow on minimal medium. One clone was retained and designated pKD105.

The finding that removal of the 1.45 kb BglII fragment from pKD101 does not affect the ability of the plasmid to complement aroB implies that aroB does not lie between the two BglII sites and does not span either of them.

### 3.2.5 Construction of pKD106

In a similar way, one final construction was made. The 3.5 kb BglII-EcoRI fragment of pKD101 was inserted into BamHI and EcoRI cleaved pAT153. Several colonies were grown up and plasmid DNA extracted. In each case digests showed the presence of the 3.5 kb insert expected and the loss of the BamHI site (the BamHI/BglII hybrid site is cut by neither enzyme). This plasmid is able to complement an aroB mutant.

The sub-cloning constructions described in this section are summarised in Figure 3.2.

More detailed restriction enzyme analysis of pKD106 was carried out with a view to identifying restriction sites which might be used to clone fragments for DNA sequence analysis. A combination of single and double digests was

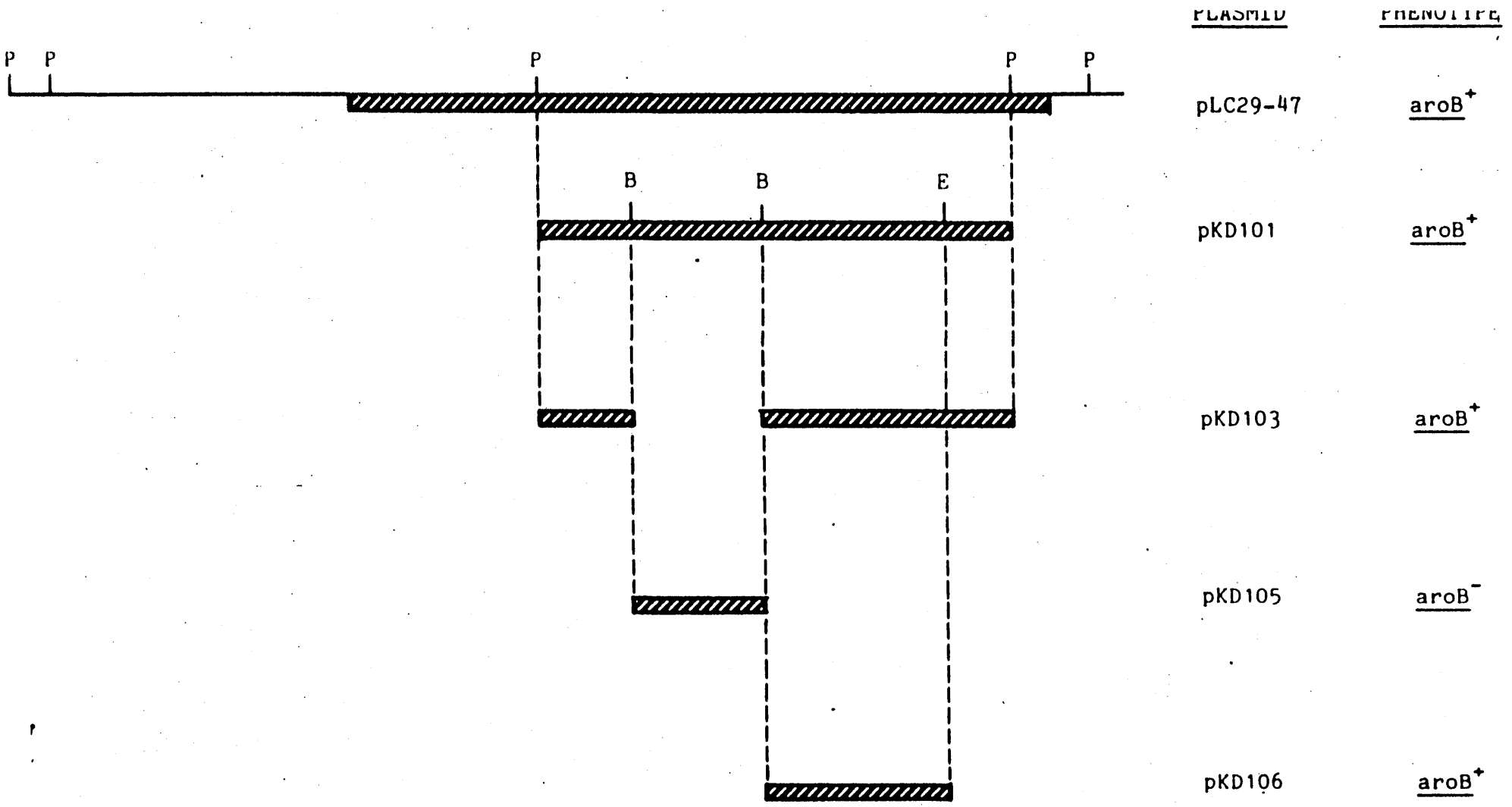



Figure 3.2: Summary of the sub-cloning of the E.coli aroB gene from pLC29-47. Hatching indicates E.coli DNA.

was used and the digestion products analysed on 1.5% or 2% agarose gels. The data obtained is shown in Figure 3.1c.

### 3.2.6 Activity of 3-dehydroquinate synthase in crude extracts

The specific activity of 3-dehydroquinate synthase was determined in crude extracts of E.coli K12, E.coli AB2826 and E.coli AB2826 harbouring each of the plasmids described in the preceding sections.

Crude extracts were prepared from stationary phase cultures as described in Section 2.18. Cells were grown in 100 ml minimal medium, except in the case of E.coli AB2826 and E.coli AB2826/pKD105 which were grown in minimal medium containing the aromatics supplement. Sonication buffer 'A' was used.

DHQ synthase was measured by using 3-dehydroquinase as a coupling enzyme; this enzyme converts 3-dehydroquinate to 3-dehydroshikimate which absorbs strongly at 234 nm (see Section 2.19). The results are shown in Table 3.1.

Extracts of E.coli K12 contain a very low concentration of DHQ synthase. In the crude extract there is a high background $\Delta A_{234}$ in the absence of added DAHP and this rate may contribute to the measured DHQ synthase rate. It is therefore difficult to quantitate the DHQ synthase level in wild type cells and so the estimates of the over-expression factor in the table can vary according to the value taken for the basal level in E.coli K12. We can be sure that the increase in the rate of absorbance change

following DAHP addition is the authentic DHQ synthase activity for three reasons:

(i) The reaction is inhibited by NADH, a known inhibitor of the enzyme. This was checked in two ways. Firstly, addition of NADH to a final concentration of 50 μM during an assay causes the rate to drop. Secondly, addition of NADH to 50 μM before initiation with DAHP gives a diminished DHQ synthase rate.

(ii) The DHQ synthase reaction is dependent on $Co^{2+}$. Crude extracts are partially active without added $Co^{2+}$. This $Co^{2+}$ can be chelated by the addition of EDTA/NaOH pH 7.0 to 10 mM and incubation at 25° for 3 min. Following the standard assay procedure, but omitting $Co^{2+}$ there is no response to the addition of DAHP. Cobalt can then be added back to the reaction mix and DHQ synthase activity is seen when $[Co^{2+}]$ is in excess of [EDTA].

(iii) The reaction is dependent on the presence of 3-dehydroquinase. A time lag in the increase in the rate of change of absorbance following DAHP initiation is observed and this decreases as more coupling enzyme is added. When no purified 3-dehydroquinase is added to the mix and the reaction is dependent on endogenous 3-dehydroquinase, the lag is greater but the same steady state rate is reached.

The results shown in Table 3.1 indicate that in all the crude extracts of cells carrying recombinant plasmids, except *E. coli* AB2826/pKD105, there is an increase in the

| Extract | Specific activity of 3-dehydroquinate synthase | Relative activity |
|---|---|---|
| K12 | 0.07 | 1 |
| AB2826 | no activity | - |
| AB2826/pLC29-47 | 0.072 | 10 |
| AB2826/pKD101 | 0.168 | 24 |
| AB2826/pKD102 | 0.154 | 22 |
| AB2826/pKD103 | 0.136 | 19 |
| AB2826/pKD104 | 0.140 | 20 |
| AB2826/pKD105 | no activity | - |
| AB2826/pKD106 | 0.136 | 19 |

Table 3.1 Specific activity of 3-dehydroquinate synthase in E.coli crude extracts.

specific activity of DHQ synthase. This correlates with the ability of the plasmids to complement the aroB mutation in E.coli AB2826.

E.coli AB2826/pLC29-47 produces approximately 10 fold more DHQ synthase than does E.coli K12. This strain, however, grows poorly on the minimal medium required to maintain the plasmid.

All the other aroB complementing plasmids overexpress DHQ synthase approximately 20-25 fold. This indicates that in each case, the entire aroB gene and any flanking sequences involved in its expression have been cloned. These derivatives of pLC29-47, cloned in pAT153 can now be maintained and grown in hosts which are not aroB; the plasmids may be maintained by antibiotic selection. It is important to maintain the integrity of these plasmids by growing them in recA host strains. E.coli AB2826 is $recA^+$ and there is a danger that the aroB gene on the plasmid might recombine with the defective chromosomal aroB marker. For this reason, all plasmids have been maintained in E.coli HB101 which is $recA^-$, $aroB^+$.

## 3.3 Relationship to other work on aroB

Frost et al. (1984) have independently sub-cloned the aroB gene from pLC29-47 by a different route. Their strategy was to clone fragments of pLC29-47, produced by limited digestion with tetranucleotide recognising enzymes into the vector pKK223-3 (J. Brosius, unpublished results),

prior to selection for the aroB gene by relief of auxotrophy of a suitable E.coli mutant.

pKK223-3 is a pBR322 derivative, designed for very high level overexpression of foreign genes in E.coli. Its structure is shown in Figure 3.3. Fragments of DNA from any source may be cloned into the 'polylinker' site. The 'polylinker' is flanked by a 'tac' promoter and a terminator sequence. The 'tac' promoter is a hybrid of the E.coli trp operon promoter -35 region and the E.coli lac operon -10 region, with the lac UV5 mutation which renders the promoter insensitive to catabolite repression. This structure gives very high level initiation of transcription from the trp -35 region but control from the lac -10 region. Growth of the plasmid in a $lacI^q$ strain switches off expression of any gene downstream of the promoter, but this repression can be overcome by the addition of IPTG to log phase cultures. The terminator sequence of the E.coli rrnB operon is also necessary for the efficient expression of a gene located in the 'polylinker' site.

The plasmid pJB14, which carries aroB cloned into pKK223-3 was derived by the following route: pKK223-3 was prepared by digestion by EcoRI, filling in the ends produced and ligation of ClaI linker molecules. Digestion with ClaI then provided free ClaI ends. pLC29-47 was digested with TaqI under conditions that would give incomplete cleavage of the plasmid. The fragments were separated by electrophoresis on an agarose gel and those corresponding to the size range 2 - 2.5 kb were excised and cloned into

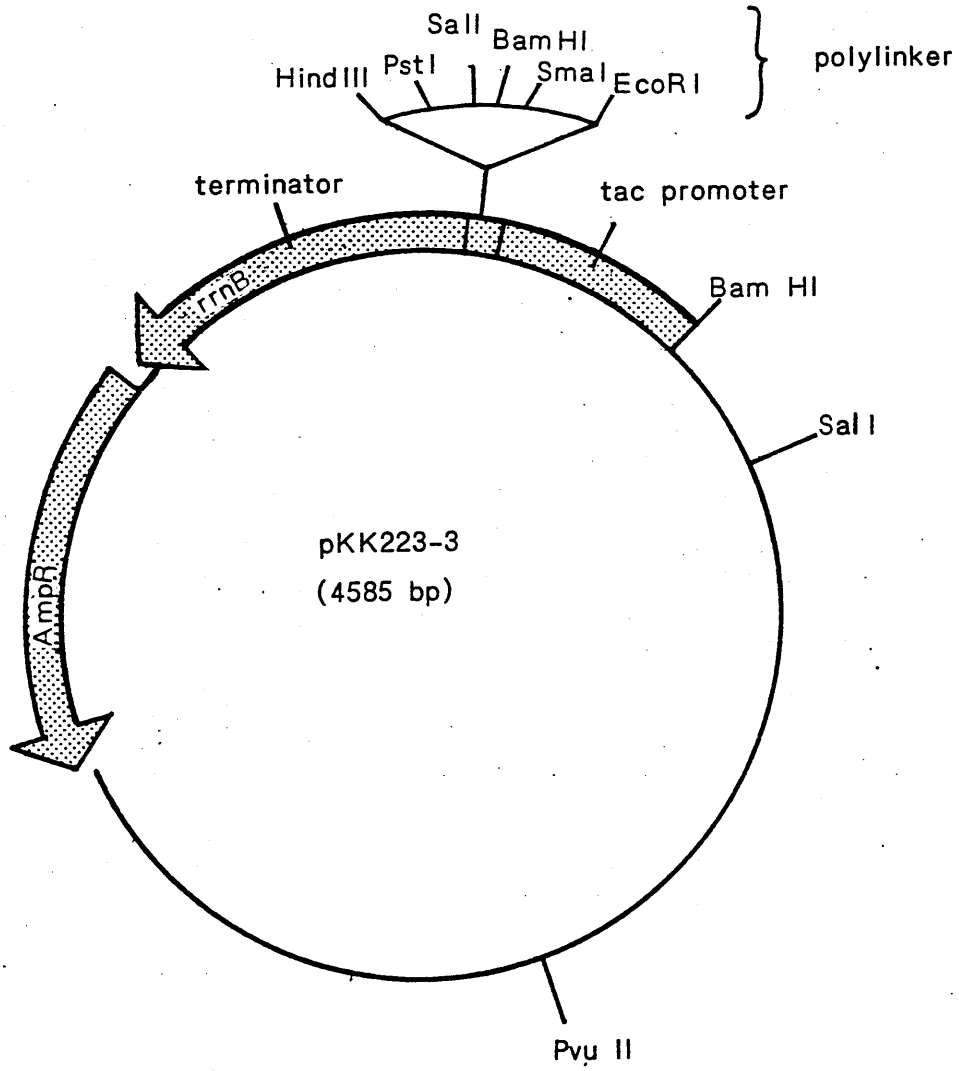



Figure 3.3: Structure of pKK223-3.

pKK223-3. The aroB carrying recombinant (pJB14) was identified by relief of auxotrophy of E.coli AB2826. Frost et al. claim that cells harbouring this plasmid contain 1 000 times more DHQ synthase activity when fully induced than do E.coli K12.

Comparison of the detailed restriction map of the insert in pJB14 (G. Millar, unpublished results) with that of pKD106 shows a region which they have in common (Figure 3.1d). This allows the direction of transcription of the aroB gene to be determined since the gene is transcribed from the tac promoter of pKK223-3.

## 3.4 Further studies on aroB

The long term aim of studies on DHQ synthase is to determine the mechanism of action of this unusual reaction. Ultimately this will require the determination of the three-dimensional structure of the enzyme. Gene cloning can assist structural studies in two ways. Firstly, the crystallisation of an enzyme, prior to X-ray crystallography requires large quantities of pure enzyme which can now be obtained from overexpressing strains. Secondly the amino acid sequence of the enzyme can most easily be derived by determining the DNA sequence of the gene. An accurate knowledge of the amino acid sequence is a prerequisite to the model building associated with three-dimensional structure determination.

In furtherance of this aim, a simple two column purification of homogeneous DHQ synthase has been developed

(Frost *et al.*, 1984). The starting material for this is a cell lysate of *E.coli* harbouring the plasmid pJB14 in which *aroB* expression was induced by IPTG during the log phase of growth. It is now possible by this method to purify approximately 100 mg of enzyme from 50 g of *E.coli*.

The complete amino acid sequence of DHQ synthase has recently been determined by DNA sequencing of the *aroB* gene (G. Millar, unpublished), but is not discussed in this thesis.

# CHAPTER 4 STUDIES ON THE aroD GENE OF E.coli

## 4.1 Introduction

### 4.1.1 The aroD gene

The E.coli aroD gene encodes the enzyme 3-dehydroquinase. A 5.6 kb PstI fragment of the E.coli genomic DNA has been cloned into pBR322 and shown to carry aroD (Kinghorn et al., 1981). This recombinant (pJKK12) encodes a 3-dehydroquinase activity which is indistinguishable from that isolated from wild type E.coli on the basis of heat stability, ammonium sulphate fractionation and sucrose density gradients. Evidence from the sucrose density gradients, and from minicell labelling of plasmid encoded proteins suggested that the native enzyme is a dimer composed of two 31 500 dalton subunits. Recently, this quarternary structure has been confirmed by studies on the purified E.coli enzyme (Chaudhuri & Coggins, 1984b).

### 4.1.2 Analysis of pJKK12: (i) Kinghorn et al. (1981)

The 5.6 kb insert in pJKK12 is sufficiently large to encode several E.coli genes. Restriction mapping and sub-cloning were carried out in order to identify the location of aroD on the clone.

The insert contains two BglII sites and single sites for HpaI, SalI and XhoI. No EcoRI, BamHI or HindIII sites were detected (Figure 4.1).

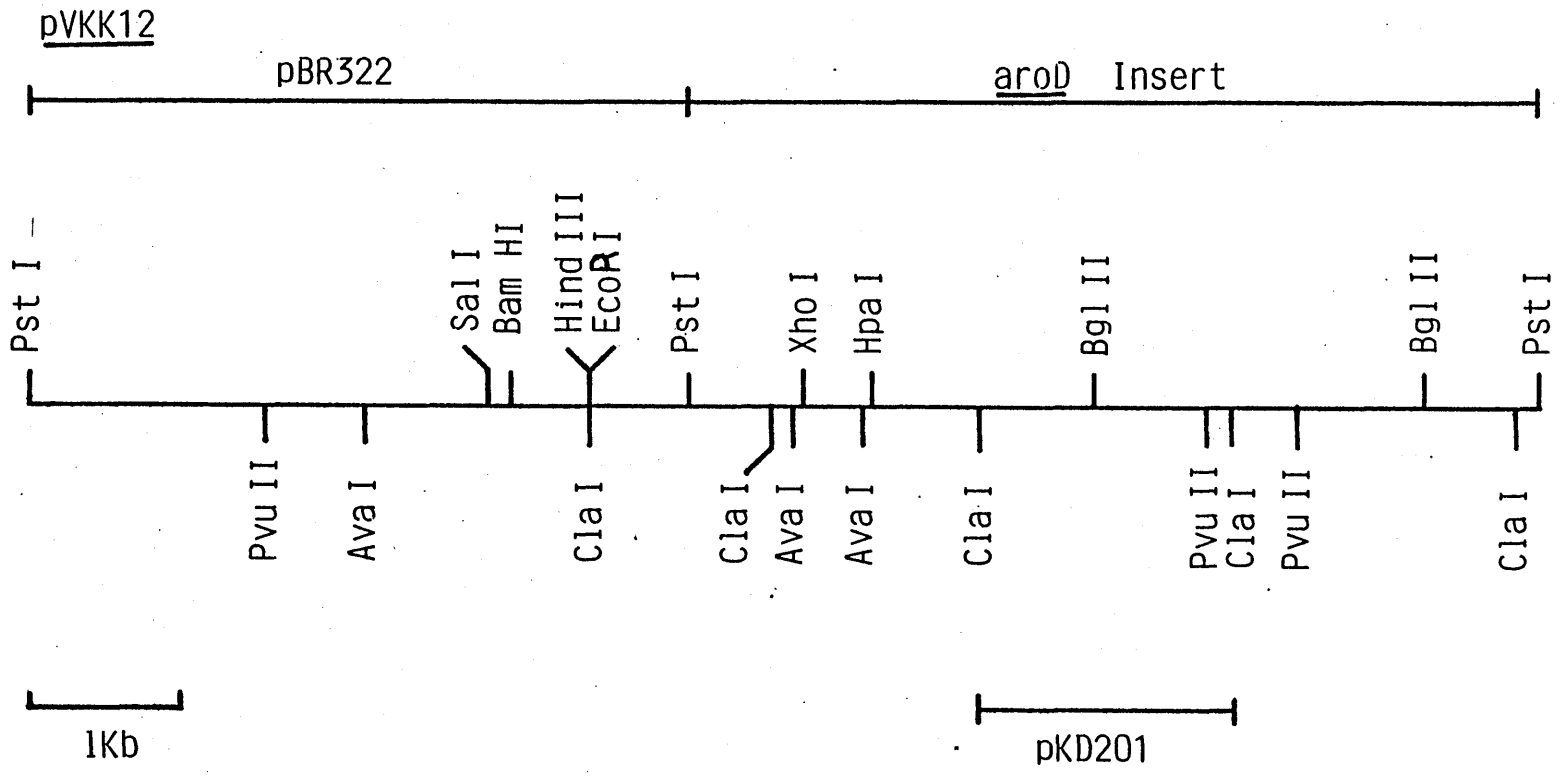



Figure 4.1: Restriction map of pJKK12. Enzyme cleavage sites for the insert shown above the line were determined by Kinghorn et al. (1981). The sites below the line were determined in this work. The bar labelled 'pKD201' shows the location of the ClaI fragment, sub-cloned in pKD201.

pJKK12 was digested with BglII and the smaller fragment produced was subcloned into the BamHI site of pBB322 (recombinant pJKK26). After deleting the BglII fragment, the plasmid was recircularised (pJKK20). Neither pJKK26 nor pJKK20 complement aroD mutants of E.coli. This, together with evidence from $^{35}S$ labelling of proteins in mini cells suggests that BglII cleaves within the aroD gene or within a region essential for aroD expression.

### 4.1.3 Analysis of pJKK12: (ii) This study

Previous workers (Kinghorn et al., 1981; see above) had suggested that the aroD gene spanned one of the two BglII sites located in the insert of pJKK12 which encoded only aroD and would have been of suitable size for DNA sequence analysis.

Restriction enzyme digests of pJKK12 were carried out with three enzymes: AvaI, ClaI and PvuII. A series of double digests were used to locate the sites at which these enzymes cut the 5.6 kb insert. The results are illustrated in Figure 4.1.

It was shown that ClaI cuts the insert four times and has a site midway between the two BglII sites. Therefore it is possible to identify ClaI fragments of 1.8 kb and 2.2 kb which each carry a different portion of pJKK12, surrounding a BglII site.

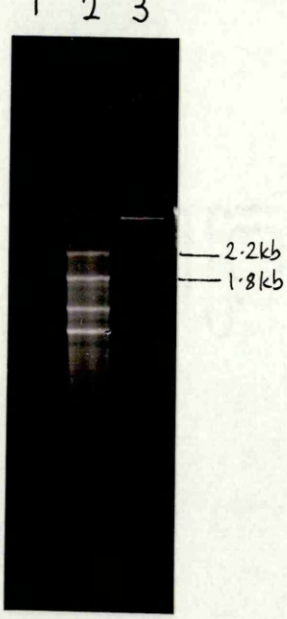



Figure 4.2a: ClaI sub-clones of pJKK12

Track 1 pKD202/ClaI
2 pJKK12/ClaI
3 pKD201/ClaI

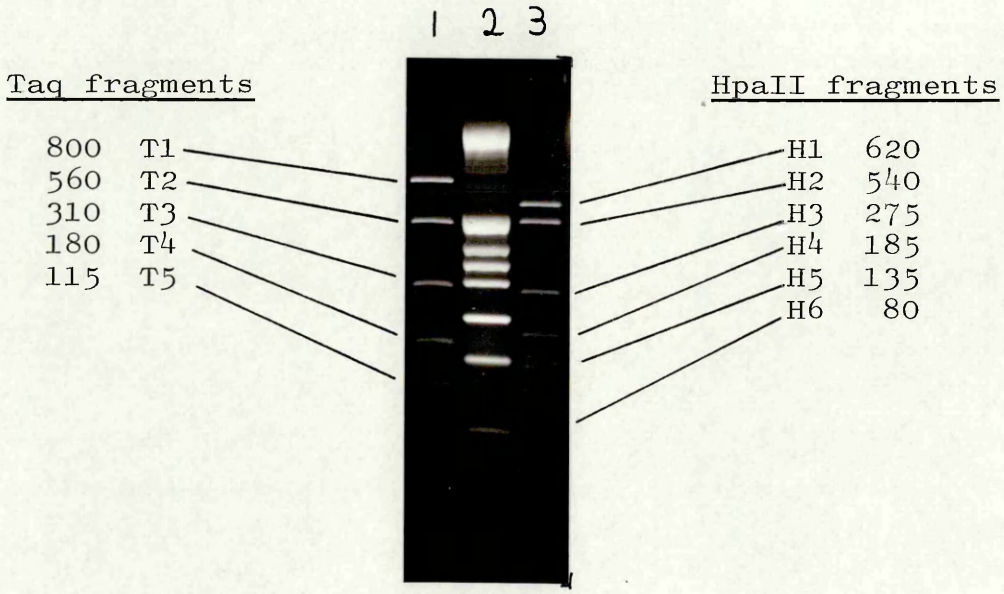



Figure 4.2b: TaqI and HpaII digests of the ClaI insert of pKD201 (sizes are in bp)

Tract 1 TaqI digest of the 1.8 kb ClaI insert of pKD201
2 pBR322/HinfI
3 HpaII digest of the 1.8 kb ClaI insert of pKD201

## 4.2 Sub-cloning and analysis of the aroD gene

### 4.2.1 Construction of pKD201 and pKD202

2 μg of pJKK12 was digested with ClaI and separated by electrophoresis on 1% LMT agarose. The 1.8 kb and 2.2 kb bands were excised and the DNA extracted. Each DNA sample was then ligated into the ClaI site of phosphatase-treated pAT153. Ligation mixes were used to transform *E.coli* AB2848 (*aroD*) and plated on LA/amp. 100 transformants from each ligation/transformation were gridded onto LA/amp, LA/tet and MM. All those colonies arising from ligation of the 1.8 kb ClaI fragment which were $amp^R$, $tet^S$ were also able to grow on MM. By contrast, all those transformants arising from ligation of the 2.2 kb ClaI fragment which were $amp^R$, $tet^S$ were found to be incapable of growth on MM. Plasmid DNA was extracted from five $amp^R$, $tet^S$ colonies of each type and in each case a ClaI insert was found which was of the expected size. pAT153 recombinants carrying the 1.8 kb and 2.2 kb inserts were named pKD201 and pKD202 respectively. An ethidium bromide stained agarose gel of a ClaI digest of these plasmids is shown in Figure 4.2a.

### 4.2.2 3-Dehydroquinase activity in crude extracts

Crude extracts of *E.coli* K12, *E.coli* AB2848/pKD201 and *E.coli* AB2848 were prepared as described in Section 2.18. 100 ml cultures were grown in minimal medium or 'aromatics' medium (*E.coli* AB2848 only). Sonication buffer 'B' was used. The results obtained are shown in Table 4.1.

| Extract | Specific activity of 3-dehydroquinase (U/mg) |
|---|---|
| *E.coli* | 0.032 |
| *E.coli* AB2848 | no activity |
| E.coli AB2848/pKD201 | 2.955 |

Table 4.1 Specific activity of 3-dehydroquinase in crude extracts.

These results confirm the pattern of aroD complementation. The 3-dehydroquinase specific activity in cells harbouring pKD201 is approximately 3 U/mg, a 90-100 fold increase over that found in wild type E.coli. From the estimated $M_r$ of 3-dehydroquinase, the aroD gene would be expected to be approximately 1 kb in length. Since aroD is being over-expressed at the copy number level of pAT153, this implies that the entire aroD coding region and any sequences necessary for its expression are located on the 1.8 kb ClaI fragment.

## 4.3 Determination of the DNA sequence of the aroD gene

### 4.3.1 Sequencing strategy for aroD

A strategy was required for sequencing the aroD gene which would give the entire sequence of the 1.8 kb ClaI fragment on both strands and then allow identification of the aroD gene within this sequence. The logic behind the development of this strategy is set out below:

(i) Since pKD201 was constructed by insertion of a ClaI fragment into pAT153, the insert is flanked by an EcoRI and a HindIII site. Recovery of the insert from EcoRI + HindIII cleaved pKD201 and cloning into M13mp8 and M13mp9 followed by DNA sequence analysis will give the sequence at either end of the insert.

(ii) Cleavage of pKD201 at the unique BglII site within the insert and at the unique BamHI site within the vector yields two fragments. Each of these may be cloned into BamHI cleaved M13mp8 and the sequence

either side of the BglII site derived.

(iii) These sequences will provide a framework onto which the remainder of the sequence can be built. The majority of the sequence information can now be obtained by cloning smaller fragments of the insert DNA and sequencing these.

(iv) The insert DNA can be fragmented by the use of restriction enzymes which recognise tetranucleotide sequences. For the ClaI clone, the most useful enzymes for this purpose are TaqI and HpaII. These enzymes produce the same 'sticky ends' as ClaI, and may be cloned into the AccI site of M13mp8 or M13mp9 (Figure 4.3).

The successful outcome of this approach to sequencing the aroD gene relies on two factors. Firstly, the enzyme used must have sites distributed in such a way as to give fragments with lengths up to 400 bp. The entire sequence on both strands of the DNA can then be determined for each of these fragments. Secondly, the distribution of the sites must be such that the sequences obtained from one set of cleavages will overlap the restriction sites for the second set of cleavages.

It is important that the entire DNA sequence is derived on both strands of the DNA as a confirmation of the sequence on the complementary strand. This is especially true in regions of DNA which encode proteins where an omission or insertion of a single base pair can lead to mis-interpretation of an open reading frame, with disastrous consequences.


Figure 4.3: Restriction enzyme sites used in cloning of TaqI and HpaII fragments for DNA sequence analysis into the AccI site of M13mp8 and M13mp9. The AccI site in these vectors has the sequence GTCGAC.

Mistakes leading to an alteration in the sequence of the C-terminal region of a protein might go unnoticed because a polypeptide of similar molecular weight and composition might be predicted.

All restriction sites must also be overlapped in order that no small fragments are missed during the cloning procedures.

4.3.2 Size distribution of TaqI and HpaII fragments

8 µg of pKD201 was digested with ClaI and the resulting DNA fragments separated by electrophoresis on a preparative 1% LMT gel. The 1.8 kb ClaI insert was excised and prepared for digestion with a second enzyme as described in Section 2.14. The second enzyme digestion was with HpaII (10 U, 37°, 2 h) or TaqI (20 U, 54° oven, 2 h); this was followed by electrophoresis on 2% agarose (Figure 4.2b).

HpaII cleaves the fragment at least five times yielding six fragments, ranging in size from 80 bp to 620 bp. These were designated H1 to H6 as shown on the diagram. There are at least four TaqI sites on the insert, yielding five fragments designated T1 to T5. Only three of these fragments are below 400 bp in length. Altogether four out of the eleven TaqI and HpaII fragments are greater than 400 bp in length and if the restriction sites are such that any of the large fragments are substantially overlapped then it may not be possible to obtain the entire sequence of the ClaI fragment on both strands of the DNA from these two sets of fragments. For this reason,

it was necessary to locate the TaqI and HpaII sites on the insert DNA.

4.3.3 Distribution of TaqI and HpaII sites on pKD201

The TaqI and HpaII sites on the insert DNA were detected by mapping using the partial restriction enzyme digestion technique of Smith & Bernstiel (1976). The principle of this technique is as follows: The DNA in question is enzymatically labelled with $^{32}P$ at one end. It is then digested with the chosen enzyme(s) so as to produce a partial digest. A large spectrum of partial digestion products may be produced, but the labelled fragments form a simple overlapping series, all with a common labelled terminus. These are fractionated according to molecular weight by gel electrophoresis and detected by autoradiography. The relative mobility of each labelled fragment allows determination of its molecular weight by comparison with a set of DNA standards and this in turn locates the distance of the respective restriction sites from the labelled terminus. The order of the fragments and their lengths thus correspond directly to the order of the restriction sites along the DNA molecule.

5 µg of pKD201 (approximately 2.5 pmol) was digested with EcoRI and the DNA purified free of restriction enzyme by phenol/chloroform extraction and ethanol precipitation. The 5'-phosphoryl groups were removed by treatment with phosphatase (Section 2.15). The termini were then labelled using $\gamma$-$^{32}P$ ATP and T4 polynucleotide kinase (Section 2.38).

After completion of the labelling reaction the number of TCA precipitable counts showed that approximately 14% of the label had been incorporated into DNA.

The labelled material was again purified by phenol/chloroform extraction and ethanol precipitation. The DNA was resuspended in 50 µl 1 x high salt restriction enzyme buffer containing BSA (100 µg/ml). 10 U HindIII was added and digestion carried out for 1 h at 37°. The DNA fragments in the digest were then separated by electrophoresis on 1% LMT agarose. The 1.8 kb insert band was excised and prepared for further enzyme digestion (Section 2.14). This fragment is now labelled uniquely at the EcoRI end.

A series of secondary restriction enzyme digests was set up as detailed in Table 4.2. The digests were stopped after the appropriate time by the addition of EDTA to 10 mM and by placing the reaction tube on ice, causing the agarose to solidify. 5 µl of gel loading dye was added to each tube and the tube heated to 65° for 5 min before loading onto a 2% agarose gel. The standards used on the same gel were:

Track 1 : pBR322 HinfI fragments 5' end labelled ($^{32}$P)

Track 12: λ EcoRI/HindIII fragments 5' end labelled ($^{32}$P)

Following electrophoresis, the gel was blotted onto nitrocellulose (Section 2.41). The nitrocellulose blot was autoradiographed at -70° with an intensifying screen. Several exposures of the blot were made with the exposure time varied from 2 h to 2 days, in order to clearly visualise as many bands as possible. Figures 4.4a and 4.4b show

| Track | Approximate quantity of labelled DNA | Enzyme and number of units | Time of incubation |
|---|---|---|---|
| 1 | (pBR322, HinfI fragments) | | |
| 2 | 0.5 µg | 0.4 U HpaII | 20 min |
| 3 | 0.5 µg | 0.3 U HpaII | 20 min |
| 4 | 0.5 µg | 0.2 U HpaII | 20 min |
| 5 | 0.5 µg | 0.1 U HpaII | 20 min |
| 6 | 0.5 µg | 5 U BglII | 1 h |
| 7 | 0.5 µg | 5 U PvuII | 1 h |
| 8 | 0.1 µg | 0.25 U TaqI | 20 min |
| 9 | 0.2 µg | 0.5 U TaqI | 20 min |
| 10 | 0.5 µg | 1.0 U TaqI | 20 min |
| 11 | 0.5 µg | 1.25 U TaqI | 20 min |
| 12 | (λ, EcoRI + HindIII fragments) | | |

Table 4.2 Enzymes used in partial digestion of labelled pKD201 fragment. Track refers to the gel track on Figure 4.4.

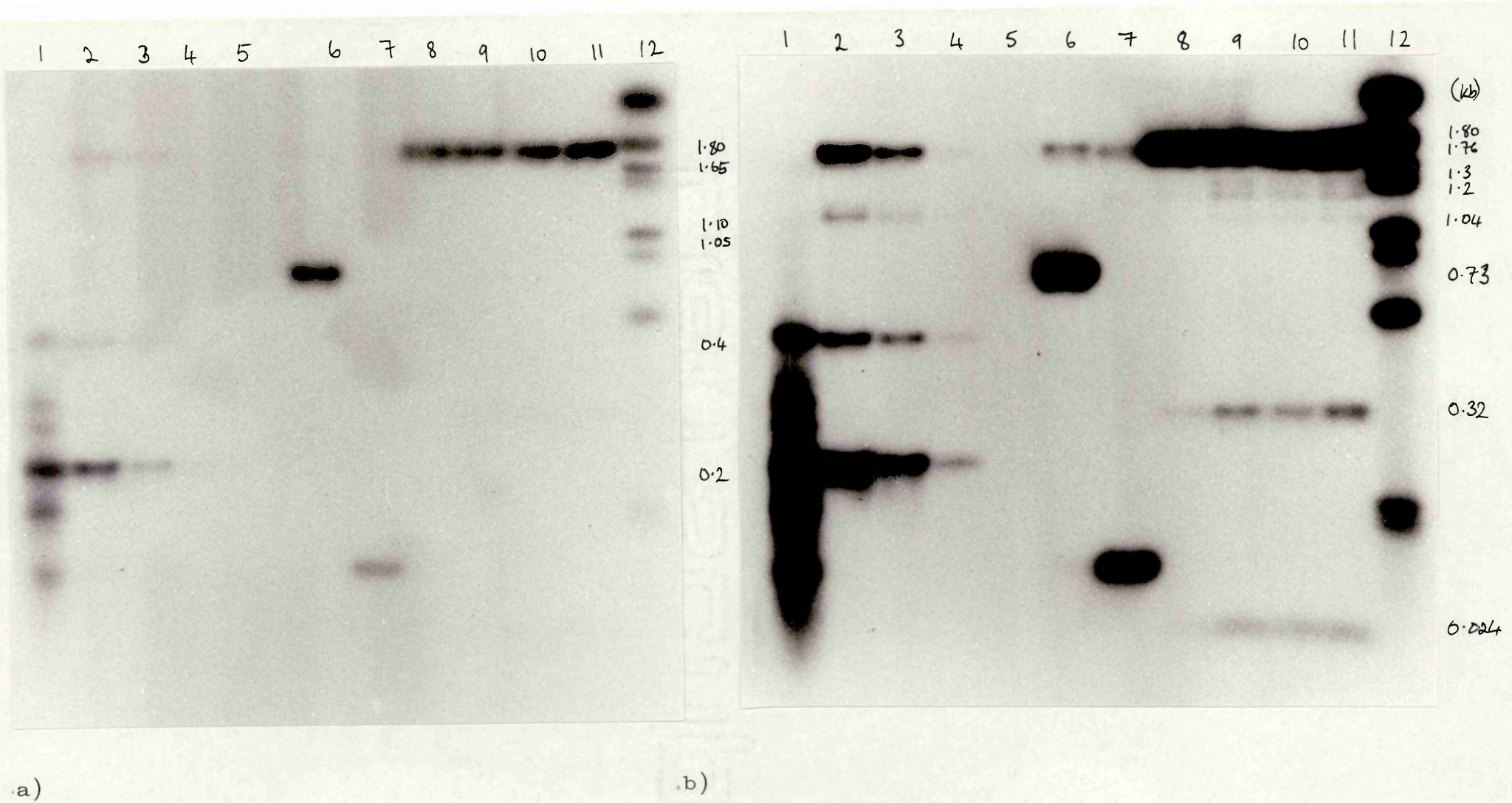



Figure 4.4: Autoradiographs from partial mapping experiment (Section 4.3.3). Fiture 4.4a shows a 2 h exposure; Figure 4.4b shows a 7 h exposure. The key to each track is given in Table 4.2.

2 h and 7 h exposures respectively.

Ideally, it is hoped that by varying the amount of restriction enzyme and/or DNA in the partial digest one will hit upon conditions under which each of the possible partial digestion products are equally well represented. The autoradiographs shown in Figures 4.4a and 4.4b show that this was not achieved - some of the bands are very strong and others are very weak, (hence the need to show two different exposures). However, it is possible to locate all of the restriction sites on the DNA using the information obtained from the autoradiographs.

The distance from the labelled EcoRI site to each restriction site was calculated from the relative mobilities of the fragments, compared to the standard λ EcoRI/HindIII fragments. The data obtained was used to construct the restriction map shown in Figure 4.5.

Analysis of the restriction map showed that it was possible to cover the entire sequence on both strands of the DNA. This is despite the fact that the two longest fragments, T1 and H1 overlap. The unique BglII site cuts both fragments approximately in the centre of each and so sequencing out from this site on either side will complete the sequences built up from each set of TaqI and HpaII clones. The autoradiographs also show that it is very unlikely that there are any very small TaqI or HpaII fragments which would not have been seen on an ethidium bromide gel such as that in Figure 4.4. Any such fragments would have been seen on the autoradiograph as an intense, wide band.

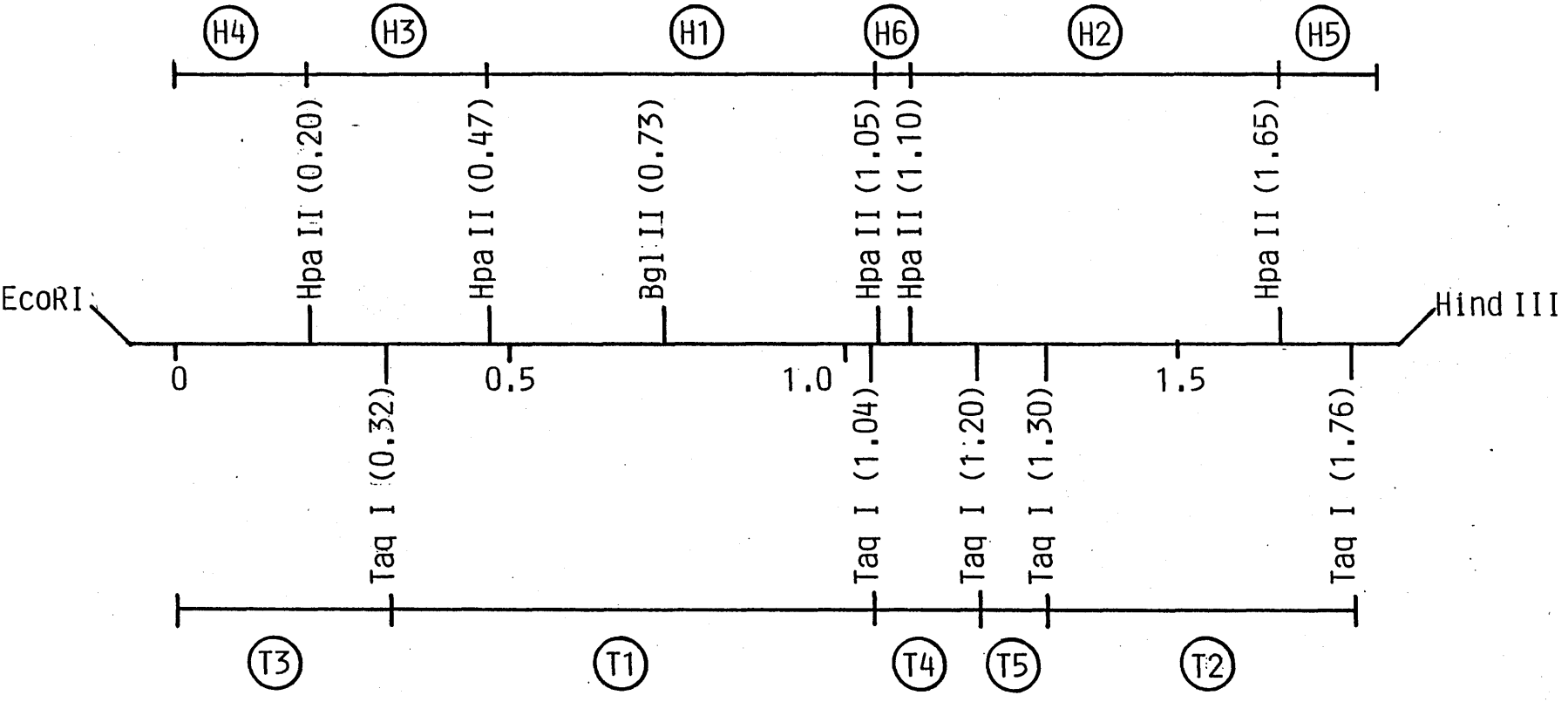



Figure 4.5: Restriction map of the ClaI insert in pKD201. Distances are in kb from the labelled EcoRI end (see Section 4.3.3). The fragments H1-H6 and T1-T6 are shown in Figure 4.2.

4.3.4 Preparation of M13 clones for sequence analysis

(a) The EcoRI - HindIII fragment

1 µg of pKD201 was digested with EcoRI and HindIII, purified by electrophoresis through 1% LMT agarose and the 1.8 kb insert fragment identified. The band was excised and the DNA extracted and purified by phenol/chloroform extraction and ethanol precipitation, and then resuspended in 20 µl 1 x ligase reaction buffer. 10 µl was placed in each of two tubes and ligation mixes set up in a final volume of 20 µl (see Section 2.25) and containing 20 ng M13mp8 (EcoRI + HindIII cleaved) or 20 ng M13mp9 (EcoRI + HindIII cleaved). 0.1 U of T4 ligase was added and incubation was continued at 14° for 4 h.

Ligation mixes were used to transform E. coli JM101. After overnight growth, two clear plaques were picked from each plate and inoculated into 1.5 ml 2TY (containing 0.1 ml/10 ml broth of fresh JM101). After 5 h growth at 37°, single strand template DNA was purified (Section 2.27). The template DNA was checked by electrophoresis on 1% agarose. The same check was performed on all the templates purified for DNA sequencing, and this is discussed in Section 4.5.4.

(b) The BglII - BamHI fragments

1 µg of pKD201 was digested with BglII and BamHI. The two resulting DNA fragments were purified by electrophoresis on 1% LMT agarose. Both bands were excised and the DNA extracted. Then both DNA fragments were cloned

into BamHI cleaved M13mp8 and two templates purified for each ligation/transformation.

(c) The TaqI and HpaII fragments

8 µg pKD201 was digested with ClaI and electrophoresed on a preparative 1% LMT agarose gel. The 1.8 kb ClaI fragment was excised and prepared for further digestion. Secondary digests were performed with TaqI and HpaII and these were electrophoresed on a 2% LMT agarose gel. A pattern similar to that in Figure 4.2b was observed. Each TaqI and each HpaII band was cut out of the gel and the DNA purified. Separate ligation mixes were set up for each fragment; in each case 20 ng of AccI cleaved M13mp8 was used. Each ligation mix was transformed into E.coli JM101 and after overnight growth, four clear plaques were picked and grown.

After the 5 h growth period, all of the 1.5 ml cultures were centrifuged (5 min, microfuge). One of each set of four templates was further manipulated to produce single strand template DNA. Each of the other three supernatants was stored at -20°.

4.3.5 Checks on single-stranded template DNA

Each single stranded template purified for DNA sequencing was checked by electrophoresis on a 1% agarose gel. This was performed prior to sequencing for two reasons:

(i) to confirm that the template preparation contained DNA, and that the yield of DNA was 'normal'. Initially,

templates were run alongside a control amount of DNA, but latterly experience enabled those template preparations containing reduced amounts of DNA to be identified. If the DNA concentration in a preparation is low, then steps can be taken to increase the DNA content of the subsequent annealing reaction. Visualisation of the DNA also reveals very approximately the insert size in a recombinant. Recombinants with insert sizes less than about 300 bp are difficult to distinguish from control M13 template. However, it is occasionally possible to distinguish between different cloned fragments by the mobility of the templates on agarose gels.

(ii) to check for contamination of the preparation. If two bands are seen in one track then it is likely that the DNA from two different phages has been co-purified. This may arise from picking two plaques from a crowded region of a plate. This is especially a problem when very few clear plaques have been produced. The two phages can be separated by retransformation and re-purification.

During template preparation, the final ethanol precipitated DNA is resuspended in a volume of 50 µl. A 4 µl sample of this is taken for gel electrophoresis and is added to 4 µl gel loading buffer before loading. An example of the type of gel produced is shown in Figure 4.6;

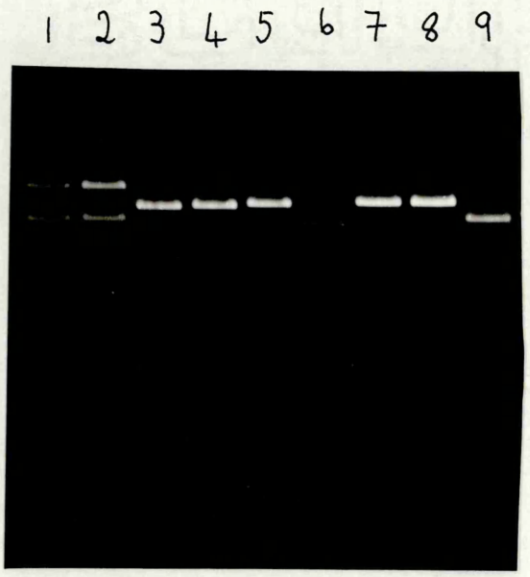



Figure 4.6: Agarose gel electrophoresis of single-stranded template DNA

| Track | | | |
|---|---|---|---|
| 1) 2) | M13mp8/BglII-BamHI | large fragment |
| 3) 4) | M13mp8/BglII-BamHI | small fragment |
| 5) 6) | M13mp8/EcoRI-HindIII | fragment |
| 7) 8) | M13mp9/EcoRI-HindIII | fragment |
| 9 | M13mp9 | |

it clearly demonstrates the points made above.

Figure 4.6 is a gel of the templates described in Sections 4.3.4(a) and 4.3.4(b). The following points should be noted:

(i) The quantity of DNA in each track is similar, with the exception of track 6.

(ii) Comparison with track 9 clearly shows those templates which are recombinants, and the mobility on the gel is related to the known size of the inserts.

(iii) Tracks 5 and 6 are different. Track 5 has, as expected the same size of insert as do the templates in tracks 7 and 8. Track 6 has the same mobility as track 9, indicating that a blue plaque was picked rather than a clear plaque.

(iv) Tracks 1 and 2 show that two different DNA's have been co-purified. This has probably resulted from accidentally picking a blue plaque near to a clear plaque.

### 4.3.6 Sequencing reactions

Recombinant templates which were known to have inserts of less than 250 bp in length (T4, T5, H4, H5, H6) were sequenced using the standard recipe sequencing mixes (Section 2.30) followed by electrophoresis on a buffer gradient polyacrylamide gel (Section 2.31).

For all the other clones, sequencing reactions were performed using the 'low dideoxy' mixes (Section 2.30) and separated by electrophoresis on two gels. A buffer

gradient gel gave the first 200-250 nucleotides of sequence and this was extended to 350-400 nucleotides by a second gel - a non-gradient 6% polyacrylamide gel (Section 2.31), run for approximately 6 h.

After electrophoresis, the gels were dried and autoradiographed overnight (Section 2.32).

### 4.3.7 'First round' of DNA sequencing

DNA sequence data was obtained from each of the EcoRI-HindIII clones and from the BglII-BamHI clones, ensuring that the sequence was not that of pAT153. Each of the TaqI and HpaII clone templates was sequenced. Each sequence was entered into a separate file on a Digital PDP11-34 computer using the BATIN program (Section 2.36). Each file was then compared, in turn, with each other file in two ways: firstly by aligning the sequences to be compared and looking for matches along the sequence and secondly by aligning the complementary sequence to that which is being compared. This analysis was carried out by the TTEM program (Section 2.36).

### 4.3.8 'Second round' of DNA sequencing

The data derived from running TTEM allowed a preliminary map of overlaps and complementary sequences to be built up. The preliminary map for the aroD clone showed those regions which could be completed by sequencing the complementary strand/other end of the TaqI or HpaII clones. In those cases, the supernatants which had been stored at -20$^{o}$ were

thawed and template DNA purified in the usual manner. These were then sequenced until the entire sequence was complete, on both strands of the DNA, and all the restriction sites used in cloning had been overlapped.

### 4.3.9 The complete sequence of the 1.8 kb ClaI fragment

Each nucleotide in the sequence was checked against the complementary nucleotide by reference to the original autoradiographs, until all ambiguities were resolved. The complete sequence of the 1.8 kb ClaI fragment was built up using the program DBUTIL (see Section 2.36). The entire sequence is illustrated in Figure 4.7, and the sequencing strategy summarised in Figure 4.10.

## 4.4 Analysis of the DNA sequence

### 4.4.1 Identification of open reading frames on the sequence

#### (a) Translation of both strands of the DNA sequence

Both strands of the DNA sequence were translated in all three reading frames using the TRN TRP program (see Section 2.36). The positions of stop codons were plotted and the open reading frames identified. It was expected that the *aroD* gene would be located near the unique BglII site in the sequence (Kinghorn *et al.*, 1981).

Only one open reading frame was found which was sufficiently long to encode an enzyme of the correct molecular weight. This open reading frame spans the BglII site. All the other reading frames are blocked many times.

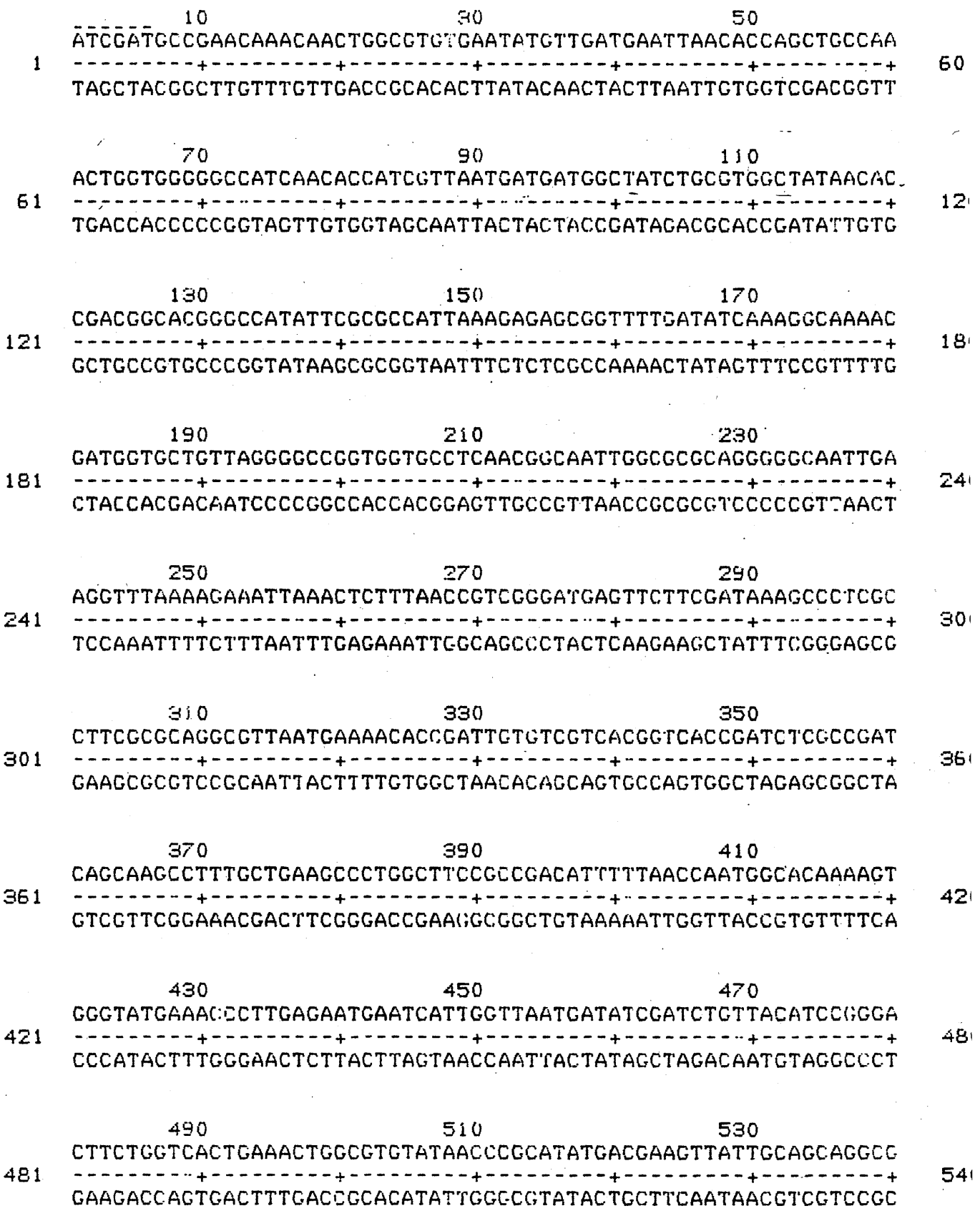



Figure 4.7: The complete DNA sequence of the 1.8Kb ClaI fragment inserted in pKD201, and the predicted AroD coding region.

```
      GTTGTTCGACCAACGTTTTGCTAACTACCTATGCCGTACCACAACACCGTTCCCCGACTT

             610                   630                   650
      CAGTTCACATTATGGACTGGCAAAGATTTCCCTCTCGAATATGTTAAACAGGTCATGGGC
601   ---------+---------+---------+---------+---------+---------+   660
      GTCAAGTGTAATACCTGACCGTTTCTAAAGGGAGACCTTATACAATTTGTCCAGTACCCC

                                                  AroD coding sequence ->

             670                   690                   710
      TTCGGTGCCTGACAGGCTGACCGCGTGCAGAAAGGGTAAAAAATGAAAACCGTAACTGTA
661   ---------+---------+---------+---------+---------+---------+   720
      AAGCCACGGACTGTCCGACTGGCGCACGTCTTTCCCATTTTTTACTTTTGGCATTGACAT
                                                 MetLysThrValThrVal
                                                  M  K  T  V  T  V
                                                 [1]

        BglII
        ------                    750                   770
      AAAGATCTCGTCATTGGTACGGGCGCACCTAAAATCATCGTCTCGCTGATGGCGAAAGAT
721   ---------+---------+---------+---------+---------+---------+   780
      TTTCTAGAGCAGTAACCATGCCCGCGTGGATTTTAGTAGCAGAGCGACTACCGCTTTCTA
      LysAspLeuValIleGlyThrGlyAlaProLysIleIleValSerLeuMetAlaLysAsp
       K  D  L  V  I  G  T  G  A  P  K  I  I  V  S  L  M  A  K  D
      [7]

             790                   810                   830
      ATCGCCAGCGTGAAATCCGAAGCTCTCGCCTATCGTGAAGCGGACTTTGATATTCTGCAA
781   ---------+---------+---------+---------+---------+---------+   840
      TAGCGGTCGCACTTTAGGCTTCGAGAGCGGATAGCACTTCGCCTGAAACTATAAGACGTT
      IleAlaSerValLysSerGluAlaLeuAlaTyrArgGluAlaAspPheAspIleLeuGlu
       I  A  S  V  K  S  E  A  L  A  Y  R  E  A  D  F  D  I  L  E
      [27]

             850                   870                   890
      TGGCGTGTGGACCACTATGCCGACCTCTCCAATGTGGAGTCTGTCATGGCGGCAGCAAAA
841   ---------+---------+---------+---------+---------+---------+   900
      ACCGCACACCTGGTGATACGGCTGGAGAGGTTACACCTCAGACAGTACCGCCGTCGTTTT
      TrpArgValAspHisTyrAlaAspLeuSerAsnValGluSerValMetAlaAlaAlaLys
       W  R  V  D  H  Y  A  D  L  S  N  V  E  S  V  M  A  A  A  K
      [47]

             910                   930                   950
      ATTCTCCGTGAGACCATGCCAGAAAAACCGCTGCTGTTTACCTTCCGCAGTGCCAAACAA
901   ---------+---------+---------+---------+---------+---------+   960
      TAAGAGGCACTCTGGTACGGTCTTTTTGGCGACGACAAATGGAAGGCGTCACGGTTTCTT
      IleLeuArgGluThrMetProGluLysProLeuLeuPheThrPheArgSerAlaLysGlu
       I  L  R  E  T  M  P  E  K  P  L  L  F  T  F  R  S  A  K  E
      [67]
```

Figure 4.7:(cont)

```
      CCGGCGCTCGTCCGCTAAAGGTGGCTCCGAATAATAACGTGAGTAGCACGTCGGTAGCTG
      GlyGlyGluGlnAlaIleSerThrGluAlaTyrTyrCysThrHisArgAlaAlaIleAsp
       G  G  E  Q  A  I  S  T  E  A  Y  Y  C  T  H  R  A  A  I  D
      [87]

            1030                 1050                 1070
      AGCGGCCTGGTTGATATGATCGATCTGGAGTTATTTACCGGTGATGATCAGGTTAAAGAA
1021  ---------+---------+---------+---------+---------+---------+  1080
      TCGCCGGACCAACTATACTAGCTAGACCTCAATAAATGGCCACTACTAGTCCAATTTCTT
      SerGlyLeuValAspMetIleAspLeuGluLeuPheThrGlyAspAspGlnValLysGlu
       S  G  L  V  D  M  I  D  L  E  L  F  T  G  D  D  Q  V  K  E
      [107]

            1090                 1110                 1130
      ACCGTCGCCTACGCCCACGCGCATGATGTGAAAGTAGTCATGTCCAACCATGACTTCCAT
1081  ---------+---------+---------+---------+---------+---------+  1140
      TGGCAGCGGATGCGGGTGCGCGTACTACACTTTCATCAGTACAGGTTGGTACTGAAGGTA
      ThrValAlaTyrAlaHisAlaHisAspValLysValValMetSerAsnHisAspPheHis
       T  V  A  Y  A  H  A  H  D  V  K  V  V  M  S  N  H  D  F  H
      [127]

            1150                 1170                 1190
      AAAACGCCGGAAGCCGAAGAAATCATTGCCCGTCTGCGCAAAATGCAATCCTTCGACGCC
1141  ---------+---------+---------+---------+---------+---------+  1200
      TTTTGCGGCCTTCGGCTTCTTTAGTAACGGGCAGACGCGTTTTACGTTAGGAAGCTGCGG
      LysThrProGluAlaGluGluIleIleAlaArgLeuArgLysMetGlnSerPheAspAla
       K  T  P  E  A  E  E  I  I  A  R  L  R  K  M  Q  S  F  D  A
      [147]

            1210                 1230                 1250
      GATATTCCTAAGATTGCGCTGATGCCGCAAAGTACCAGCGATGTGCTGACGTTGCTTGCC
1201  ---------+---------+---------+---------+---------+---------+  1260
      CTATAAGGATTCTAACGCGACTACGGCGTTTCATGGTCGCTACACGACTGCAACGAACGG
      AspIleProLysIleAlaLeuMetProGlnSerThrSerAspValLeuThrLeuLeuAla
       D  I  P  K  I  A  L  M  P  Q  S  T  S  D  V  L  T  L  L  A
      [167]

            1270                 1290                 1310
      GCGACCCTGGAGATGCAGGAGCAGTATGCCGATCGTCCAATTATCACGATGTCGATGGCA
1261  ---------+---------+---------+---------+---------+---------+  1320
      CGCTGGGACCTCTACGTCCTCGTCATACGGCTAGCAGGTTAATAGTGCTACAGCTACCGT
      AlaThrLeuGluMetGlnGluGlnTyrAlaAspArgProIleIleThrMetSerMetAla
       A  T  L  E  M  Q  E  Q  Y  A  D  R  P  I  I  T  M  S  M  A
      [187]

            1330                 1350                 1370
      AAAACTGGCGAAATTTCTCGTCTGGCTGGTGAAGTATTTGGCTCTGGCGGCAACTTTTGG
1321  ---------+---------+---------+---------+---------+---------+  1380
      TTTTGACCGCTTTAAAGAGCAGACCGACCACTTCATAAACCGAGACCGCCGTTGAAAACC
      LysThrGlyGluIleSerArgLeuAlaGlyGluValPheGlySerGlyGlyAsnPheTrp
       K  T  G  E  I  S  R  L  A  G  E  V  F  G  S  G  G  N  F  W
      [207]
```

Figure 4.7:(cont)

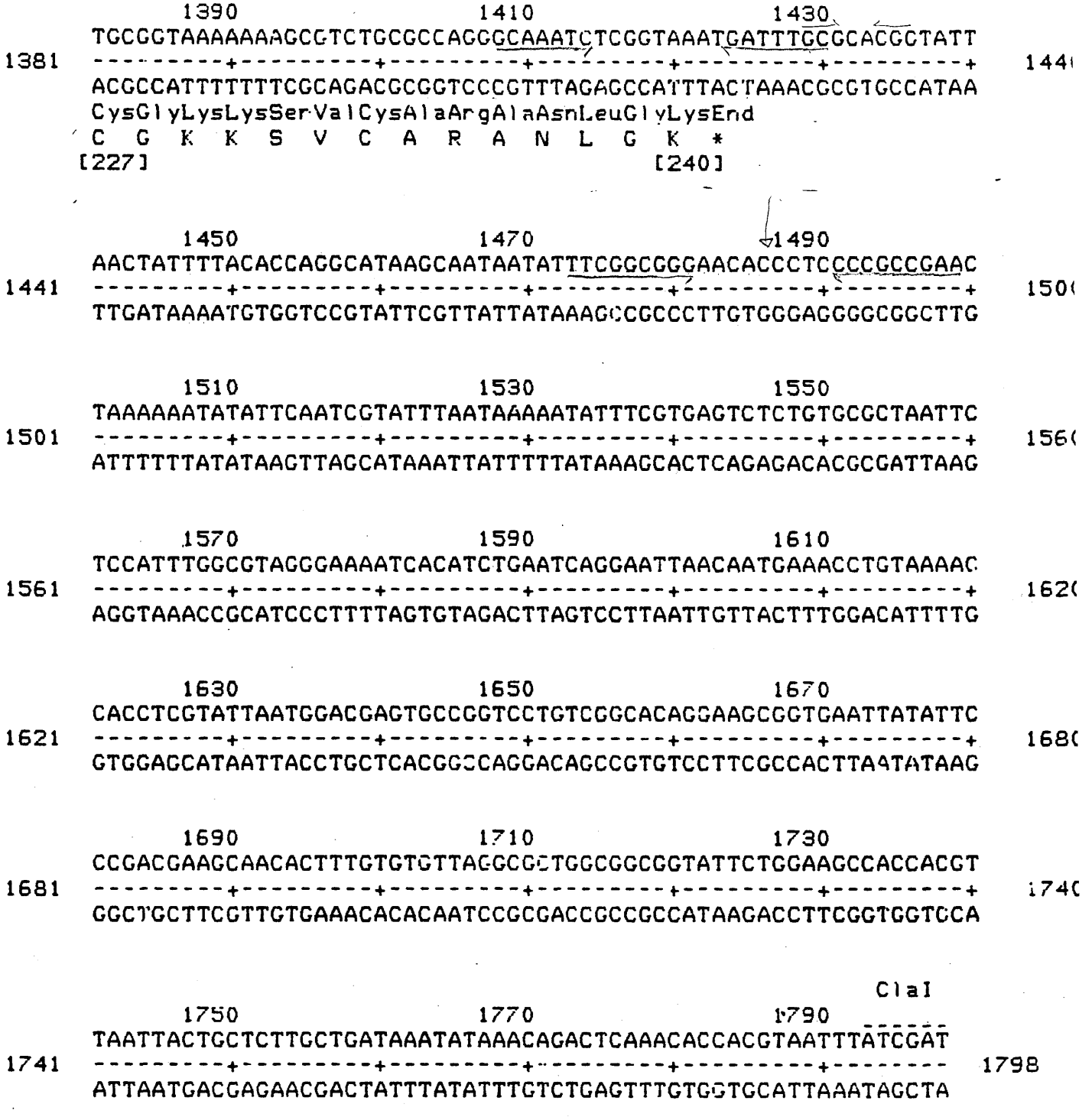



Figure 4.7:(cont)

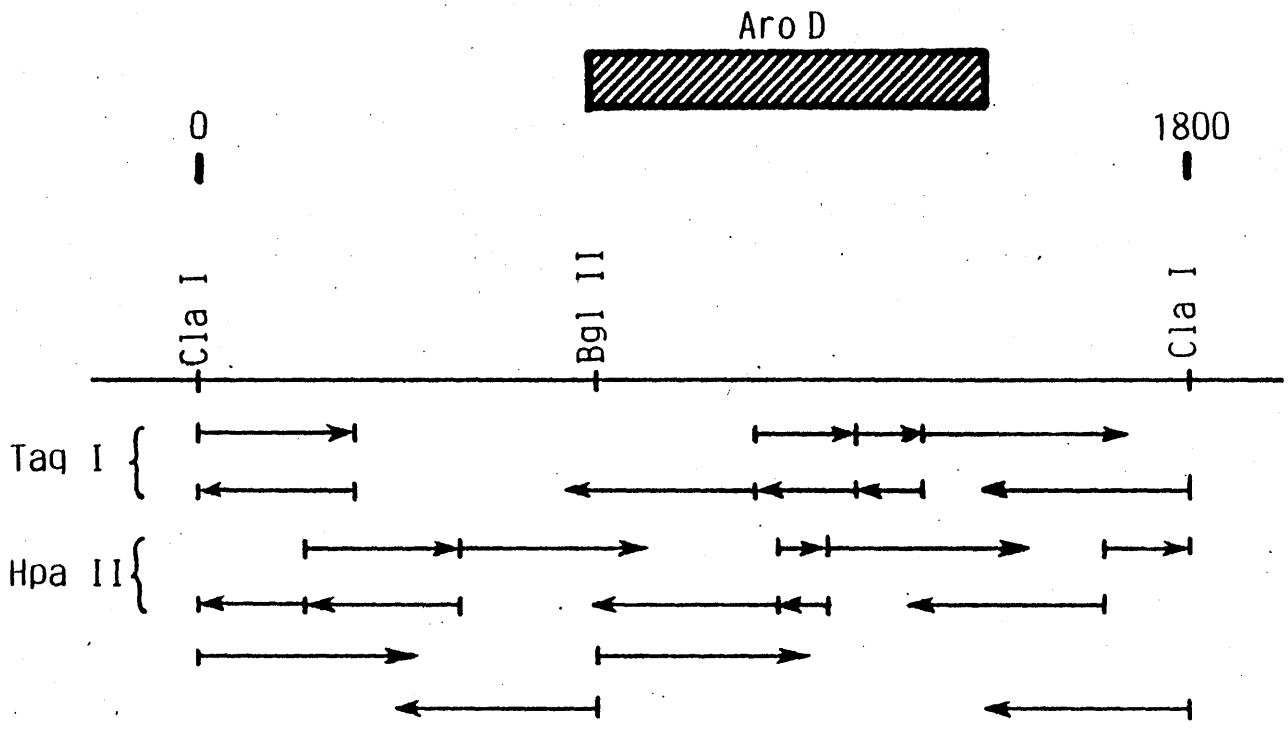



Figure 4.10: Sequencing strategy for the 1.8 kb ClaI insert in pKD201, and location of the aroD coding sequence. Arrows indicate direction and extent of sequences.

(b) Application of Fickett's "Testcode"

The program TESTCODE in the Wisgen package (Section 2.36) was applied to both strands of the DNA sequence. TESTCODE searches for genes by plotting the running value of Fickett's "Testcode" statistic (Fickett, 1982) along the length of the sequence. The testcode statistic is a measure of the 'period three constraint' of each region of a DNA molecule; it is independent of the reading frame and is based on measurements of the period three constraints in an entire database for regions thought to be coding and non-coding.

The resulting plots (Figures 4.8a and 4.8b) are divided into three regions for which the statistic makes a prediction. The top is the region within which 95% of coding regions fall and within which 95% of the non-coding regions do not fall. The bottom region is the opposite. The middle ground is the 'window of vulnerability' where the statistic does not have a significant ability to predict coding and non-coding regions.

Figure 4.8a shows that the open reading frame falls into the region where 95% of coding regions fall, confirming that this probably represents the aroD gene sequence. It is interesting to note that a second open reading frame (data not shown) which begins at the ClaI site and runs to position 516 (Figure 4.7) is also likely to be protein coding i.e. the 3' end of a gene. The identity of this gene is unknown.

Figure 4.8b shows that there are no significant protein coding regions on the other strand.

Figure 4.8: Application of Fickett's Testcode to the sequence of the 1.8 kb ClaI fragment.

(a) The DNA strand shown as 0-1800 in Figure 4.7, including the predicted aroD coding region.

(b) The complementary strand; no significant open reading frames were observed in this sequence.

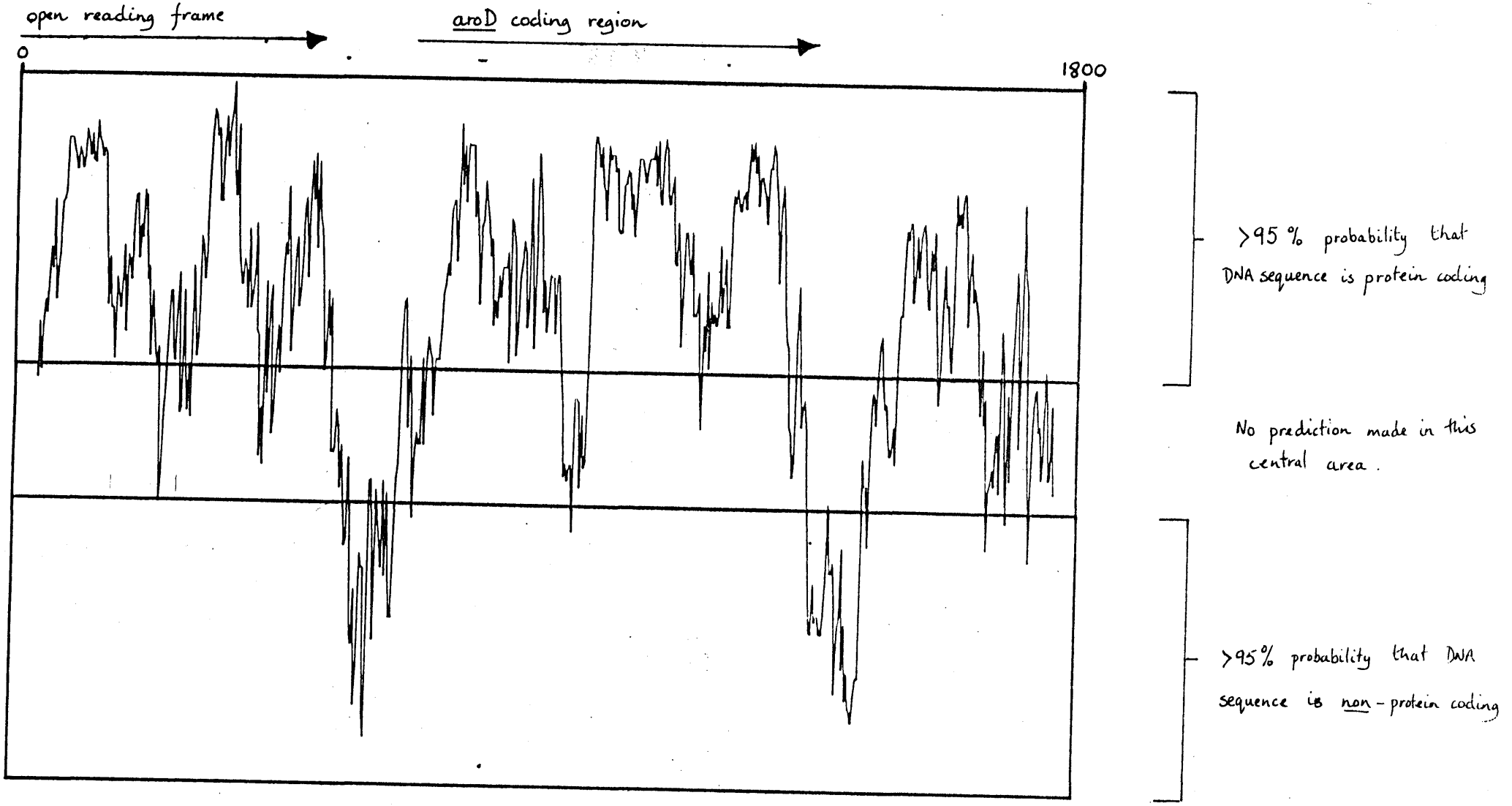

open reading frame
aroD coding region
0
1800
>95 % probability that DNA sequence is protein coding
No prediction made in this central area.
>95% probability that DNA sequence is non-protein coding

Figure 4.8a.

Non-coding strand — no open reading frames

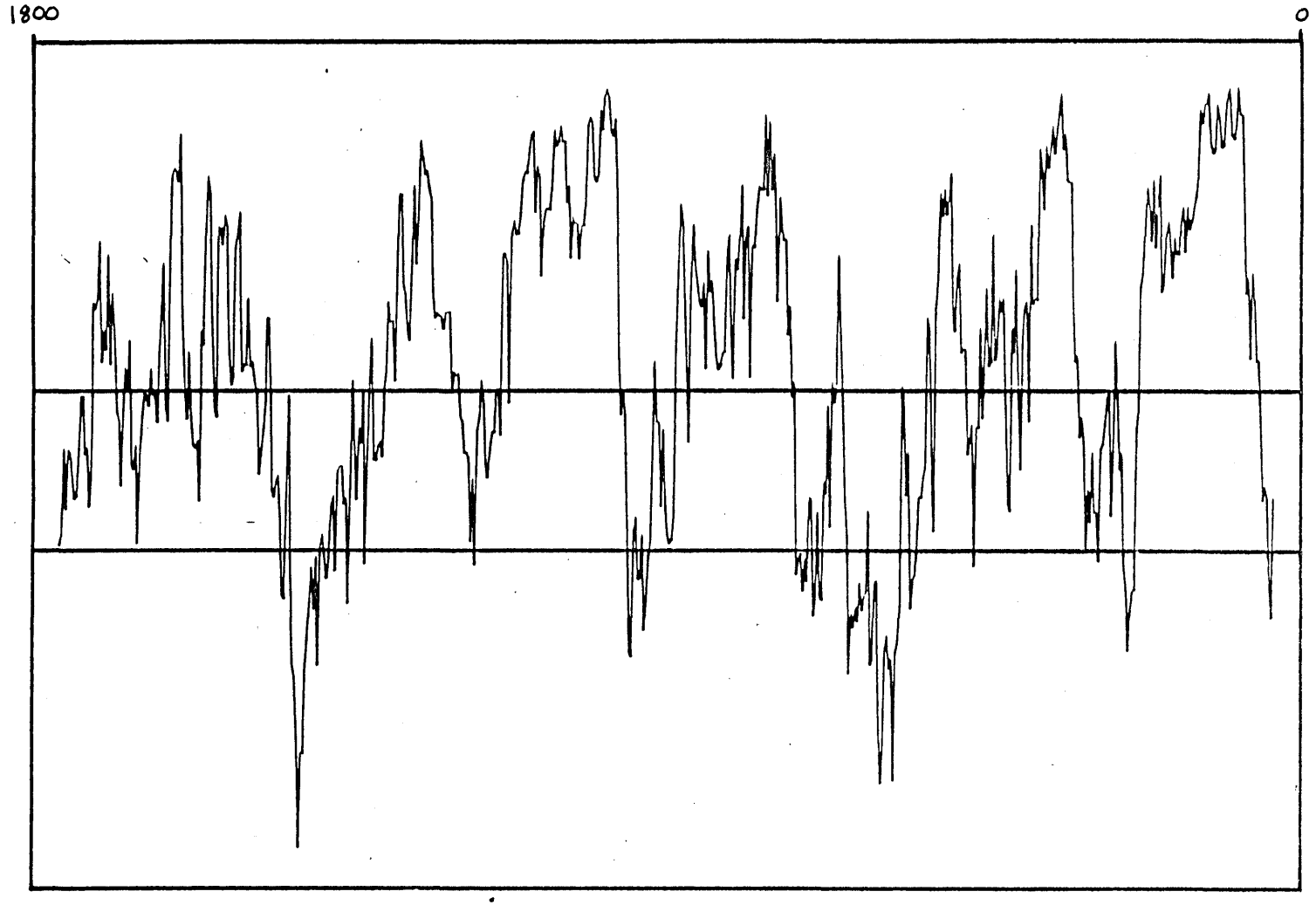



Figure 4.8b

### 4.4.2 Does the open reading frame encode 3-dehydroquinase?

#### (a) Size and composition of the polypeptide encoded in the open reading frame

Taking the ATG/Met codon at position 703 in the sequence as the start of the open reading frame, then the gene encodes a polypeptide of 240 amino acids, ending with a UGA/stop codon at position 1423.

The amino acid sequence of the putative *aroD* gene is shown in Figure 4.7. From this sequence, the amino acid composition of 3-dehydroquinase has been predicted, and is shown in Table 4.3. From this data a molecular weight of 28117 can be calculated for the protein. This compares favourably with the estimated $M_r$ of 29000 obtained for the purified *E.coli* enzyme by SDS-PAGE (Chaudhuri & Coggins, 1984b).

#### (b) Codon utilisation in the open reading frame

Recent comparisons of a number of *E.coli* structural genes have shown that there is a non-random utilisation of codons in protein coding sequences. The bias in genetic code usage has two main components. Firstly there is a correlation with tRNA availability in the cell and secondly the choice of 'codewords' generally favours codon-anticodon interactions of intermediate strength over those involving very strong or very weak interactions (Grosjean & Fiers, 1982; Gouy & Gautier, 1982). This bias in codon usage can be used to predict whether or not an open reading frame in the DNA sequence is likely to be a protein coding sequence. The codon utilisation for the open reading frame/putative

| Amino acid | Number predicted by DNA sequence |
|---|---|
| Asp | 17 |
| Thr | 15 |
| Ser | 16 |
| Glu | 18 |
| Pro | 7 |
| Gly | 13 |
| Ala | 29 |
| Cys | 3 |
| Val | 17 |
| Met | 11 |
| Ile | 16 |
| Leu | 19 |
| Tyr | 6 |
| Phe | 8 |
| His | 6 |
| Lys | 17 |
| Arg | 10 |
| Trp | 2 |
| Asn | 4 |
| Gln | 6 |

Table 4.3 Amino acid composition of 3-dehydroquinase predicted from the DNA sequence.

aroD sequence (Figure 4.7) is listed in Table 4.4. The analogous data for four enzymes of the tryptophan biosynthetic pathway, namely trpE, trpD, trpC and trpB has been pooled and is listed in Table 4.5, for comparison (data taken from Table 2 of Gouy & Gautier, 1982). The two tables show a very similar pattern of codon utilisation, which suggests very strongly that the putative aroD sequence actually encodes a protein. One might expect the utilisation patterns to be similar since all the above mentioned genes encode proteins which fulfil a similar biosynthetic role within the cell and will be expressed at approximately the same level during the synthesis of tryptophan.

Examples of bias include: avoidance of the ATA codon for isoleucine; preference for CGT/C over CGA/G or AGA/G for arginine; preference for GGT/C over GGA/G for glycine (see Table 4.4).

Further aspects of the codon utilisation of aroD will be discussed in Chapter 6.

### 4.4.3 Other features associated with a monocistronic E.coli gene

(a) Ribosome binding site

The initiation of protein synthesis in bacteria involves the specific binding of the small ribosome subunit to a region of the mRNA containing the initiation codon. The initiating subunit selects only the AUG or GUG codon at the beginning of a cistron, ignoring the many internal AUG or GUG codons. Recognition involves the untranslated

| | T | C | A | G | |
|---|---|---|---|---|---|
| T | PHE 5 | SER 3 | TYR 5 | CYS 0 | T |
| | PHE 3 | SER 5 | TYR 1 | CYS 3 | C |
| | LEU 1 | SER 0 | Term - | Term - | A |
| | LEU 1 | SER 2 | Term - | TRP 2 | G |
| C | LEU 1 | PRO 2 | HIS 4 | ARG 7 | T |
| | LEU 5 | PRO 0 | HIS 2 | ARG 2 | C |
| | LEU 0 | PRO 2 | GLN 2 | ARG 0 | A |
| | LEU 11 | PRO 3 | GLN 4 | ARG 2 | G |
| A | ILE 9 | THR 3 | ASN 2 | SER 2 | T |
| | ILE 7 | THR 8 | ASN 2 | SER 4 | C |
| | ILE 0 | THR 0 | LYS 16 | ARG 0 | A |
| | MET 11 | THR 4 | LYS 1 | ARG 1 | G |
| G | VAL 2 | ALA 3 | ASP 11 | GLY 5 | T |
| | VAL 6 | ALA 13 | ASP 6 | GLY 8 | C |
| | VAL 4 | ALA 6 | GLU 11 | GLY 0 | A |
| | VAL 5 | ALA 7 | GLU 7 | GLY 0 | G |

Table 4.4 Codon utilisation in the E.coli aroD gene.

| | T | C | A | G | |
|---|---|---|---|---|---|
| T | PHE 68 | SER 30 | TYR 76 | CYS 27 | T |
| | PHE 67 | SER 39 | TYR 40 | CYS 34 | C |
| | LEU 57 | SER 21 | Term - | Term - | A |
| | LEU 52 | SER 42 | Term - | TRP 12 | G |
| C | LEU 43 | PRO 20 | HIS 62 | ARG 77 | T |
| | LEU 43 | PRO 25 | HIS 62 | ARG 122 | C |
| | LEU 21 | PRO 28 | GLN 70 | ARG 9 | G |
| | LEU 244 | PRO 115 | GLN 121 | ARG 7 | G |
| A | ILE 113 | THR 22 | ASN 59 | SER 33 | T |
| | ILE 83 | THR 99 | ASN 91 | SER 68 | C |
| | ILE 2 | THR 27 | LYS 115 | ARG 7 | A |
| | MET 91 | THR 40 | LYS 26 | ARG 2 | G |
| G | VAL 62 | ALA 71 | ASP 133 | GLY 116 | T |
| | VAL 51 | ALA 154 | ASP 81 | GLY 127 | C |
| | VAL 31 | ALA 70 | GLU 187 | GLY 27 | A |
| | VAL 113 | ALA 162 | GLU 62 | GLY 47 | G |

Table 4.5 Codon utilisation for a number of the *trp* biosynthetic enzyme structural genes from *E.coli*, namely *trpE*, *trpD*, *trpC* and *trpB*.

sequence of the mRNA to the 5' side of the initiator codon. Shine & Dalgarno (1975) showed that there is a purine-rich tract in the ribosome binding site of different bacterial mRNAs which shows complementarity to the 3' end of 16S ribosomal RNA, and that it is the specific pairing of mRNA to 16S rRNA which allows the initiation of translation to take place.

It has been proposed that the sequence 5'-ACCUCCU-3' in 16S rRNA is the site at which pairing takes place. A consensus sequence for the recognition region has been derived from a knowledge of a number of E.coli mRNAs; the sequence 5'-AGGAG-3' (Shine-Dalgarno or S-D sequence) precedes structural genes at a distance of 4-6 bp from the initiation AUG or GUG codon. The control of initiation of translation may be mediated by the degree of complementarity of the mRNA and rRNA and by the accessibility of the purine-rich segment of the message to the terminus as determined by the secondary structure of the initiation region.

The DNA sequence upstream of the proposed aroD coding region is illustrated in Figure 4.9. The region to the 5'-side of the putative AUG/Met initiation codon is very purine rich. A possible S-D sequence is located 8 nucleotides before the start of the gene and is indicated on Figure 4.9. This provides further evidence that the open reading frame encodes a gene and that the AUG/Met codon is the correct initiation point. The next Met codon is located at position 769 and is not preceded by a purine rich tract.

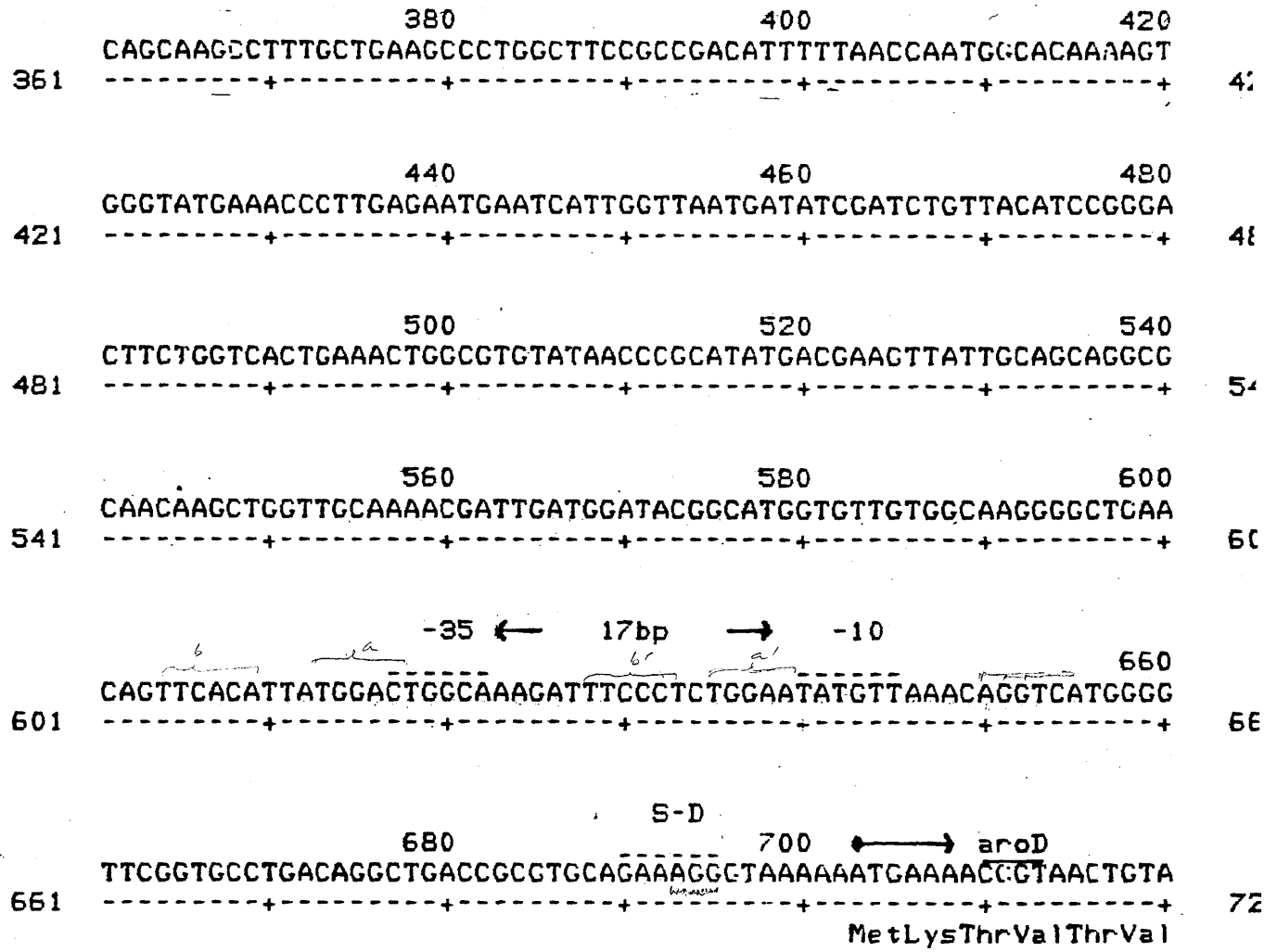



Figure 4.9: The DNA sequence upstream of the aroD coding region showing the location of a possible promoter sequence.

(b) Promoter

Gene expression in E.coli is controlled, in part, at the level of RNA transcription. Transcriptional regulation can be achieved by modulating the efficiency with which RNA polymerase can recognise and interact with initiation sites (promoters) on the DNA.

Promoters are expected to share common structural features, reflecting a similar interaction with RNA polymerase. Comparison of a large number of E.coli promoters has shown that they contain two regions of conserved DNA sequence, located about 10 and 35 base pairs upstream of the transcription startpoint (Rosenberg & Court, 1979; Hawley & McClure, 1983). However there is also considerable sequence diversity which is related both to the wide range of initiation frequencies and to the partial overlap of binding sites for transcriptional control proteins.

The first conserved region is located about 10 bp upstream of the mRNA start point and is known as the '-10 region' or 'Pribnow box' (Pribnow, 1975). The consensus Pribnow box sequence is TATAAT. Within this sequence, however, some residues are more highly conserved than others. The final T is very well conserved. Four and five bp upstream, two other strongly conserved bases (T and A respectively) are found. The three residues between these sites, TAA, are also conserved, though less stringently.

The second conserved region is located about 35 bp upstream of the mRNA start point (the '-35 region' or recognition region). In order to align -35 regions with

-10 regions, it is necessary to allow a $\pm$2 residue flexibility in the spacer region between them. The homology between promoters consists of a highly conserved TTG trinucleotide upstream and adjacent to a less conserved ACA sequence.

Comparison of many mRNA startpoints indicates that initiation of transcription takes place predominantly from a single position some 6 to 9 bp beyond the final T of the -10 region. The starting nucleotide is usually an A, sometimes a G, but only very rarely a C or a T.

The region upstream of the putative aroD coding sequence (Figure 4.9) was scanned by eye for promoter like sequences. Initially, a search was made for the highly conserved hexanucleotide TA---T, as expected for the -10 region. In the 300 bp preceding the coding region, only one such sequence was found, at position 640. This sequence TATGTT is similar to the consensus TATAAT. 17 bp further upstream the sequence CTGGCA occurs; this is homologous to the consensus -35 region, TTGACA. The mRNA startpoint could then be either of the G residues at positions 651 and 652, both of which are within the allowed 6 to 9 bp from the final T of the -10 region. Some startpoints outwith this spacing are known, and so the actual transcription start could be the A residue at position 650, or the A residue at position 655. The latter A residue is in fact at the centre of the weakly conserved triplet CAT, often found around the A initiation point.

This search for the aroD promoter is by no means exhaustive. More rigorous methods of searching for a promoter have been developed, making use of a weighted analysis of each residue in a sequence (e.g. Staden, 1984), but the computer programs were not available to us.

There is as yet no evidence, either in vitro or in vivo to support the prediction outlined above for the location of the aroD promoter. Further experiments such as S1 nuclease mapping of the plasmid encoded transcripts would provide hard evidence for the proposals.

(c) Terminator

Comparisons of the site of termination of transcription have been made for a number of E.coli genes (Rosenberg & Court, 1979). Three common features are found at termination sites:

(i) an inverted repeat sequence

(ii) U residues are found at the 3' terminus of the transcript

(iii) G/C rich sequences are found preceding the stopsite.

The sequence to the 3'-side of the putative aroD coding region was scanned, again by eye, for the presence of an inverted repeat. One inverted repeat was located centred at position 1486. This has a stem of 9 bp and a 9-base loop. If this forms part of the terminator of aroD transcription, then this sequence fulfils two of the features of a terminator detailed above. The repeat is not followed by a run of T residues. Again, experimental evidence is

required to reveal the true point of termination of the aroD transcript.

## 4.5 Future prospects

It is now possible to compare the sequence of 3-dehydroquinase from E.coli with the sequence of the catabolic dehydroquinase of Neurospora, the product of the qa-2 gene (Alton et al., 1982; Dr M. Case, unpublished results), and with the peptide sequence around the active site lysine of the Neurospora arom complex (Dr S. Choudhuri, unpublished results). This will be discussed in Chapter 7.

The availability of 3-dehydroquinase in large amounts from an overproducing strain of E.coli (Dr S. Choudhuri, unpublished results) will allow the protein chemistry of the active site of the enzyme to be investigated. A knowledge of the amino acid sequence of the enzyme will assist this investigation.

## 4.6 Note added in proof

Since the time of writing, the amino-terminal sequence and the amino acid composition of 3-dehydroquinase have been determined (Murray Campbell, unpublished results). The N-terminal sequence confirms that the aroD coding region begins with the Met at position 703, and the amino acid composition is in good agreement with that predicted by the DNA sequence (data not shown).

# CHAPTER 5 SUB-CLONING AND ANALYSIS OF THE E.coli aroA GENE

## 5.1 Introduction

### 5.1.1 The source of the aroA gene for this study

The E.coli ribosomal protein S1 is encoded by the gene rpsA, which is located at minute 20.5 on the E.coli genetic map, close to both aroA and serC (Bachmann, 1983; Figures 1.3 & 1.4). As part of a study of the structure and function of this protein, the rpsA gene has been isolated and its entire DNA sequence has been determined (Kitakawa et al., 1980; Schnier & Isono, 1982).

The rpsA gene has no readily selectable phenotype and so its proximity to the gene serC was utilised to obtain a collection of phage which carried the region of the E.coli chromosome around min 20. Plaque-forming λ phage were isolated after induction of an E.coli K12 λ lysogen. Those phage which, because of abherrant excision, had picked up a region of the E.coli chromosome were identified by their ability to complement the serC mutation, following infection of E.coli KL282. SerC is a selectable marker; mutants require serine and pyridoxine for growth and so relief of auxotrophy was used to isolate a series of transducing phage carrying serC.

In vitro transcription of non-λ genes and two dimensional gel electrophoresis of the products was used to show that one of the phage, hereafter named λ pserC also synthesised a 65kdal protein with identical properties to ribosomal protein S1. Further proof that the phage carried the

*serC-rpsA* region of the *E.coli* chromosome was provided by infection of an *E.coli aroA* mutant - *E.coli* AB2829. Again, relief of auxotrophy was used to show that *aroA* was present on the phage, as expected, since *aroA* is located between *serC* and *rpsA* (Kitakawa *et al.*, 1980).

### 5.1.2 The genomic DNA insert in λpserC

The genomic DNA inserted in λ*pserC* has been extensively mapped with a number of restriction enzymes (Schnier & Isono, 1982; Figure 5.1). λ*pserC* DNA was digested with a number of enzymes in turn, which were known to cleave the genomic DNA insert, and the DNA was transcribed and translated *in vitro*. Translation products were subjected to gel electrophoresis and enzymes identified which abolished a band corresponding to protein S1. In this way, the *rpsA* gene was located. It was then sub-cloned and the entire DNA sequence of the gene determined.

Figure 5.1 shows the restriction map of the *serC-aroA* region and the location of the *rpsA* gene. This diagram also illustrates the conventional view of the map adopted throughout this work i.e. 'left' or 'right' is with respect to this map. Attention is also drawn to the regions designated A-C.

*E.coli* strain KL282, lysogenic for λ*pserC* was obtained from Dr K. Isono, Max-Planck-Institut fur Molekulare Genetik, Berlin, FRG. In this lysogen, the λ phage is present at one copy per chromosome, and hence is providing only one extra copy of the *aroA* gene. To increase the

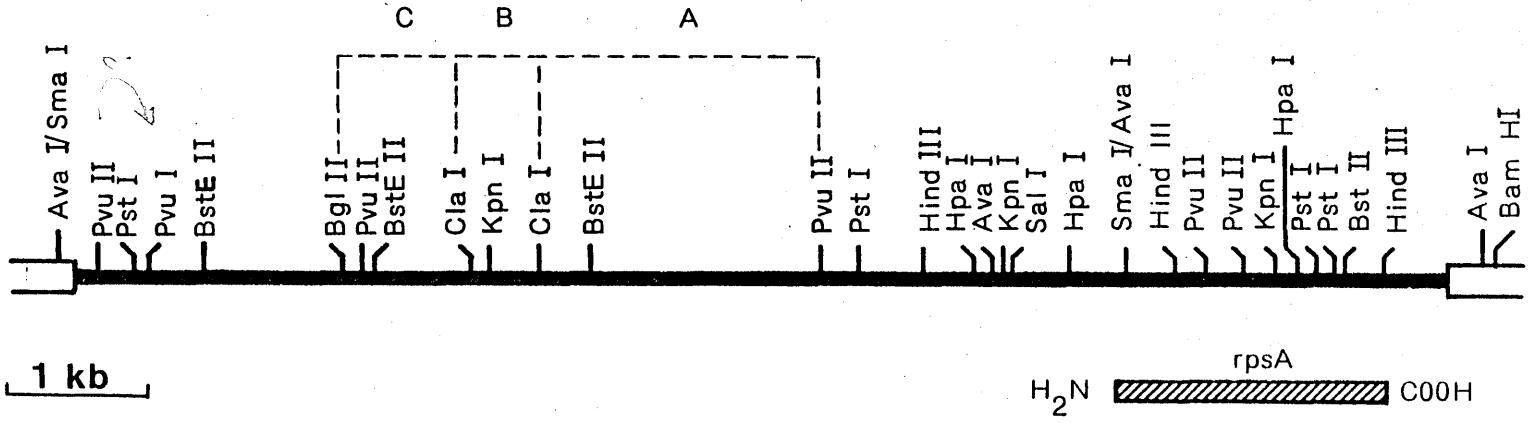



Figure 5.1: Restriction map of the genomic insert in λpserC (from Schnier & Isono, 1982), and location of the rpsA gene. Note the regions designated A, B and C; and the orientation of the map - left and right in the text refer to this map.

copy number of aroA and hence to overexpress EPSP synthase, the gene must be sub-cloned onto multi-copy plasmid vectors; pAT153 and pBR322 have been used for this purpose.

## 5.2 Sub-cloning and analysis of aroA

### 5.2.1 Construction of pKD501 and pKD502

Examination of the restriction map in Figure 5.1 reveals that there is a PstI fragment of approximately 4.5 to 5.0 kb in size which carries most of the genomic material of λpserC but which does not encode rpsA. Sub-cloning this fragment gave the best chance of transferring aroA onto a plasmid, without disrupting the gene in any way.

λpserC was purified after heat induction of the E.coli KL282 lysogen, as described in Section 2.9. DNA extracted from the phage was digested with PstI and the fragments separated by electrophoresis on a 0.7% LMT agarose gel (Figure 5.2). Three bands were observed in positions corresponding to the size range 4.5 - 5.0 kb. The individual bands were not completely resolved and so all three were excised as a group. DNA was extracted and ligated to PstI cleaved and phosphatase treated pAT153. The ligation mix was used to transform E.coli AB2829 and the plasmid transformed cells selected by growth overnight on LA/tet. Colonies obtained in this way were replica plated onto minimal medium and incubated a further 24 h at 37°. Because of the auxotrophic mutation in E.coli AB2829, only colonies transformed with recombinant plasmids carrying aroA would grow under these conditions. From the many colonies which grew

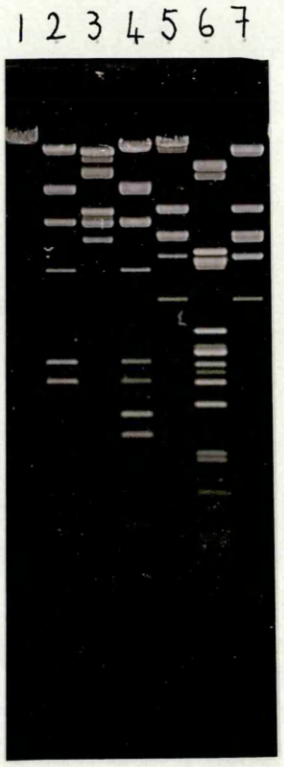



Figure 5.2: Restriction enzyme digestion of λpserC

| Track | DNA | Enzyme | |
|---|---|---|---|
| 1 | λpserC | undigested | |
| 2 | λ | HindIII | (wild type) |
| 3 | λpserC | BamHI | |
| 4 | λpserC | HindIII | |
| 5 | λpserC | EcoRI | |
| 6 | λpserC | PstI | |
| 7 | λ | EcoRI | (wild type) |

on minimal medium, 100 were picked and checked for ampicillin sensitivity; all were found to be sensitive, indicating an insertion at the PstI site. Plasmid DNA was extracted from 10 $tet^R$, $amp^S$ colonies by the mini-prep technique and digestion with various restriction enzymes established that all the clones carried the anticipated 4.6 kb PstI fragment. Two clones (pKD501 and pKD502) were retained for further examination; each carried the same fragment, but in opposite orientations (Figure 5.3).

Large scale plasmid DNA preparations of pKD501 and pKD502 were made (Section 2.7) and the DNA digested with various mixtures of restriction enzymes in order to confirm the restriction map shown in Figure 5.1, and to establish that the PstI insert originated from λpserC. Figure 5.4 shows one of the gels obtained. The pattern obtained for each enzyme confirmed that expected from the restriction map and from the map of pAT153, except in the case of PvuII. Figure 5.1 locates two PvuII sites within the PstI fragment, but the digests clearly showed that three fragments were produced. There are no PvuII sites in pAT153. Further mapping has shown that there is a PvuII site to the right of the leftward PstI site; it is possible that the PvuI site on Figure 5.1 at this location should be a PvuII site.

### 5.2.2 Overexpression of EPSP synthase in cells carrying pKD501

The expression of EPSP synthase from cells carrying

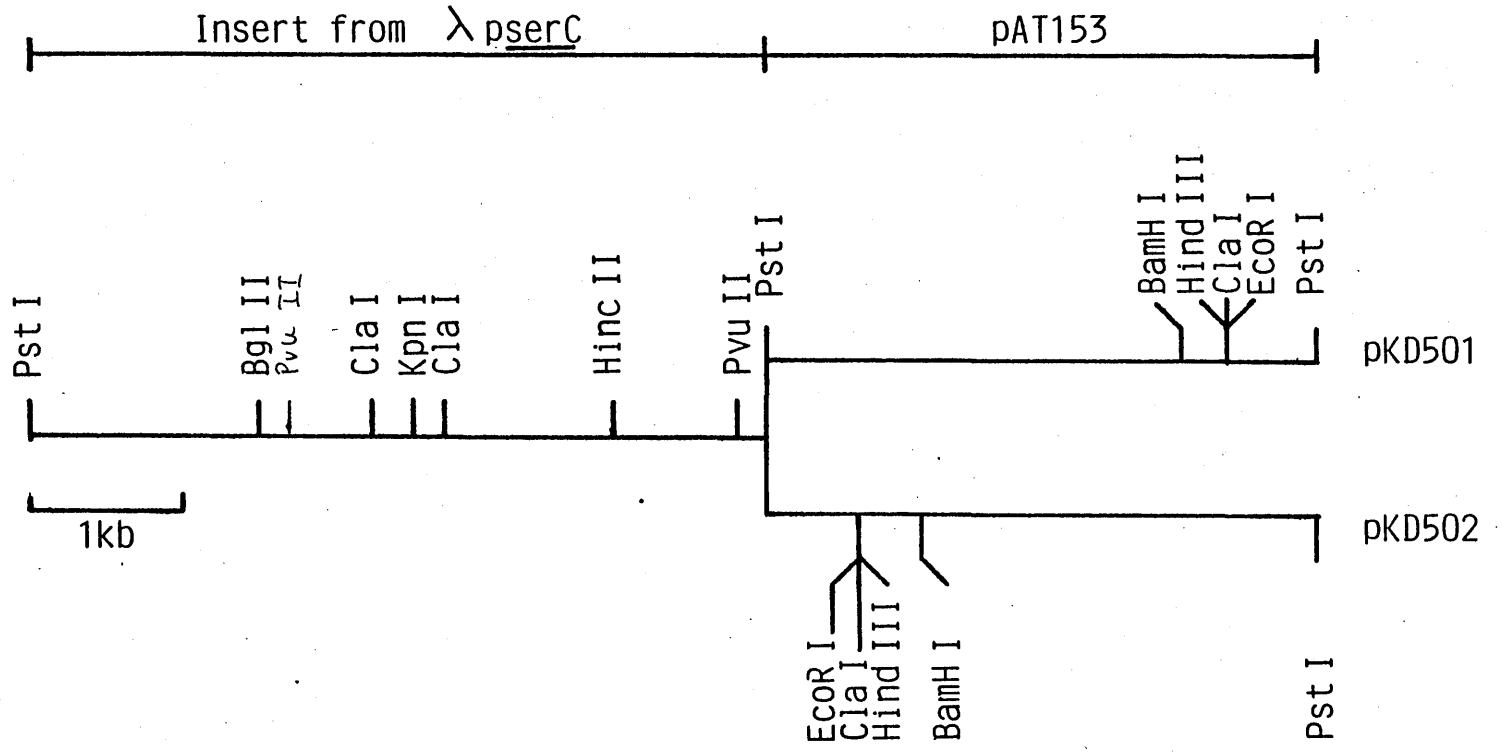



Figure 5.3: Orientation of the PstI insert in pKD501 and pKD502. The detailed restriction map of the insert is shown in Figure 5.1; the map of pAT153 is derived from Twigg & Sherratt (1980) and the DNA sequence of pBR322 (Sutcliffe, 1979).

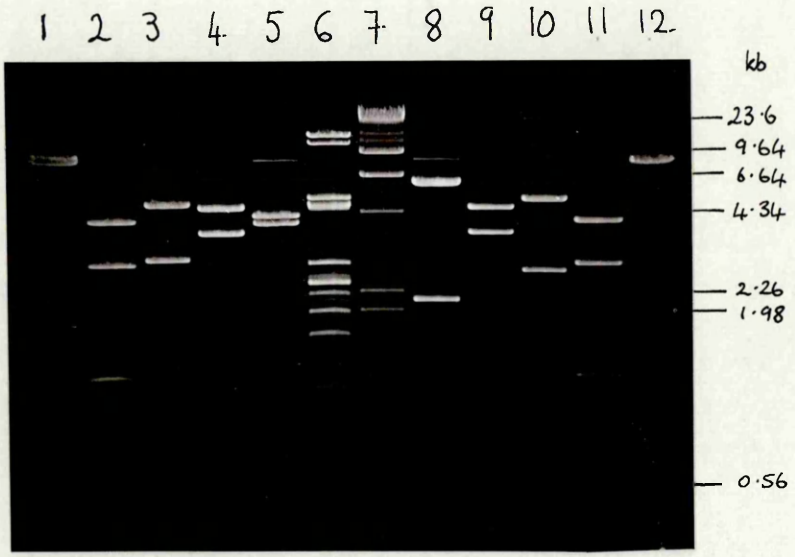



Figure 5.4: Restriction enzyme digestion of pKD501 and pKD502

| Track | DNA | Enzyme |
|---|---|---|
| 1 | pKD502 | HindIII |
| 2 | pKD502 | PvuII |
| 3 | pKD502 | ClaI |
| 4 | pKD502 | PstI |
| 5 | pKD502 | HindIII + BglII |
| 6 | λpserC | PstI |
| 7 | λ | HindIII |
| 8 | pKD501 | HindIII + BglII |
| 9 | pKD501 | PstI |
| 10 | pKD501 | ClaI |
| 11 | pKD501 | PvuII |
| 12 | pKD501 | HindIII |

various aroA recombinant plasmids will be discussed in greater detail in Section 5.2.7. In order to establish that E.coli AB2829/pKD501 can be used to overproduce the enzyme, EPSP synthase activity was measured in crude extracts of E.coli K12 and E.coli AB2829/pKD501. Cells were grown in liquid minimal medium to saturation ($A_{650}$= 1.3-1.5) and sonicated in sonication buffer 'B'. The specific activity of EPSP synthase in crude extracts of E.coli AB2829/pKD501 was 0.42 U/mg, which is over 100-fold higher than that found in E.coli K12 strain ATCC 14948 (0.004 U/mg).

The high specific activity of EPSP synthase in E.coli AB2829/pKD501 has led to the use of this strain for the preparation of EPSP synthase. On a large scale, cells are grown in 500 ml batches in 2.5 l conical flasks on an orbital shaker. Minimal medium is always used; cells are grown to saturation before harvesting.

A loopful of -20° glycerol stock E.coli AB2829/pKD501 is inoculated into 10 ml LA/tet and shaken overnight. 1 ml is transferred into 100 ml minimal medium and shaken a further 24 h. 5 ml is then used to inoculate the 500 ml flasks.

### 5.2.3 Is plasmid encoded EPSP synthase identical to the 'wild type' enzyme?

Since the overproduced enzyme was required for protein chemistry and enzymology it was essential to demonstrate that it was identical to that purified from the wild type E.coli K12 strain.

EPSP synthase has been purified to homogeneity from both E.coli K12 (Lewendon & Coggins, 1983) and from E.coli AB2829/pKD501 (Duncan, Lewendon & Coggins, 1984a). The enzyme purified from E.coli AB2829/pKD501 is indistinguishable from that isolated from untransformed E.coli on the basis of:

(i) P.A.G.E. under native conditions and in the presence of SDS.

(ii) one-dimensional peptide mapping after partial S.aureus V8 protease digestion.

(iii) fractionation on a TSK G2000 SW gel permeation column.

(iv) kinetic parameters.

(data obtained by Ann Lewendon, and described in detail in Duncan, Lewendon & Coggins, 1984a).

It is now possible to produce milligram quantities of homogeneous EPSP synthase. The availability of large quantities of pure, active protein has facilitated further mechanistic and structural studies of this enzyme (Ann Lewendon, Ph.D. Thesis, University of Glasgow, 1984).

### 5.2.4 Growth of E.coli in the presence of Glyphosate

It has previously been shown that the growth of E.coli is inhibited in the presence of the herbicide glyphosate and that the enzymatic step involved is that catalysed by EPSP synthase (Section 1.4.6). This suggests that glyphosate tolerant cells might be obtained as follows:

(i) by the production of a structurally altered form of the enzyme, so that the herbicide has a lowered affinity for the enzyme, binding less efficiently than substrate;

(ii) by overproduction of the EPSP synthase from an amplified aroA gene, requiring more herbicide to inhibit growth.

The aroA gene is expressed constitutively in E.coli (Tribe et al., 1976) and so the gene will be overexpressed in cells carrying a multi-copy plasmid by a factor similar to the copy number of the plasmid. This has already been shown to be the case with aroA cloned into pAT153 (pKD501) where a 100 fold higher level of EPSP synthase is found.

In order to test the hypothesis that amplification of aroA will make E.coli cells glyphosate tolerant, E.coli K12 was tested for growth on minimal agar containing a range of concentrations of glyphosate, with and without the aroA recombinant plasmid.

E.coli K12 and E.coli K12/pKD501 were grown in LB and LB/tet respectively to an $A_{650}$ = 0.5 (mid log phase). Cultures were harvested, washed in M9 salts and resuspended in the original volume of M9 salts. 100 µl aliquots (cell density approximately $10^8$ cells/ml) were spread onto MM, MM containing glyphosate and MM + aromatics + glyphosate plates (analytical grade glyphosate was sterilised by filtration before adding the appropriate amount to medium at 55°. Glyphosate was a gift from Dr S. Ridley, I.C.I. Jealott's Hill).

The concentrations of glyphosate used and the results obtained are shown in Table 5.1. These results show that on MM, growth of E.coli K12 is substantially inhibited by 3 mM-glyphosate and is completely stopped by 5 mM-glyphosate. Transformation of E.coli with pKD501 increases tolerance so that near normal growth is obtained even on 50 mM-glyphosate.

Minimal medium + aromatics contains the end products of aromatic biosynthesis and so growth of E.coli K12 and E.coli AB2829 on this medium should be unaffected by the presence of glyphosate. Clearly this is not so. Growth is affected in both cases, though to a greater extent for E.coli AB2829. This may be due to secondary effects of glyphosate on the metabolism of E.coli and unconnected with the effect on EPSP synthase.

The results presented here are qualitative, but they clearly illustrate that amplification of the aroA gene confers resistance to glyphosate. In this respect, the aroA gene can be used as a selectable marker on plasmids in the same way as ampicillin or tetracycline resistance. This might have its uses in larger scale growth of plasmid carrying cells where a selection must be maintained. It is both difficult and costly to remove antibiotics from growth media before disposal. Glyphosate poses no such problems. It has no effects on mammalian systems and is bio-degradable.

In order to quantify the effects of glyphosate on the growth of E.coli it is necessary to measure growth rates

| (glyphosate) (mM) | Minimal Plates | | Minimal + aromatics Plates | |
|---|---|---|---|---|
| | *E.coli* K12 | *E.coli* K12/pKD501 | *E.coli* K12 | *E.coli* AB2829 |
| 0 | +++ | +++ | +++ | +++ |
| 1 | +++ | +++ | +++ | +++ |
| 2 | +++ | +++ | +++ | +++ |
| 3 | ++ | +++ | +++ | +++ |
| 4 | + | +++ | +++ | ++ |
| 5 | - | +++ | +++ | ++ |
| 20 | - | +++ | ++ | +/- |
| 30 | - | +++ | | |
| 40 | - | +++ | | |
| 50 | - | ++ | | |

Table 5.1: Growth of *E.coli* in the presence of glyphosate

Key:
- +++ confluent growth overnight
- ++ slight growth overnight
- + growth after 2 day incubation
- +/- slight growth after 2 day incubation
- - no growth

in liquid medium under constant conditions and varying glyphosate concentrations. Some experiments of this kind have been carried out by workers at Monsanto Co., St. Louis, USA (Rogers et al., 1983). They have shown that E.coli cells harbouring a recombinant aroA carrying plasmid, pMON4, are able to grow in the presence of 40 mM-glyphosate. This is eight times the concentration required to give a very much reduced growth rate for E.coli cells carrying a plasmid lacking the aroA gene.

5.2.5 Construction of pKD503, pKD504 and pKD505

The PstI fragment inserted in pKD501 is 4.6 kb in length. This is considerably larger than the size required to encode the aroA gene (calculated to be 1.2 - 1.4 kb, from the $M_r$ of 49 000 estimated for EPSP synthase by SDS-PAGE (Lewendon & Coggins, 1983)).

If the aroA gene could be more precisely located within the 4.6 kb DNA fragment then the amount of DNA sequencing required to obtain the gene sequence would be considerably reduced.

The 'shotgun' cloning approach was used to produce a series of derivatives of λpserC carrying the aroA gene. λpserC DNA was cleaved with a pair of different enzymes and the DNA fragments randomly cloned into pAT153 which had been cut with the same enzymes, or enzymes producing the same complementary ends (N.B. since the recognition sequence of AvaI is G↓PyCGPuG, the internal redundancy means that there is only a one in four chance of an AvaI end cloning into the AvaI site of pAT153).

For each enzyme combination: 1µg of λ*pserC* DNA was digested with a pair of enzymes and mixed with 100 ng of pAT153, digested with enzymes producing the same complementary ends. The mixture was immediately phenol/chloroform extracted and ethanol precipitated. Ligation mixes were set up and incubated overnight at 14° before being used to transform *E.coli* AB2829. Transformed cells were selected by growth on LA/amp, replica plated onto MM and incubated a further 24 h. Six colonies from the MM plate were each inoculated into 10 ml LB/amp and after overnight growth, plasmid DNA was prepared by the mini-prep technique.

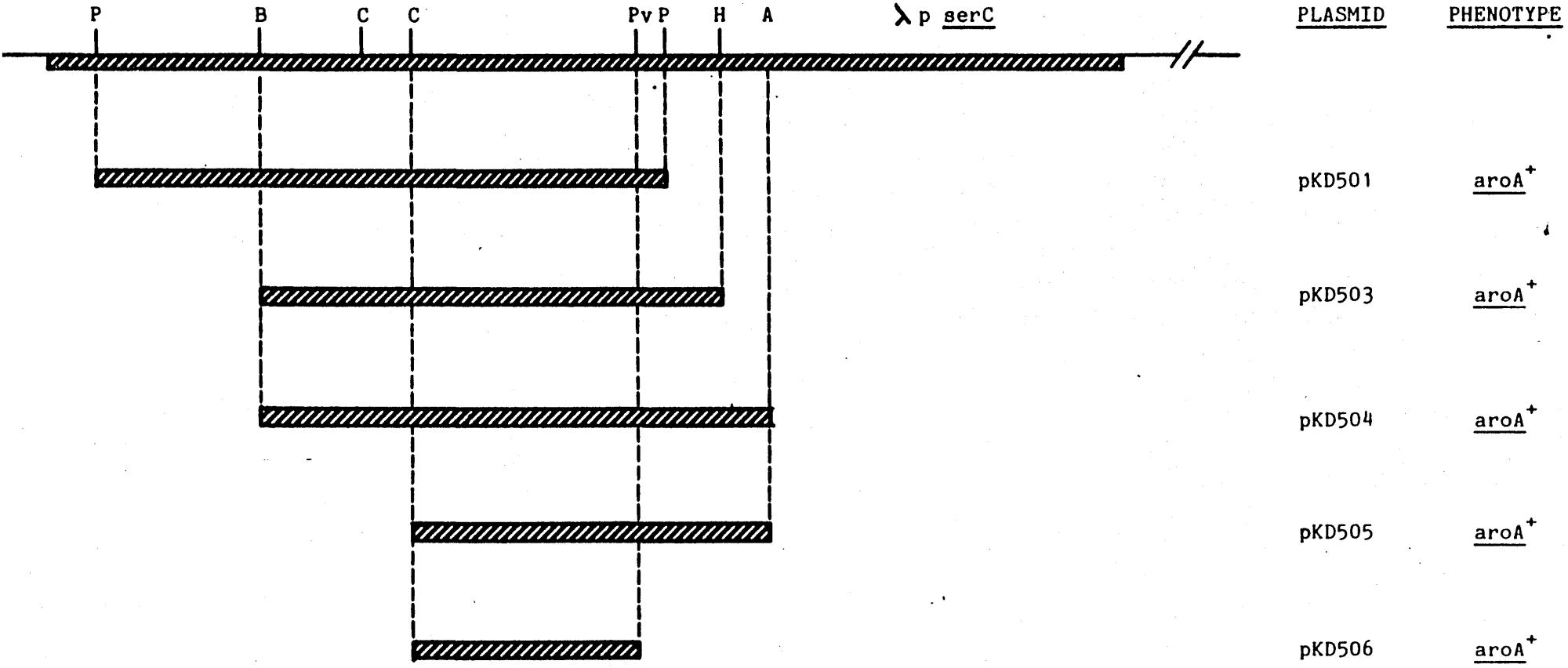
The enzyme combinations used and the resulting recombinants obtained are detailed in Table 5.2.

For each combination of enzymes, the six mini-preps were found to contain the same recombinant plasmid. A series of single- and double-enzyme digests were carried out on each plasmid in order to locate the region of the phage carried on the plasmid. The information obtained from this, and the deduced location of each cloned fragment is shown in Figure 5.5.

Combining the sub-cloning results so far, the boundaries of the *aroA* complementing region of the phage are the ClaI site used in the construction of pKD505 and the PstI site used in constructing pKD501.

| λ *pserC* digested by: | pAT153 digested by: | Resultant recombinant: |
|---|---|---|
| BglII + HindIII | BamHI + HindIII | pKD503 |
| BglII + AvaI | BamHI + AvaI | pKD504 |
| ClaI + AvaI | ClaI + AvaI | pKD505 |

Table 5.2 Restriction enzyme combinations used in 'shotgun' cloning of *aroA* from λ*pserC*.




Pst I B- Bgl II C- Cla I Pv- Pvu II H- Hind III A- Ava I

Figure 5.5: Summary of the sub-cloning of the E.coli aroA gene from λ pserC. Hatching indicates E.coli DNA

5.2.6 Construction of pKD506

The cloning of a ClaI-PstI fragment into pAT153 results in the disruption of both the ampicillin and the tetracycline resistance genes. For this reason, it was decided to clone a fragment extending from the ClaI site to a PvuII site approximately 200 bp to the left of the PstI site. When this 1.9 kb fragment was cloned into pBR322, the ampicillin resistance gene remained intact, and it was hoped that the recombinant plasmid would still be *aroA* complementing.

The 1.9 kb ClaI-PvuII fragment was isolated after electrophoresis of a ClaI/PvuII digest of pKD501 on 1% LMT agarose. It was ligated (overnight, $0^{o}$) to ClaI/PvuII cleaved pBR322 and used to transform *E.coli* AB2829. Six colonies were obtained after overnight growth on LA/amp and colonies were picked and transferred onto LA/tet and MM. All were capable of growth on MM. Plasmid DNA was extracted from one colony; restriction enzyme digests confirmed that this recombinant plasmid contained the correct 1.9 kb ClaI-PvuII fragment, and it was designated pKD506.

The sub-cloning results are summarised in Figure 5.5. Each plasmid is phenotypically *aroA*$^{+}$ on the basis that it can relieve the auxotrophic mutation in *E.coli* AB2829. All the plasmids are equally capable of complementing *aroA*. None of the transformed cells show a slow growing phenotype which might be indicative of either weak complementation from an altered EPSP synthase, or intra-molecular

complementation between plasmid encoded enzyme and mutant chromosomally encoded enzyme. These results suggest that the whole *aroA* gene lies on pKD506 between the ClaI and PvuII sites. However, confirmation that the gene is intact and that any sequences necessary for its expression are also present can only be gained through determination of the specific activity of EPSP synthase in crude extracts of cells harbouring each of the plasmids. As stated earlier, the *aroA* gene is expressed constitutively and so one can predict that EPSP synthase will be overexpressed at the level of the copy number of the vector. This would suggest that each clone will overexpress approximately 100-fold, as the copy number of pAT153 is approximately 100. This level of expression has already been shown to occur with pKD501 (see Section 5.2.2). One of the plasmids, pKD506, since it was obtained from pBR322, would be expected to have a lower copy number.

### 5.2.7 Specific activity of EPSP synthase in crude extracts of cells carrying the recombinant aroA plasmids

It has already been shown that *E.coli* AB2829 cells harbouring pKD501 are an excellent source of EPSP synthase. The high copy number of the plasmid results in a 100-fold increase in the specific activity of EPSP synthase in these cells as compared with the activity found in wild type cells. In this section the levels of EPSP synthase in cells harbouring other recombinant plasmids carrying the *aroA* gene are compared.

Crude extracts were prepared as detailed in Section 2.18 from 100 ml minimal medium cultures grown to saturation ($A_{650}$ 1.3 - 1.5). Sonication buffer 'B' was used throughout. In each case, the values given for specific activities are an average of at least two independent duplicate experiments. All the assays were performed in the reverse direction, following the formation of PEP by coupling to the pyruvate kinase and lactate dehydrogenase reactions (Figure 5.6). The results obtained are shown in Table 5.3.

In the crude extracts there is a steady absorbance change at 340 nm in the absence of added EPSP, which is due to an NADH oxidase activity. This blank rate makes the accurate determination of the specific activity of EPSP synthase in extracts of wild type cells difficult. The blank rate poses no problems in cells where the aroA gene is overexpressed from a plasmid because the observed rates are so high. Despite the difficulties in measuring low levels of EPSP synthase activity, we can be confident that the relative expression values quoted in Table 5.3 are accurate for the following reason: during full scale preparation of the enzyme from E.coli K12 or from E.coli AB2829/pKD501, the specific activity of EPSP synthase can be determined very accurately after ammonium sulphate fractionation of the crude extract (the first step in the purification procedure), and the overexpression factor obtained is 100, in good agreement with the value obtained in Table 5.3.

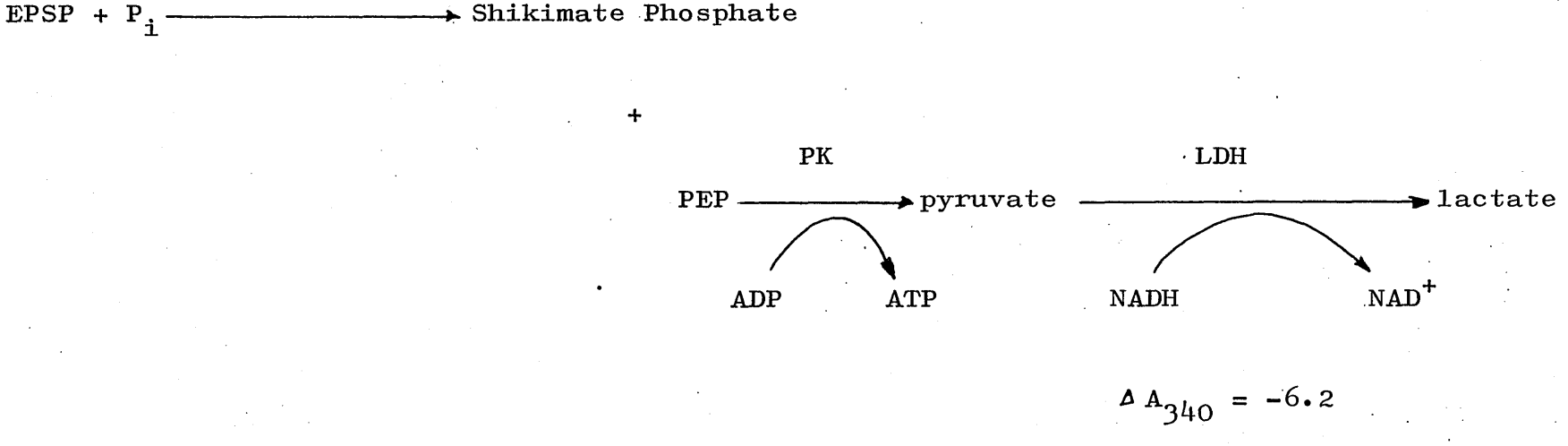



Figure 5.6. Coupling of EPSP synthase to the pyruvate kinase and lactate dehydrogenase reactions. PK = pyruvate kinase, LDH = lactate dehydrogenase (PK/LDH was purchased from Boehringer).

| Extract | Specific activity of EPSP synthase (U/mg) | Relative activity |
|---|---|---|
| K12 | 0.004 | 1 |
| AB2829 | no activity | - |
| AB2829/pKD501 | 0.420 | 105 |
| AB2829/pKD502 | 0.425 | 106 |
| AB2829/pKD503 | 0.425 | 106 |
| AB2829/pKD504 | 0.356 | 89 |
| AB2829/pKD505 | 0.0044 | 1.1 |
| AB2829/pKD506 | 0.0048 | 1.2 |

Table 5.3: Specific activity of EPSP synthase in E.coli crude extracts.

All the plasmids derived from pAT153 should have similar high copy numbers and therefore give similar high levels of EPSP synthase. For cells harbouring pKD501, pKD502, pKD503 and pKD504, the specific activity of EPSP synthase in the crude extract is indeed similar, and is approximately 100 fold higher than in E.coli K12. In contrast, the crude extracts of cells harbouring pKD505 and pKD506 have EPSP synthase specific activities which are very much less than expected. The values obtained are approximately the same as the level found in wild type cells. This amount of enzyme will be sufficient to allow mutant cells to grow in unsupplemented medium. Thus all the plasmids are capable of relieving the auxotrophic requirement of E.coli AB2829, and are indistinguishable phenotypically. However, an explanation is required for the occurrence of two classes of aroA containing plasmid, one of which gives 'high' levels of EPSP synthase activity and one of which gives unexpectedly 'low' levels of activity.

### 5.2.8 Expression of aroA on pKD505 and pKD506

Both pKD505 and pKD506 were constructed by insertion of fragments obtained by cleavage at a common ClaI site. pKD506 extends 1.9 kb to a PvuII site, but pKD505 extends beyond this site, and beyond the PstI site used in the construction of pKD501. It is possible, therefore, that cloning from the ClaI site is responsible for a reduction in the level of expression of the aroA gene. Two simple

models which would explain the low levels of enzyme activity are shown in Figure 5.7.

Figure 5.7a shows the removal of the 3' end of the aroA coding region during the cloning procedure. This would result in the production of large quantities of a truncated polypeptide which might have very little enzyme activity.

Figure 5.7b illustrates a second possibility, the removal or alteration of the promoter of the aroA gene. The low level of expression observed might then be due either to an altered, ineffective aroA promoter, or, if the aroA promoter has been completely removed, expression might be due to a minor plasmid promoter.

### 5.2.9 SDS-P.A.G.E. of crude extracts

In an attempt to distinguish between the two possibilities outlined in the previous section, crude extracts were analysed by SDS-P.A.G.E. In extracts where the enzyme is overexpressed, it ought to be possible to see a Coomassie Blue stained band corresponding to EPSP synthase providing there are no highly abundant E.coli proteins migrating to the same position in the gel. If the reduction in specific activity is due to production of a truncated polypeptide then a shift in the position of this band should be observed. However, if the problem is the disruption of the promoter, then a heavily staining band should be absent from the low activity extracts.

Crude extracts were prepared and analysed by electrophoresis on a 12% polyacrylamide gel in the presence of SDS, as described in Section 2.21. The gel was stained

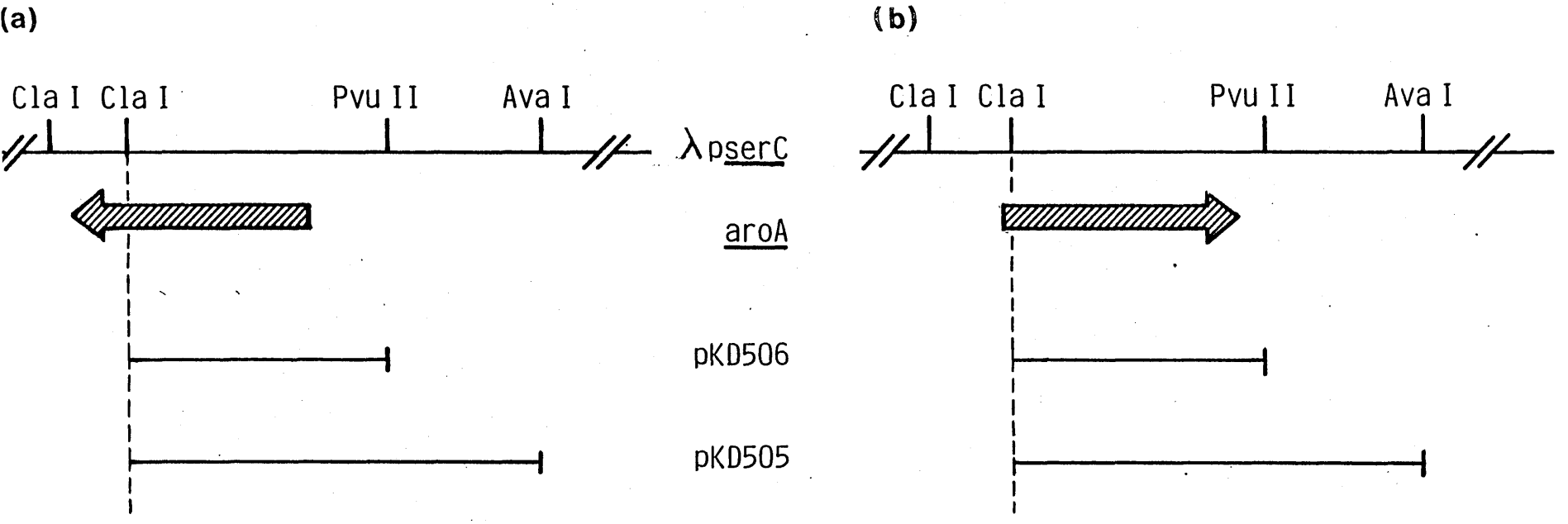



Figure 5.7: Possible explanation for the low specific activity of EPSP synthase in cells harbouring plasmids pKD505 and pKD506.

(a) The ClaI site is within the aroA coding sequence. Expression of aroA results in the production of a truncated polypeptide, with lowered specific activity.

(b) The promoter of aroA is interrupted at the ClaI site, and so expression is reduced.

with Coomassie Blue, destained overnight and photographed (Figure 5.8).

A heavily staining band of the same mobility as EPSP synthase can be seen in tracks a - d which correspond to the 'high activity' extracts. However, the same band is observed in the 'low activity' extracts (tracks f and g) and also in E.coli K12 (track h) and E.coli AB2829 (track i). Thus, no conclusions can be drawn from this gel regarding the expression of aroA in 'low activity' and 'high activity' extracts. A highly abundant E.coli protein is masking EPSP synthase throughout the gel: the likeliest candidate for this is Elongation Factor Tu, which is expressed at a very high level (89 000 molecules/cell, data from Gouy & Gautier, 1982) and has an $M_r$ of 46 000.

This negative result is presented here, however, because the gel shown in Figure 5.8 shows that in four of the crude extracts, tracks a-d, there is a very prominent band ($M_r$ 40 000) which is not present in the other extracts. This protein is presumably also plasmid encoded since it is not present in extracts of E.coli K12 but is only present in the 'high activity' extracts. This would imply that it is encoded by sequences to the left of the ClaI site used in the construction of pKD505 and pKD506. The significance of the prominent band will be discussed in detail in Chapter 6.

### 5.2.10 The 0.5 kb ClaI fragment

One way of restoring 'high activity' to the 'low activity' plasmids might be to reclone and include the

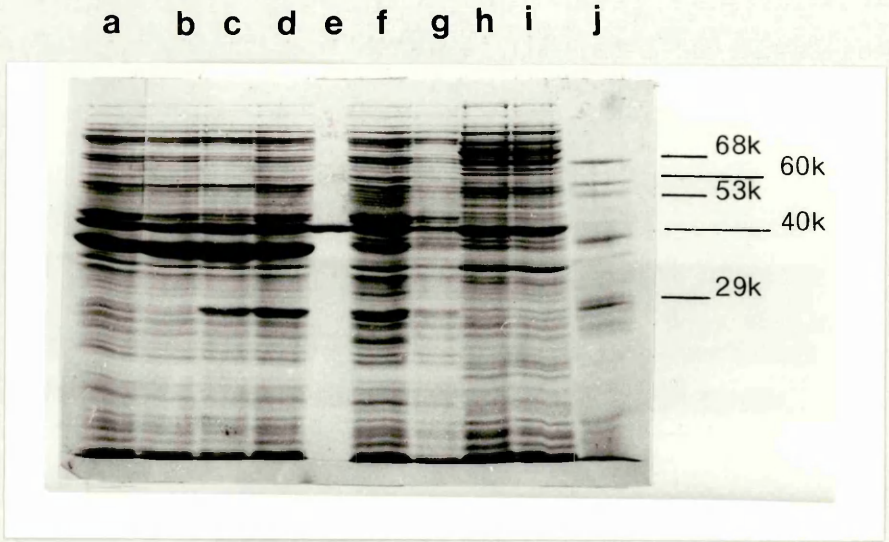



Figure 5.8: SDS-PAGE of E.coli crude extracts

| Track | | |
|---|---|---|
| | a | E.coli AB2829/pKD501 |
| | b | E.coli AB2829/pKD502 |
| | c | E.coli AB2829/pKD503 |
| | d | E.coli AB2829/pKD504 |
| | e | EPSP synthase |
| | f | E.coli AB2829/pKD505 |
| | g | E.coli AB2829/pKD506 |
| | h | E.coli AB2829 |
| | i | E.coli K12 |
| | j | Molecular weight markers |

sequences to the left of the ClaI site in the plasmids. This is feasible as there is a unique ClaI site in pKD506 and the 0.5 kb ClaI fragment ('B' on Figure 5.1) to the left of the ClaI site can be readily isolated from pKD501.

Fragment B can be inserted into pKD506 in either orientation, but it will be expected to restore 'high activity' only when it is in the correct orientation relative to Fragment A. Glyphosate tolerance would be expected to select any recombinants with 'high activity' phenotype. E.coli AB2829/pKD506 shows the same pattern of sensitivity to glyphosate as does E.coli K12, but restoration of 'high activity' will change the pattern to that of E.coli AB2829/pKD501.

Fragment B was recovered after digestion of pKD501 with ClaI and electrophoresis through LMT agarose. This was then ligated to ClaI cleaved and phosphatase treated pKD506. Extra care was taken in this case to ensure that the phosphatase reaction had removed all 5'-phosphate groups. A control ligation after phosphatase treatment showed that essentially no recircularisation of pKD506 was possible and hence that any transformants obtained after ligation of fragment B would be recombinants. It was then predicted that approximately 50% of recombinants would be glyphosate tolerant.

The ligation mix was used to transform E.coli AB2829 and transformants were selected by overnight growth on LA/amp. As a selection for the restoration of 'high activity', 100 $amp^R$ colonies were transferred and gridded onto an MM plate and an MM + glyphosate (20 mM) plate.

Under these conditions, only colonies carrying recombinant clones with inserts in the correct orientation would be expected to grow.

None of the colonies screened were able to grow on the plate containing glyphosate (AB2829/pKD501 did grow on this plate). The experiment was repeated and 800 colonies were screened by replica plating onto MM + glyphosate (20 mM). Again, none were able to grow on glyphosate. Plasmid DNA was prepared from 20 transformants and digested with ClaI. All were found to have an insert at the ClaI site of the correct size. It was not possible, however, to show conclusively that any of these had fragment B in the correct orientation relative to fragment A of pKD506, as the only known restriction site on fragment B (KpnI; Figure 5.1) is very close to the centre.

Crude extracts were prepared from cells carrying six of the recombinant clones and the specific activity of EPSP synthase was determined. In each case, the activity was close to wild type, i.e. the clones were of the 'low activity' type.

It is exceedingly unlikely that all of the transformants screened had the 0.5 kb ClaI fragment introduced into pKD506 in the 'wrong' orientation. This leads to the conclusion that insertion of this DNA and restoration of the correct fragment B - fragment A interface has no effect on expression of the <u>aroA</u> gene. 0.5 kb of DNA is enough to encode more than a third of the <u>aroA</u> gene. Addition of this much DNA to the inserts in pKD505 and pKD506

should improve the levels of EPSP synthase activity if these had been reduced by truncation of the gene. Since there is no improvement, an alternative explanation for the 'low activity' plasmids is required. One possibility is that aroA is not a 'normal' monocistronic E.coli gene, but is co-ordinately expressed with the 40 000 $M_r$ gene product. To obtain evidence for this hypothesis, the entire DNA sequence of the aroA region has been determined.

## 5.3 Determination of the DNA sequence of the aroA gene

### 5.3.1 DNA sequencing strategy for aroA

DNA sequencing studies on the aroA gene were concentrated initially on fragment A of λpserC - the 1.9 kb ClaI-PvuII fragment subcloned in pKD506. The M13/dideoxy sequencing method was used throughout. Full details of the techniques employed to construct clones and produce templates, and the dideoxy sequencing reactions are given in Sections 2.22 to 2.30.

The strategy which was followed for providing clones and templates for sequencing is outlined below. It was hoped that this would give the entire DNA sequence on both strands and overlaps at all the restriction sites used in cloning.

(a) left end of the sequence: use a ClaI fragment from pKD502 to define the left end of the sequence.

(b) Right end of the sequence: there is a unique HincII site within fragment A which is located approximately 1.1 kb to the right of the ClaI site; cloning of a

0.8 kb HincII-PvuII fragment defines the right end of the sequence and also gives some internal sequence.

(c) Internal sequence: use a set of fragments produced by digestion of fragment A with the restriction enzymes TaqI and HpaII (the rationale for this has been outlined in Section 4.4.1). The sequence of these fragments will give most of the internal sequence.

### 5.3.2 The ClaI site

pKD502 was digested with ClaI and the DNA fragments separated by electrophoresis on 1% LMT agarose. A 2.5 kb fragment spanning the ClaI-PvuII region and extending to the ClaI site in the vector was excised and cloned into the AccI site of M13mp8. Four recombinant templates were purified and the sequence to the right of the ClaI site was determined. The sequence was entered into a file on a PDP11-34 computer using the BATIN program (Section 2.36).

### 5.3.3 The PvuII site

In a similar way, a 0.8 kb HincII-PvuII fragment of pKD506 was isolated and ligated to SmaI cleaved M13mp8. Six templates were purified and sequenced in turn till the sequence at both ends had been obtained. Again, the data was entered into the computer. The sequence at the PvuII end showed that the nearest HpaII and TaqI sites were located 12 and 1 bp respectively away from the PvuII site - this indicated that in cloning HpaII and TaqI fragments very little information would be lost through the inability to clone the TaqI-PvuII and HpaII-PvuII fragments.

### 5.3.4 Number and size of HpaII and TaqI fragments in region A

The ClaI-PvuII fragment of pKD506 was isolated by digestion of 8 µg of pKD506 with ClaI and PvuII, followed by electrophoresis on a preparative 1% LMT agarose gel. This fragment was then digested (without further purification) by the enzymes HpaII and TaqI.

The digestion products were separated by electrophoresis on a 5% polyacrylamide gel (Section 2.12) and the gel stained with ethidium bromide (Figure 5.9). Unfortunately, the quantity of DNA used was too little to allow the bands to be seen under the normal exposure conditions. The photograph is overexposed to compensate for this. Examination of the original negative gives more information than the photograph. Clearly, at least nine bands are present in the HpaII digest and six bands are seen in the TaqI digest.

### 5.3.5 Preparation of HpaII and TaqI clones for sequencing

An aliquot of each digestion mix from the experiment detailed in the preceding section and containing approximately 100 ng of fragments was purified by phenol/chloroform extraction and ethanol precipitation. After recovery, the fragments were ligated to AccI cleaved M13mp8 and transformed into *E.coli* JM101. Single strand template DNA was produced from sixty clear plaques.

### 5.3.6 Characterisation of the HpaII and TaqI clones

As these clones have been constructed using a 'shotgun' cloning approach, many will contain the same fragment

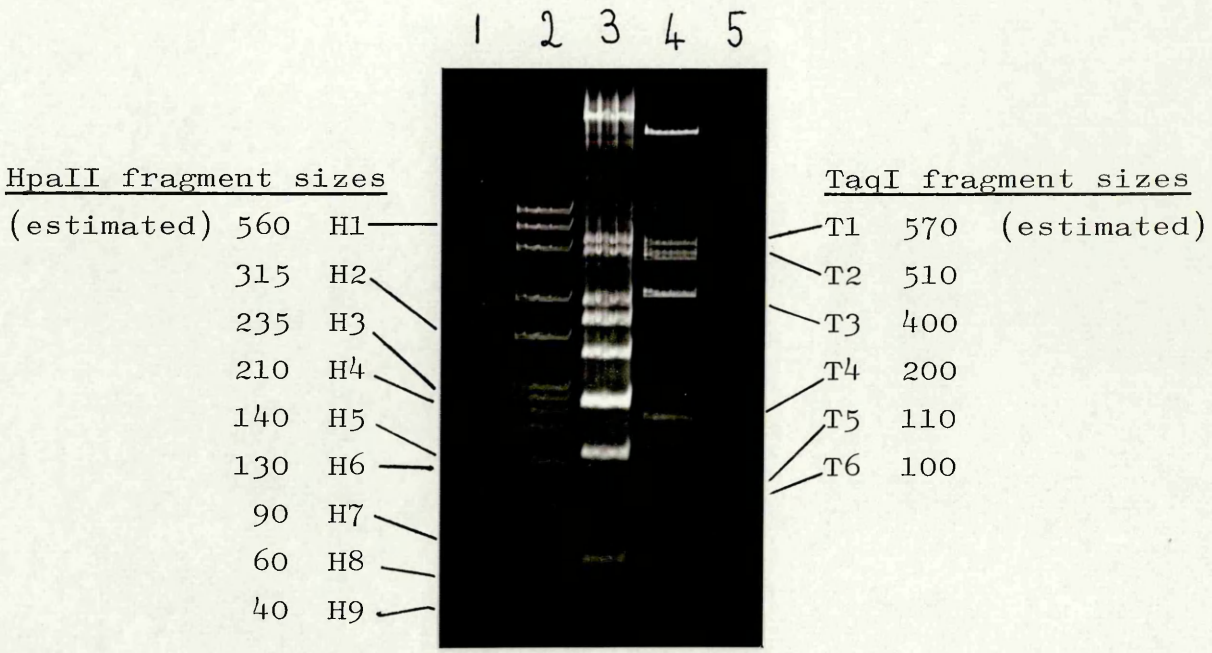



Figure 5.9: The TaqI and HpaII fragments within the 1.9 kb ClaI-PvuII region (sizes are in bp)

| Track | DNA | Enzyme |
|---|---|---|
| 1 | 1.9 kb ClaI-PvuII fragment | HpaII |
| 2 | pKD506 | HpaII |
| 3 | pBR322 | HinfI |
| 4 | pKD506 | TaqI |
| 5 | 1.9 kb ClaI-PvuII fragment | TaqI |

inserted in the same or in the opposite direction. All the clones were subjected to T-track analysis (see Section 2.33) and those showing the same pattern were grouped. In some cases it was possible to identify complementary strands of the same fragment from the length of the cloned segment. A number of clones, thought to represent all the possible sequences, were selected for complete sequence analysis.

### 5.3.7 Sequence analysis of HpaII and TaqI clones

Sequencing reactions were performed as described in Sections 2.28 - 2.30, using the 'normal' sequencing reaction mixes. Electrophoresis was on 6% polyacrylamide gels and the gels were dried before overnight autoradiography (Section 2.32). In most cases, sequencing reaction products were electrophoresed on two linear gels; one was run for $4\frac{3}{4}$ h and the other for approximately $1\frac{3}{4}$ h. Clones which T-tracking had shown to have an insert of less than 200 bp were electrophoresed on a single buffer gradient gel.

The sequence data showed that many of the clones which were sequenced had more than one fragment inserted into the vector. This may have been caused by having too high a ratio of fragment ends to vector ends in the ligation mixes, or by having too high a concentration of ligase, causing concatenation of fragments before cloning. This problem was avoided in later experiments by phosphatase treatment of the fragments, rather than the vector.

Where a TaqI site was found in the sequence of a TaqI clone, the sequence was divided into two parts, A and B, referring to before and after that site. In only a very few cases were the fragments linked as they would be *in vivo* indicating that most multiple cloning events were due to spurious linkage and not partial digestion of the original DNA.

Each sequence was entered into a computer file.

### 5.3.8 Building up the sequence

The complete sequence of the ClaI-PvuII region was built up using the method outlined in detail in Chapter 4 for the *aroD* gene. Briefly, each sequence was compared with all the other sequences and a map of overlaps and complementary sequences built up. Each nucleotide was then checked against its complementary strand partner and all doubtful sequences resolved. Corrected sequences were then built into a single complete sequence.

The complete sequence obtained in this way contained several gaps. Not every TaqI or HpaII fragment was present in the collections of clones.

### 5.3.9 Filling in the gaps

In order to complete the sequence, two further experiments were needed:

(a) Clone turn-around (see Section 2.35). HpaII fragment H2 (Figure 5.9) was turned around in order to complete the complementary strand in this region.

(b) The largest TaqI fragment (T1 in Figure 5.9) was missing from the set of TaqI clones. This fragment was isolated after TaqI digestion of pKD506 and electrophoresis through 2% LMT agarose. After ligation to AccI cleaved M13mp8 and transformation of E.coli JM101, six recombinant templates were purified as described previously. These were sequenced in turn till the sequence at both ends of the fragment was obtained. Two clones carrying the fragment in opposite orientation were re-sequenced using the 'low dideoxy' mixes (see Section 2.30) and electrophoresed on a 6% linear polyacrylamide gel for 6 h. This allowed the sequence to be read far enough to overlap a HpaII site. This in fact showed that there were two HpaII sites, separated by 63 bp. The 63 bp HpaII fragment was missing from the set of HpaII clones.

### 5.3.10 Complete sequence of the ClaI-PvuII fragment

The sequencing strategy, including the location of the sites used for cloning and the extent of sequencing for the entire ClaI-PvuII fragment is summarised in Figure 5.10. The complete DNA sequence is shown in Figure 5.11. The ClaI-PvuII region is 1959 bp in length and the entire sequence is covered on both strands of the DNA with all the restriction enzyme sites used for cloning overlapped by 34 sequences. Note that the numbering system puts the ClaI site at position 791 in the sequence, not at position 1. The reason for this is that later experiments (see

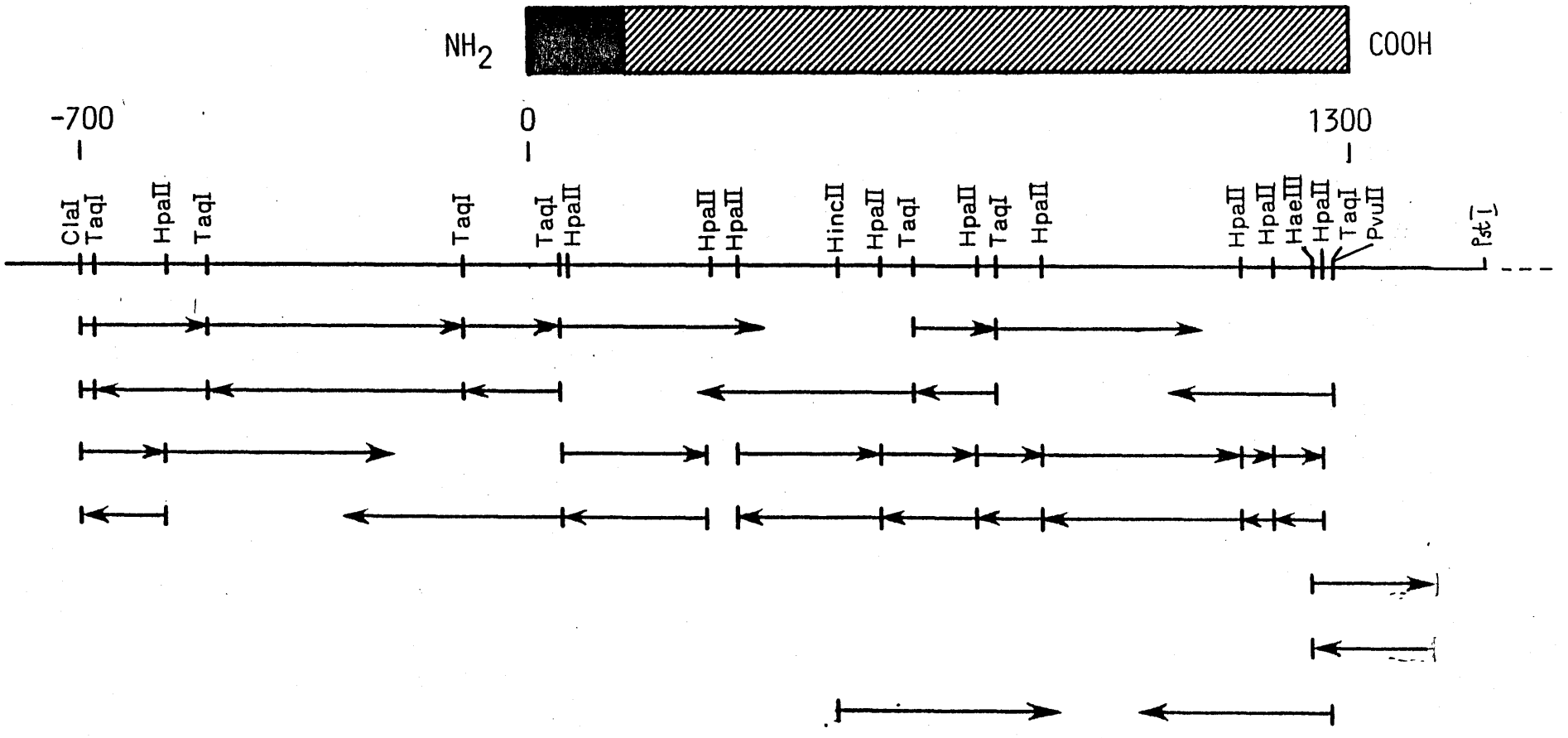



Figure 5.10: DNA sequencing strategy for the ClaI-PvuII region. The sequence corresponding to the aroA coding region is indicated by the box; the solid part corresponds to the N-terminal amino acid sequence.

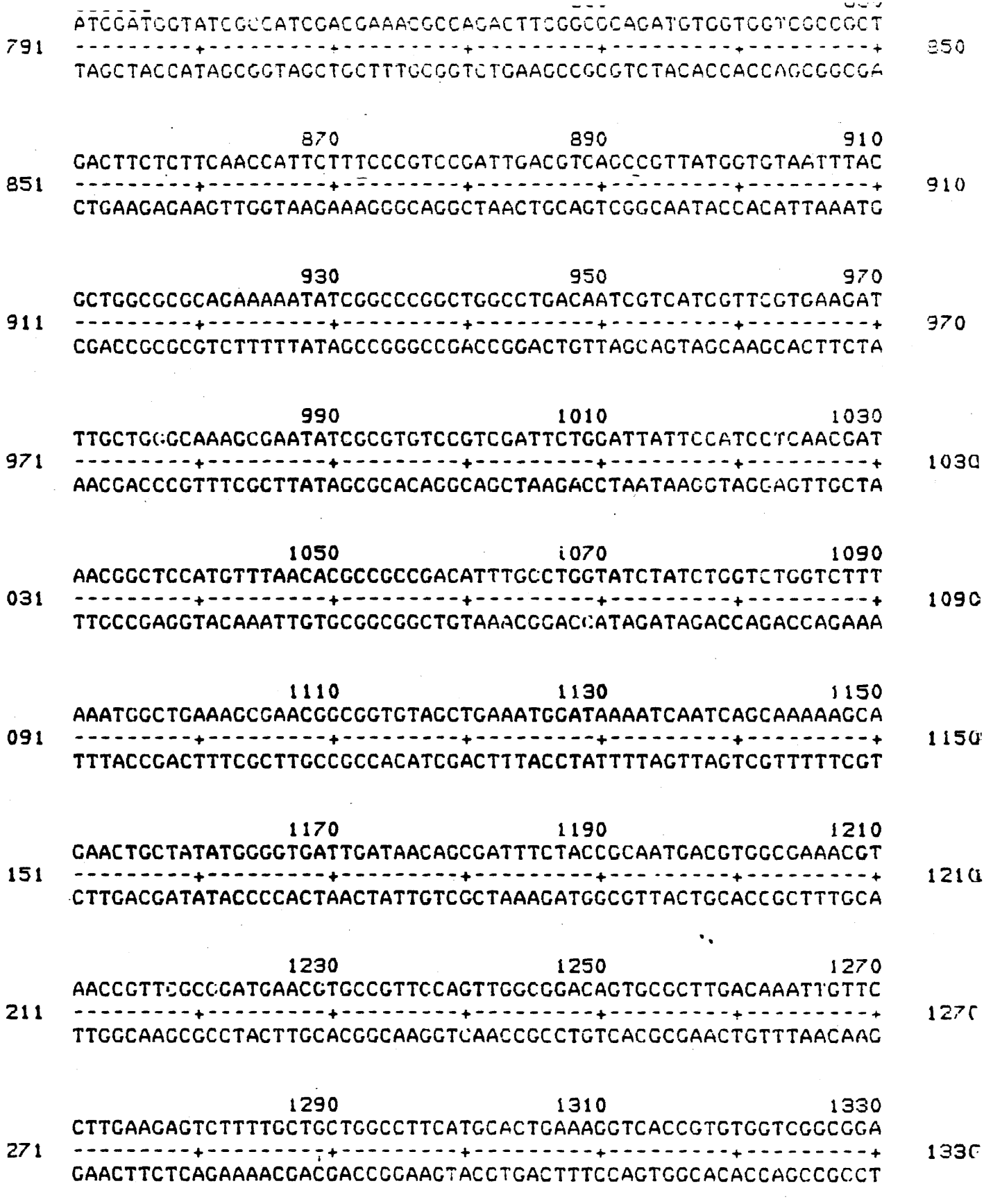
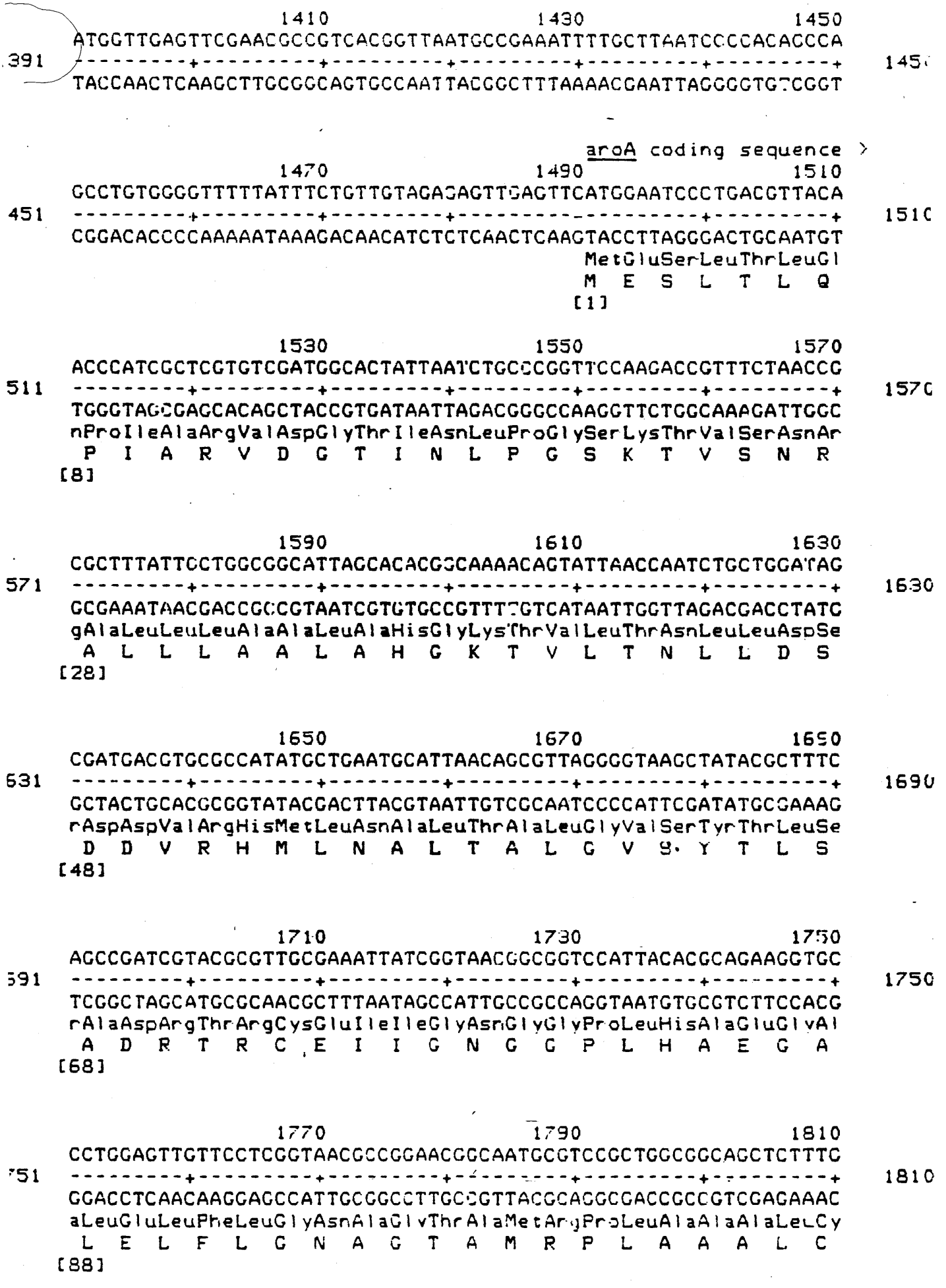



Figure 5.11:Complete DNA sequence of the ClaI-PstI region, and the coding sequence of the *E.coli* *aroA* gene.


1410
1430
1450
ATGGTTGAGTTCGAACGCCGTCACGGTTAATGCCGAAATTTTGCTTAATCCCCACACCCA
391
145
TACCAACTCAAGCTTGCGGCAGTGCCAATTACGGCTTTAAAACGAATTAGGGGTGTCGGT
aroA coding sequence
1470
1490
1510
GCCTGTGGGGTTTTTATTTCTGTTGTAGAGAGTTGAGTTCATGGAATCCCTGACGTTACA
451
151
CGGACACCCCAAAAATAAAGACAACATCTCTCAACTCAAGTACCTTAGGGACTGCAATGT
MetGluSerLeuThrLeuGl
M E S L T L Q
[1]
1530
1550
1570
ACCCATCGCTCGTGTCGATGGCACTATTAATCTGCCCGGTTCCAAGACCGTTTCTAACCG
511
157
TGGGTAGCGAGCACAGCTACCGTGATAATTAGACGGGCCAAGGTTCTGGCAAAGATTGGC
nProIleAlaArgValAspGlyThrIleAsnLeuProGlySerLysThrValSerAsnAr
P I A R V D G T I N L P G S K T V S N R
[8]
1590
1610
1630
CGCTTTATTCCTGGCGGCATTAGCACACGGCAAAACAGTATTAACCAATCTGCTGGATAG
571
163
GCGAAATAACGACCGCCGTAATCGTGTGCCGTTTTGTCATAATTGGTTAGACGACCTATC
gAlaLeuLeuLeuAlaAlaLeuAlaHisGlyLysThrValLeuThrAsnLeuLeuAspSe
A L L L A A L A H G K T V L T N L L D S
[28]
1650
1670
1690
CGATGACGTGCGCCATATGCTGAATGCATTAACAGCGTTAGGGGTAAGCTATACGCTTTC
631
169
GCTACTGCACGCGGTATACGACTTACGTAATTGTCGCAATCCCCATTCGATATGCGAAAG
rAspAspValArgHisMetLeuAsnAlaLeuThrAlaLeuGlyValSerTyrThrLeuSe
D D V R H M L N A L T A L G V S Y T L S
[48]
1710
1730
1750
AGCCGATCGTACGCGTTGCGAAATTATCGGTAACGGCGGTCCATTACACGCAGAAGGTGC
691
175
TCGGCTAGCATGCGCAACGCTTTAATAGCCATTGCCGCCAGGTAATGTGCGTCTTCCACG
rAlaAspArgThrArgCysGluIleIleGlyAsnGlyGlyProLeuHisAlaGluGlyAl
A D R T R C E I I G N G G P L H A E G A
[68]
1770
1790
1810
CCTGGAGTTGTTCCTCGGTAACGCCCGGAACGGCAATGCGTCCGCTGGCGGCAGCTCTTTC
751
181
GGACCTCAACAAGGAGCCATTGCGGCCTTGCCGTTACGCAGGCGACCGCCGTCGAGAAAC
aLeuGluLeuPheLeuGlyAsnAlaGlyThrAlaMetArgProLeuAlaAlaAlaLeuCy
L E L F L G N A G T A M R P L A A A L C
[88]

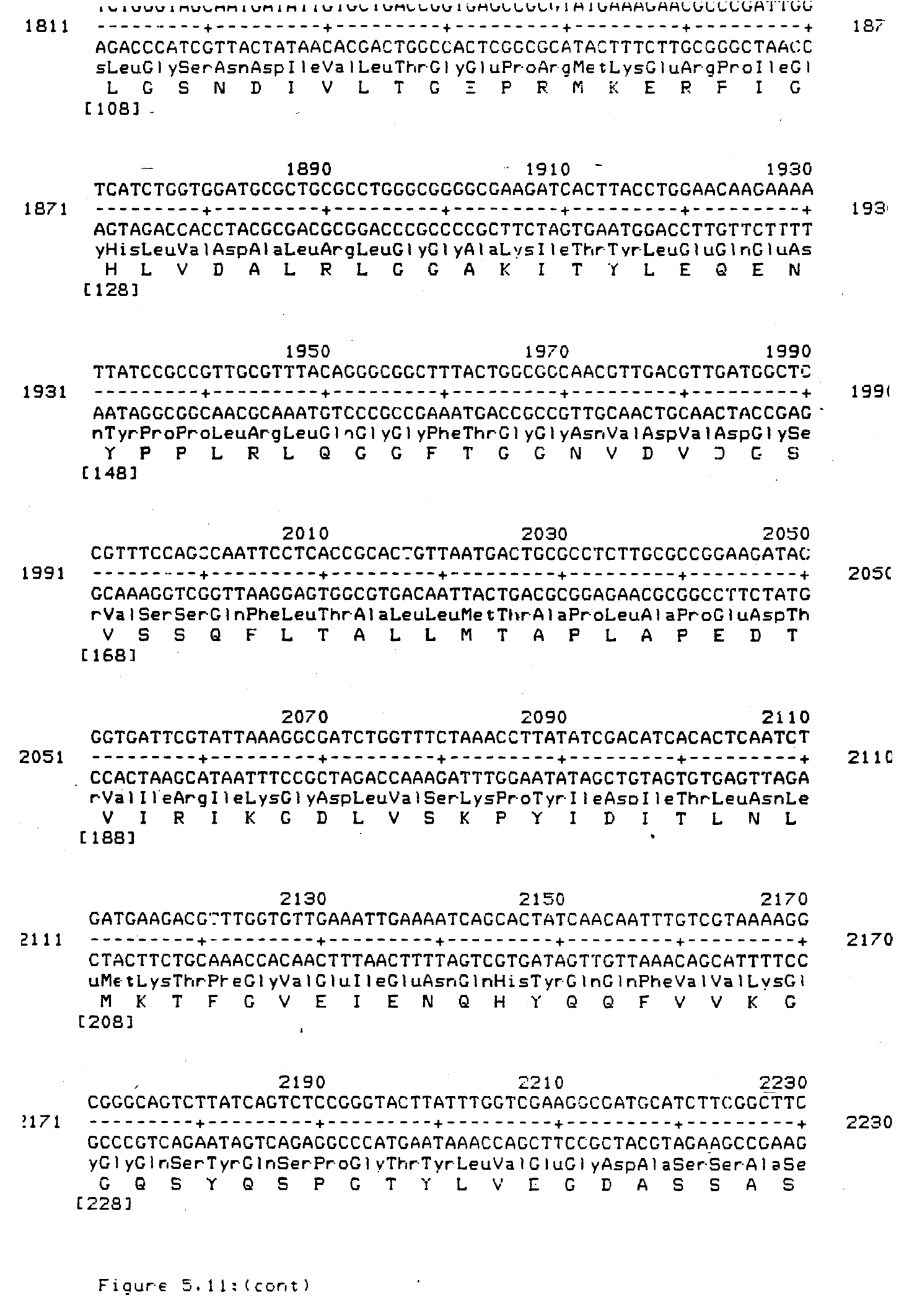



Figure 5.11:(cont)

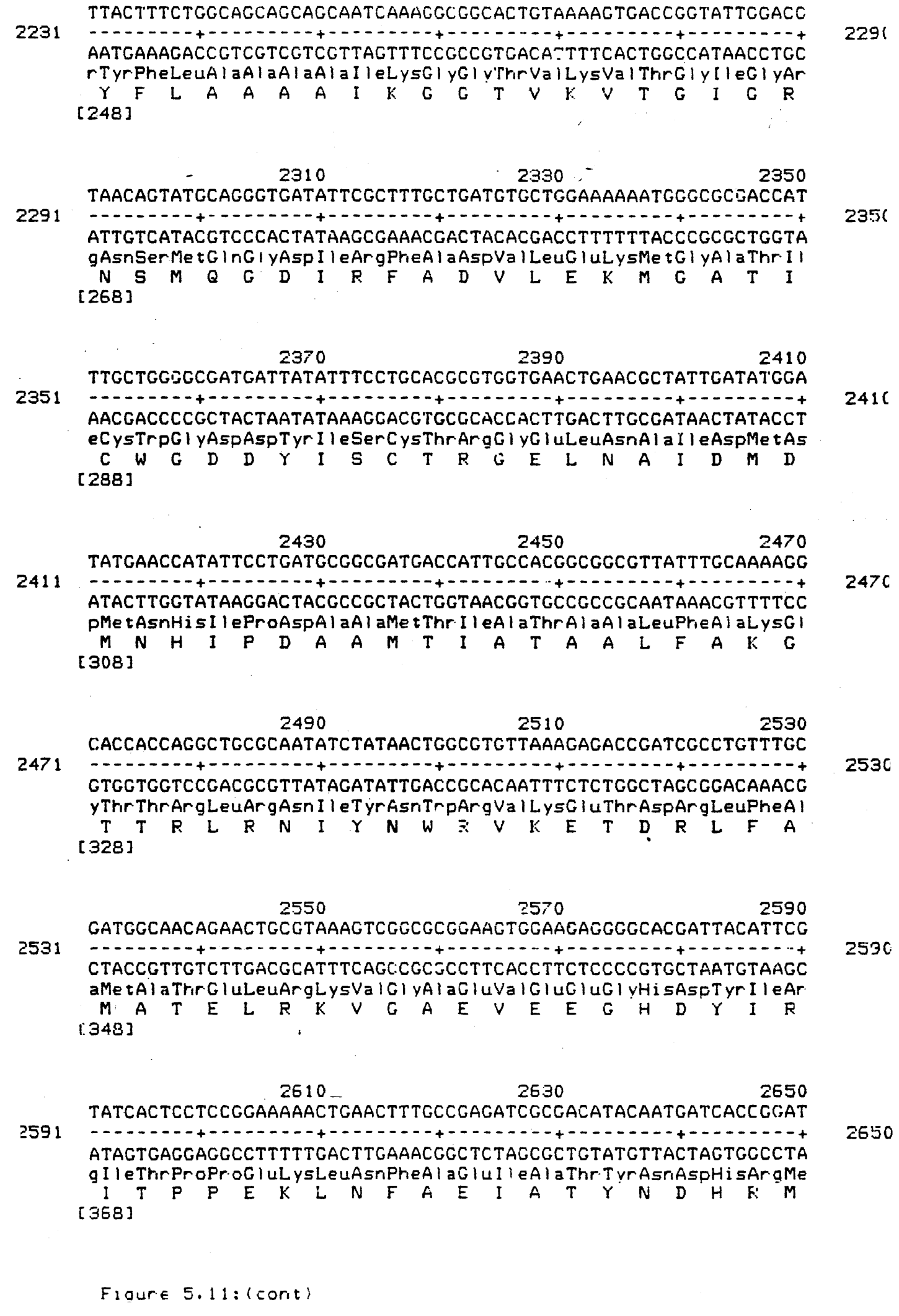



Figure 5.11:(cont)

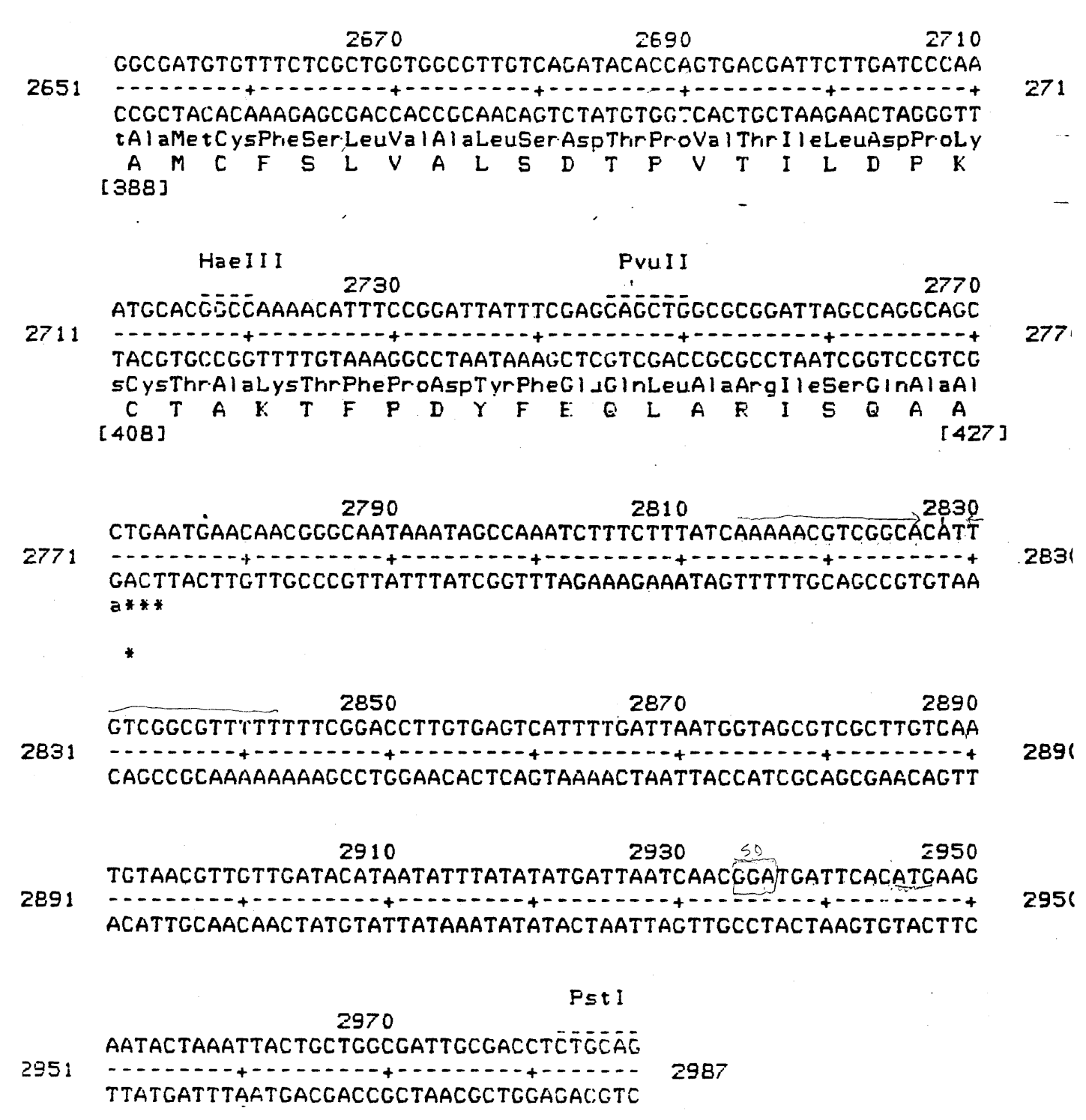



Figure 5.11:(cont)

Chapter 6) provided sequence data to the left of this ClaI site and the use of these numbers keeps the same system throughout this thesis, starting at the leftmost sequence obtained.

### 5.4.1 Identification of the aroA coding region on the sequence

(a) Amino-terminal sequence of EPSP synthase

As part of the overall strategy for obtaining the entire EPSP synthase sequence, the amino-terminal sequence of the enzyme was established. This was carried out using a Beckman Model 890 liquid phase sequencer at the protein sequencing facility at Aberdeen University. This sequence was determined by Ann Lewendon and is illustrated in Figure 5.12. Technical details of the derivation of the sequence and an explanation of the anomalies and gaps in this sequence can be found in Duncan, Lewendon and Coggins (1984b).

(b) Identification of the open reading frame encoding EPSP synthase

The open reading frame encoding EPSP synthase could now be identified by comparison of the amino-terminal protein sequence with sequences downstream of methionine codons in the DNA sequence. Both strands of the DNA sequence were translated in all three possible reading frames using the TRNTRP program (Section 2.36).

The start of the open reading frame was found to be at position 1491 in the sequence shown in Figure 5.11.

1 10

Met - Glu - Ser - Leu - Thr - Leu - Gln - Pro - Ile - Ala -

11 20

Arg - Val - X - Gly - Thr - Ile - Asn - Leu - Pro - Gly -
(Asp)

21 30

Ser - Lys - Ser - Val - X - Asn - His - Ala - Leu - Leu -

31 40

Leu - Ala - Ala - Leu - Ala - His - Gly - Val - X - Val -
(Val)

41 44

Leu - X - Asn - Leu -

Figure 5.12. Amino-terminal sequence of EPSP synthase.

The location of the open reading frame is illustrated on Figure 5.10. The solid portion corresponds to that part which has been confirmed by protein sequence; the hatched portion is the remainder of the sequence. In the amino-terminal region there is an exact agreement between the experimentally determined amino acid sequence and the amino acid sequence deduced from the DNA sequence.

The open reading frame continues to the PvuII site; cloning from this site, as in the construction of pKD506 had, in fact, clipped off the 3' end of the gene and removed an unknown number of residues from the carboxy-terminus of the protein.

## 5.4.2 Isolation of a HaeIII fragment to complete the sequence

In order to complete the sequence of the aroA gene it was necessary to determine the DNA sequence beyond the PvuII site. Examination of the sequence derived so far (Figure 5.11) revealed that there was a HaeIII site at position 2717. It was hoped that it would be possible to clone a fragment extending from this position to beyond the PvuII site in pKD502, and that the sequence of this fragment might complete the aroA sequence.

pKD502 was digested with ClaI and a 2.5 kb fragment isolated after electrophoresis in LMT agarose. This fragment spans the ClaI-PvuII region and extends through the PstI site used to construct pKD502 to the ClaI site in the vector. This fragment was digested (without further

purification) with HaeIII. Two small aliquots were removed. One was digested with PvuII and the other with PstI. Digests were separated by electrophoresis on a second LMT gel and a HaeIII fragment, which is cleaved by both PvuII and PstI, was identified. This fragment was isolated and the DNA purified. It was cloned into SmaI cleaved M13mp8 and the sequence of both strands determined.

The sequence extends from position 2717 in Figure 5.11 to position 3757 in the sequence of pBR322 (Sutcliffe, 1979). The sequence of the entire ClaI-PstI region is now complete and hence the entire sequence of the aroA gene.

### 5.4.3 The predicted amino acid sequences of EPSP synthase

The DNA sequence of the coding region of the aroA gene predicts a 427 amino acid polypeptide chain of calculated $M_r$ 46112 (Figure 5.11). This is in good agreement with the subunit $M_r$ value of 49 000 determined by SDS-PAGE (Lewendon & Coggins, 1983) and with the native $M_r$ of 42 000 determined by gel permeation chromatography on a TSK G2000 SW column (Duncan, Lewendon & Coggins, 1984a).

### 5.4.4 The amino acid composition of EPSP synthase

A comparison of the experimentally determined amino acid composition of EPSP synthase (Duncan, Lewendon & Coggins, 1984b) with the amino acid composition predicted from the nucleotide sequence is shown in Table 5.4. The overall agreement is very good.

| Amino acid | Relative amino acid composition based on Leu = 48 residues | Theoretical amino acid composition predicted from the DNA sequence |
|---|---|---|
| Asx | 41.9 | 44 |
| Thr | 31.1 | 34 |
| Ser | 19.7 | 21 |
| Glx | 38.8 | 34 |
| Pro | 18.1 | 18 |
| Gly | 42.8 | 37 |
| Ala | 44.1 | 46 |
| Cys | 4.9 | 6 |
| Val | 21.7 | 24 |
| Met | 13.6 | 14 |
| Ile | 24.2 | 26 |
| Leu | 48.0 | 48 |
| Tyr | 13.1 | 13 |
| Phe | 13.2 | 13 |
| His | 8.1 | 8 |
| Lys | 17.0 | 17 |
| Arg | 17.2 | 22 |
| Trp | nd | 2 |
| Gln | nd | 12 |
| Asn | nd | 18 |

Table 5.4: Amino acid composition of EPSP synthase compared with the composition predicted by the DNA sequence of the aroA gene.

The protein chemical results quoted here are entirely consistent with the predicted amino acid sequence and there can be no doubt that the sequenced gene is aroA.

### 5.4.5 Codon utilisation in the aroA gene

The codon utilisation for the open reading frame corresponding to EPSP synthase is shown in Table 5.5. A brief comparison with the analogous data for four trp biosynthetic genes (Table 4.5) and with the putative aroD gene (Table 4.4) shows that they all follow a similar pattern of utilisation, providing further evidence that this is the EPSP synthase protein coding region. Further aspects of the codon utilisation of the aroA gene will be discussed in Chapter 6.

### 5.4.6 Altered expression of the aroA gene

The results presented in Sections 5.2.7 - 5.2.10 of this chapter provide evidence that the expression of the aroA gene is altered by cloning from the ClaI site used in the construction of pKD505 and pKD506. The orientation of the gene on the clone, is running left to right, ruled out the first of the two simple interpretations of the 'low activity' clones illustrated in Figure 5.7a.

This suggested that the altered expression must be related to either the absence or the disruption of the promoter of the gene which was presumed to be near the ClaI site (Figure 5.7b). The 5' end of the aroA protein coding region has been located 700 bp away from the ClaI

| | T | | C | | A | | G | |
|---|---|---|---|---|---|---|---|---|
| T | PHE | 9 | SER | 6 | TYR | 9 | CYS | 2 T |
| | PHE | 4 | SER | 5 | TYR | 4 | CYS | 4 C |
| | LEU | 10 | SER | 2 | Term | - | Term | - A |
| | LEU | 5 | SER | 2 | Term | - | TRP | 2 G |
| C | LEU | 4 | PRO | 4 | HIS | 3 | ARG | 12 T |
| | LEU | 3 | PRO | 3 | HIS | 5 | ARG | 7 C |
| | LEU | 0 | PRO | 2 | GLN | 5 | ARG | 0 A |
| | LEU | 26 | PRO | 9 | GLN | 7 | ARG | 2 G |
| A | ILE | 17 | THR | 7 | ASN | 9 | SER | 1 T |
| | ILE | 9 | THR | 10 | ASN | 9 | SER | 5 C |
| | ILE | 0 | THR | 7 | LYS | 14 | ARG | 0 A |
| | MET | 14 | THR | 10 | LYS | 3 | ARG | 1 G |
| G | VAL | 7 | ALA | 6 | ASP | 23 | GLY | 13 T |
| | VAL | 4 | ALA | 7 | ASP | 3 | GLY | 18 C |
| | VAL | 4 | ALA | 15 | GLU | 16 | GLY | 2 A |
| | VAL | 9 | ALA | 18 | GLU | 6 | GLY | 4 G |

Table 5.5: Codon utilisation in the E.coli aroA gene.

site (Figure 5.10). In Section 5.2.10 it was shown that the addition of fragment B (Figure 5.1) to the insert in pKD506 had no effect on aroA expression. This implies that expression is controlled by a sequence over 1200 bp upstream of the first codon of the aroA gene. The expression of aroA will be discussed in Chapter 6.

### 5.4.7 The aroA gene is truncated in pKD506

The other explanation for the low levels of EPSP synthase activity obtained with pKD506 transformed cells was that the aroA polypeptide made in these cells was prematurely terminated (Figure 5.7a). In Figure 5.13, the normal C-terminal sequence of EPSP synthase is shown alongside the sequence that would be produced from pKD506 transformed cells. In this case, the open reading frame runs to an 'in phase' stop codon in the sequence of pBR322.

EPSP synthase from pKD506 would be two residues shorter than the wild type enzyme and the last five residues would be, except for a single arginine, quite different from those found at the normal C-terminus. This difference could contribute to the 'low activity' of pKD506, but unless there is a crucial catalytic residue in this region it is unlikely to be responsible for such a low specific activity.

```
   GATTATTTCGAGCAGCTGGCGCGGATTAGCCAGGCAGCCTGA
1  ---------+---------+---------+---------+--  42
   AspTyrPheGluGlnLeuAlaArgIleSerGlnAlaAlaEnd
```

Figure 5.13a: C-terminal sequence of EPSP synthase synthesised from pKD501.

```
       2740     PvuII
          ¦
   GATTATTTCGAGCAGCTGCCTCGCGCGTTTCGGTGA
1  ---------+---------+---------+------  36
   AspTyrPheGluGlnLeuProArgAlaPheArgEnd
```

Figure 5.13b: C-terminal sequence of EPSP synthase synthesised from pKD506.

Figure 5.13: Carboxy-terminal sequence of EPSP synthase.

# CHAPTER 6 EXPRESSION OF THE E.COLI aroA GENE

## 6.1 Introduction

### 6.1.1 E.coli gene structure

A typical E.coli protein coding gene consists of three parts:

(i) the coding region - an open reading frame beginning with an ATG/Met or GTG/Val codon and ending with a TGA, TAG or TAA codon, specifying the product of the gene.

(ii) the 5'-flanking region - several important features are located in the sequence upstream of the coding region, most of which are concerned with the control of expression of the gene. Immediately before the coding region is a ribosome binding site or Shine-Dalgarno sequence. Upstream of this is the promoter or RNA polymerase binding site, which defines the startpoint for transcription of the gene. Related modulators of transcription and translation such as an operator (Jacob & Monod, 1961) or an attenuator (Yanofsky, 1981) may also be found here.

(iii) the 3'-flanking region - normally a stop signal for transcription (terminator) is found here.

Figure 6.1 shows the structure of such a gene and the approximate locations of these features, some of which have been described in detail in Chapter 4.

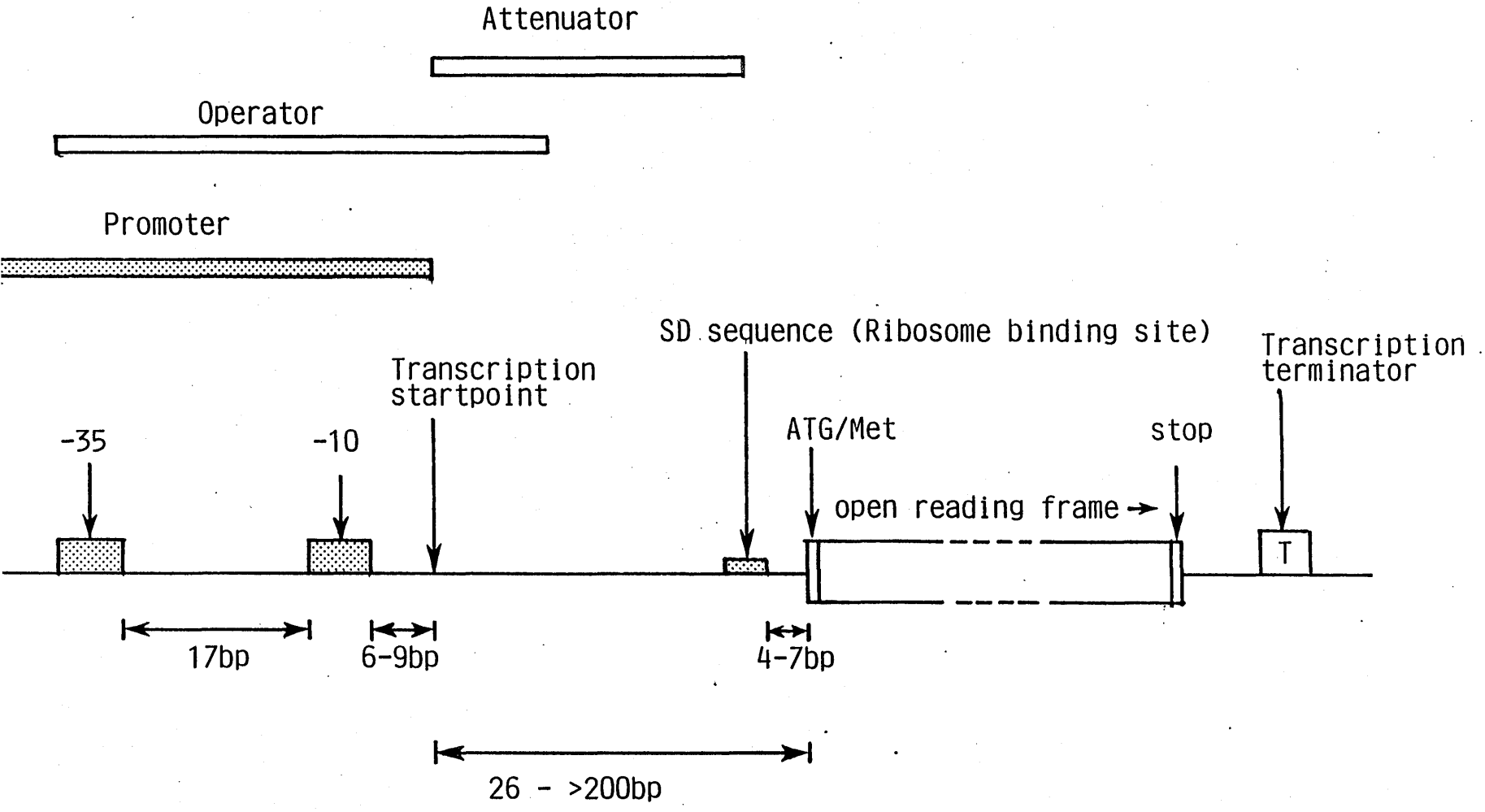



Figure 6.1: Features associated with a 'typical' E.coli gene. The operator and attenuator are only found in some genes.

Most E.coli genes display these features, but a large number are organised into operons, in which a number of cistrons are co-transcribed from a single promoter into a single message which may or may not be processed.

The experiments described in Chapter 5 suggest that the aroA gene is not organised in the manner described in Figure 6.1. Is there any evidence that aroA forms part of a larger transcriptional unit, or operon? This question may be addressed initially by a study of the sequences flanking the aroA coding region.

### 6.1.2 Examination of the 5'-flanking sequences of the aroA protein coding region - the 700 bp ClaI - ATG/Met sequence.

(a) Shine-Dalgarno sequence

The sequence upstream of the aroA coding region is illustrated in Figure 6.2. A possible S-D sequence is shown; it is located 4 bp before the ATG start codon. This GAG is likely to be the ribosome binding site for translation of aroA.

(b) Promoter

As described in Section 4.4.3, the consensus promoter -10 region contains three highly conserved nucleotides out of six, occurring in the sequence TA ---T. The 700 bp was searched by eye for this sequence and it was found four times. In none of these cases is the sequence preceded 17 bp upstream by the consensus sequence TTGACA. It is therefore


Figure 6.2: The DNA sequence between the open reading frame and the aroA gene.

unlikely that any of these sequences provides the promoter for *aroA*, although it is possible that one or other functions as a very weak promoter, allowing low levels of expression such as that seen in cells carrying pKD505 and pKD506 (Section 5.2.7).

(c) Open reading frames

Translation of the sequence with the TRNTRP program shows that there is an open reading frame running from left to right, from the ClaI site to a stop codon at position 1418 (Figure 5.11). This is 70 nucleotides before the start codon of *aroA*, and may represent the carboxy-terminal portion of a protein coding gene - an open reading frame of this length would not be found in the DNA by chance.

### 6.1.3 The 70 bp sequence between the open reading frames

The DNA sequence between the open reading frames is illustrated in Figure 6.2. Within this region is a sequence which resembles that of a rho-independent terminator of transcription (Rosenberg & Court, 1979), an inverted repeat followed by a run of T residues. The stem-loop structure which can form from this inverted repeat is illustrated in Figure 6.3. Using the rules of Tinoco *et al.* (1973), the free energy of this RNA structure was calculated to be -18.8 kcal/mol.

### 6.1.4 Examination of the 3'-flanking sequences of the aroA protein coding region

Approximately 45 bp downstream of the *aroA* coding region

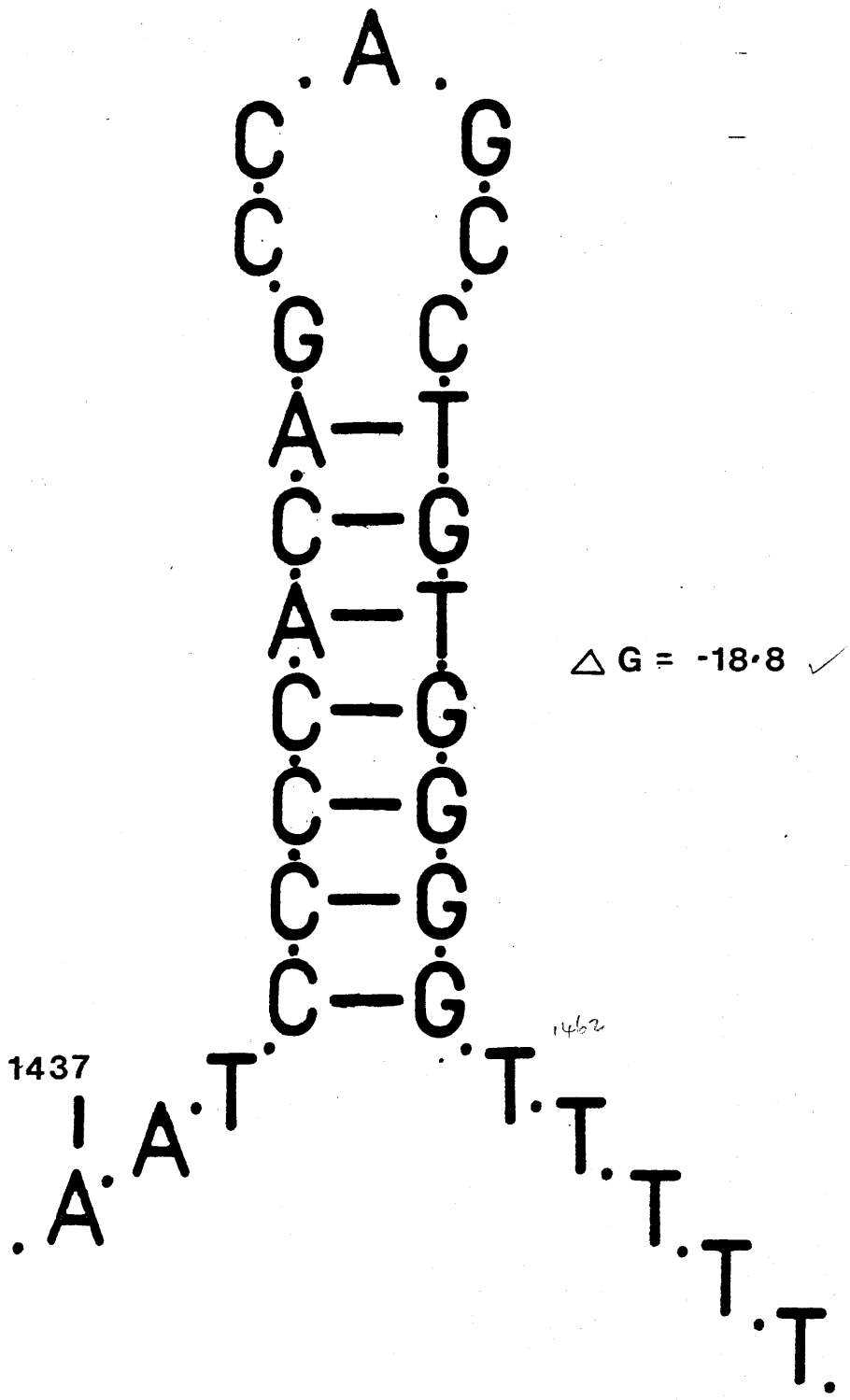



Figure 6.3: The stem-loop structure which may form downstream of the open reading frame. It is typical of the structure of a rho-independent transcription terminator.

(centred at position 2828) is a second possible rho-independent terminator. It is not as G/C rich, nor is it as long as the inverted repeat preceding aroA.

### 6.1.5 What is the relevance of the features observed in the 5'-flanking sequence of the aroA coding region?

The features observed in the 5'-flanking sequence of the aroA coding region suggest that there is another E.coli gene adjacent to and upstream of aroA. The coding region of this gene ends at a position 70 bp upstream of aroA, and is followed by a rho-independent transcription terminator. There is no space between the terminator of the putative upstream gene and the start of the aroA coding sequence for a promoter, nor is there any evidence from the sequence that a promoter exists. This makes aroA expression dependent on expression from the promoter of the upstream gene and read-through by RNA polymerase at the terminator. Thus, aroA forms part of an E.coli operon.

This explains why cloning from the ClaI site destroys expression of the aroA gene. The upstream gene is cut and aroA is separated from its promoter. The low level of expression from pKD505 and pKD506 is probably due to RNA polymerase binding to a weak promoter-like sequence or to a promoter on the vector.

It was also observed in Chapter 5 that cloning the 0.5 kb ClaI fragment to the left of the ClaI site back into pKD506 has no effect on aroA expression. From this, one can predict that the upstream open reading frame will

originate to the left of or around the second ClaI site and will run through the ClaI fragment continuous with that to the left of *aroA*. The promoter of both genes will be located to the left of this ClaI site i.e. in region C of Figure 5.1. Only by determining the DNA sequence of the above mentioned regions can these predictions be investigated.

## 6.2 Further DNA sequence analysis

### 6.2.1 Determination of the sequence of the 0.5 kb ClaI fragment (fragment B)

Fragment B was isolated after digestion of pKD501 with ClaI, followed by electrophoresis through 1% LMT agarose. An aliquot containing approximately 100 ng of DNA was ligated to 20 ng AccI-cleaved M13mp8. After transformation of *E.coli* JM101, six template DNAs were produced. The sequence at both ends of the fragment was determined.

Analysis of this sequence revealed three HpaII sites. More of the ClaI fragment was digested with HpaII and the fragments cloned into the AccI site of M13mp8. Templates were produced and sequenced, allowing the complete sequence to be built up on both strands of the DNA, with the exception of a 12 bp sequence. However, this sequence was very clear on the one strand from the sequence at one end of the ClaI fragment.

### 6.2.2 DNA sequence of part of region C

pKD501 was digested with BglII and KpnI. These enzymes cleave the DNA at the sites shown in Figure 5.1, yielding a 0.8 kb fragment including region C, spanning the second ClaI site and ending at the KpnI site in the centre of region B. The BglII-KpnI fragment was isolated after electrophoresis on 1% LMT agarose and digested with HpaII. An aliquot was also digested with ClaI. The digestion products were separated by electrophoresis on 2% LMT agarose and the HpaII fragment which contains the ClaI site excised. This fragment was cloned into AccI cleaved M13mp8 and the sequence of both strands determined.

The sequence of this fragment allowed the orientation of the sequence of the ClaI fragment relative to region A and region C to be deduced. It also yielded 286 bp of sequence to the left of the ClaI site.

### 6.2.3 The sequences described above

The DNA sequencing strategy described in the preceding sections is illustrated as part of Figure 6.16, and the sequence given in Figure 6.4. This sequence was linked to that of the ClaI-PstI region and translated in all three reading frames using TRNTRP. An open reading frame was found which is continuous with the open reading frame in the 700 bp ClaI-ATG/Met region (Figure 6.5). The open reading frame begins at position 326 and therefore spans most of the ClaI fragment. This is in line with the prediction made in Section 6.1.5 that the protein coding region would

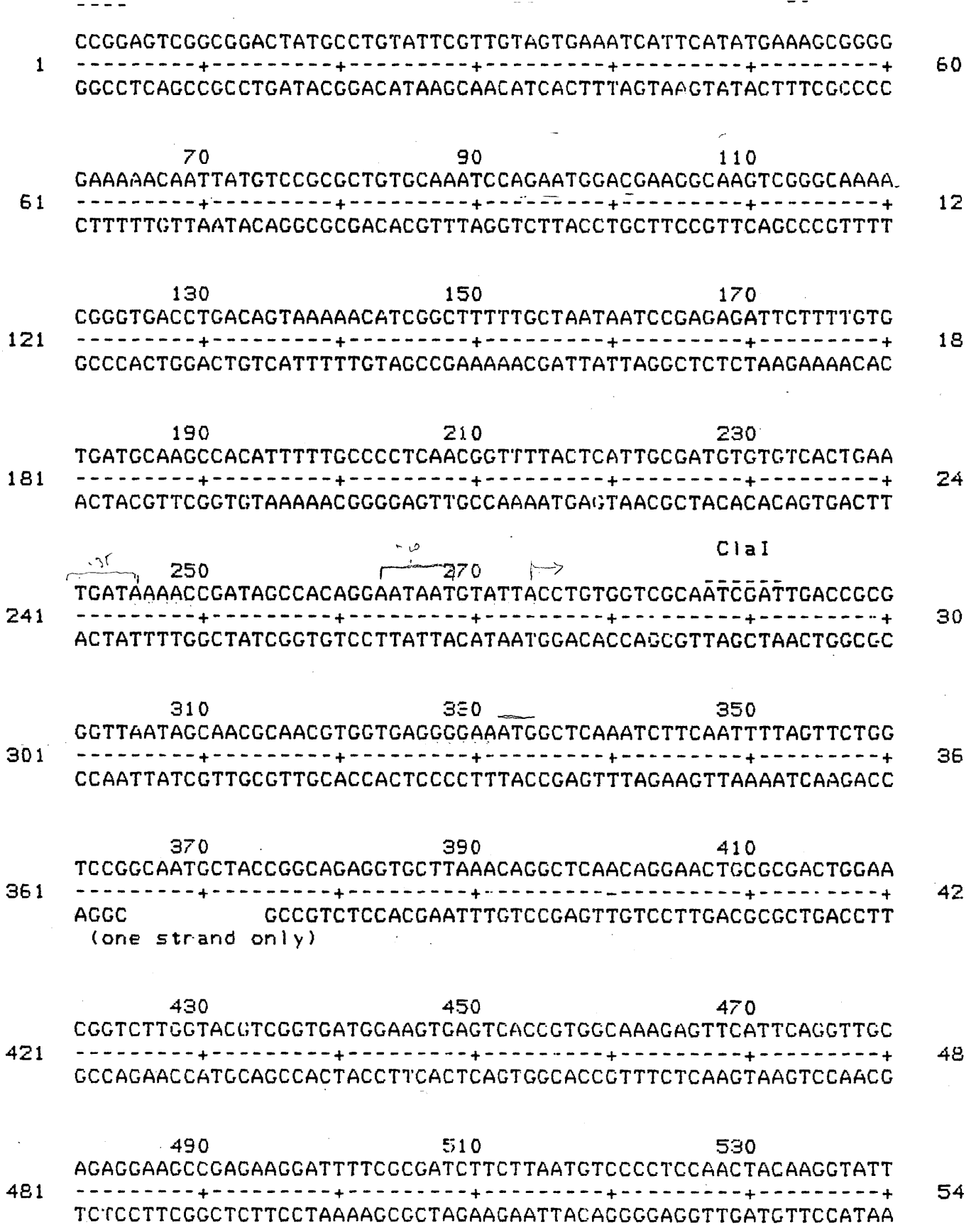



Figure 6.4:DNA sequence of the HpaII-ClaI-ClaI region.

```
        550                   570                   590
    ATTCTGCCATGGCGGTGGTCGCGGTCAGTTTGCTGCGGTACCGCTGAATATTCTCGGTGA
541 ---------+---------+---------+---------+---------+---------+   600
    TAAGACGGTACCGCCACCAGCGCCAGTCAAACGACGCCATGGCGACTTATAAGAGCCACT

        610                   630                   650
    TAAAACCACCGCAGATTATGTTGATGCCGGTTACTGGGCGGCAAGTGCCATTAAAGAAGC
601 ---------+---------+---------+---------+---------+---------+   660
    ATTTTGGTGGCGTCTAATACAACTACGGCCAATGACCCGCCGTTCACGGTAATTTCTTCG

        670                   690                   710
    GAAAAAATACTGCACGCCTAATGTCTTTGACGCCAAAGTGACTGTTGATGGTCTGCGCGC
661 ---------+---------+---------+---------+---------+---------+   720
    CTTTTTTATGACGTGCGGATTACAGAAACTGCGCTTTCACTGACAACTACCAGACGCGCG

        730                   750                   770
    GGTTAAGCCAATGCGTGAATGGCAACTCTCTGATAATGCTGCTTATATGCATTATTGCCC
721 ---------+---------+---------+---------+---------+---------+   780
    CCAATTCGGTTACGCACTTACCGTTGAGAGACTATTACGACGAATATACGTAATAACGGG

             ClaI

        790______
    GAATGAAACCATCGAT
781 ---------+------  796
    CTTACTTTGGTAGCTA
```

Figure 6.4:(cont)

start at the left of this fragment. There is also not enough room for a promoter between the ClaI site and the first ATG/Met in the open reading frame, implying, again as predicted, that the promoter lies further upstream than the second ClaI site.

### 6.2.4 Where is the start of the upstream coding region?

The most likely startpoint for the coding region of an *E.coli* gene is an ATG/Met codon. These codons occur at positions 332, 368, 440, then at 731 and further positions throughout the sequence. Considering the first three, are any preceded by an S-D sequence indicative of the startpoint of translation?

Figure 6.5 shows the three ATG codons; an S-D sequence is likely to be located 4-7 bp upstream of the startpoint. Only in the case of the ATG at position 332 is there a likely S-D sequence at the correct distance, suggesting that this represents the amino-terminus of the protein coding region. Only 43 bp separates the ClaI site from this ATG/Met.

### 6.2.5 Entire sequence of the protein coding region

The entire sequence of the protein coding region is illustrated in Figure 6.6. The startpoint is the ATG/Met at position 332 and the open reading frame runs for 362 amino acids to the TAA/stop codon at position 1418. The molecular weight of the protein encoded in this sequence is 39834. This is in agreement with the estimate of 40 000

Start of open reading frame ⟶

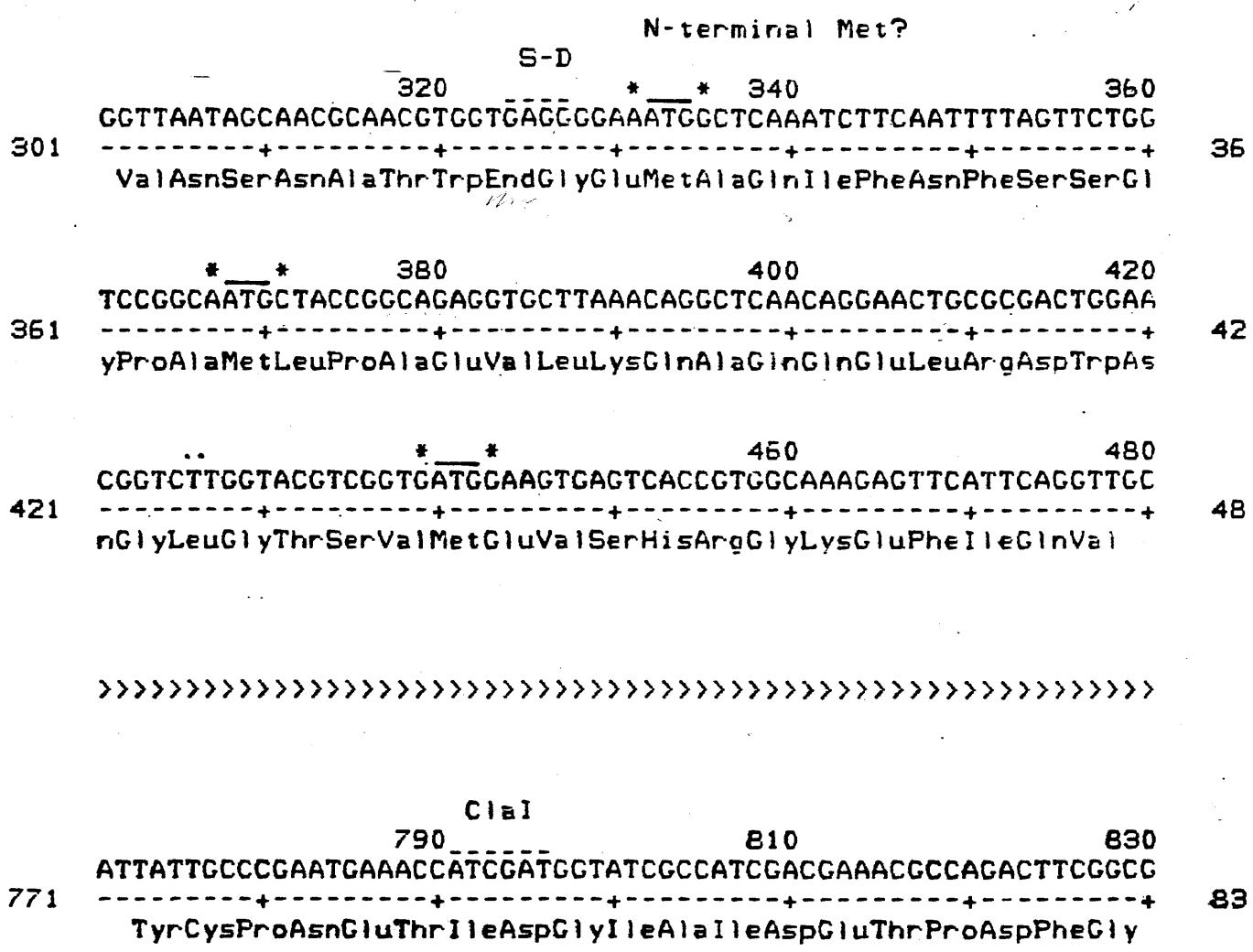



Figure 6.5: Location of the upstream open reading frame and possible S-D sequence.

```
ATGGCTCAAATCTTCAATTTTAGTTCTGGTCCGGCAATGCTACCGGCAGAGGTGCTTAAA
  1  ---------+---------+---------+---------+---------+---------+   60
   MetAlaGlnIlePheAsnPheSerSerGlyProAlaMetLeuProAlaGluValLeuLys
    M  A  Q  I  F  N  F  S  S  G  P  A  M  L  P  A  E  V  L  K
   [1]

          70                  90                 110
   CAGGCTCAACAGGAACTGCGCGACTGGAACGGTCTTGGTACGTCGGTGATGCAAGTCAGT
 61 ---------+---------+---------+---------+---------+---------+  120
   GlnAlaGlnGlnGluLeuArgAspTrpAsnGlyLeuGlyThrSerValMetGluValSer
    Q  A  Q  Q  E  L  R  D  W  N  G  L  G  T  S  V  M  E  V  S
   [21]

         130                 150                 170
   CACCGTCGCAAAGAGTTCATTCAGGTTGCAGAGGAAGCCGAGAAGGATTTTCGCGATCTT
121 ---------+---------+---------+---------+---------+---------+  180
   HisArgGlyLysGluPheIleGlnValAlaGluGluAlaGluLysAspPheArgAspLeu
    H  R  G  K  E  F  I  Q  V  A  E  E  A  E  K  D  F  R  D  L
   [41]

         190                 210                 230
   CTTAATGTCCCCTCCAACTACAAGGTATTATTCTGCCATGGCGGTGGTCGCGGTCAGTTT
181 ---------+---------+---------+---------+---------+---------+  240
   LeuAsnValProSerAsnTyrLysValLeuPheCysHisGlyGlyGlyArgGlyGlnPhe
    L  N  V  P  S  N  Y  K  V  L  F  C  H  G  G  G  R  G  Q  F
   [61]

         250                 270                 290
   GCTGCGCTACCGCTGAATATTCTCGGTGATAAAACCACCGCAGATTATGTTGATGCCGGT
241 ---------+---------+---------+---------+---------+---------+  300
   AlaAlaValProLeuAsnIleLeuGlyAspLysThrThrAlaAspTyrValAspAlaGly
    A  A  V  P  L  N  I  L  G  D  K  T  T  A  D  Y  V  D  A  G
   [81]

         310                 330                 350
   TACTGGCCGGCAAGTGCCATTAAAGAAGCGAAAAAATACTGCACGCCTAATGTCTTTGAC
301 ---------+---------+---------+---------+---------+---------+  360
   TyrTrpAlaAlaSerAlaIleLysGluAlaLysLysTyrCysThrProAsnValPheAsp
    Y  W  A  A  S  A  I  K  E  A  K  K  Y  C  T  P  N  V  F  D
   [101]

         370                 390                 410
   GCCAAAGTGACTGTTGATGGTCTGCGCGCGGTTAAGCCAATGCGTGAATGGCAACTCTCT
361 ---------+---------+---------+---------+---------+---------+  420
   AlaLysValThrValAspGlyLeuArgAlaValLysProMetArgGluTrpGlnLeuSer
    A  K  V  T  V  D  G  L  R  A  V  K  P  M  R  E  W  Q  L  S
   [121]
```

Figure 6.6: Coding sequence of the E.coli serC gene.

```
          430                   450                   470
     GATAATGCTGCTTATATGCATTATTCCCCGAATCAAACCATCGATGGTATCGCCATCGAC
421  ---------+---------+---------+---------+---------+---------+  48
     AspAsnAlaAlaTyrMetHisTyrCysProAsnGluThrIleAspGlyIleAlaIleAsp
      D   N   A   A   Y   M   H   Y   C   P   N   E   T   I   D   G   I   A   I   D
     [141]

          490                   510                   530
     GAAACGCCAGACTTCGGCGCAGATGTGGTGGTCGCCGCTGACTTCTCTTCAACCATTCTT
481  ---------+---------+---------+---------+---------+---------+  54(
     GluThrProAspPheGlyAlaAspValValValAlaAlaAspPheSerSerThrIleLeu
      E   T   P   D   F   G   A   D   V   V   V   A   A   D   F   S   S   T   I   L
     [161]

          550                   570                   590
     TCCCGTCCGATTGACGTCAGCCGTTATGGTGTAATTTACGCTGGCGCGCAGAAAAATATC
541  ---------+---------+---------+---------+---------+---------+  600
     SerArgProIleAspValSerArgTyrGlyValIleTyrAlaGlyAlaGlnLysAsnIle
      S   R   P   I   D   V   S   R   Y   G   V   I   Y   A   G   A   Q   K   N   I
     [181]

          610                   630                   650
     GGCCCGGCTGGCCTGACAATCGTCATCGTTCGTGAAGATTTGCTGGGCAAAGCGAATATC
601  ---------+---------+---------+---------+---------+---------+  660
     GlyProAlaGlyLeuThrIleValIleValArgGluAspLeuLeuGlyLysAlaAsnIle
      G   P   A   G   L   T   I   V   I   V   R   E   D   L   L   G   K   A   N   I
     [201]

          670                   690                   710
     GCGTGTCCGTCGATTCTGGATTATTCCATCCTCAACGATAACGGCTCCATGTTTAACACG
661  ---------+---------+---------+---------+---------+---------+  720
     AlaCysProSerIleLeuAspTyrSerIleLeuAsnAspAsnGlySerMetPheAsnThr
      A   C   P   S   I   L   D   Y   S   I   L   N   D   N   G   S   M   F   N   T
     [221]

          730                   750                   770
     CCGCCGACATTTGCCTGGTATCTATCTGGTCTGGTCTTTAAATGGCTGAAAGCGAACGGC
721  ---------+---------+---------+---------+---------+---------+  780
     ProProThrPheAlaTrpTyrLeuSerGlyLeuValPheLysTrpLeuLysAlaAsnGly
      P   P   T   F   A   W   Y   L   S   G   L   V   F   K   W   L   K   A   N   G
     [241]

          790                   810                   830
     GGTGTAGCTGAAATGGATAAAATCAATCAGCAAAAAGCAGAACTGCTATATGGGGTGATT
781  ---------+---------+---------+---------+---------+---------+  840
     GlyValAlaGluMetAspLysIleAsnGlnGlnLysAlaGluLeuLeuTyrGlyValIle
      G   V   A   E   M   D   K   I   N   Q   Q   K   A   E   L   L   Y   G   V   I
     [261]
```

Figure 6.6 (cont.)

```
          850                  870                  890
     GATAACAGCGATTTCTACCGCAATGACGTGGCGAAACGTAACCGTTCGCGGATGAACGTG
841  ---------+---------+---------+---------+---------+---------+  900
     AspAsnSerAspPheTyrArgAsnAspValAlaLysArgAsnArgSerArgMetAsnVal
      D  N  S  D  F  Y  R  N  D  V  A  K  R  N  R  S  R  M  N  V
     [281]

          910                  930                  950
     CCGTTCCAGTTGGCGGACAGTGCGCTTGACAAATTGTTCCTTGAAGAGTCTTTTGCTGCT
901  ---------+---------+---------+---------+---------+---------+  960
     ProPheGlnLeuAlaAspSerAlaLeuAspLysLeuPheLeuGluGluSerPheAlaAla
      P  F  Q  L  A  D  S  A  L  D  K  L  F  L  E  E  S  F  A  A
     [301]

          970                  990                 1010
     GGCCTTCATGCACTGAAAGGTCACCGTGTGGTCGGCGGAATGCGCGCTTCTATTTATAAC
961  ---------+---------+---------+---------+---------+---------+ 1020
     GlyLeuHisAlaLeuLysGlyHisArgValValGlyGlyMetArgAlaSerIleTyrAsn
      G  L  H  A  L  K  G  H  R  V  V  G  G  M  R  A  S  I  Y  N
     [321]

         1030                 1050                 1070
     GCCATGCCGCTGGAAGGCGTTAAAGCGCTGACAGACTTCATGGTTGAGTTCGAACGCCGT
021  ---------+---------+---------+---------+---------+---------+ 1080
     AlaMetProLeuGluGlyValLysAlaLeuThrAspPheMetValGluPheGluArgArg
      A  M  P  L  E  G  V  K  A  L  T  D  F  M  V  E  F  E  R  R
     [341]


     CACGGTTAA
081  ---------  1089
     HisGlyEnd
      H  G  *
     [361]
```

Figure 6.6 (cont)

for the heavily staining band on tracks a-d of the SDS-PAGE gel illustrated in Figure 5.8. It is therefore likely that the band on the gel corresponds to the product of this gene.

### 6.2.6 Codon utilisation for the open reading frame

Table 6.1 lists the codon utilisation for the sequence described above.

The overall pattern of codon utilisation in the table is similar to that for the *aroD*, *aroA* and the *trp* biosynthetic genes (Tables 4.4, 4.5 & 5.5). This suggests two things: firstly, the open reading frame is protein coding; secondly, the protein encoded is expressed at a similar level to the other gene products mentioned here. If the gene had been a repressor, expressed at only a few molecules/cell then the codon utilisation would have been quite different.

### 6.2.7 Application of Fickett's Testcode

The DNA sequence of the open reading frame was analysed by running the program TESTCODE of the WISGEN package (Sections 2.36 and 4.4.1; Fickett, 1982). The resulting plot (Figure 6.7a) was compared to a similar plot for the *aroA* coding region (Figure 6.7b) and this suggests that the open reading frame is protein coding.

| | T | C | A | G | |
|---|---|---|---|---|---|
| T | PHE 8 | SER 6 | TYR 8 | CYS 1 | T |
| | PHE 10 | SER 4 | TYR 5 | CYS 3 | C |
| | LEU 1 | SER 1 | Term – | Term – | A |
| | LEU 3 | SER 3 | Term – | TRP 5 | G |
| C | LEU 8 | PRO 1 | HIS 3 | ARG 9 | T |
| | LEU 3 | PRO 1 | HIS 3 | ARG 7 | C |
| | LEU 3 | PRO 2 | GLN 4 | ARG 0 | A |
| | LEU 12 | PRO 11 | GLN 7 | ARG 1 | G |
| A | ILE 9 | THR 1 | ASN 10 | SER 3 | T |
| | ILE 9 | THR 1 | ASN 10 | SER 3 | C |
| | ILE 0 | THR 3 | LYS 17 | ARG 0 | A |
| | MET 11 | THR 4 | LYS 3 | ARG 0 | G |
| G | VAL 7 | ALA 13 | ASP 15 | GLY 14 | T |
| | VAL 7 | ALA 8 | ASP 10 | GLY 12 | C |
| | VAL 4 | ALA 8 | GLU 13 | GLY 1 | A |
| | VAL 10 | ALA 12 | GLU 6 | GLY 1 | G |

Table 6.1: Codon utilisation for the serC gene.

Figure 6.7: Application of Fickett's Testcode to the DNA sequence of the open reading frame.

(a) the open reading frame; numbers refer to the sequence in Figure 6.6. This open reading frame is clearly protein coding, when compared with:

(b) the *aroA* coding region; numbers refer to the sequence in Figure 5.11.

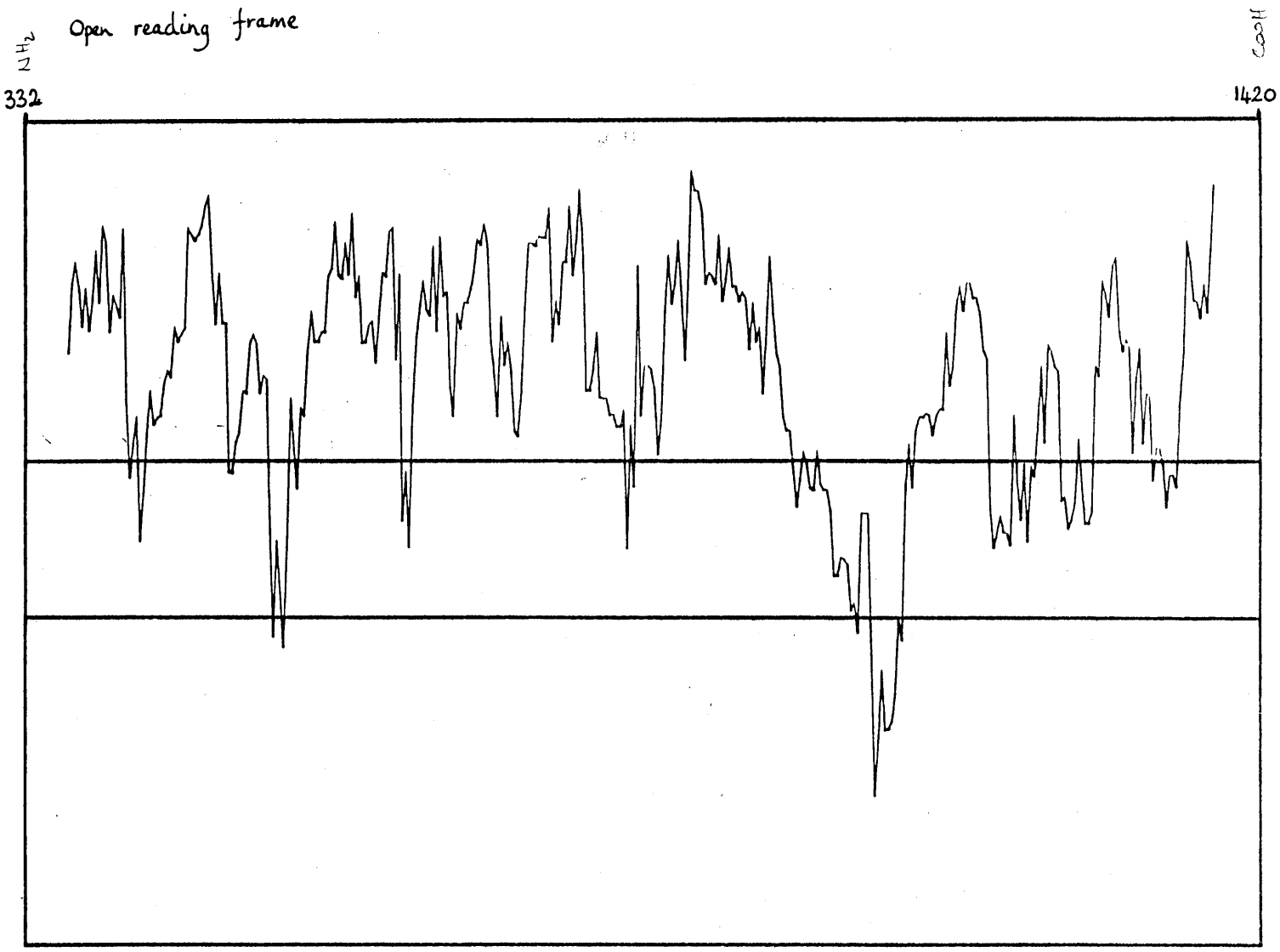



Figure 6.7a

aroA coding region

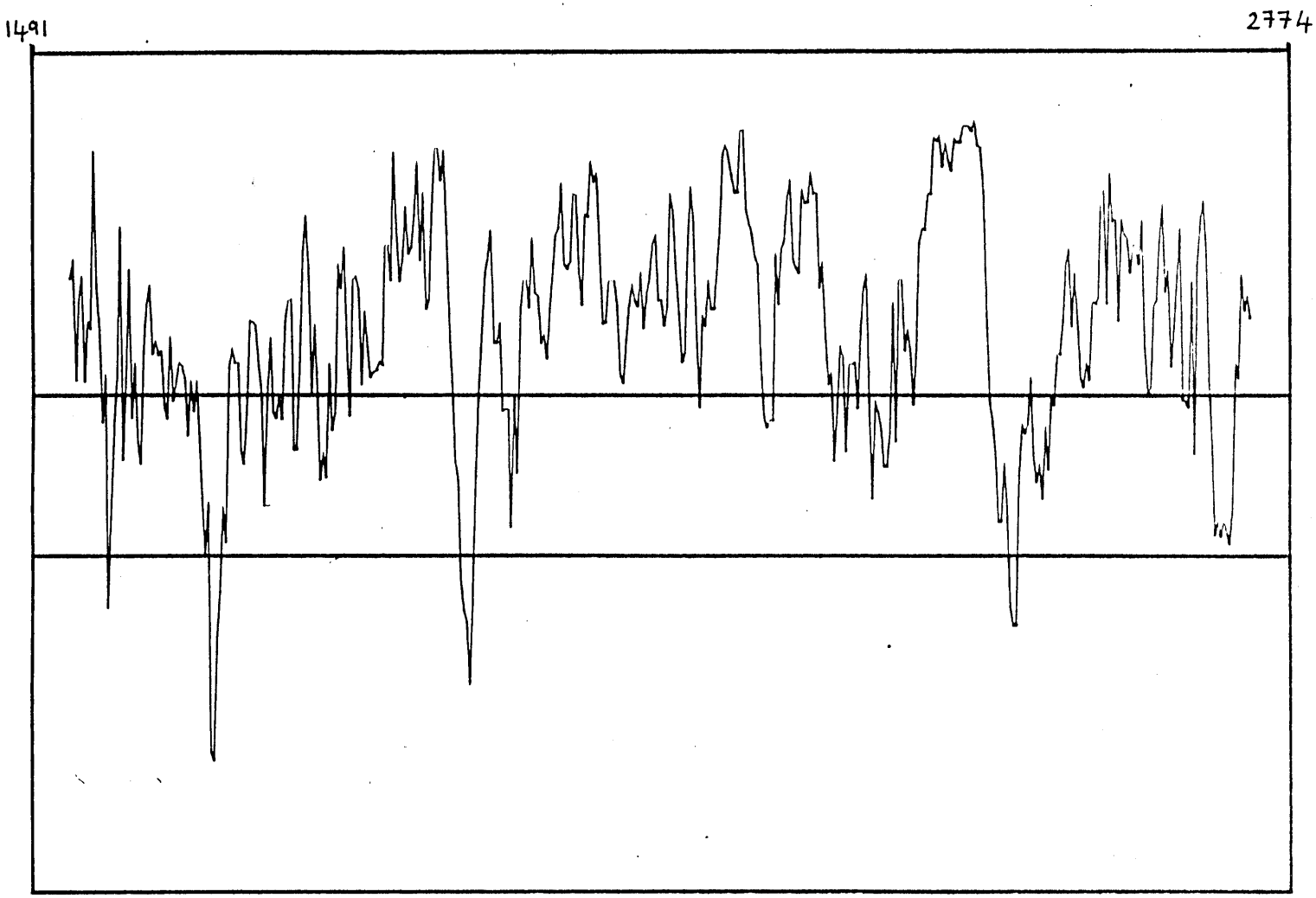



Figure 6.7b

### 6.2.8 Further aspects of the codon utilisation, relevant to expression of aroA and the other genes sequenced in this work

The codon utilisation patterns of 'highly' expressed E.coli genes (e.g. for ribosomal proteins) have been compared with those of 'weakly' expressed genes (e.g. for repressors) (Grosjean & Fiers, 1982). It has been proposed that the efficiency of translation of an mRNA is affected by two parameters. The first is the frequency of use of 'modulating' codons i.e. codons recognised by tRNAs which are present in relatively low molar amounts. The codon composition of the genes sequenced in this study have been analysed and the number of modulating codons determined (Table 6.2). For comparison, the data for 'strongly' (S) and 'moderately to weakly' (W) expressed genes of E.coli has been collated from Table III of Grosjean & Fiers (1982). The data suggests that the aro genes come into the 'moderately to weakly' expressed category (the frequency for aroD is lower than 0.02, but this is based on a single modulating codon in 59 possible positions and it is unlikely that this implies that aroD is 'strongly' expressed).

The second parameter is the frequency of use of codons in a degenerate set that optimises the codon-anticodon interaction. The data for 'optimal' and 'suboptimal' codon-anticodon interactions is shown in Table 6.3, again alongside data from Table III of Grosjean & Fiers (1982). In this case, the ratio of optimal:suboptimal codons is very close for the aro genes, the trp biosynthetic genes and

| Modulating codons | aroD | aroA | aroE | serC | trpE,D,C,B | S | W |
|---|---|---|---|---|---|---|---|
| CUA | 0 | 0 | 1 | 3 | 21 | 3 | 22 |
| AUA | 0 | 0 | 3 | 0 | 2 | 2 | 27 |
| CGA/G | 0 | 2 | 4 | 1 | 16 | 4 | 69 |
| AGA/G | 1 | 1 | 1 | 0 | 9 | 4 | 45 |
| GGA/G | 0 | 6 | 5 | 1 | 74 | 18 | 108 |
| Total | 1 | 9 | 14 | 5 | 122 | 31 | 271 |
| Total possible | 59 | 143 | 81 | 93 | 1220 | 1516 | 1612 |
| Frequency | 0.017 | 0.06 | 0.17 | 0.05 | 0.1 | 0.02 | 0.17 |

Table 6.2: Frequency of 'modulating' codons in the aro genes.

| Optimal | aroD | aroA | aroE | serC | trpE,D,C,B | S | W |
|---|---|---|---|---|---|---|---|
| UUC | 3 | 4 | 4 | 10 | 67 | 113 | 102 |
| AUC | 7 | 9 | 6 | 10 | 83 | 262 | 118 |
| CCU | 2 | 4 | 4 | 1 | 20 | 21 | 29 |
| ACU | 3 | 7 | 4 | 1 | 22 | 103 | 48 |
| UAC | 1 | 4 | 0 | 5 | 40 | 98 | 65 |
| AAC | 2 | 9 | 2 | 10 | 91 | 159 | 98 |
| CGU | 7 | 12 | 3 | 9 | 77 | 223 | 99 |
| GGU | 5 | 13 | 13 | 14 | 116 | 226 | 124 |
| Total | 30 | 62 | 36 | 60 | 516 | 1205 | 683 |
| Suboptimal | | | | | | | |
| UUU | 5 | 9 | 10 | 8 | 68 | 39 | 151 |
| AUU | 9 | 17 | 9 | 9 | 113 | 67 | 156 |
| CCC | 0 | 3 | 2 | 1 | 25 | 2 | 46 |
| ACC | 8 | 10 | 3 | 4 | 99 | 137 | 119 |
| UAU | 5 | 9 | 5 | 8 | 76 | 34 | 96 |
| CGC | 2 | 7 | 5 | 7 | 122 | 101 | 133 |
| GGC | 8 | 18 | 7 | 12 | 127 | 174 | 140 |
| Total | 37 | 73 | 41 | 49 | 630 | 554 | 841 |
| Ratio | 0.81 | 0.85 | 0.88 | 1.22 | 0.82 | 2.17 | 0.81 |

Table 6.3: Use of optimal and suboptimal codons in the aro genes.
'Ratio' refers to the ratio of optimal to suboptimal codons.

for the 'moderately to weakly' expressed class (W). The open reading frame has a higher ratio, probably indicating that its product is synthesised in greater quantity within the cell than the other gene products mentioned here.

### 6.3.1 What is the protein encoded by this open reading frame?

The region of the *E.coli* chromosome around minute 20 is illustrated in Figure 1.4 (from Bachmann, 1983). The map shows the marker loci flanking *aroA*. This cloning work has located *aroA* with respect to *rpsA*.

In the opposite direction from *rpsA*, the most closely mapped gene is *pdxC*. According to Bachmann (1983), *pdxC* is defined as a requirement of pyridoxine. Pyridoxine (Vitamin $B_6$) and its coenzyme forms function in a large number of different enzymatic reactions in which amino acids or amino groups are transformed or transferred. The most common type of enzymatic reaction requiring pyridoxal phosphate as a coenzyme is transamination. Little is known about the biosynthesis of pyridoxine in micro-organisms and it is assumed that the *pdxC* mutation is in the structural gene for a pyridoxine biosynthetic enzyme.

*pdxC* was first described in Taylor (1970). This mutation represented a third class of non-allelic mutations leading to a pyridoxine requirement. Preliminary results indicated that it was very closely linked to *aroA* (95-97% cotransduction frequency with phage P1), but it was not known whether the gene lay clockwise or anticlockwise to *aroA*.

The gene order was established by Cronan et al. (1972) during experiments aimed at mapping the fabA locus (unsaturated fatty acid biosynthesis). These indicated that

-pdxC - aroA - (rpsA) - pyrD - fabA -

is the correct orientation.

Since the product of the pdxC gene is not known, it is not possible to assay the enzyme, either in wild type E.coli or in plasmid carrying strains which will be over-expressing the product.

### 6.3.2 The serC gene

There is evidence suggesting that the mutation which defines pdxC is an allele of the serC gene. The serine biosynthetic pathway brings about the conversion of 3-phosphoglycerate to serine in three steps (Figure 6.8). The enzymes catalysing the steps are encoded in the serA, serB and serC genes, which have been located on the E.coli map (Bachmann, 1983). The serC genotype, a mutation in the gene which encodes 3-phosphoserine amino-transferase (PSAT), was first described in E.coli B (Dempsey, 1969) and subsequently in E.coli K12 (Clarke et al., 1973) where it was shown to be closely linked to aroA (P1 cotransduction frequency 93%). serA and serB mutants require only serine for growth, but serC mutants display a double requirement for serine and pyridoxine.

While characterising pyridoxine auxotrophs of E.coliB, a number of serine and pdxF mutants were studied (Dempsey & Itoh, 1970). Five classes of pyridoxineless mutants were

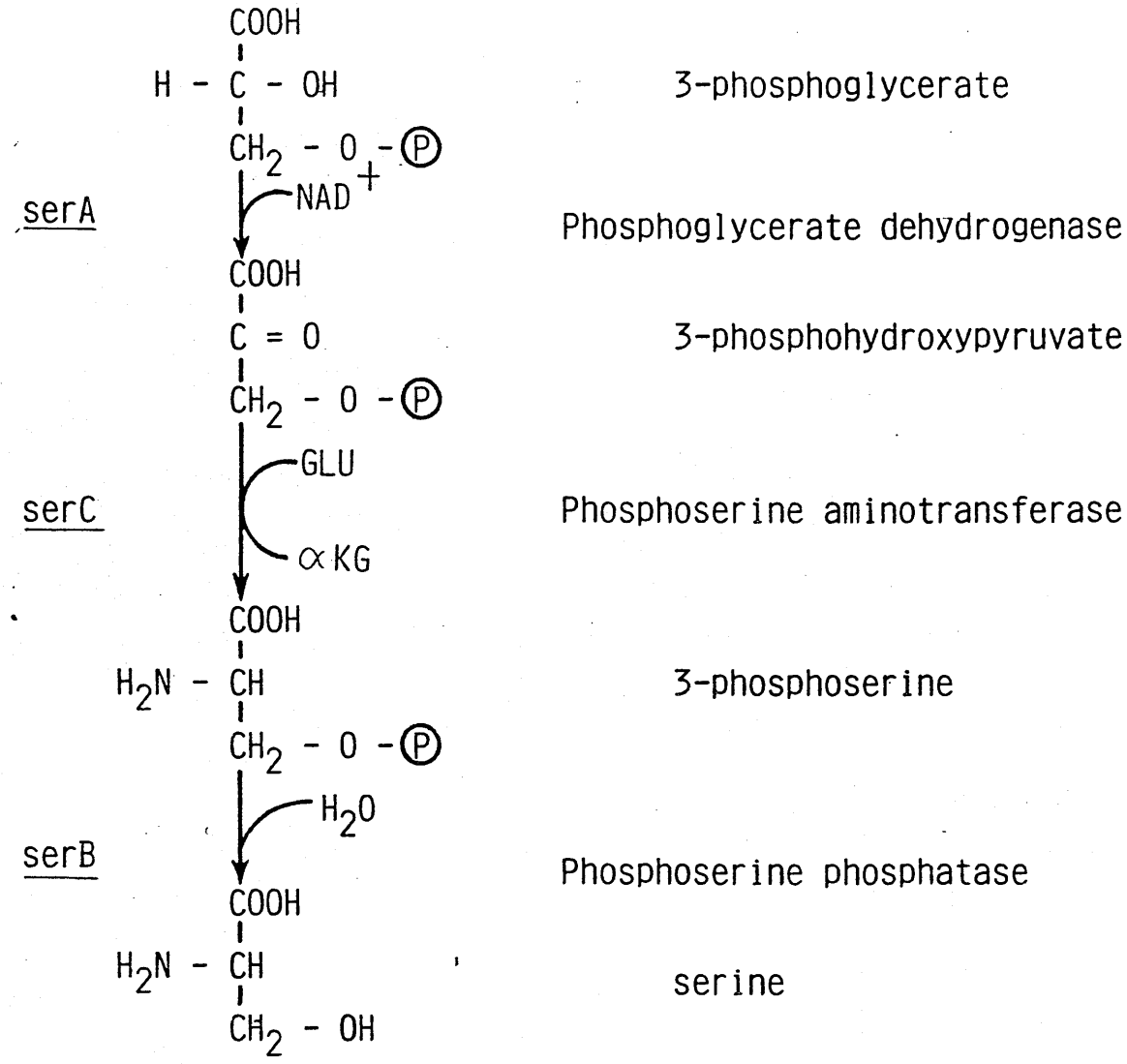



Figure 6.8: The biosynthesis of serine from 3-phosphoglycerate.

identified, one of which, pdxF lacked the enzyme PSAT. It was shown that pdxF is a special class of serC mutant. Two schemes were proposed to explain this finding and are illustrated in Figures 6.9a and 6.9b.

E.coli K12 serC mutants also require both serine and pyridoxine for growth. Shimizu and Dempsey (1978) showed that mutants revert easily to pyridoxine independence without regaining PSAT activity. When 3-hydroxypyruvate is used as a supplement, both the revertants and the parental types synthesise pyridoxine in normal amounts, but neither can use this compound to satisfy their serine requirement. As serine alone is inadequate to provide the nutritional requirement of serC mutants, these mutants must be unable to synthesise 3-hydroxypyruvate from serine. This suggests that in normal E.coli, PSAT serves as a catalyst for transaminating small amounts of serine to 3-hydroxypyruvate, which is then used in pyridoxine biosynthesis (Figure 6.9c).

Transductional analysis was used to locate the secondary mutations in the revertant strains. The results are summarised in Figure 6.10. The authors suggest that the secondary mutation in KL281B may be either in a gene which lies between serC and aroA or it may be a mutation in the serC allele itself which allows more efficient conversion of serine to 3-hydroxypyruvate. The mutation in KL281B20 which converts KL281B to wild phenotype must be so closely linked to aroA that recombination between it and aroA is extremely unlikely.

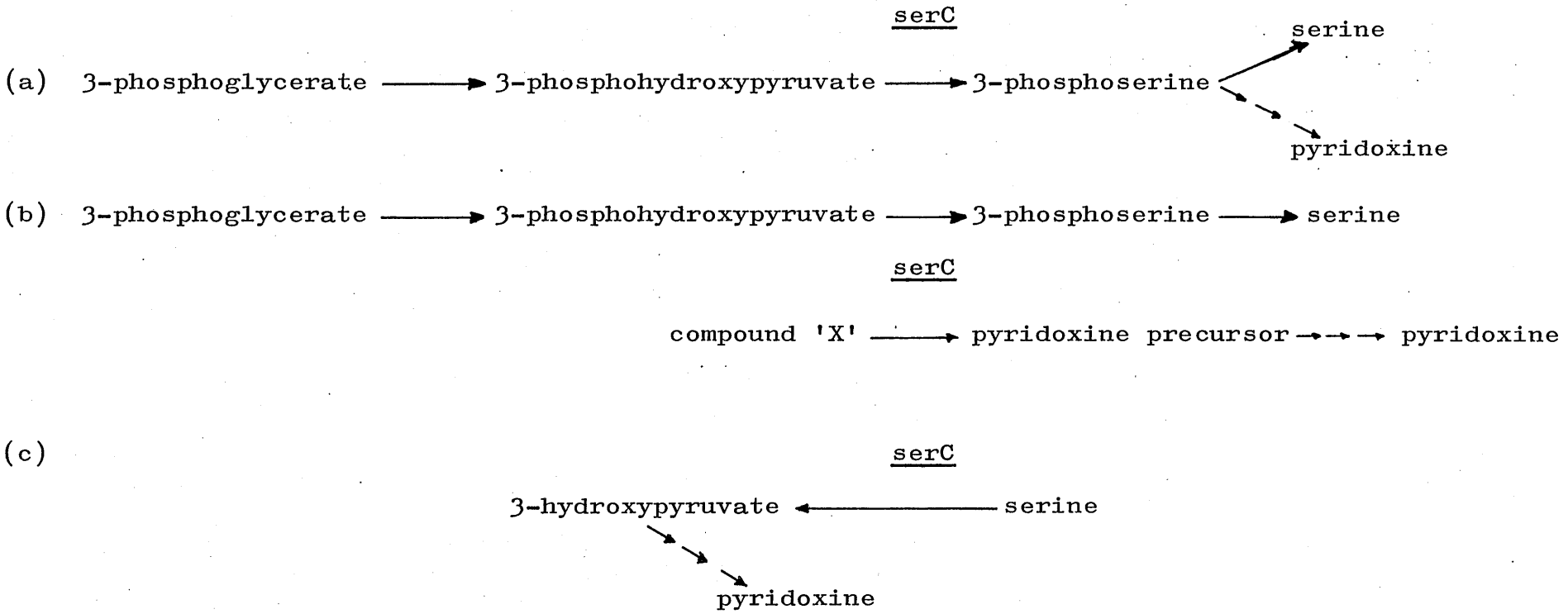



Figure 6.9. Proposed scheme for the involvement of PSAT (serC) in pyridoxine biosynthesis.

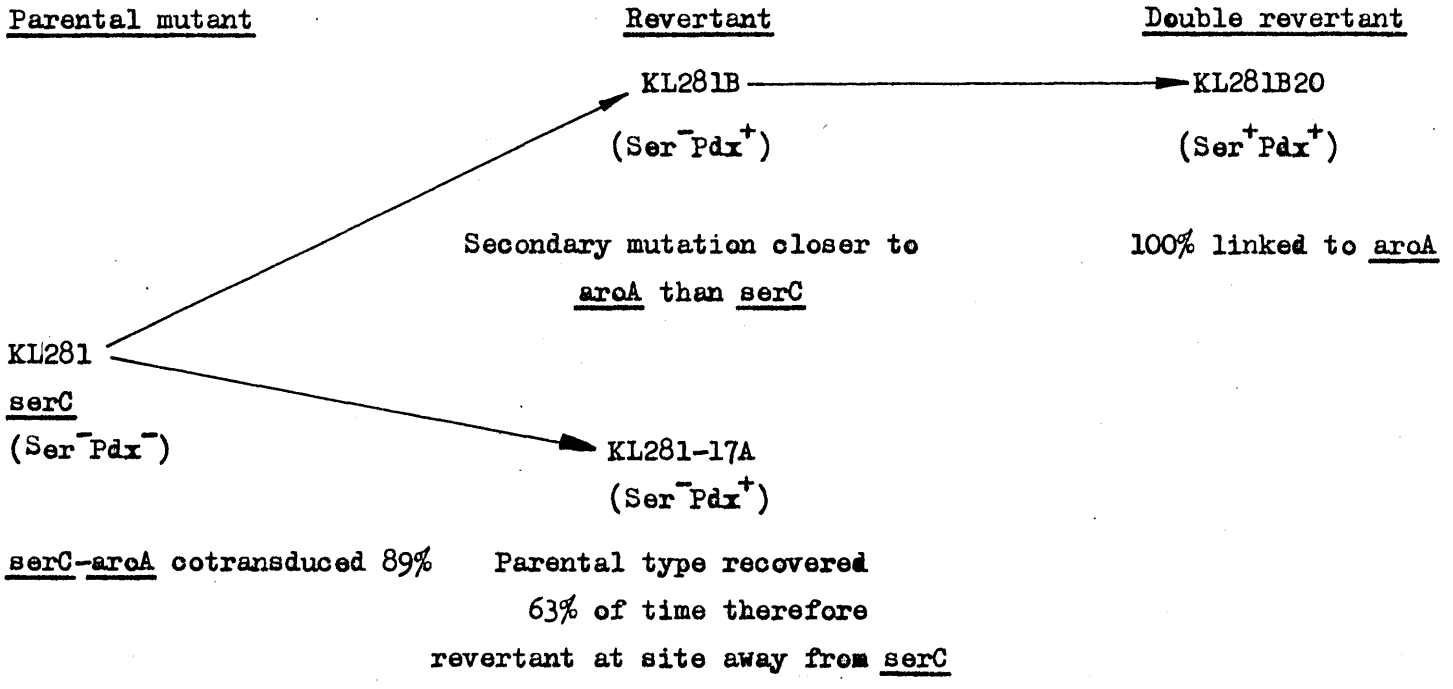



Figure 6.10: *E.coli* *serC* revertants: Linkage relationship to *aroA*

Nowhere in the literature are there any reports of transduction analysis using both serC and pdxC mutants. The observations described above suggest that serC mutants and pdxC mutants are different classes of mutants within the same gene. Any mutations in the gene upstream of aroA would map extremely close to aroA or would be inseparable from it in cotransductional analysis, as was found for the revertants described above. The results suggest that the product of the open reading frame may be PSAT.

Itoh & Dempsey (1970) have reported the purification of PSAT from E.coliB. The enzyme was nearly homogeneous (one minor contaminant on SDS-PAGE); the native $M_r$ was estimated to be 80 500 (sucrose density centrifugation). They also confirmed that the enzyme used pyridoxal 5'-phosphate as a co-factor, in common with other transaminases.

### 6.3.3 Transformation of an E.coli serC mutant

The first confirmation that the upstream gene is serC was provided by genetic complementation of an E.coli K12 serC mutant E.coli KL282. This mutant was obtained from the E.coli Genetic Stock Centre as a glycerol stock. The filter disk carrying the bacteria was dropped into 10 ml Nutrient broth (Difco, 13 g/l) and incubated overnight. This culture was then streaked out onto LA and single colonies obtained. A single colony was further purified and its growth properties investigated. This confirmed that it required supplementation with both serine (20 µg/ml)

and pyridoxine (75 μg/l) in order to grow on minimal medium.

E.coli KL282 was transformed with a number of plasmids and transformants selected by plating on LA, supplemented with the appropriate antibiotic. Several colonies from each plate were then transferred onto one minimal medium and one minimal medium plate supplemented with serine and pyridoxine. The plasmids used and the results obtained are detailed in Table 6.4. As expected, only those clones carrying the entire upstream gene are capable of complementing the serC mutation in E.coli KL282.

### 6.3.4 Specific activity of PSAT in crude extracts

More direct evidence that the upstream gene is serC was provided by measuring the specific activity of PSAT in crude extracts of E.coli K12 and of plasmid carrying strains. Once again, one can predict a 100-fold higher level in cells carrying serC on pAT153 than in wild type cells, assuming that the gene is expressed constitutively.

PSAT can be assayed by coupling the formation of α-ketoglutarate to the glutamate dehydrogenase reaction and following the decrease in $A_{340}$ as NADH is consumed (Figure 6.11). The assay conditions used were modified from Hirsch-Kolb & Greenberg (1971) (Section 2.19).

Crude extracts were prepared from 100 ml batches of E.coli cells grown in minimal medium to $A_{650}$ = 1.3. Sonication buffer 'B' was used (Section 2.18). Three extracts were assayed for PSAT activity - E.coli K12, E.coli AB2829/pKD501 and E.coli AB2829/pKD506. The specific

| Plasmid | Growth on: minimal medium | minimal + serine + pyridoxine |
|---|---|---|
| pAT153 | 0 | 1 |
| pKD501 | 1 | 1 |
| pKD503 | 1 | 1 |
| pKD506 | 0 | 1 |
| - | 0 | 1 |

Table 6.4: Growth of E.coli KL282, transformed with a number of different plasmids.
1 - growth 0 - no growth

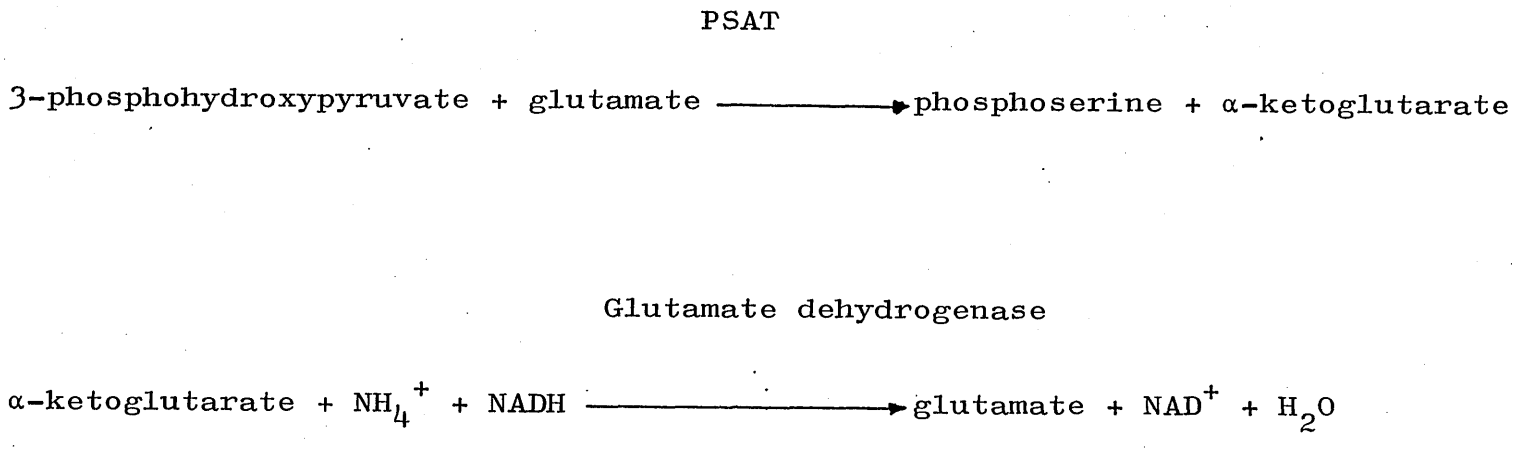



Figure 6.11. Assay reaction for PSAT.

activities of PSAT and EPSP synthase are shown in Table 6.5.

These results show that PSAT is overexpressed approximately 100-fold in cells harbouring pKD501 and that a wild type level is seen in cells harbouring pKD506. This confirms the serC complementation results.

### 6.3.5 Purification of PSAT

It has been suggested that a heavily stained band on an SDS-PAGE of crude extracts of E.coli carrying certain aroA recombinant plasmids (Figure 5.8) represents a plasmid encoded enzyme. The experiments described above imply that this band corresponds to PSAT. The $M_r$ of the protein in the band is estimated from Figure 5.8 to be 40 000; this is approximately half the native molecular weight estimate for PSAT from E.coli B (Itoh & Dempsey, 1970).

During the purification of EPSP synthase from E.coli AB2829/pKD501 the band corresponding to PSAT can be seen to persist through the first two steps, but is seen only as a minor component in the pool of EPSP synthase after the next step (Figure 1 in Duncan, Lewendon & Coggins, 1984a). EPSP synthase is eluted from phenyl-Sepharose by a decreasing salt gradient. A large protein peak is observed before the peak corresponding to EPSP synthase. This peak was assayed for PSAT and it was found that this activity was present. Fractions over the peak were pooled and the enzyme was further purified to homogeneity by chromatography on an FPLC mono-Q ion-exchange column (Ann Lewendon, unpublished results).

| Extract | Specific activity of PSAT (U/mg) | Relative activity | Specific activity of EPSP tynthase (U/mg) |
|---|---|---|---|
| E.coli K12 | 0.011 | 1 | 0.004 |
| E.coli AB2829/pKD501 | 1.06 | 96 | 0.42 |
| E.coli AB2829/pKD506 | 0.010 | 1 | 0.004 |

Table 6.5: Specific activity of PSAT in crude extracts of E.coli.

Approximately 8 mg of purified PSAT were obtained from 20 g of E.coli. The purification factor from the crude extract was 6.9 -fold. It was shown that the enzyme has a sub-unit $M_r$ of 39 000 and a native $M_r$ of 68 000, similar to the dimeric enzyme isolated from E.coli B (Itoh & Dempsey, 1970).

### 6.3.6 The amino acid composition of PSAT

The amino acid composition of purified PSAT has been determined (Ann Lewendon, unpublished results). The experimentally determined composition and the composition derived from the DNA sequence are illustrated in Table 6.6. There is a good agreement between both sets of data, confirming that the open reading frame is PSAT. Note that there is one fewer methionine residue in the experimental value than in the predicted value.

### 6.3.7 The amino-terminal sequence of PSAT

The amino-terminal sequence of PSAT was determined on the automated protein sequencer facility at Aberdeen, by Ann Lewendon. The sequence obtained is illustrated in Figure 6.12. The sequence agrees with that predicted from the DNA, starting with the Ala residue at position 335 in the DNA sequence, with the exception of Leu-19. This was identified as Phe by HPLC of the PTH-amino acid derivative at that cycle. However, there was some doubt as to whether this assignment was correct (see A. Lewendon, Ph.D. Thesis, University of Glasgow, 1984). The DNA sequence in this

| Amino acid | Relative amino acid composition based on Leu = 30 residues | Theoretical amino acid composition predicted from the DNA sequence |
|---|---|---|
| Asx | 39.6 | 45 |
| Thr | 12.5 | 12 |
| Ser | 20.2 | 20 |
| Glx | 33.3 | 30 |
| Pro | 15.7 | 15 |
| Gly | 31.4 | 28 |
| Ala | 43.2 | 41 |
| Cys | 4.7 | 4 |
| Val | 28.1 | 28 |
| Met | 10.4 | 11 |
| Ile | 18.8 | 19 |
| Leu | 30 | 30 |
| Tyr | 13.1 | 13 |
| Phe | 18.4 | 18 |
| His | 6.2 | 6 |
| Lys | 19.8 | 20 |
| Arg | 16.6 | 17 |
| Trp | nd | 5 |
| Gln | nd | 11 |
| Asn | nd | 20 |

Table 6.6: Amino acid composition of PSAT, compared with the composition predicted by the DNA sequence of the *serC* gene.

1 10

Ala - Gln - Ile - Phe - Asn - Phe - Ser - Ser - Gly - Pro -

11 20

Ala - Met - Leu - Pro - X - Glu - Val - Phe - Lys - Gln

21 30

Ala - Gln - Gln - X - Leu - Arg - Asp - Trp - Asn - Gly -

31 37

Leu - X - X - X - X - Met - Glu

Figure 6.12. Amino-terminal sequence of PSAT.

region is quite clearly correct from both strands of the DNA. The codons for Leu and Phe are:

Leu: TTA, TTG, CTT, CTC, CTA, CTG

Phe: TTT, TTC

The Leu codon used at this position is CTT; only a change to TTT could alter the DNA sequence to give Phe. The nucleotide sequence is unambiguously CTT.

The amino-terminal methionine residue predicted from the PSAT coding sequence is not found in the native enzyme, indicating that it has been removed by post-translational modification. This agrees with the finding that the experimentally derived amino acid composition has one fewer methionine than the DNA sequence predicts.

### 6.3.8 Is the linked expression of aroA and serC a cloning artefact?

It is conceivable that *aroA* and *serC* expression is not linked *in vivo* and that the observations which have been made during this work are an artefact introduced during manipulation of the genomic DNA. For example, the deletion of a piece of DNA between the transcription terminator located downstream of *serC* and the start of the *aroA* coding region may have removed the *in vivo* promoter of the *aroA* gene. It is not possible that this has happened in going from the phage DNA (λ*pserC*) to the plasmids because pKD505 and pKD506 were constructed in a different way and yet show the same properties regarding *aroA* expression.

pKD505 was derived from the phage DNA directly, but pKD506 was constructed from DNA isolated from a digest of pKD501 DNA.

This leaves the possibility that a region was deleted during the aberrant excision event which led to the formation of λ*pserC*, or during propagation of the phage, prior to selection. Deletions are not uncommon among phage which carry a portion of foreign DNA. Constraints on the amount of DNA which can be packaged into a phage head mean that deletions which make the DNA a more packagable size will be selected for.

It is very unlikely, however, that any cloning artefacts are present. The plasmid pMON1 (Rogers *et al.*, 1983), which also carries the *E.coli aroA* gene, was constructed by subcloning a HindIII fragment of a λ-recombinant phage carrying a segment of the *E.coli* chromosome around minute 20. Rogers *et al.* state that the restriction map of pMON1 is identical to the independently derived maps of λ*pserC* and another phage λ*aspC* (Christiansen & Pederson, 1981), both of which carry DNA from the minute 20 region.

Nevertheless, it was decided to confirm these findings by isolating the *aroA* gene from *E.coli* genomic DNA. High molecular weight DNA was prepared from *E.coli* K12 (ATCC 14948) by a modification of the method of Marmur (1961) (Section 2.8), followed by centrifugation to equilibrium in a CsCl gradient. The DNA was digested with PstI and purified by phenol/chloroform extraction and ethanol precipitation. The fragments were then ligated to PstI

cleaved and phosphatase treated pAT153 and the ligation mix used to transform E.coli AB2829. Transformants were selected by plating on LA/tet then replica plated onto minimal medium. Of approximately 2 000 $tet^R$ colonies screened, 8 were capable of growth on minimal medium. Plasmid DNA was isolated from these colonies and restriction enzyme digestion revealed an insert at the PstI site in only one of them. This plasmid was designated pKD508.

Further digests showed that pKD508 is formally equivalent to pKD502 in that it has an insert at the PstI site in the same orientation as in pKD502.

### 6.3.9 Are pKD508 and pKD502 identical?

A series of restriction enzyme digests has shown that pKD508 and pKD502 are not identical. Figure 6.13, tracks 10 and 11, shows that the PstI fragment inserted in pKD508 is larger (4.9 kb) than that inserted in pKD502 (4.6 kb). Digestion with ClaI (tracks 13 and 14) revealed an extra site in pKD508 which is not found in pKD502. A restriction map of pKD508 was built up (Figure 6.14). The fact that the two clones are not identical may simply be due to a different strain of E.coli K12 being used as a source of the DNA in the initial cloning/genetic manipulation experiments.

More importantly, the fragment sizes in the ClaI-PvuII region are the same. Tracks 3 and 4 (Figure 6.13) show that the same 1.9 kb ClaI-PvuII fragment is present in both plasmids. In tracks 1 and 2, the lower triplet of

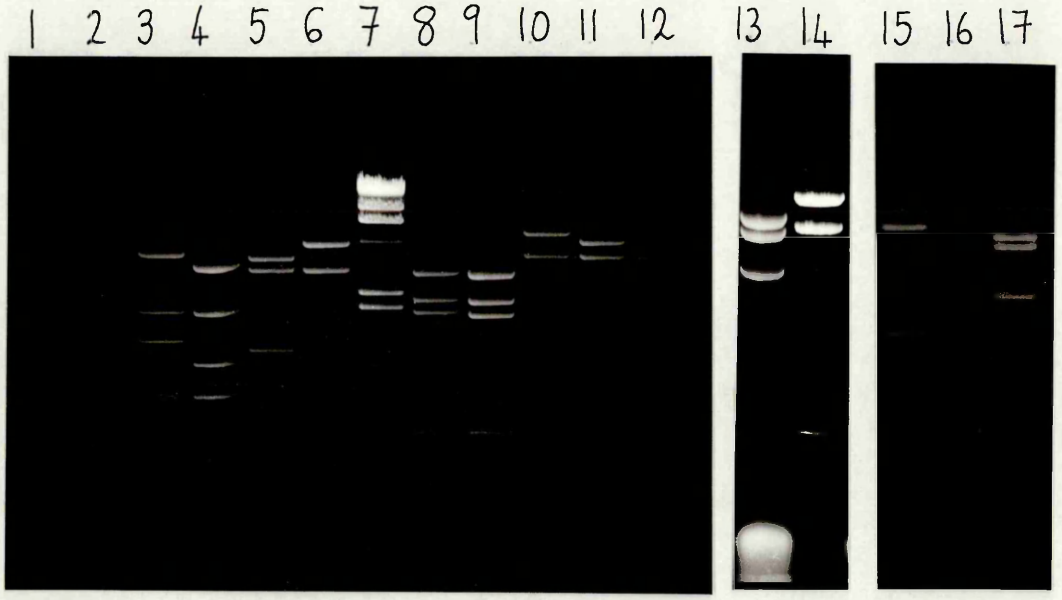



Figure 6.13: Restriction enzyme digests of pKD508 and pKD502

| Track | DNA | Enzyme(s) |
|---|---|---|
| 1 | pKD508 | BstEII + HincII |
| 2 | pKD502 | BstEII + HincII |
| 3 | pKD508 | ClaI + PvuII |
| 4 | pKD502 | ClaI + PvuII |
| 5 | pKD508 | BglII + ClaI |
| 6 | pKD502 | BglII + ClaI |
| 7 | λ | HindIII |
| 8 | pKD508 | PstI + ClaI |
| 9 | pKD502 | PstI + ClaI |
| 10 | pKD508 | PstI |
| 11 | pKD502 | PstI |
| 12 | pAT153 | PstI |
| 13 | pKD508 | ClaI |
| 14 | pKD502 | ClaI |
| 15 | 1.9 kb ClaI-PvuII | fragment of pKD502, digested with HpaII |
| 16 | 1.9 kb ClaI-PvuII | fragment of pKD508, digested with TaqI |
| 17 | 1.9 kb ClaI-PvuII | fragment of pKD502, digested with TaqI |

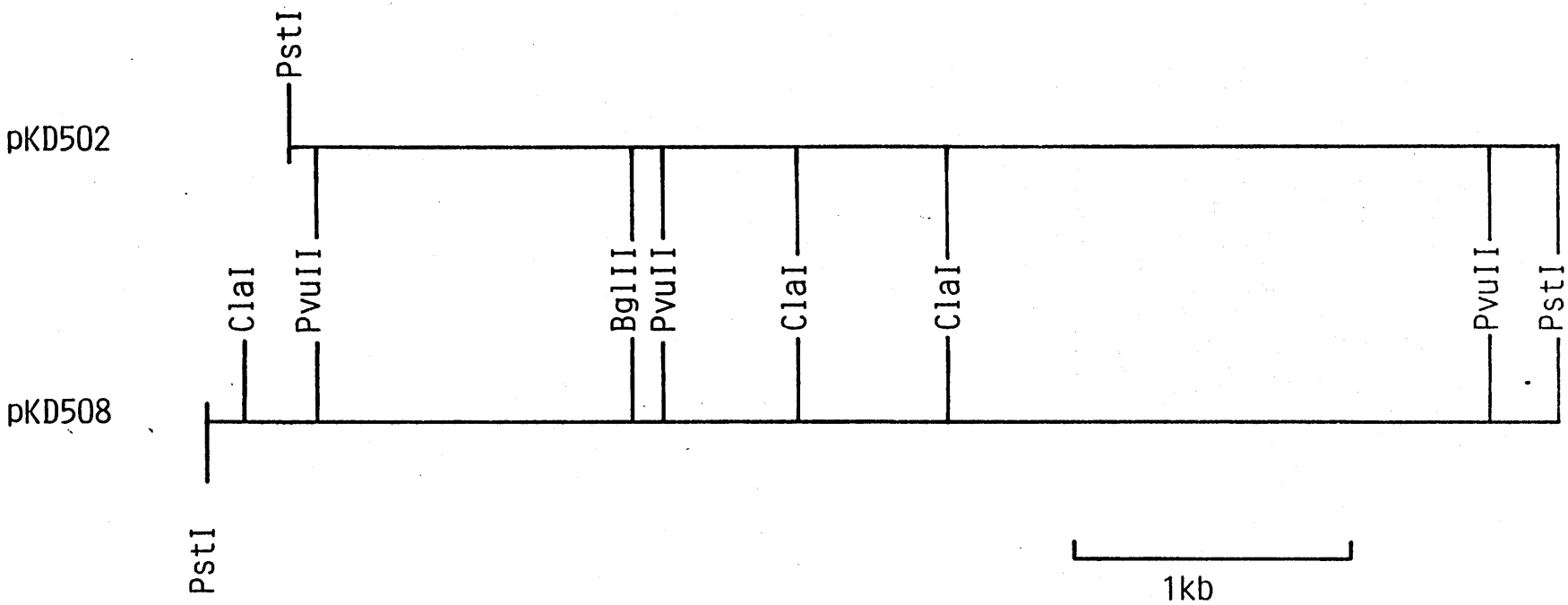



Figure 6.14: Comparative restriction maps of plasmids pKD502 and pKD508.

bands is the same; the middle band of this triplet is the 0.6 kb BstEII-HincII fragment which spans the serC-aroA boundary.

These digests might not reveal minor deletions or insertions which could alter the DNA between the genes. A ClaI/PvuII digest of pKD508 and of pKD502 was separated by electrophoresis on 1% LMT agarose. In each case, the 1.9 kb ClaI-PvuII band was excised and digested (without further purification) with TaqI. Half of the fragment DNA from pKD502 was digested with HpaII. Digestion products were separated by electrophoresis on 2% agarose and the ethidium bromide stained gel is shown in Figure 6.13. It is difficult to see the TaqI fragments in the digest of pKD508 ClaI-PvuII fragment, but the negative of this photograph clearly shows that the patterns are identical, and in particular, no difference in the mobility of the 124 bp TaqI fragment spanning the serC-aroA interface can be observed.

These restriction enzyme digests suggest very strongly that the structure of the intergenic region in the DNA cloned from λpserC is identical to that cloned from genomic DNA, and hence that the expression of serC and aroA are linked in vivo. Proof that the two intergenic sequences are identical could be obtained by cloning and sequencing the 124 bp TaqI fragment from pKD508. To date, this has not been carried out.

## 6.4 The serC-aroA operon

### 6.4.1 Model of the operon

The most obvious conclusion from the experiments described in Chapter 5 and in this Chapter is that the *serC* and *aroA* genes form an operon. This conclusion suggests a model for the expression of the *aroA* gene in *E.coli* which is shown in Figure 6.15.

In this model, *serC* and *aroA* are both transcribed from the same promoter (or promoters) which is located upstream of *serC*. Two messages are synthesised. mRNA1 is a polycistronic message which extends from the promoter to a point on the 3' side of *aroA*. mRNA2 is monocistronic, encoding only *serC* and extending from the promoter to the transcription terminator between *serC* and *aroA*. mRNA2 is likely to be the more abundant message, as the intracellular concentration of the *serC* product is higher than that of the *aroA* product, both when the cloned operon and the genomic operon are expressed. The model also predicts that transcription terminates *in vivo* frequently at the intercistronic terminator and that *aroA* expression is dependent on read-through at this terminator.

There are several ways in which this hypothesis can be tested experimentally. For example:

(a) Northern blot analysis: this involves blotting RNA prepared from *E.coli* AB2829/pKD501 to nitrocellulose and probing with various parts of the *serC-aroA* operon. Probes from the *serC* region should identify two mRNA species, while those from the *aroA* region should identify a single message.

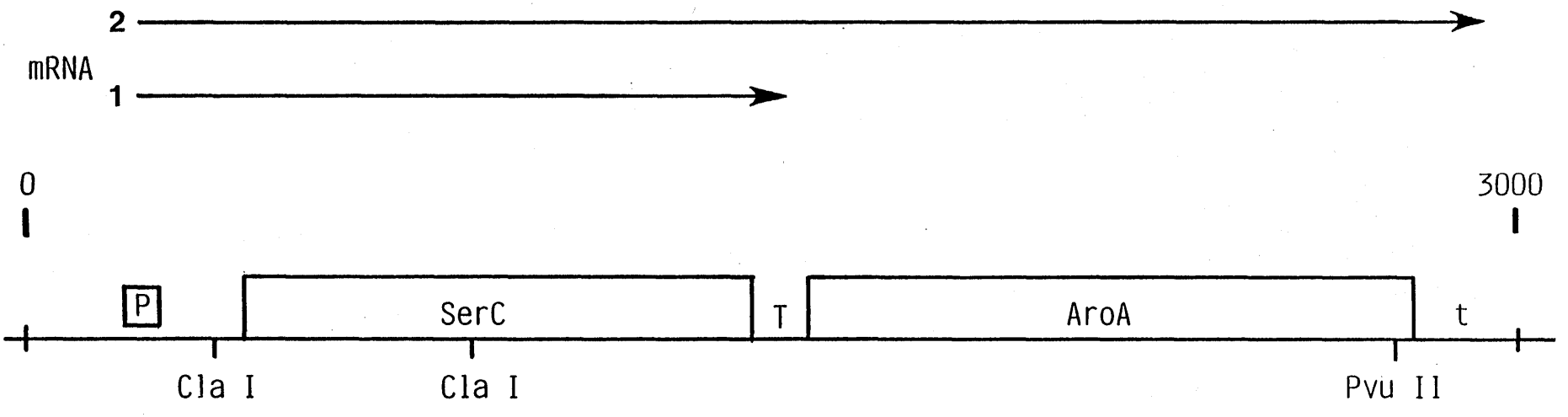



Figure 6.15: Proposed model of the serC-aroA operon. mRNA1 is monocistronic and encodes only the serC gene. mRNA2 is polycistronic, encoding both serC and aroA. P indicates the common promoter of the genes; T indicates the strong transcription terminator between the genes and t indicates the weaker stem-loop after aroA.

(b) S1 nuclease mapping: This can be used to identify transcription start and stop sites.

(c) Hybrid select/translate: by binding a DNA fragment from the aroA gene to nitrocellulose, it should be possible to select the aroA message. Translation of the message in a cell-free system would then be expected to give two products, one corresponding to EPSP synthase, and one to PSAT.

At the time of writing, the first two experiments have been carried out and the preliminary data obtained is presented in the sections which follow.

### 6.4.2 Preparation of RNA from E.coli AB2829/pKD501

RNA was prepared from E.coli AB2829/pKD501 by the hot phenol procedure (Section 2.39). It was then subjected to electrophoresis on 1% agarose. A smear was observed with bands corresponding to 16S and 23S rRNA.

### 6.4.3 Northern blot of RNA

The RNA was subjected to electrophoresis on a 1% agarose gel containing formaldehyde (2.2M final concentration) as described in Maniatis et al. (1982). Approximately 40 µg of RNA was loaded in each of four tracks and electrophoresis carried out till the dye front had migrated about 2/3 way down the gel. The running buffer was recirculated during electrophoresis.

The formaldehyde - denatured RNA was transferred to nitrocellulose as follows: the gel was soaked in 20 x SSC

for 10 min. A blot was set up as described in Section 2.41 and left overnight. The nitrocellulose filter was then washed briefly in 3 x SSC, dried in air (1 h) and baked at 80° under vacuum for 2 h.

### 6.4.4 Preparation of probes for Northern blot

Radioactively labelled probes for the RNA blots were prepared by synthesising the second strand of M13 single-stranded template DNA corresponding to parts of the *serC-aroA* operon. ($\alpha^{32}$P)dCTP was incorporated into the newly synthesised strand; templates were chosen so that the labelled strand would be complementary to the mRNA.

Four probes were chosen; they are illustrated in Figure 6.16 as P1 to P4. The arrows indicate the labelled strand. The rationale behind the use of these particular portions of the sequence is as follows:

P1 - ClaI fragment: if two messages are synthesised, then this probe should locate both.

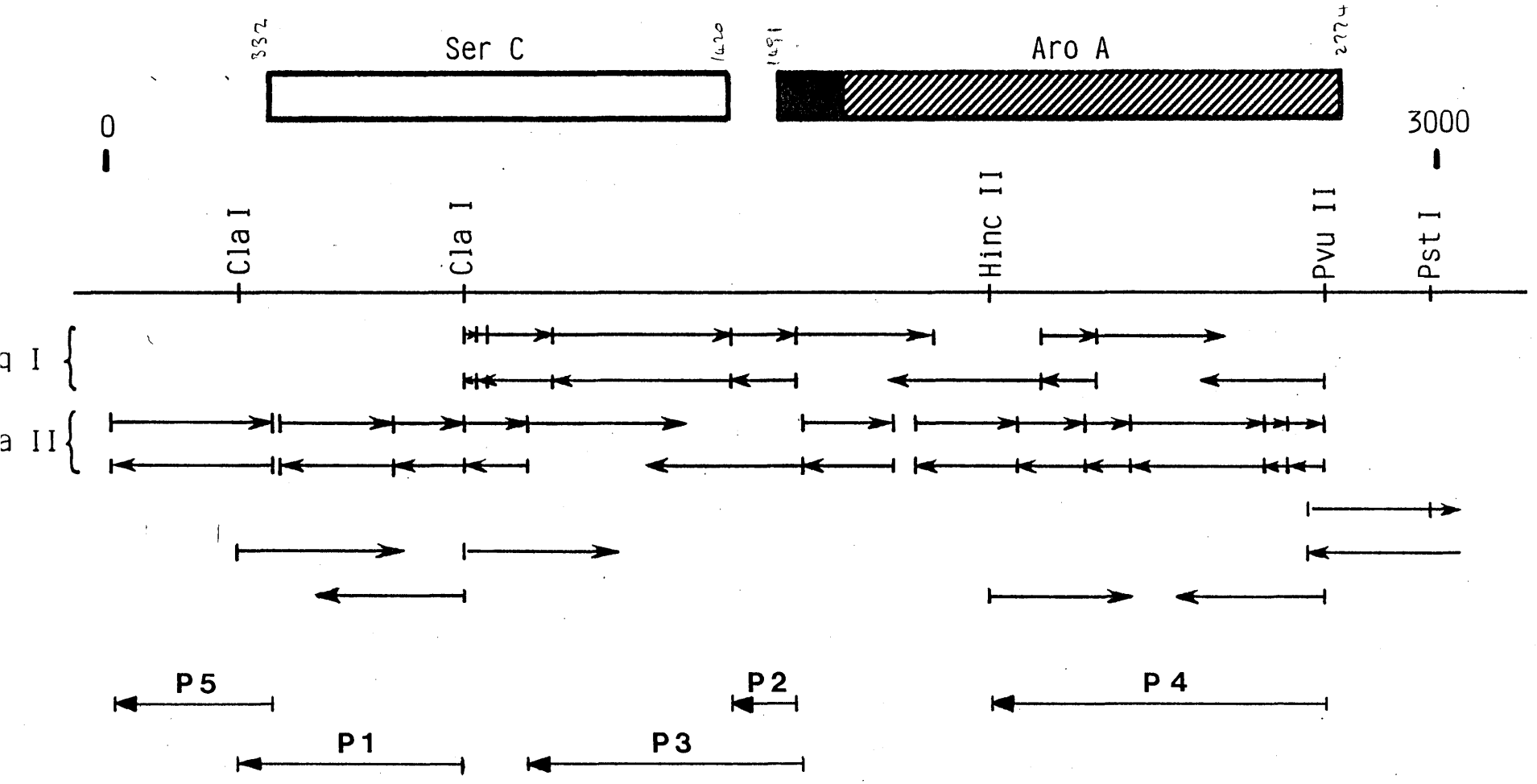
P2 - TaqI fragment: this fragment overlaps both genes and includes the terminator sequence. It should again hybridise to both mRNAs.

P3 - HpaII fragment: same reasoning as for P2, but a longer probe extending beyond the 5' and 3' ends of the TaqI fragment.

P4 - HincII-PvuII fragment: this probe should detect only *aroA* encoding sequences, i.e. only the longer of the two predicted mRNAs.

Figure 6.16: Summary of the sequencing strategy for the entire serC-aroA operon, and location of the probes used in Northern blot analysis, and S1 mapping experiments.

| | | | |
|---|---|---|---|
| Probes: | P1 | 0.5 kb | ClaI fragment |
| | P2 | 124 bp | TaqI fragment |
| | P3 | 614 bp | HpaII fragment |
| | P4 | 0.8 kb | HincII-PvuII fragment |
| | P5 | 364 bp | HpaII fragment |


332
Ser C
1420
1491
Aro A
2274
0
3000
Cla I
Cla I
Hinc II
Pvu II
Pst I
q I
a II
P5
P2
P4
P1
P3

Single-stranded templates were annealled to primer as described in Section 2.28. To this was added 4 µl $C^{o}$ (*CTP recipe, Section 2.30), 1.25 µl ($\alpha^{32}$P)dCTP (Amersham PB10165; >400 Ci/mmol) and 1 U Klenow fragment of DNA polymerase. After incubation at room temperature for 30 min, 2 µl of 10 x restriction enzyme buffer, 2 µl BSA (1 mg/ml) and 5 U each of HindIII and EcoRI were added. Incubation was continued a further 1 h at 37° and digestion products were separated by electrophoresis on 2% LMT agarose. Bands corresponding to the labelled inserts were excised and the DNA purified by phenol/chloroform extraction and ethanol precipitation. The probes were now resuspended in 100 µl 6 x SSC and the radioactivity counted:

| | |
|---|---|
| P1 | $1.5 \times 10^5$ cpm/10 µl |
| P2 | $0.8 \times 10^5$ cpm/10 µl |
| P3 | $1.1 \times 10^5$ cpm/10 µl |
| P4 | $1.4 \times 10^5$ cpm/10 µl |

6.4.5 Prehybridisation of nitrocellulose filters

The nitrocellulose was cut into four identical strips, each having a blot of the E.coli AB2829/pKD501 RNA. The strips were sealed in polythene along with prehybridisation solution (0.2 ml/$cm^2$), taking great care to exclude air bubbles (Section 2.42). Prehybridisation was at 68° for 6 h in a shaking water bath.

### 6.4.6 Hybridisation of nitrocellulose filters

The corner of the bag was cut and the prehybridisation solution was removed. This was replaced by hybridisation solution (50 $\mu l/cm^2$) and the bag resealed. Hybridisation solution was the same as prehybridisation solution, except that the remaining 90 $\mu l$ of probe was added before it was placed in the bag.

Hybridisation was overnight at $68^{\circ}$ in a shaking water bath.

Filters were then removed and the hybridisation solution released to radioactive waste. Before drying, the filters were treated as described in Maniatis *et al.* (1982). Autoradiography was at $-70^{\circ}$; an intensifying screen was used.

### 6.4.7 Autoradiography of Northern blot

The autoradiograph resulting from the blot described above is shown in Figure 6.17. The four probes used show a variation in the pattern of mRNA molecules identified. Probes P1, P2 and P3 identify two mRNA molecules whereas P4 identifies one (the longer one). This is the result which was predicted for this experiment. It suggests that two mRNAs are synthesised from a promoter upstream of *serC*.

### 6.4.8 Nuclease S1 mapping of the serC-aroA transcripts

The enzyme Nuclease S1 (from *Aspergillus oryzae*) will degrade RNA or denatured DNA into 5'-mononucleotides, but

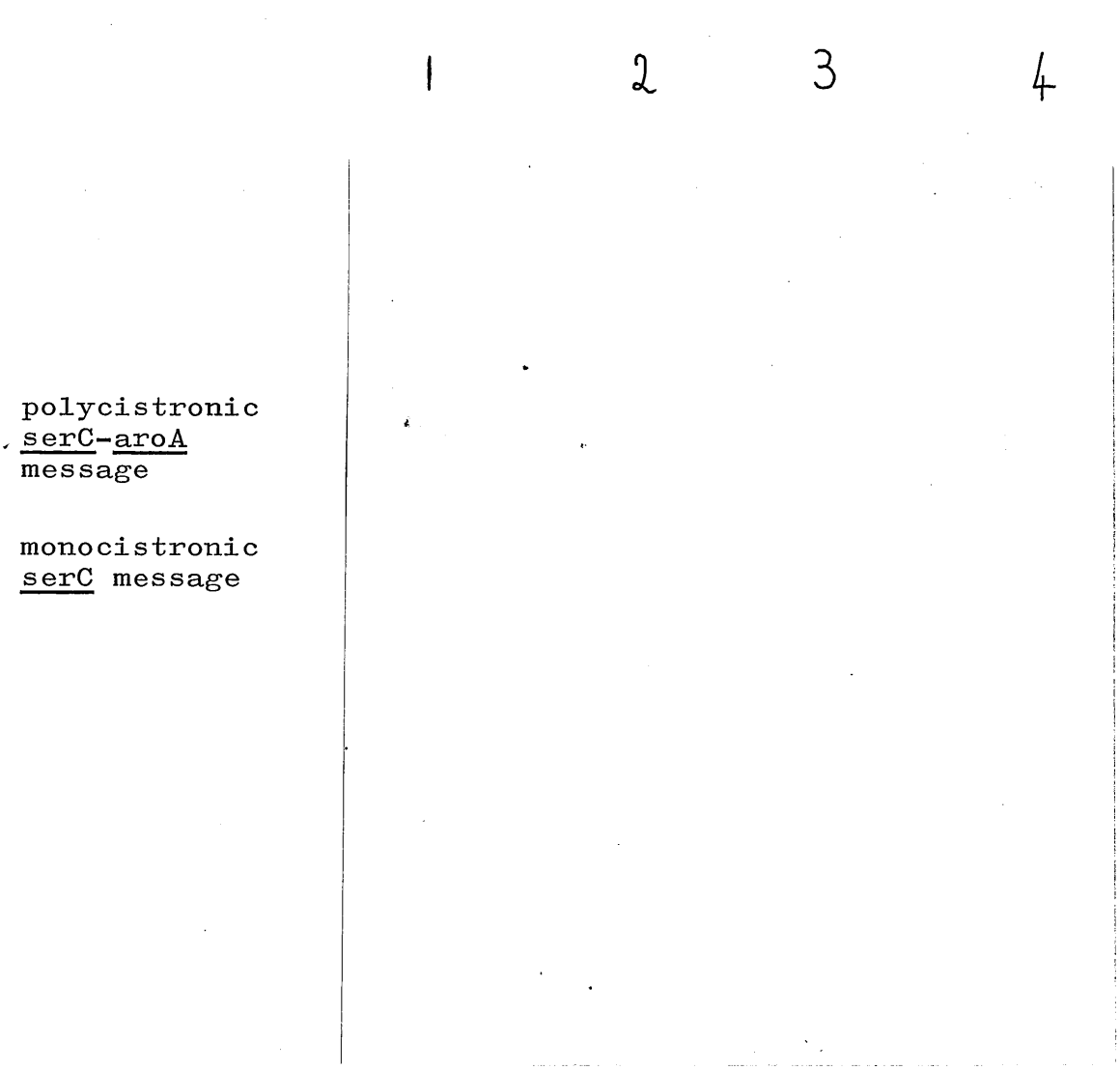



Figure 6.17: Northern blot analysis of the serC-aroA operon

The probes used in each lane were:

Track 1 P4 (aroA coding sequences only)
2 P1 (serC coding sequences only)
3 P2 (aroA and serC coding sequences and intergenic region)
4 P3 (aroA and serC coding sequences and intergenic region)

will not degrade duplex DNA or RNA/DNA hybrids under appropriate conditions. A fragment of DNA encoding a promoter can be hybridised to mRNA and the RNA/DNA hybrid digested by nuclease S1. The digestion product, the protected DNA sequence, can then be identified by gel electrophoresis, and from its length the location of the transcription startpoint can be deduced.

The probes used in this work were labelled throughout one strand of the DNA (the strand complementary to the mRNA) in the same way as the probes for the Northern blot (Section 6.4.4). This has the disadvantage that the protected lengths which result could originate from either end of the fragment. However, it is technically easier to follow this approach and the results presented here are preliminary.

Three probes were used (Figure 6.16). The logic behind their use is as follows:

P2 - TaqI fragment: this should give full length protection and/or protection of a shorter length corresponding to mRNA2 to the terminator site.

P3 - HpaII fragment: confirmation of the results from P2; also initiation from a promoter upstream of <u>aroA</u> and upstream of P2 would be seen.

P5 - HpaII fragment: this should map the transcription startpoint.

The probes were prepared exactly as described for the Northern blot probes, except that they were finally resuspended in 50 µl hybridisation buffer (Section 2.42). The radioactivity in each sample was:

P2 $0.6 \times 10^5$ cpm/5 µl

P3 $1.4 \times 10^5$ cpm/5 µl

P5 $0.9 \times 10^5$ cpm/5 µl

To each probe was added 50 µg RNA in 50 µl hybridisation buffer. The sample was incubated for 3 h at 53°. 300 µl ice cold 1 x S1 buffer (Section 2.40), containing 2 000 U S1 nuclease was added and incubation continued at 43° for 45 min. Digests were phenol/chloroform extracted and ethanol precipitated then resuspended in 10 µl TE buffer. 4 µl of formamide dye mix was added and the sample boiled for 3 min before loading 3 µl onto a DNA sequencing polyacrylamide gel, alongside a sequencing 'ladder' as length markers. The gels obtained are shown in Figure 6.18.

## 6.4.9 Results of S1 mapping

The S1 mapping results presented here are very preliminary. However, they do confirm the predictions made for the mRNA startpoint in the model and also can be used to define accurately the transcription startpoint for the *serC* gene. Results are shown in Figure 6.18.

Fragment P3 was protected over its full length from S1 degradation. This could be due either to protection by mRNA spanning the intergene region or to protection by the complementary DNA strand. A second protected region has a length of approximately 525 nucleotides. If this represents protection by the 3' end of an mRNA then the message terminates at position 1462 in the sequence, at the T residues which form part of the terminator structure

Figure 6.18: S1 nuclease mapping of the serC-aroA transcript

A - electrophoresis for 2 h on a buffer gradient 6% acrylamide gel.

1 A )
2 C ) DNA sequencing reactions as size markers
3 G )
4 T )

5 mapping of fragment P2

6 mapping of fragment P3

7 mapping of fragment P5

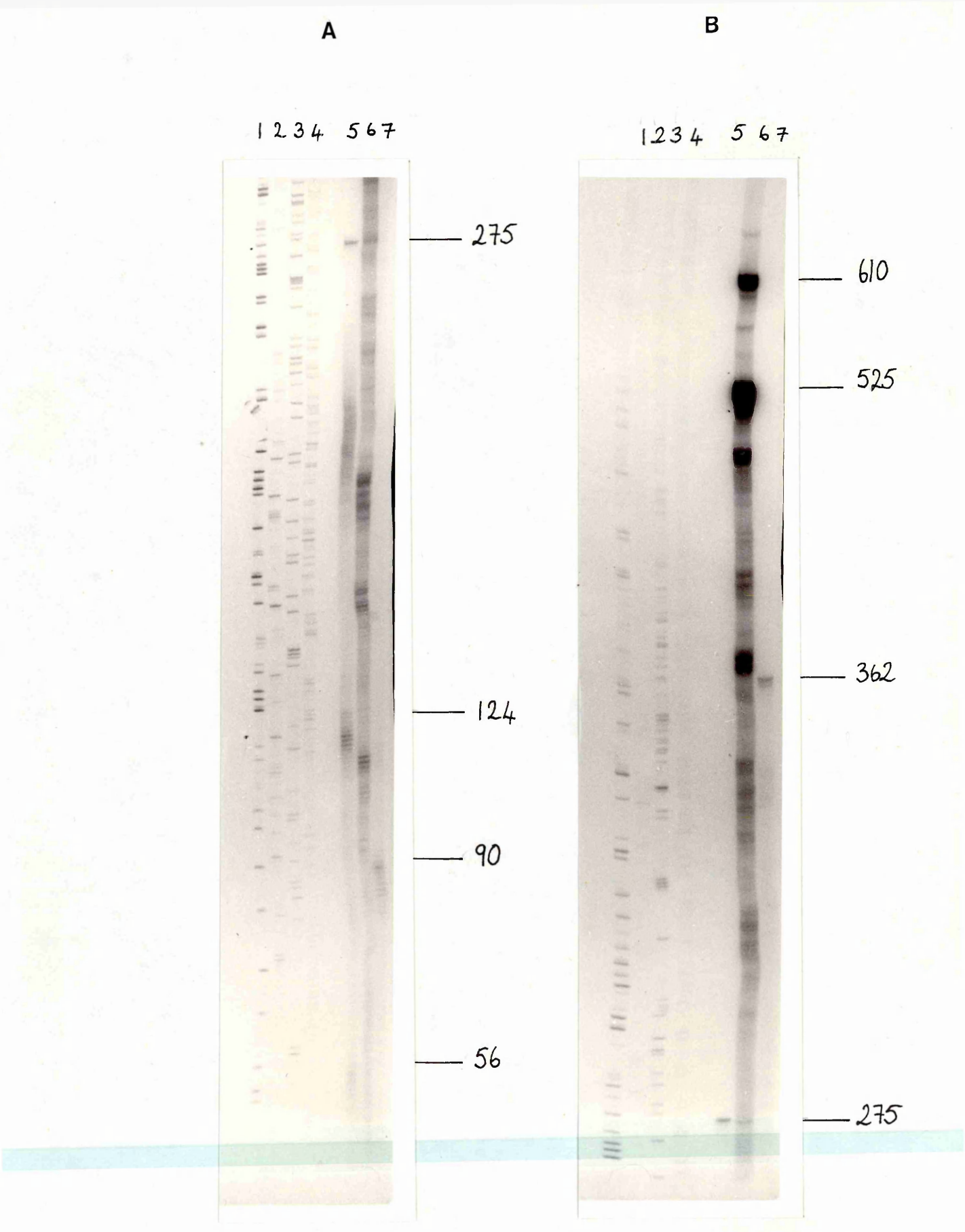
B - electrophoresis for 5 h on a non-gradient 6% acrylamide gel.

8 A )
9 C )
10 G ) DNA sequencing reactions as size markers
11 T )

12 mapping of fragment P2 124

13 mapping of fragment P3 1214

14 mapping of fragment P5 364


A
1 2 3 4 5 6 7
275
124
90
56
B
1 2 3 4 5 6 7
610
525
362
275

following serC. The alternative explanation that this protected region is the 5' end of the aroA message is unlikely as the transcription startpoint would be at position 1023, and no promoter-like sequences are present in this region.

Fragment P2 shows a more complex pattern. The M13 clone from which fragment P2 was obtained carries two TaqI fragments, the second of which is not complementary to the predicted mRNA and hence will not be protected by mRNA. The pattern obtained for S1 mapping of P2 has three protected lengths:-

275 bp corresponds to protection of the full tandem clone by the opposite strand of the DNA.

124 bp corresponds to protection of the TaqI fragment containing the intergenic region, by an mRNA which spans this region and encodes both serC and aroA.

56 bp corresponds to protection by a transcript which either stops or starts within the 124 bp intergene TaqI fragment. If protection is by the 3' end of mRNA then the transcript stops at position 1458. This correlates well with the above finding that P3 is protected from digestion by a transcript which ends at position 1462.

The pattern obtained for P2 therefore provides conclusive evidence that there are two mRNA transcripts, one of which encodes the serC and aroA genes and the other terminates at the rho-independent terminator sequence between the genes.

Fragment P5 is protected over its full length by its complementary strand. There is, however, a ladder of protected lengths, the longest of which is 90 nucleotides. The ladder effect is probably due to degradation at the 5' end of the mRNA prior to hybridisation. Working back from the HpaII site, the transcription startpoint was determined to be at position 274.

### 6.4.10 Transcription startpoint and promoter of the serC-aroA operon

The transcription startpoint determined by S1 mapping of P5 is shown in Figure 6.19. Upstream of this is a sequence which corresponds to the consensus *E.coli* promoter -10 and -35 regions.

### 6.4.11 Other features in the promoter region

A search for inverted repeats in the sequence of the promoter region revealed a 26 bp sequence which can form a stem-loop structure similar to that shown in Figure 6.3. The -10 region of the promoter is then located at the top of the stem-loop. This is a common feature of the operator sites of genes which are regulated by a repressor. Examples of this in aromatic biosynthesis include the *trp* operon (Platt, 1978), the *aroG* gene (Davis & Davidson, 1982) and the *aroH* gene (Zurawski et al., 1981). The -10 region of the *trp* operon promoter is located at the top of a stem loop to which the *trp* repressor is known to bind, preventing RNA polymerase from binding to the promoter. These

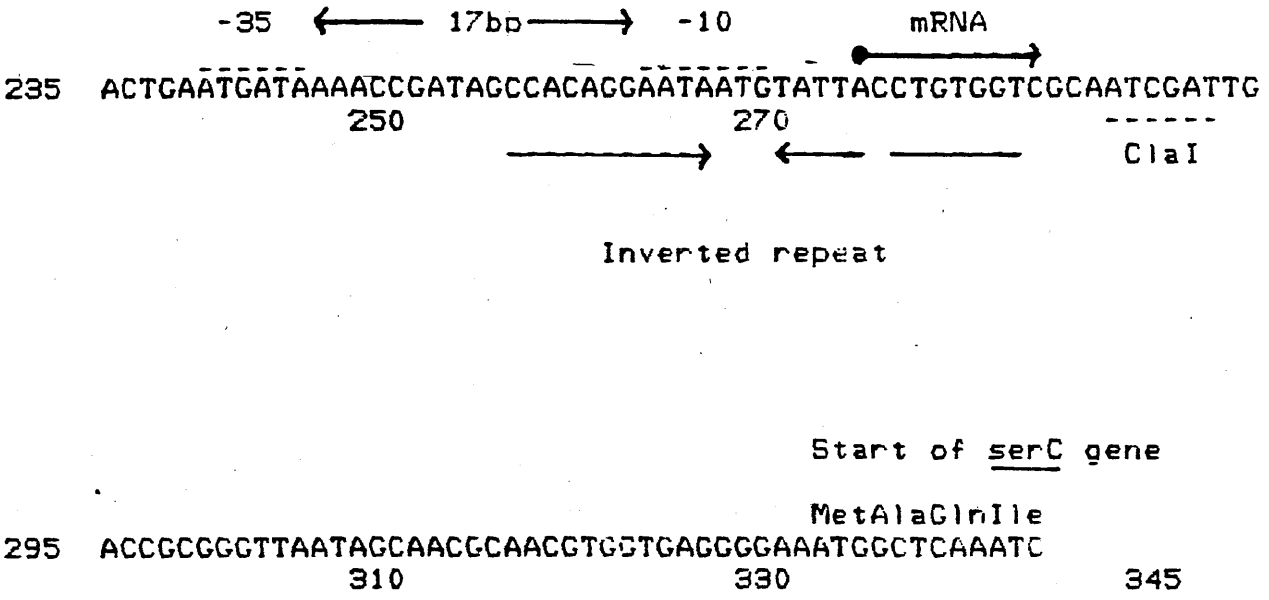



Figure 6.19: The promoter of the serC-aroA operon.

observations suggest that expression of the serC-aroA operon is regulated by a repressor. There is no evidence to support this hypothesis as no regulatory mutants of either aroA or serC have been described. Further discussion of the expression is based on speculation and cannot at present be substantiated by experimental evidence.

## 6.5 Examples of linked gene expression systems in E.coli

Investigations of the expression of several sets of linked E.coli genes have shown that the serC-aroA operon structure is not uncommon. The U.V. - inducible lexA and lexD genes are co-transcribed, but are separated by a transcription terminator (Miki et al., 1984). The majority of constitutive transcription from the lexA promoter is terminated at a terminator immediately following the gene. However, after induction of lexA, a fraction of the transcription passes through the terminator, allowing the lexD protein to be synthesised.

A second example is the E.coli pyrE operon. pyrE is the structural gene for orotate phosphoribosyl transferase, an enzyme which participates in the biosynthesis of UMP. Expression of pyrE is regulated by a UTP-modulated relief of an attenuator situated in front of the gene (Poulsen et al., 1984). The gene in front of pyrE, and which forms the other member of this operon, remains to be identified.

In both examples, the operon structure is similar: promoter-gene-transcription terminator-gene.

The terminator can be regarded as an attenuator of expression of the second gene. However, both examples differ from the serC-aroA case in that there is clearly a functional reason for the co-ordination of expression and attenuation in these systems. In the first case, lexA is constitutively expressed at a low level and is auto-regulated. On U.V.-induction, the lexA product is cleaved by recA and transcription is no longer repressed leading to the synthesis of large quantities of lexA product, and by the attenuation mechanism, to smaller quantities of lexD product, both of which are required for the 'SOS response'. In the second case, the function of the gene preceding pyrE is not known, but the important point to note is that the attenuation mechanism is relieved only in response to a low level of UTP within the cell, and hence pyrE expression takes place.

In the biosynthetic operons which have been studied in detail in E.coli, e.g. the trp and his operons, the enzymes encoded in the genes of the operon are related in that they catalyse reactions on the same pathway. Induction of expression of the operon is necessary in order that the end product of the pathway may be synthesised, and each one of the gene products must be produced, otherwise the pathway will not function. For this reason, it is widely believed that co-ordination of gene expression in response to the needs of the cell is the principal function of the operon.

The same cannot be said of the serC-aroA operon. EPSP synthase and PSAT are enzymes which catalyse reactions

on completely separate biosynthetic pathways. There is evidence in the literature that both aroA and serC are constitutively expressed (Tribe et al., 1976; McKitrick & Pizer, 1980). This begs the question: is there any way in which co-ordinate control of the expression of these genes might be of value to the cell, possibly in a way which would be masked in the experiments which showed constitutive expression?

There are two biosynthetic points of contact between the end products of the serine and shikimate biosynthetic pathways. The first is in the synthesis of tryptophan. At the final step, indolglycerol phosphate and serine are converted to tryptophan and glyceraldehyde 3-phosphate by tryptophan synthase. Tryptophan is the second least abundant of the common amino acids, constituting 1% or less of the average protein mass. Serine, however, plays a much more central physiological role. Apart from protein synthesis, it is also used in cysteine and phospholipid biosynthesis, and conversion of serine to glycine provides most of the 'one-carbon' pool of the cell, which is required for the synthesis of many important intermediates. The biosynthesis of tryptophan will therefore not be expected to perturb significantly the serine pool of the cell, and so co-ordinate control of the synthesis of tryptophan precursors is extremely unlikely.

The second point of contact is in the biosynthesis of enterochelin (enterobactin), which is responsible for sequestering iron from the environment. Iron is an essential

requirement for bacterial growth, but under normal environmental conditions it exists as an insoluble ferric hydroxide polymer or in the human and animal most is bound to transferrin in the serum or lactoferrin in secretory fluids. In response to such iron-limiting conditions, most aerobic and facultative anaerobic organisms have evolved a high affinity pathway for assimilation of Fe(III) which consists of low-molecular-weight carriers (siderophores) and the cognate membrane receptors for binding and uptake. Enterochelin is the siderophore manufactured by E.coli and forms the 'high affinity' iron uptake system. A second 'low affinity' system is available to E.coli. If citrate is present in the growth medium, induction of a system for citrate-mediated iron transport occurs (Neilands, 1982).

Enterochelin synthesis begins with chorismate which is converted in three steps to 2,3-dihydroxybenzoic acid. Enterochelin synthetase, which is composed of four enzymes, then converts 2,3-dihydroxybenzoic acid, and serine to enterochelin, a cyclic trimer of 2,3-dihydroxybenzoylserine. The enterochelin iron uptake system consists of at least seven biosynthetic genes (entA - entG), a transport gene (fep) and a gene (fes) involved in the breakdown of the ferric enterochelin complex, all of which map in a cluster around minute 13. The expression of the nine genes is regulated by the intracellular iron concentration; the genes have been cloned and the genetic organisation of the cluster investigated (Laird et al., 1980; Laird & Young, 1980). Recently, it has been shown that five of

the genes form an operon (*entA*(CGB)E) and that all the genes are transcribed from iron-regulated promoters (Fleming *et al.*, 1983).

Buck & Griffiths (1981) have shown that in media where ferric iron is not freely available, *E.coli* contains a population of specifically undermodified tRNAs. The undermodified $tRNA^{phe}$, $tRNA^{tyr}$ and $tRNA^{trp}$ can function as positive regulatory elements of the aromatic amino acid transport system. An increased ability to take up aromatic amino acids from the medium might give the cell a growth advantage. The synthesis of enterochelin is likely to impose an imbalance upon the aromatic amino acid biosynthetic pathway by depleting the chorismate pool; the enhanced ability to acquire rather than synthesise these amino acids may be an important feature of cellular economy where enterochelin synthesis is necessary for growth. It has also been shown that the undermodified tRNAs have lowered translation efficiencies and can relieve transcription termination at the *trp* and *phe* operons (Buck & Griffiths, 1982). In *Salmonella*, modification is related to growth in aerobic or anaerobic conditions. Iron deficiency is usually only a problem during aerobiosis, and it is under these conditions that tRNA is undermodified (Buck & Ames, 1984).

McCray & Herrmann (1976) investigated the derepression of certain aromatic amino acid biosynthetic enzymes in *E.coli* grown in medium deficient in iron. The results obtained suggested that the increased synthesis of enterochelin

during iron starvation could be accounted for by de-repression of DAHP synthase (tyr) and the branch pathway to dihydroxybenzoate. Derepression of DAHP synthase would increase the biosynthesis of chorismate in the cell, thereby maintaining the chorismate pool for aromatic amino acid biosynthesis. The increased expression of aroF is thought to be due to relief of attenuation at the tyr operon (O'Connell et al., 1983, cited in Herrmann, 1983).

The specific activity of EPSP synthase was not determined under the conditions used in McCray & Herrmann (1976). The only clue in the literature to the involvement of iron in the expression of aroA comes from the work which showed that aroA is constitutively expressed (Tribe et al., 1976). Close examination of Table 2 of this paper reveals that the specific activity of EPSP synthase was consistently lower in cells grown in iron citrate medium than in cells grown in either minimal medium or minimal medium supplemented with aromatics. The specific activity of the other enzymes in the pathway remained constant.

The biosynthesis of enterochelin requires serine and chorismate in equimolar amounts. Studies by McKitrick & Pizer (1980) have indicated that there is a constant, but rapidly turning over pool of serine within the cell. Under the experimental conditions used, it was shown that the transaminase (PSAT) was in large excess. It is likely, however, that in the medium used to grow the cells (M9 salts + glucose), the bacteria would be actively scavenging

iron and so would require a large amount of serine for enterochelin biosynthesis. Any reduction in PSAT activity due to repression of gene expression would therefore be masked.

The observations described in this section provide evidence that iron limitation has a large effect on gene expression in E.coli. For the systems which have been studied in detail, the modulating effect is due to under-modification of tRNA, which alters the attenuation of transcription at a number of biosynthetic operons. The repressor which controls expression of the ent, fes and fep genes is not known. It is possible that these genes are regulated in the same way as the aroF gene; examination of the DNA sequences upstream of aroF and the ent genes may answer this.

There is no evidence that attenuation plays a role in the expression of the serC-aroA operon. The sequences in the 5' leader region cannot be folded to form structures analogous to these found in the trp operon (Yanofsky, 1981); nor is there a 'leader peptide' located in this region. The inverted repeat at the -10 region of the promoter may be involved in a different type of regulatory mechanism. This may be the operator site for an as yet unidentified repressor. It is possible that the genes are expressed at a low constitutive level under conditions where there is sufficient iron and that this amount of each enzyme can satisfy the requirements for biosynthesis of both serine and chorismate. Only under iron stress are the pools of

serine and chorismate depleted rapidly and so increasing the synthesis of the two enzymes might help to replenish these intermediates.

Many more straightforward interpretations of the linked gene situation are possible. A deletion of the promoter of *aroA* may have made *aroA* expression dependent on expression of *serC*. Alternatively, a transposition event may have moved one of the genes intact to within the other gene. Either *serC* and its terminator was moved to lie between *aroA* and its promoter, or *aroA* was moved to the site downstream of *serC*.

## 6.6 Suggestions for further work

Definitive proof that the *aroA* and *serC* genes are translated from the same mRNA may be obtained by *in vitro* translation of message selected by an *aroA* sequence probe.

The factors which affect expression of *serC* and *aroA* *in vivo* must be studied. In particular, the effects of the presence or absence of iron in the growth medium will show if iron is a modulator of the expression of this operon.

# CHAPTER 7 GENERAL DISCUSSION

## 7.1 Sequence homology searching

### 7.1.1 Introduction

One of the aims of the work presented in this thesis was to determine the DNA sequence of one or more of the *E.coli aro* genes, deduce the corresponding amino acid sequence and then to compare both gene and protein sequences with the sequences of the *N.crassa* (or any other) *arom* cluster gene and the corresponding multifunctional protein sequence. As the sequence of an *arom* gene is not yet available, this discussion will necessarily be restricted to making comparisons of two *E.coli* genes that have been sequenced with the sequences of related genes. The preliminary results of these two searches are presented here as examples of the searches which will be instigated in the near future.

It is possible to look for homologies between the primary structures of proteins and/or their gene sequences in a number of ways. The simplest is to scan and compare sequences by eye for highly conserved features or to line up sequences which are known to be closely related. However, the comparison and matching of sequences which are distantly related, or unrelated, is extremely difficult; it requires a computer to compare the sequences and to collate information. A number of programs have been written for this purpose and also for searching databases for homologous or related sequences. For example, the program

COMPARE in the WISGEN package (Section 2.36) is derived from the matrix analysis of Maizel & Lenk (1981). The review by Doolittle (1981) provides a valuable insight into the pitfalls of comparing sequences and deducing relatedness.

### 7.1.2 3-dehydroquinase

The complete sequences of two 3-dehydroquinase enzymes are available: the *E.coli aroD* gene product (Chapter 4) and the *N.crassa* inducible *qa*-2 gene product (Alton *et al.*, 1982; Dr M.E. Case, unpublished results). Although these enzymes catalyse the same reaction, they play a different role in the cell; the *E.coli* enzyme has a biosynthetic function, whereas the inducible *N.crassa* enzyme has a function in catabolism. Are the enzymes derived from a common ancestor or are they the result of convergent evolution of function? Are either of them related in sequence to the 3-dehydroquinase domain of the *arom* polypeptide? These questions can be addressed by comparing the sequence around the active site lysine of *arom* 3-dehydroquinase with the sequences around the 17 lysine residues in *E.coli* 3-dehydroquinase and the two lysine residues in the catabolic 3-dehydroquinase.

It is known that the mechanism of 3-dehydroquinase involves the formation of a Schiff base intermediate with a lysine residue at the active site of the enzyme. This can be trapped and labelled with $^{3}$H-sodium borohydride (Butler *et al.*, 1974). The active site lysine of *arom* was labelled in this way and following proteolysis a labelled

peptide was isolated and its sequence determined by a micro-protein sequencing technique (Dr S. Chaudhuri, unpublished results). The sequences flanking all of the lysine residues are aligned in Table 7.1. The sequence most homologous to the arom active site sequence is Lys-138 of the E.coli enzyme, where 5/15 residues are identical. In Neurospora the lysine is part of the tetrapeptide VKLV and in E.coli the sequence is VKVV, a single conserved change. One can predict that if both biosynthetic 3-dehydroquinases are truly homologous, then lys-138 represents the most likely candidate for the active site of the E.coli enzyme. Neither of the lysines in the catabolic 3-dehydroquinase are homologous to the arom sequence or to the sequence around lys-138. Experiments are in progress in our laboratory aimed at labelling and identifying the active site lysine from both the E.coli and the catabolic enzyme.

This result can be interpreted as implying that the catabolic and biosynthetic enzymes have arisen separately, or that they have diverged substantially from a common ancester. Evidence in support of this comes from consideration of the $M_r$ of the enzymes. The catabolic enzyme has an $M_r$ of 20,000 and exists as a dodecamer whereas the E.coli enzyme has an $M_r$ of 28,000 and is a dimer. Following limited proteolysis, it is possible to isolate a fragment of the arom complex which carries the 3-dehydroquinase and shikimate dehydrogenase activities (Smith & Coggins, 1983). This has an $M_r$ of 68 000 and is comparable in size to the bifunctional 3-dehydroquinase/shikimate dehydrogenase of P.mungo which has an $M_r$ of 59 000 (Koshiba, 1979).

| | | No. of matches |
|---|---|---|
| | * | |
| arom | ALQHGDVVKLVGVAR | |
| | | |
| AroD | | |
| Lys-2 | -------MKTVTVKD | 3 |
| -7 | --MKTVTVKDLVIGT | 2 |
| -17 | LVIGTGAPKIIVSLM | 1 |
| -25 | KIIVSLMAKDIASVK | 1 |
| -31 | MAKPIASVKSEALAY | 3 |
| -66 | VESVMAAAKILRETM | 1 |
| -75 | ILRETMPEKPLLFTF | 1 |
| -85 | LLFTFRSAKEGGEQA | 3 |
| -125 | LFTGDDQVKETVAYA | 3 |
| -138 | AYAHAHDVKVVMSNH | 5 |
| -147 | VMSNHDFHKTPEAEE | 2 |
| -160 | EEIIARLRKMQSFDA | 1 |
| -170 | QSFDADIPKAALMPQ | 2 |
| -207 | PIITMSMAKTGEISR | 2 |
| -229 | SGGNFWCGKKSVCAR | 3 |
| -230 | GGNFWCGKKSVCARA | 2 |
| -240 | VCARANLGK------ | 1 |
| | | |
| qa-2 | | |
| Lys-95 | EAGLGPGDKVSAIII | 2 |
| -147 | RHHSYLSDKAVAVIC | 3 |

Table 7.1: Sequence around the active site lysine of arom 3-dehydroquinase, and the lysine residues in the E.coli and catabolic N.crassa enzymes.

### 7.1.3 Phosphoserine aminotransferase

The primary sequence of a number of transaminases has been determined (Barra *et al.*, 1980; Doonan *et al.*, 1975; Huynh *et al.*, 1980; Kondo *et al.*, 1984; Shlyapnikov *et al.*, 1979; Dr M. Edwards, unpublished results). These sequences have been very highly conserved during evolution; comparison of bacterial and mammalian sequences show that they are approximately 30% conserved. This begs the question: Is the sequence of the *serC* gene product also so highly conserved?

The sequences around the lysine to which pyridoxal 5'-phosphate binds in the transaminases described above are listed in Table 7.2, along with the sequences flanking the lysine residues in the sequence of PSAT. This comparison shows that none of the lysines in PSAT is homologous to the sequence around the lysine in the other enzymes.

The program COMPARE in the WISGEN package (Section 2.36) was used to compare the protein sequence of PSAT with that of *E.coli* tyrosine aminotransferase (the *tyrB* gene product). Parameters were set so that each time an identical match of two residues was found in a window of three residues, a point was plotted. A plot of the data obtained (using DOTPLOT on a Hewlett-Packard HP 7221 plotter) is shown in Figure 7.1. There is no clear pattern to the spread of points. A similar comparison between the tyrosine amino-transferase sequence and the sequence of chicken heart cytoplasmic aspartate aminotransferase is shown in Figure 7.2. In this case the homology is striking; the points along the diagonal show that the sequences are highly

Aspartate aminotransferase

| | | |
|---|---|---|
| * | | |
| QSFSKNFGLYNERVGN | Chicken cyto. | Shlyapnikov et al (1979 |
| QSFSKNFGLYNERVGN | Pig cyto. | Doonan et al (1975) |
| QSYAKNMGLYGERVGA | Pig mito. | Barra et al (1980) |
| QSYAKNMGLYGERVGA | Rat mito | Hyunh et al (1980) |
| SSYSKNFGLYNERVGA | E.coli | Kondo et al (1984) |

Tyrosine aminotranferase

| | | |
|---|---|---|
| NSFSKIFSLYGERVGG | E.coli | Dr. M.Edwards (unpub.) |

Phosphoserine aminotransferase - lysine residues

| | |
|---|---|
| Lys-20 | AEVLKQAQQELRDWNG |
| -44 | SHRGKEFIQWAEEAEK |
| -55 | EEAEKDFRDLLNVPSN |
| -68 | PSNYKVLFEHGGERGN |
| -91 | ILGDKTTADYVDAGYN |
| -108 | ASAIKEAKKYCTPNVF |
| -111 | IKEAKKYCTPNVFDAK |
| -112 | KEAKKYCTPNVFDAKV |
| -122 | VFDAKVTVDGCRAVKP |
| -132 | LRAVKPMREWQLSDNA |
| -198 | AGAQKNIGPAGLTIVI |
| -219 | DLLGKANIACPSILDY |
| -254 | GLVFKWLKANGGVAEM |
| -257 | FKWLKANGGVAEMDKI |
| -267 | AEMDKINQQKAELLYG |
| -272 | INQQKAELLYGVIDNS |
| -292 | NDVAKRNRSRMNVPFQ |
| -311 | SALDKLFLEESFAAGL |
| -326 | LHALKGHRVVGGMRAS |
| -348 | LEGVKALTDFMVEFER |

Table 7.2 The lysine residue to which pyridoxal 5-phosphate binds in a number of aminotransferases, and the lysine residues in PSAT.

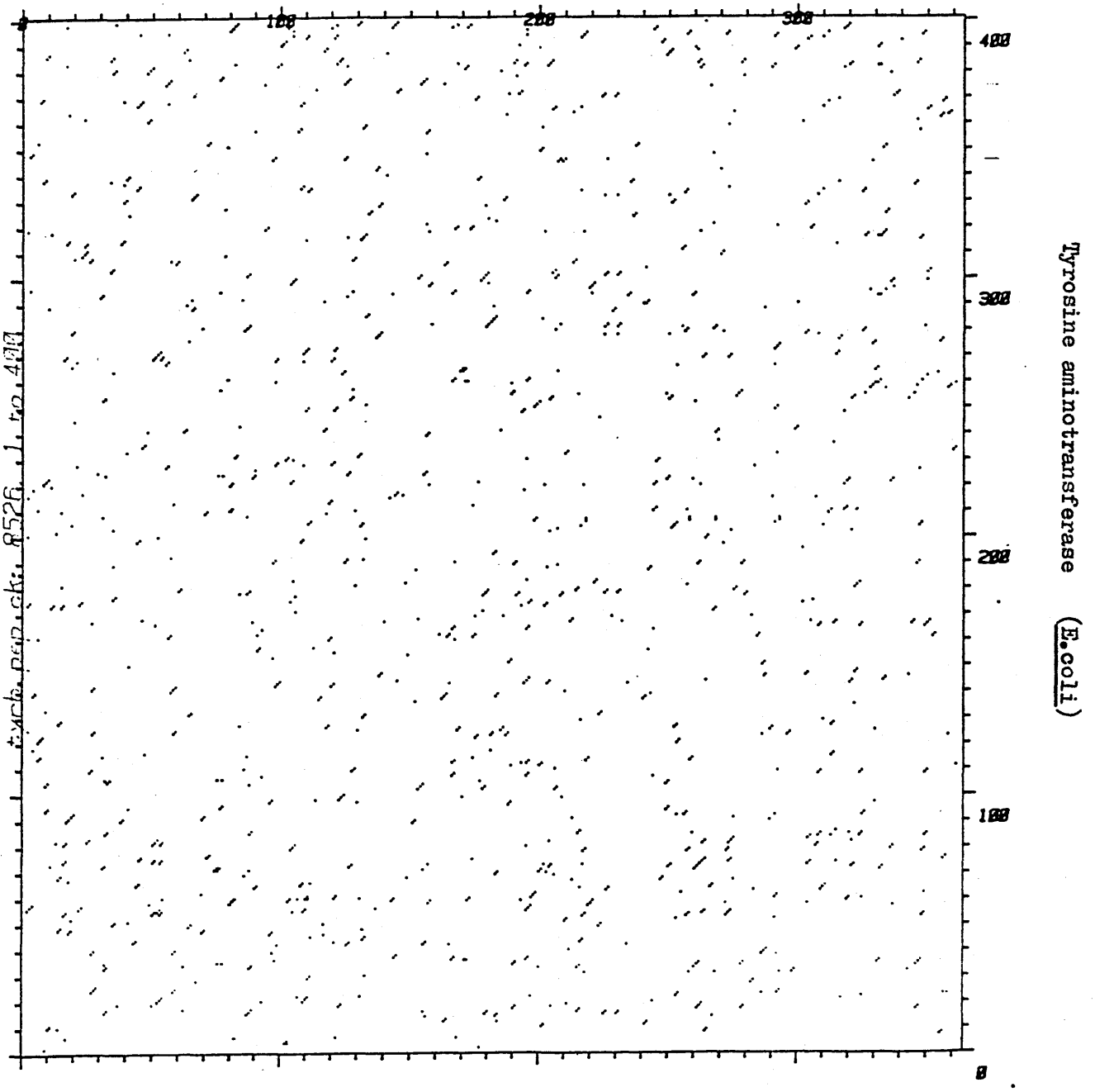



Figure 7.1: Dotplot of the PSAT amino acid sequence vs. the amino acid sequence of E.coli tyrosine aminotransferase. A dot was plotted at each position where there were two identical amino acids in a group of three.

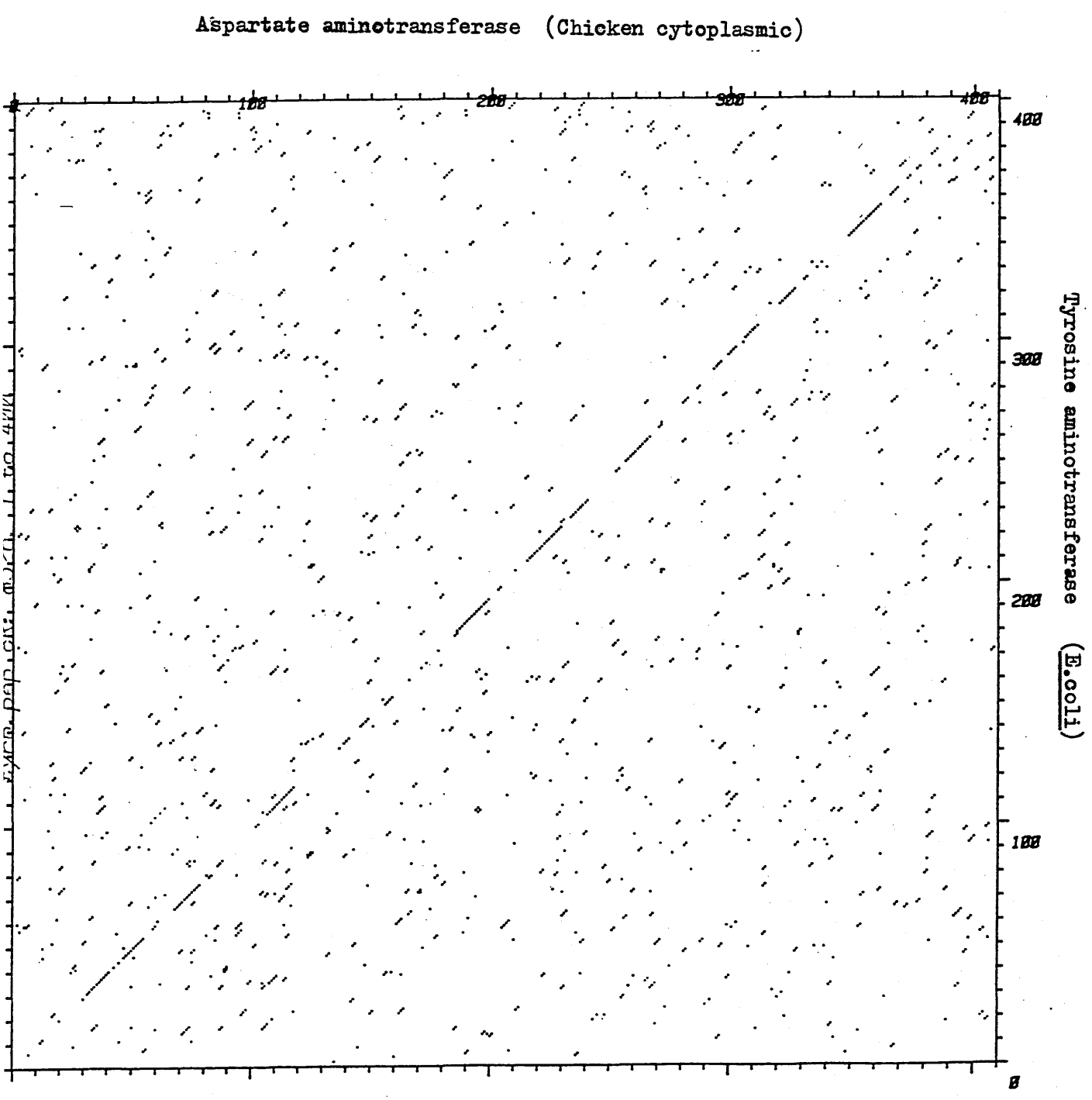



Figure 7.2: Dotplot of Chicken cytoplasmic aspartate aminotransferase vs E.coli tyrosine aminotransferase. A dot was plotted at each position where there were two identical amino acids in a group of three.

conserved. This is despite the fact that one sequence is from a prokaryotic and one from a eukaryotic organism.

These results show that PSAT does not belong to the same family of transaminases as the other highly conserved members described in Table 7.2. Since the other transaminases are so well conserved and PSAT is so completely different it is likely that the two classes of enzyme have evolved completely independently of one another - an exceptionally good example of convergent evolution of function. Only a knowledge of the sequences of the other aminotransferases of *E.coli* will tell if PSAT is unique among transaminases or if there are others in this second class. The nucleotide sequences of the 5'-coding regions of two other aminotransferase genes from bacteria are known. The predicted amino-terminal protein sequences of transaminase B from *E.coli* (the *ilvE* gene product; Lawther *et al.*, 1979) and imidazolyacetolphosphate aminotransferase from *S.typhimurium* (the *hisC* gene product; Riggs & Artz, 1984) are not homologous to each other, or to the amino-terminal sequences of PSAT, or to the tyrosine and aspartate aminotransferases from *E.coli*. However, without knowing the entire sequences it is not possible to say that they are not homologous to PSAT.

A final point is that the difference between PSAT and the other transaminases may be related to the function or evolution of the *serC-aroA* operon. It is possible that the ancestral *serC* gene had a different function, but one

which was related to the function of the aroA gene. The linked expression of the genes might at that time have been important to the cell. As the function of the gene changed, the linked expression may have become unnecessary, but there would be no evolutionary pressure to uncouple expression, and in any case, uncoupling would require the evolution of a separate promoter for aroA.

## 7.2 Evolution of the arom gene

The sequences of the aroD and aroA genes, along with the sequence of aroE (I. Anton, Ph.D. Thesis, University of Glasgow, 1984) and aroB (G. Millar, unpublished results) will be used to deduce in detail their evolutionary relationship to the arom gene, when this sequence becomes available. At present, the sequences and complementary protein chemistry data allow us to give hard numbers to the $M_r$'s of the E.coli shikimate pathway enzymes. The $M_r$'s are shown in Table 7.3.

With the exception of 3-dehydroquinase, the E.coli enzymes corresponding to the arom activities exist as monomers. This was unexpected and is unusual for E.coli enzymes, particularly for shikimate dehydrogenase, which represents the only example of a monomeric dehydrogenase in E.coli. It is interesting to note that when the subunit $M_r$'s are added, the total for one of each type of enzyme is 163 000 (see Table 7.3). This is very close to the subunit $M_r$ of 165 000 for the N.crassa arom polypeptide (Lumsden & Coggins, 1977).

| Pathway step | E.coli sub-unit $M_r$ | N.crassa sub-unit $M_r$ |
|---|---|---|
| 1 (trp) | 39 000 (2) | 52 000 (4) |
| 2 | 40 000 (1) | |
| 3 | 28 000 (2) | |
| 4 | 29 000 (1) | |
| 5 | 20 000 (1) | 165 000 (2) |
| 6 | 46 000 (1) | |
| 7 | not known | 50 000 (?) |

Table 7.3. The quarternary structure of the shikimate pathway enzymes. The number of subunits in the native enzyme is shown in brackets.

Together with the comparative 3-dehydroquinase sequences and the comparison with the bifunctional 3-dehydroquinase/ shikimate dehydrogenase plant enzyme, this provides more evidence that the arom gene evolved by fusion of ancestral E.coli-like genes.

## 7.3 Future prospects for the aroA gene

The identification of EPSP synthase as the primary target for the herbicide glyphosate has stimulated interest in the use of a cloned aroA gene in gene replacement experiments in crop plants. A mutationally altered form of EPSP synthase might be resistant to glyphosate and if the gene encoding this form of the enzyme was inserted and expressed in a plant, then it might confer on the plant resistance to the herbicide. This would allow the crop to be sown at the same time as the application of the herbicide, increasing the length of the growing season.

Towards this end, a strain of S.typhimurium has been isolated which has a mutant aroA gene and which synthesises EPSP synthase that is resistant to inhibition by glyphosate (Comai et al., 1983). DNA sequence analysis of both the wild type Salmonella aroA allele and the glyphosate resistant aroA allele has shown two things. Firstly, the sequence of the wild type enzyme is very homologous to the E.coli aroA sequence. Secondly, a point mutation, resulting in a single amino acid substitution is responsible for the glyphosate resistant phenotype (D. Stalker, W. Hiatt, L. Comai, unpublished results). The change involved is:

Position 101 wild type TCG/Ser

mutant CCG/Pro

The expression of an altered aroA gene in plants depends on two factors. It is necessary to insert the gene into the plant genome in such a way as to ensure that it is stably inherited. The gene must then be capable of being transcribed and translated in the plant system. That both of these are possible has been shown in a number of cases. A chimaeric gene composed of the promoter of the Pisum sativum ribulose 1,5-bisphosphate carboxylase gene along with the coding sequence of a bacterial chloramphenicol transacetylase gene was introduced into the genome of Nicotiana tabacum using a Ti plasmid of Agrobacterium tumefaciens. It was demonstrated that this gene was expressed and that the level of expression was light-inducible in chloroplast-containing transformed tissue (Herrera-Estrella et al., 1984). A second example is the introduction of a chimaeric gene containing the coding sequence of the bacterial gene for neomycin phosphotransferase II (the Kanamycin resistance enzyme) to plant cells. Ti plasmid mediated transformation of Nicotiana plumbaginifolia was used to introduce this gene to cells in culture and the cells regenerated to morphologically normal plants. These plants carry a functional Kanamycin resistance gene which is inherited and expressed as a dominant Mendelian trait (Horsch et al., 1984).

Ahmed, S.I. & Giles, N.H. (1969). Organisation of enzymes of the common aromatic synthetic pathway: Evidence for aggregation in fungi. Journal of Bacteriology, 99, 231-237.

Alton, N.K., Buxton, F., Patel, V., Giles, N.H. & Vapnek, D. (1982). 5'-Untranslated sequences of two structural genes in the qa gene cluster of Neurospora crassa. Proceedings of the National Academy of Sciences (USA), 79, 1955-1959.

Anton, I.A. & Coggins, J.R. (1983). Subcloning of the Escherichia coli shikimate dehydrogenase gene (aroE) from the transducing bacteriophage λspc1. Biochemical Society Transactions, 12, 275-276.

Artz, S.W. & Holzschu, D. (1983). Hitidine biosynthesis and its regulation. In: Amino acids: biosynthesis and genetic regulation. eds, Herrmann, K.M. & Somerville, R.L., Ch. 21, p.379-404. London: Addison-Wesley.

Bachmann, B. (1983). Linkage map of Escherichia coli K12, edition 7. Microbiological Reviews, 44, 180-230.

Barra, D., Bossa, F., Doonan, S., Fahmy, H.M.A., Hughes, G.J., Martini, F., Petruzzelli, R. & Wittmann-Leobold, B. (1980). The cytosolic and mitochondrial aspartate aminotransferases from pig heart. European Journal of Biochemistry, 108, 405-414.

Berlyn, M.B. & Giles, N.H. (1969). Organisation of enzymes in the polyaromatic synthetic pathway: separability in bacteria. Journal of Bacteriology, 99, 222-230.

Berylyn, M.B., Ahmed, S.I. & Giles, N.H. (1970). Organisation of polyaromatic biosynthetic enzymes in a variety of photosynthetic organisms. Journal of Bacteriology, 104, 768-774.

Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research, 7, 1513-1523.

Bisswanger, H. & Schimcke-Ott, E. (1980). Multifunctional Proteins. New York: John Wiley & Sons.

Bolivar, F. & Backman, K. (1979). Plasmids of Escherichia coli as cloning vectors. Methods in Enzymology, 68, 245-267.

Bonner, D.M., De Moss, J.A. & Mills, S.E. (1965). The evolution of an enzyme. In: Evolving Genes and Proteins, eds. Bryson, V. & Vogel, H.J. p. 305-318. London: Academic Press.

Boocock, M.R. & Coggins, J.R. (1983). Kinetics of 5-enol-pyruvylshikimate 3-phosphate synthase inhibition by glyphosate. FEBS Letters, 154, 127-133.

Boudet, A.M. & Lécusson, R. (1974). Studies on the association of 5-Dehydroquinate hydro-lyase and shikimate: $NADP^+$-oxidoreductase in higher plants. Planta, 119, 71-79.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. Analytical Biochemistry, 72, 248-254.

Brown, K.D. & Somerville, R.L. (1971). Repression of aromatic amino acid biosynthesis in Escherichia coli K12. Journal of Bacteriology, 108, 386-399.

Buck, M. & Ames, B.N. (1984). A modified nucleotide in tRNA as a possible regulator of aerobiosis: synthesis of cis-2-methylthioribosylzeatin in the tRNA of Salmonella. Cell, 36, 523-531.

Buck, M. & Griffiths, E. (1981). Regulation of aromatic amino acid transport by tRNA: role of 2-methylthio-$N^6$-($\Delta^2$-isopentynyl)-adenosine. Nucleic Acids Research, 9, 401-414.

Buck, M. & Griffiths, E. (1982). Iron mediated methylthiolation of tRNA as a regulator of operon expression in E.coli. Nucleic Acids Research, 10, 2609-2624.

Burgoyne, L., Case, M.E. & Giles, N.H. (1969). Purification and properties of the aromatic (arom) synthetic enzyme aggregate of Neurospora crassa. Biochimica et Biophysica Acta, 191, 452-462.

Butler, J.R., Alworth, W.L., & Nugent, M.J. (1974). Mechanism of dehydroquinase catalysed dehydration. Journal of the American Chemical Society, 96, 1617-1618.

Camakaris, H. & Pittard, J. (1983). Tyrosine biosynthesis. In: Amino acids: Biosynthesis and genetic regulation, eds. Herrmann, K.M. & Somerville, R.L. Ch. 19, p.339-350. London: Addison-Wesley.

Chaudhuri, S. & Coggins, J.R. (1984a). The purification of shikimate dehydrogenase from Escherichia coli. Biochemical Journal, in press.

Chaudhuri, S. & Coggins, J.R. (1984b). Purification of 3-dehydroquinase from Escherichia coli. Biochemical Journal, in press.

Chaudhuri, S. & Coggins, J.R. (1984c). The purification and catalytic properties of catabolic 3-dehydroquinase from Neurospora crassa. Biochemical Journal, in press.

Christiansen, L. & Pedersen, S. (1981). Cloning, restriction endonuclease mapping and post-transcriptional regulation of rpsA, the structural gene for ribosomal protein S1. Molecular and General Genetics, 181, 548-551.

Clarke, L. & Carbon, J. (1976). A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire E.coli genome. Cell, 9, 91-99.

Clarke, S.J., Low, B. & Konigsberg, W.H. (1973). Close linkage of the genes serC (for phosphohydroxy pyruvate transaminase) and serS (for seryl-tRNA synthetase) in Escherichia coli K12. Journal of Bacteriology, 113, 1091-1095.

Coggins, J.R. (1982). Neurospora Newsletter, 29, 8.

Coggins, J.R., Smith, D.D.S., Chaudhuri, S., Coia, A.A., Lambert, J.M. & Nimmo, G.A. The arom multifunctional enzyme of Neurospora crassa: Structural and functional studies on a pentafunctional polypeptide. Biochemical Society Transactions, 9, 44P.

Coggins, J.R., Boocock, M.R., Campbell, M.S., Chaudhuri, S., Lambert, J.M., Lewendon, A., Mousdale, D.M. & Smith, D.D.S. (1985). Functional domains involved in aromatic amino acid biosynthesis. Biochemical Society Transactions, in press.

Comai, L., Sen, L.C., Stalker, D.M. (1983). An altered aroA gene product confers resistance to the herbicide Glyphosate. Science, 221, 370-371.

Cossart, P., Katinka, M. & Yaniv, M. (1981). Nucleotide sequence of the thrB gene of E.coli and its two adjacent regions: the thrAB and thrBC junctions. Nucleic Acids Research, 9, 339-347.

Crawford, I.P. (1980). Gene fusions in the tryptophan pathway: Tryptophan synthase and Phosphoribosyl-anthranilate isomerase: Indoleglycerol-phosphate synthase. In: Multifunctional Proteins, eds. Bisswanger, H. & Schmincke-Ott, E. Ch. 5, p.151-173.

Crawford, I.P., Nichols, B.P. & Yanofsky, C. (1980). Nucleotide sequence of the trpB gene in E.coli and S.typhimurium. Journal of Molecular Biology, 142, 489-502.

Cronan, J.E., Silbert, D.F. & Wulff, D.L. (1972). Mapping of the fabA locus for unsaturated fatty acid biosynthesis in Escherichia coli. Journal of Bacteriology, 112, 206-211.

Dagert, M. & Ehrlich, S.D. (1979). Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene, 6, 23-28.

Davis, B.D. (1951). Aromatic biosynthesis. I. The role of shikimic acid. Journal of Biological Chemistry, 191, 315-325.

Davis, W.D. & Davidson, B.E. (1982). The nucleotide sequence of aroG, the gene for 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (phe) in E.coli K12. Nucleic Acids Research, 10, 4045-4058.

Dempsey, W.B. (1969). 3-phosphoserine transaminase mutants of Escherichia coli B. Journal of Bacteriology, 100, 1114-1115.

Dempsey, W.B. & Itoh, H. (1970). Characterisation of pyridoxine auxotrophs of E.coli: Serine and pdxF mutants. Journal of Bacteriology, 104, 658-667.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Research, 12, 387-395.

Donahue, T.F., Farabaugh, P.J. & Fink, G.R. (1982). The nucleotide sequence of the HIS4 region of yeast. Gene, 18, 47-59.

Doolittle, R.F. (1981). Similar amino acid sequences: Chance or common ancestry? Science, 214, 149-159.

Doonan, S., Doonan, H.J., Hanford, R., Vernon, C.A., Walker, J.M., Airold, L.P.S., Bossa, F., Barra, D., Carloni, M., Fasella, P. & Riva, F. (1975). The primary structure of aspartate aminotransferase from pig heart muscle. Biochemical Journal, 149, 497-506.

Dougan, G., Saul, M., Warren, G. & Sherratt, D. (1978). A functional map of plasmid ColE1. Molecular and General Genetics, 158, 325-327.

Doy, C.H. & Brown, K.D. (1965). Control of aromatic biosynthesis: The multiplicity of 7-phospho-2-oxo-3-deoxy-D-arabino-heptonate D-erythrose-4-phosphate-lyase (pyruvate phosphorylating) in Escherichia coli W. Biochimica et Biophysica Acta, 104, 377-389.

Duncan, K. & Coggins, J.R. (1983). Subcloning of the Escherichia coli genes aroA (5-enolpyruvylshikimate 3-phosphate synthase) and aroB (3-dehydroquinate synthase). Biochemical Society Transactions, 12, 274-275.

Duncan, K., Lewendon, A. & Coggins, J.R. (1984a). The purification of 5-enolpyruvylshikimate 3-phosphate synthase from an overproducing strain of Escherichia coli. FEBS Letters, 165, 121-127.

Duncan, K., Lewendon, A. & Coggins, J.R. (1984b). The complete amino acid sequence of Escherichia coli 5-enolpyruvyl shikimate 3-phosphate synthase. FEBS Letters, 170, 59-63.

Ely, B. & Pittard, J. (1979). Aromatic amino acid biosynthesis: Regulation of shikimate kinase in E.coli K12. Journal of Bacteriology, 138, 933-943.

Ferrara, P., Duchange, N., Zakin, M.M. & Cohen, G.H. (1984). Internal homologies in the two aspartokinase-homoserine dehydrogenases of Escherichia coli K12. Proceedings of the National Academy of Sciences (USA), 81, 3019-3023.

Fickett, J.W. (1982). Recognition of protein coding regions in DNA sequences. Nucleic Acids Research, 10, 5303-5318.

Fleming, T.P., Nahlik, M.S. & McIntosh, M.A. (1983). Regulation of enterobactin iron transport in E.coli: characterisation of ent::Mu d($Ap^{R}$Lac) operon fusions. Journal of Bacteriology, 156, 1171-1177.

Frost, J., Bender, J., Kadonaga, J.T. & Knowles, J.R. (1984). Dehydroquinate synthase from E.coli: Purification, cloning and the construction of overproducers of the enzyme. Biochemistry, in press.

Gaertner, F.H. & Cole, K.W. (1977). A cluster gene: evidence for one gene, one polypeptide, five enzymes. Biochemical and Biophysical Research Communications, 75, 259-264.

Garner, C. & Herrmann, K.M. (1983). Biosynthesis of phenylalanine. In: Amino acids: Biosynthesis and genetic regulation, eds. Herrmann, K.M. & Somerville, R.L. Ch. 18, p.323-338. London: Addison-Wesley.

Giles, N.H. (1978). The organisation, function and evolution of gene clusters in Eukaryotes. American Naturalist, 112, 641-657.

Giles, N.H., Case, M.E., Partridge, C.W.H. & Ahmed, S.I. (1967a). A gene cluster in Neurospora crassa coding for an aggregate of five aromatic synthetic enzymes. Proceedings of the National Academy of Sciences (USA), 58, 1453-1460.

Giles, N.H., Partridge, C.W.H., Ahmed, S.I. & Case, M.E. (1967b). The occurrence of two dehydroquinases in Neurospora crassa, one constitutive and one inducible. Proceedings of the National Academy of Sciences (USA), 58, 1930-1937.

Gollub, E., Zalkin, H. & Sprinson, D.B. (1967). Correlation of genes and enzymes, and studies on regulation of the aromatic pathway in Salmonella. Journal of Biological Chemistry, 242, 5323-5328.

Gorini, L., Gundersen, W. & Burger, M. (1961). Genetics of regulation of enzyme synthesis in the arginine biosynthetic pathway in E.coli. Cold Spring Harbour Symposium on Quantitative Biology, 26, 173-182.

Gouy, M. & Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Research, 10, 7055-7074.

Grosjean, H. & Fiers, W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene, 18, 199-209.

Hasan, N. & Nester, E.W. (1978a). Dehydroquinate synthase in Bacillus subtilis. Journal of Biological Chemistry, 253, 4999-5004.

Hasan, N. & Nester, E.W. (1978b). Purification and properties of chorismate synthase from Bacillus subtilis. Journal of Biological Chemistry, 253, 4993-4998.

Haslam, E. (1974). The Shikimate Pathway. London: Butterworths.

Hautala, J.A., Jacobson, J.W., Case, M.E. & Giles, N.H. (1975). Purification and characterisation of catabolic dehydroquinase, an enzyme in the inducible quinic acid catabolic pathway of Neurospora crassa. Journal of Biological Chemistry, 250, 6008-6014.

Hawley, D.K. & McClure, W.R. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Research, 11, 2237-2255.

Henner, D.J. & Hoch, J.A. (1980). The Bacillus subtilis chromosome. Microbiological Reviews, 44, 57-82.

Herrera-Estrella, L., Van den Broeck, G., Maenhaut, R., Van Montagu, M., Schell, J., Timko, M. & Cashmore, A. (1984). Light-inducible and chloroplast-associated expression of a chimaeric gene introduced into Nicotiana tabacum using a Ti plasmic vector. Nature, 310, 115-120.

Herrmann, K.M. (1983). The common aromatic biosynthetic pathway. In: Amino acids: biosynthesis and genetic regulation, ed. Herrmann, K.M. & Somerville, R.L. Ch. 17, p.301-322. London: Addison-Wesley.

Hirsch-Kolb, H. & Greenberg, D.M. (1971). Phosphoserine aminotransferase (sheep brain). Methods in Enzymology, 17B, 331-334.

Holmes, D.S. & Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Analytical Biochemistry, 114, 193-197.

Horsch, R.B., Fraley, R.T., Rogers, S.G., Sanders, P.R., Lloyd, A. & Hoffmann, N. (1984). Inheritance of functional foreign genes in plants. Science, 223, 496-498.

Huang, L., Montoya, A.L. & Nester, E.W. (1975). Purification and characterisation of shikimate kinase enzyme activity in Bacillus subtilis. Journal of Biological Chemistry, 250, 7675-7681.

Huiet, L. (1984). Molecular analysis of the Neurospora qa-1 regulatory region indicates that two interacting genes control qa gene expression. Proceedings of the National Academy of Sciences (USA), 81, 1174-1178.

Huynh, Q.K., Sakakibara, R., Watanabe, T. & Wada, H. (1980). Primary structure of mitochondrial oxaloacetic transaminase from rat liver: comparison with that of the pig heart isozyme. Biochemical and Biophysical Research Communications, 97, 474-479.

Itoh, H. & Dempsey, W.B. (1970). Purification of 3-phosphoserine-α-ketoglutarate transaminase from E.coli B. Life Sciences, 9, 1289-1294.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology, 3, 318-356.

Katinka, M., Cossart, P., Sibilli, L., Saint-Girons, I., Chahignac, M.A., Le Bras, G., Cohen, G.N. & Yaniv, M. (1980). Nucleotide sequence of the thrA gene of Escherichia coli. Proceedings of the National Academy of Sciences (USA), 77, 5730-5733.

Kinghorn, J.R. & Hawkins, A.R. (1982). Cloning and expression in E.coli K12 of the biosynthetic dehydroquinase function of the arom cluster gene from the Eukaryote, Aspergillus nidulans. Molecular and General Genetics, 186, 145-152.

Kinghorn, J.R., Schweizer, M., Giles, N.H. & Kushner, S.R. (1981). The cloning and analysis of the aroD gene of E.coli K12. Gene, 14, 73-80.

Kirschner, K. & Bisswanger, H. (1976). Multifunctional proteins. Annual Review of Biochemistry, 45, 143-166.

Kitakawa, M., Blumenthal, L. & Isono, K. (1980). Isolation and characterisation of specialised transducing λ phages carrying ribosomal protein genes of Escherichia coli. Molecular and General Genetics, 180, 343-349.

Kondo, K., Wakabayashi, S., Yagi, T. & Kagamiyama, H. (1984). The complete amino acid sequence of aspartate aminotransferase from Escherichia coli: Sequence comparison with pig isoenzymes. Biochemical and Biophysical Research Communications, 122, 62-67.

Koshiba, T. (1979). Organisation of enzymes of the shikimate pathway of Phaseolus mungo seedlings. Plant and Cellular Physiology, 20, 667-670.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227, 680-685.

Laird, A.J., Ribbons, D.W., Woodrow, G.C. & Young, I.G. (1980). Bacteriophage Mu-mediated gene transposition and in vitro cloning of the enterochelin gene cluster of E.coli. Gene, 11, 347-357.

Laird, A.J. & Young, I.G. (1980). Tn5 mutagenesis of the enterochelin gene cluster of E.coli. Gene, 11, 359-366.

Lambert, J.M., Boocock, M.R. & Coggins, J.R. (1984). The 3-dehydroquinate synthase activity of the pentafunctional arom enzyme complex of Neurospora crassa is zinc-dependent. Biochemical Journal, in press.

Larimer, F.W., Morse, C.C., Beck, A.K., Cole, K.W. & Gaertner, F.H. (1983). Isolation of the ARO1 cluster gene of Saccharomyces cerevisiae. Molecular and Cellular Biology, 3, 1609-1614.

Lawther, R.P., Nichols, B., Zurawski, G. & Hatfield, G.W. (1979). The nucleotide sequence preceding and including the beginning of the ilvE gene of the ilvGEDA operon of Escherichia coli K12. Nucleic Acids Research, 7, 2289-2301.

de Leeuw, A. (1967). Gene-enzyme relationships in polyaromatic auxotrophic mutants in Saccharomyces cerevisiae. Genetics, 56, 554-555.

Levin, J.G. & Sprinson, D.B. (1964). The enzymatic formation and isolation of 5-enolpyryvylshikimate 3-phosphate. Journal of Biological Chemistry, 239, 1142-1150.

Lewendon, A. & Coggins, J.R. (1983). Purification of 5-enolpyruvylshikimate 3-phosphate synthase from Escherichia coli. Biochemical Journal, 213, 187-191.

Lumsden, J. & Coggins, J.R. (1977). The subunit structure of the arom multienzyme complex of Neurospora crassa. A possible pentafunctional polypeptide chain. Biochemical Journal, 161, 599-607.

Lumsden, J. & Coggins, J.R. (1978). The subunit structure of the arom multienzyme complex of Neurospora crassa. Evidence from peptide maps for the identity of the subunits. Biochemical Journal, 169, 441-444.

Maitra, U.S. & Sprinson, D.B. (1978). 5-dehydro-3-deoxy-D-arabino-heptulosonic acid 7-phosphate: An intermediate in the 3-dehydroquinate synthase reaction. Journal of Biological Chemistry, 253, 5426-5430.

Maizel, J.V. & Lenk, R.P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proceedings of the National Academy of Science (USA), 78, 7665-7669.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbour Publications.

Marmur, J. (1961). A procedure for the isolation of DNA from micro-organisms. Journal of Molecular Biology, 3, 208-218.

Mass, W.K. (1961). Studies on repression of arginine biosynthesis in E.coli. Cold Spring Harbour Symposium on Quantitative Biology, 26, 183-191.

McCandliss, R.J., Poling, M.J. & Herrmann, K.M. (1978). 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase: Purification and molecular characterisation of the phenylalanine-sensitive isoenzyme from Escherichia coli. Journal of Biological Chemistry, 253, 4259-4265.

McCarthy, A.D. & Hardie, D.G. (1984). Fatty acid synthase - an example of protein evolution by gene fusion. Trends in Biochemical Sciences, 9, 60-63.

McCray, J.W. & Herrmann, K.M. (1976). Derepression of certain aromatic amino acid biosynthetic enzymes of Escherichia coli K12 by growth in $Fe^{3+}$-deficient medium. Journal of Bacteriology, 125, 608-615.

McKitrick, J.C. & Pizer, L.I. (1980). Regulation of phosphoglycerate dehydrogenase levels and effect on serine synthesis in E.coli K12. Journal of Bacteriology, 141, 235-245.

Messing, J. (1983). New M13 vectors for cloning. Methods in Enzymology, 101, 20-78.

Messing, J., Crea, R. & Seeburg, P.H. (1981). A system for shotgun DNA sequencing. Nucleic Acids Research, 9, 309-321.

Messing, J. & Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene, 19, 269-276.

Miki, T., Shirabe, K., Ebina, Y. & Nakazawa, A. (1984). Transcriptional analysis of the lexA gene of Escherichia coli: attenuation and cotranscription with the neighbouring region. Nucleic Acids Research, 12, 1203-1217.

Morell, H., Clark, M.J., Knowles, P.F. & Sprinson, D.B. (1967). The enzymatic synthesis of chorismic and prephenic acids from 5-enolpyruvylshikimate 3-phosphate. Journal of Biological Chemistry, 242, 82-90.

Mousdale, D.M. & Coggins, J.R. (1984a). Purification and properties of 5-enolpyruvylshikimate 3-phosphate synthase from seedlings of Pisum sativum L. Planta, 160, 78-83.

Mous dale, D.M. & Coggins, J.R. (1984b). Subcellular localization of the common shikimate pathway enzymes in Pisum sativum L. Planta, in press.

Nakanishi, N. & Yamamoto, M. (1984). Analysis of the structure and transcription of the aro3 cluster gene in Schizosaccharomyces pombe. Molecular and General Genetics, 195, 164-169.

Neidhart, F.C., Vaughn, V., Phillips, T.A. & Bloch, P.L. (1983). Gene-protein index of Escherichia coli K12. Microbiological Reviews, 47, 231-284.

Neilands, J.B. (1982). Microbial envelope proteins related to iron. Annual Review of Microbiology, 36, 285-309.

Nichols, B.P. & Yanofsky, C. (1979). Nucleotide sequence of trpA of Salmonella typhimurium and Escherichia coli: An evolutionary comparison. Proceedings of the National Academy of Sciences (USA), 76, 5244-5248.

Nimmo, G.A. & Coggins, J.R. (1981). The purification and properties of the tryptophan-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from Neurospora crassa. Biochemical Journal, 197, 427-436.

Patel, V., Schweizer, M., Dykstra, C.C., Kushner, S.R. & Giles, N.H. (1981). Genetic organisation and transcriptional regulation in the qa gene cluster of Neurospora crassa. Proceedings of the National Academy of Sciences (USA), 78, 5783-5787.

Pittard, J. & Wallace, B.J. (1966). Distribution and function of genes concerned with aromatic biosynthesis in Escherichia coli. Journal of Bacteriology, 91, 1494-1508.

Platt, T. (1978). Regulation of gene expression in the tryptophan operon of Escherichia coli. In: The Operon. eds. Miller, J.H. & Reznikoff, p.263-302. New York: Cold Spring Harbour Publications.

Poling, M.D., Suzich, J., Shultz, J. & Herrmann, K.M. (1981). Purification and properties of the tryptophan-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from E.coli. Federation Proceedings, 40, 1581.

Polley, L.D. (1978). Purification and characterisation of 3-Dehydroquinate hydrolyase and Shikimate oxido-reductase. Evidence for a bifunctional enzyme. Biochimica et Biophysica Acta, 526, 259-266.

Poulsen, P., Bonekamp, F. & Jensen, K.F. (1984). Structure of the Escherichia coli pyrE operon and control of pyrE expression by a UTP modulated intercistronic attenuation. The EMBO Journal, 3, 1783-1790.

Pribnow, D. (1975). Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proceedings of the National Academy of Sciences (USA), 72, 784-788.

Riggs, D. & Artz, S. (1984). The hisD-hisC gene border of the Salmonella typhimurium histidine operon. Molecular and General Genetics, 196, 526-529.

Rines, H.W., Case, M.E. and Giles, N.H. (1969). Mutants in the arom gene cluster of Neurospora crassa specific for biosynthetic dehydroquinase. Genetics, 61, 789-800.

Rogers, S.G., Brand, L.A., Holder, S.B., Sharps, E.S. & Brackin, M.J. (1983). Amplification of the aroA gene from Escherichia coli results in tolerance to the herbicide glyphosate. Applied and Environmental Microbiology, 46, 37-43.

Rosenberg, M. & Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. Annual Review of Genetics, 13, 19-53.

Sanderson, K.E. & Roth, J.R. (1983). Linkage map of Salmonella typhimurium. Microbiological Reviews, 47, 410-553.

Sanger, F. (1981). Determination of nucleotide sequences in DNA. Science, 214, 1205-1210.

Schnier, J. & Isono, K. (1982). The DNA coding sequence for ribosomal protein S1. Nucleic Acids Research, 10, 1857-1865.

Schoner, R. & Herrmann, K.M. (1976). 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase: Purification, properties and kinetics of the tyrosine-sensitive isoenzyme from Escherichia coli. Journal of Biological Chemistry, 251, 5440-5447.

Schweizer, M., Case, M.E., Dykstra, C.C., Giles, N.H. & Kushner, S.R. (1981). Identification and characterisation of recombinant plasmids carrying the complete qa gene cluster from Neurospora crassa including the qa-1$^{+}$ regulatory gene. Proceedings of the National Academy of Sciences (USA), 78, 5086-5090.

Shimizu, S. & Dempsey, W.B. (1978). 3-hydroxypyruvate substitutes for pyridoxine in serC mutants of Escherichia coli K12. Journal of Bacteriology, 134, 944-949.

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

Shlyapnikov, S.V., Myasnikov, A.N., Severin, E.S., Myagkova, M.A., Torchinsky, Y.M. & Braunstein, A.E. (1979). Primary structure of cytoplasmic aspartate amino-transferase from chicken heart and its homology with pig heart isoenzymes. FEBS Letters, 106, 385-388.

Shultz, J., Hermodson, M.A., Garner, C.C. & Herrmann, K.M. (1984). The nucleotide sequence of the aroF gene of Escherichia coli and the amino acid sequence of the encoded protein, the tyrosine-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. Journal of Biological Chemistry, 259, 9655-9661.

Smith, H.O. & Birnstiel, M.L. (1976). A simple method for DNA restriction site mapping. Nucleic Acids Research, 3, 2387-2398.

Smith, D.D.S. & Coggins, J.R. (1983). Isolation of a bi-functional domain from the pentafunctional arom enzyme complex of Neurospora crassa. Biochemical Journal, 213, 405-415.

Staden, R. (1978). Further procedures for sequence analysis by computer. Nucleic Acids Research, 5, 1013-1015.

Staden, R. (1980). A new method for the storage and manipulation of DNA gel reading data. Nucleic Acids Research, 8, 3673-3694.

Staden, R. (1984). Computer methods to locate signals in nucleic acid sequences. Nucleic Acids Research, 12, 505-519.

Steinrucken, H.C. & Amrhein, N. (1980). The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate 3-phosphate synthase. Biochemical and Biophysical Research Communications, 94, 1207-1212.

Strauss, A. (1979). The genetic fine structure of the complex locus aro3 involved in early aromatic amino acid biosynthesis in Schizosaccharomyces pombe. Molecular and General Genetics, 172, 233-241.

Sutcliffe, J.G. (1979). Complete sequence of the E.coli plasmid pBR322. Cold Spring Harbour Symposium on Quantitative Biology, 43, 77-90.

Takeda, Y., Nishimura, A., Nishimura, Y., Yamada, M., Yasuda, S., Suzuki, H. & Hirota, Y. (1981). Synthetic ColE1 plasmids carrying genes for penicillin-binding proteins of E.coli. Plasmid, 6, 86-98.

Taylor, A.L. (1970). Current linkage map of Escherichia coli. Bacteriological Reviews, 34, 155-175.

Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. & Gralla, J. (1973). Improved estimation of secondary structures in ribonucleic acids. Nature New Biology, 246, 40-41.

Tribe, D.E., Camakaris, H. & Pittard, J. (1976). Constitutive and repressible enzymes of the common pathway of aromatic biosynthesis in E.coli K12: Regulation of enzyme synthesis at different growth rates. Journal of Bacteriology, 127, 1085-1097.

Twigg, A.J. & Sherratt, D. (1980). Trans-complementable copy-number mutants of plasmid ColE1. Nature, 283, 216-218.

Tyler, B.M., Geever, R.F., Case, M.E. & Giles, N.H. (1984). Cis-acting and trans-acting regulatory mutations define two types of promoters controlled by the qa-1F gene of Neurospora. Cell, 36, 493-502.

Vapnek, D., Hautala, J.A., Jacobson, J.W., Giles, N.H. & Kushner, S.R. (1977). Expression in E.coli K12 of the structural gene for catabolic dehydroquinase of N.crassa. Proceedings of the National Academy of Sciences (USA), 74, 3508-3512.

Wallace, B.J. & Pittard, J. (1967). Genetic and biochemical analysis of the isoenzymes concerned in the first reaction of aromatic biosynthesis in Escherichia coli. Journal of Bacteriology, 93, 237-244.

Weiss, U. & Edwards, J.M. (1980). The Biosynthesis of Aromatic Compounds. New York: John Wiley & Sons.

Welch, G.R., Cole, K.W. & Gaertner, F.H. (1974). Chorismate synthase of Neurospora crassa: a flavoprotein. Archives of Biochimistry and Biophysics, 165, 505-518.

Welch, G.R. & Gaertner, F.H. (1975). Influence of an aggregated multienzyme system on transient time: kinetic evidence for compartmentation by an aromatic amino acid synthesising complex of Neurospora crassa. Proceedings of the National Academy of Sciences (USA), 72, 4218-4222.

Welch, G.R. & Gaertner, F.H. (1976). Co-ordinate activation of a multienzyme complex by the first substrate. Archives of Biochemistry and Biophysics, 172, 476-489.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. Nature, 289, 751-758.

Zakin, M.M., Duchange, N., Ferrara, P. & Cohen, G.H. (1983). Nucleotide sequence of the metL gene of Escherichia coli. Journal of Biological Chemistry, 258, 3028-3031.

Zalkin, H. (1980). Anthranilate synthase: Relationships between bifunctional and monofunctional enzymes. In: Multifunctional Proteins, ed. Bisswanger, H. & Schmincke-Ott, E. Ch. 4, p.123-149. New York: John Wiley & Sons.

Zalkin, H., Paluh, J.L., van Cleemput, M., Moye, W.S. & Yanofsky, C. (1984). Nucleotide sequence of Saccharomyces cerevisiae genes TRP2 and TRP3 encoding bifunctional Anthranilate synthase: Indole-3-glycerol phosphate synthase. Journal of Biological Chemistry, 259, 3985-3992.

Zalkin, H. & Yanofsky, C. (1982). Yeast gene TRP5: Structure, function, regulation. Journal of Biological Chemistry, 257, 1491-1500.

Zurawski, G., Gunsalus, R.P., Brown, K.D. & Yanofsky, C. (1981). Structure and regulation of aroH, the structural gene for tryptophan-repressible 3-deoxy-D-arabino-heptulosonate 7-phosphate of Escherichia coli. Journal of Molecular Biology, 145, 47-73.