Manipulation of the shikimate pathway in *Streptomyces coelicolor*

by

Philip Andrew McKernan

Thesis submitted to the Faculty of Science, University of Glasgow
for the degree of Doctor of Philosophy

Robertson Laboratory of Genetics
July, 1995

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

The research reported in this thesis is my own original work except where otherwise stated, and has not been submitted for any other degree

July 1995

Dedicated to my family, especially Mum and Dad
for all their help, support and endless patience

Acknowledgements

Many thanks to Iain Hunter, my supervisor, for his help and guidance during both the research and writing (and for coming back from Strath occasionally to make sure we're all still working!), and for slogging through this "little pamphlet".

My gratitude goes to Marion, for keeping the lab in full working order (incredibly), keeping us all sane (relatively), and make sure we all behave (nearly)...oh, and also for not bullying me TOO much about the lab coat.

To Linda, I give my office....I have really appreciated the constant access to the office, Mac, printer and radio. Hopefully you'll get some peace at last.

To the Prep Room ladies, ever helpful and happy: we don't know how lucky we are.

Also, thanks to Douglas Armstrong and Geoff Dance for their help with the scanners, and to John Greene for the dehydroquinate, etc.

This work was supported by a grant from the SERC.

To the lab, I apologise for any bouts of temper, sullenness, lunacy or outrageous cheerfulness (and for hogging the Mac). Oh, and the cricket stories. Thanks to all my friends in the lab, department, cricket and football teams, etc., for all the fun, sarcasm, insults, nights out, beer, sporting excellence, etc. I know who you all are (and am plotting my revenge).

Thanks also to my friends in Crosby for coming to see me occasionally (or not, as the case may be) and for still speaking to me when I managed to get home.

Many thanks in particular to Paul, for showing that it can be done; for his Materials (especially all the spelling mistakes); for dragging me to the pub on many occasions and generally being sociable; and for all his help and support -- a true mate (not bad for a Manc)!

Thanks (?) also to the cat for howling so much.

My eternal gratitude goes to Gill, for always being there when I needed help, a shoulder to cry on, an ego boost or confidence, for generally cheering me up constantly, and especially for getting me through the worst bits of the PhD! I will never forget, and hope I can repay the favour somehow.

The PhD student ?

CONTENTS

LIST OF CONTENTS — Page Number

LIST OF FIGURES **Page Number**

LIST OF TABLES

Abbreviations

Å	Angstroms
AMPPD	3-(2'-Spiroadamantane)-4-methoxy-4-(3"-phosphoryloxy)-phenyl-1,2-dioxetane
Arg	arginine
ATP	adenosine triphosphate
BSA	bovine serum albumin
bp	base pair
C-(terminal)	carboxy-(terminal)
CIP	calf intestinal phosphate
Cl^-	chloride anion
ds	double-stranded
Da	Daltons
DAHP	3-deoxy-D-arabino heptulosonate 7-phosphate
DEPC	diethyl pyrocarbonate
DHQ	dehydroquinate
DHS	dehydroshikimate
DMSO	dimethyl sulphoxide
(g)DNA	(genomic) deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylene diamine tetra acetate
E4P	erythrose-4-phosphate
EPSP	5-enolpyruvylshikimate 3-phosphate
EtBr	ethidium bromide
IPTG	isopropyl-ß-D-thiogalactoside
kb	kilo base pairs
MIC	minimum inhibitory concentration
MOPS	morpholino propane sulphonic acid
M_r	molecular weight
N-(terminal)	amino-(terminal)
NAD^+	nicotinamide adenine dinucleotide
$NADP^+$	nicotinamide adenine dinucleotide phosphate
NIH	National Institute of Health [USA]
nt	nucleotide

(d)dNTP	(di)deoxy nucleotide phosphate
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEP	phospho*enol* pyruvate
dH_2O	distilled water
RNA	ribonucleic acid
RNase	ribonuclease
S3P	shikimate 3-phosphate
SDS	sodium dodecyl sulphate
ss	single-stranded
SSC	standard saline citrate
TAE	Tris-acetate/EDTA buffer
TBE	Tris-borate/EDTA buffer
TE	Tris/EDTA buffer
TEMED	N,N,N',N'-tetramethylethylene diamine
TES	Tris/EDTA/sucrose buffer
TFB	transformation buffer
Tris	Tris (hydroxymethyl) aminoethane
dTTP	thymidine-5'-triphosphate
U	units of enzyme activity
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside

SUMMARY

The work in this thesis has been performed as part of an investigation into the upregulation of flux through the shikimate pathway of *Streptomyces coelicolor*, in order to produce higher levels of substrates for the synthesis of secondary metabolites.

Disruption of the *aroQ* gene, which encodes the type II dehydroquinase (the third enzyme in this pathway) was attempted to block the pathway, and allow analysis of flux through the first two steps of the pathway. However, these experiments were not successful.

Overexpression of the *aroQ* gene in *S.coelicolor* may cause an upregulation of this step, and possibly the whole pathway. To achieve such overexpression, a 3 kb fragment containing the *aroQ* gene was subcloned into a high copy-number streptomycete vector, and transformed into *S.lividans* prior to transformation into *S.coelicolor*. The plasmid produced causes a 7-fold increase in the activity of dehydroquinase. Since only a single protein appears to be overexpressed from this plasmid, it appears that the *aroQ* gene is not transcriptionally coupled to any other genes. However, this plasmid could not be isolated from *S.lividans* cells for transformation into *S.coelicolor*.

The *aroA* gene encodes EPSP synthase, the enzyme which catalyses the fifth step of the pathway. Cloning of this gene was attempted by designing oligonucleotide primers against regions of similarity from the amino acid sequences of EPSP synthases from other species, with the oligonucleotides being used to clone the gene via PCR and probing of genomic blots and a λ-library. All efforts at cloning *aroA* were unsuccessful.

The *aroD* gene from *E.coli*, which encodes the type I dehydroquinase, was subcloned onto a bifunctional vector under the control of the hybrid *tac* promoter. The plasmid produced was transformed into *S.coelicolor*. It caused an approximately 40-fold increase in the activity of dehydroquinase. The overexpressed protein (the type I dehydroquinase) was thermolabile as expected, but showed much higher resistance to inhibition by Cl^- anions than anticipated. It appears to be overexpressed only around the point of transition from exponential to stationary phase, presumably due to the protein containing the rare TTA codon (encoding a leucine). The incorporation of leucine at this point in the polypeptide is controlled at the translational level by the presence of the $tRNA_{UUA}$, which is encoded by the *bldA* gene, and is apparently only expressed at significant levels at the end of the growth phase.

CHAPTER 1

Introduction

1.1 Introduction to *Streptomyces*

Streptomyces are filamentous, aerobic, Gram-positive eubacteria which grow as a mycelial mat consisting of unseptated vegetative (substrate) hyphae with an approximate diameter of 0·5 μm, and contain numerous copies of the genome. Their natural environment is typically soil, where they exist saprophytically, feeding on decaying organic material. They therefore secrete a large number of extracellular enzymes, including proteases, cellulases, lipases and nucleases, to assist in the breakdown of this organic material.

Streptomyces show a complex life-cycle and, under conditions of substrate limitation which induce a stringent response, differentiate to produce aerial hyphae. These aerial mycelia subsequently develop to produce septa and spores (**Figure 1.1**). Each spore contains only a single copy of the chromosome, and the spores may become dispersed to allow growth under more suitable conditions.

Two distinct sets of genes have been identified as being involved in the differentiation process: the *bld* genes are involved in the production of aerial hyphae from the substrate mycelia; the *whi* genes have been implicated in the production of spores from these aerial mycelia. Certain sigma factors, including σ^{WhiG}, appear to function in the regulation of the differentiation process. These sigma factors are concerned with the activation of transcription of those genes involved in the differentiation process and its regulation, and in the production of the secondary metabolites. Some of these genes contain an extremely rare TTA codon which appears not to be present in genes involved in the primary metabolic pathways [Chater, 1989]. Although the *bldA* gene, encoding the leucine $tRNA_{TTA}$, appears to be transcribed at all developmental stages, this tRNA appears to be active only when the cell undergoes a stringent response, caused by growth in adverse conditions which induces differentiation [Leskiw *et al.*, 1991].

Although *Streptomyces* are bacteria, they appear to be more closely related to fungi than to other prokaryotes. Their habitat and growth cycle are more similar to those of fungi than of most bacteria, and the dehydroquinase enzyme of *S.coelicolor* shows no similarity to this enzyme from most bacteria (for example *E.coli*), but shows extensive sequence similarity and functional identity to the dehydroquinase from the quinate pathways of the fungi *Aspergillus nidulans* and *Neurospora crassa* [White *et al.*, 1990].

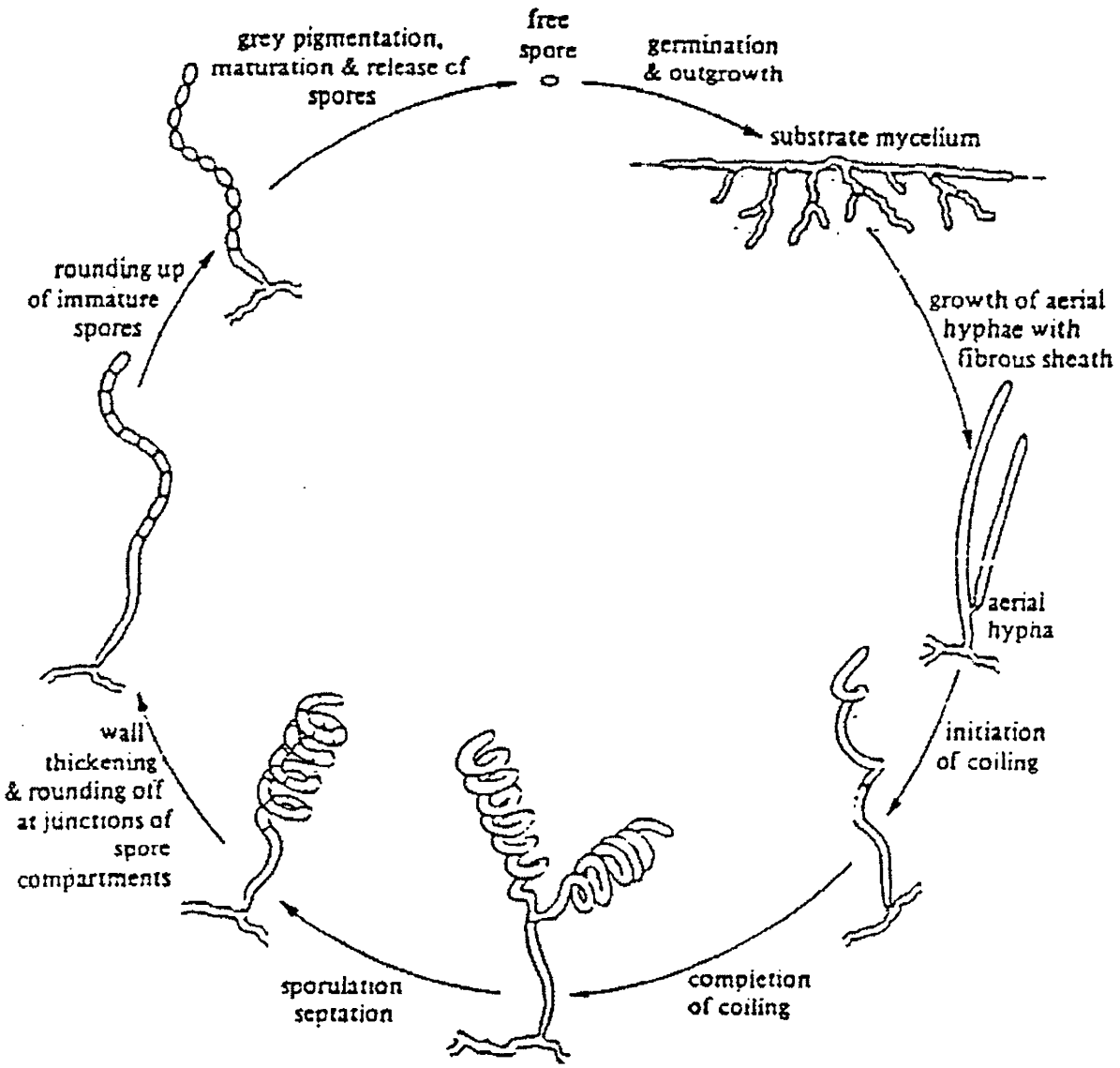


Figure 1.1 Life cycle of *S.coelicolor*, adapted from Chater & Merrick, 1979

1.1.1 Taxonomy of *Streptomyces*

According to the classical method of taxonomy, streptomycetes wee believed to be closely related to the lower eukaryotes, particularly the filamentous fungi, due to their physiology, morphology, complex growth cycle including differentiation and sporulation, and due to their synthesis of secondary metabolites (antibiotics). However, through numerical taxonomy (the comparison of a large number of characteristics between different organisms), due to their positive response to the Gram stain (implying that they are Gram-positive), their whole cell organisation (lack of organelles such as nucleus and mitochondria), absence of chromatin and lack of nucleosomes (hence absence of mitosis) and presence of bacterial ribosomes (especially rRNA (16 and 23S)), streptomycetes have been redefined as bacteria.

Molecular taxonomy investigates the degree of genetic relatedness between species by comparison of their DNA composition (percentage G + C content, the ability of similar sequences to cross-hybridise and particularly rRNA identity. rRNA genes appear to have been extremely well conserved during evolution and therefore the most suitable sequences for comparison. Little work has been performed on the ribosomal genes of *Streptomyces*, although there has been some recent activity on the topic (Mehling *et al* (1995) Microbiology **141:** 2139-2147.

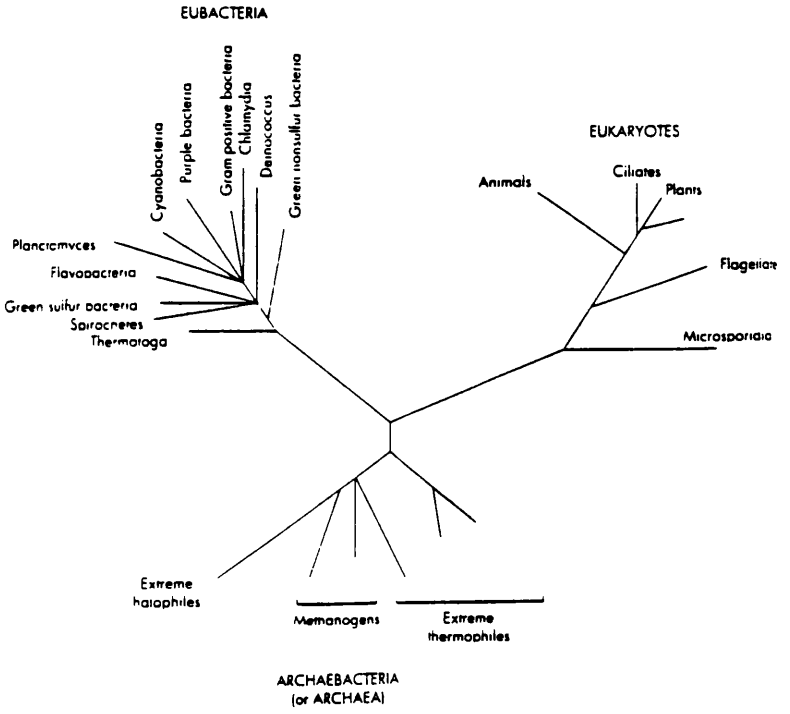


Figure 1.1.1 The Universal Phylogenetic Tree, taken from Glazer & Nikkaido, 1995

1.2 The *Streptomyces* genome

Streptomyces DNA has a high G+C content, between 70-74% (approximately 73% for *S.coelicolor*), which is not due to repetitive satellite DNA. This bias produces a non-random codon usage, with almost exclusive use of G or C in the third position [Wright & Bibb, 1992]. The *S.coelicolor* genome has been estimated at 8 Mb, and has been mapped by pulse-field gel electrophoresis (PFGE) using three "rare-cutter" restriction endonucleases (whose restriction sites consist of only A or T nucleotides) [Kieser *et al.*, 1992; van Wezel *et al.*, 1995]. The *S.coelicolor* genome consists of four quadrants, two of which (at 3 o'clock and 9 o'clock) are considered to be "silent regions", and to which very few cloned genes have been assigned [Hopwood & Kieser, 1990].

The chromosome of *S.lividans* has also been physically mapped using "rare-cutter" restriction endonucleases and PFGE, and its size estimated at approximately 8 Mb [Leblond *et al.*, 1993]. These authors were, however, unable to clone linking segments between a pair of fragments for each of the two enzymes used. They therefore suggested that the *S.lividans* chromosome may be linear, and not circular as anticipated from its circular linkage map. They also showed, by cross-hybridisation, that *S.lividans* and *S.coelicolor* are very closely-related strains.

Lin *et al.* [1993] identified an incomplete copy of the linear plasmid SLP1 integrated into the *S.lividans* chromosome, evidently at the end of a linear molecule, and identified two restriction fragments which apparently contain the free chromosomal ends, confirming that the chromosome of *S.lividans* is linear. They also showed that the free ends are bound together by terminally-bound protein molecules. Each terminus contains four inverted repeat sequences (palindromes), with all of the palindromes having a 3 nt unpaired centre. These authors proposed that these palindromes may be involved in recognition of the termini by DNA polymerase or the terminally-bound proteins. They also showed that deletions across the termini, followed by fusion of the two new chromosomal ends, could produce a circular molecule which is able to replicate stably, presumably utilising the *oriC* which has previously been identified in the *S.lividans* chromosome [Zakrzewska-Czerwinska & Schrempf, 1992]. It is therefore likely that chromosomal replication may be performed on either a circular or linear template.

The chromosomes from six other species of streptomycete (including *S.coelicolor*) were also shown to be likely to exist in a linear form [Lin *et al.*, 1993], and a number of *Streptomyces* strains contain a region homologous to the *oriC* from *S.lividans*. The *oriC* of *S.coelicolor* has since been cloned [Calcutt & Schmidt, 1992], and shows over

99% identity to the *oriC* of *S.lividans*. Linear chromosomes are typically associated with eukaryotes, and have so far been found in only one other bacterial genus, the *Borrelia* spirochetes, as shown initially in *B.burgdorferi* [Casjens & Huang, 1993].

Streptomyces chromosomes appear not to contain completely clustered genes for synthesis of the secondary metabolites, although some of the genes implicated in such a pathway are often clustered together. Genes from the same biosynthetic pathway frequently lie on opposite quarters of the chromosome when it is viewed as a circular molecule. This has been suggested as being due to ancient duplication events, followed by either loss of one of the pair of duplicated genes or evolution, with one of the pair altering its activity to become a new enzyme [Hopwood & Kieser, 1990]. The genes for secondary metabolic pathways appear, however, to be clustered in *Streptomyces*.

S.coelicolor contains a number of extra-chromosomal DNA molecules. SCP1, a giant (~350 kb) linear plasmid, may replicate autonomously, or may integrate into the chromosome. SLP1 may also exists as an integrated molecule, although it may form a stable circular molecule in *S.lividans*. It is thought to be related to the temperate bacteriophages due to its *int* and *xis* genes, encoding the enzymes which accomplish integration and excision events respectively, and its *attP* site, through which it integrates via a homologous recombination event with the chromosomal *attB* site, which overlaps with a gene encoding a $tRNA^{Tyr}$ [Vögtli & Cohen, 1992]. *S.coelicolor* also contains a minicircle which can either replicate autonomously or integrate at two distinct sites, and acts as host for the temperate bacteriophage ØC31, which can infect lytically (replicating autonomously, and lysing the cell to release its progeny phage) or lysogenically (integrating into the chromosome at the *attB* site).

1.3 *Streptomyces* secondary metabolites

Streptomyces produce numerous strain-specific compounds which are not metabolised or utilised for cell production, and are typically excreted from the cells. These compounds are termed secondary metabolites, and they are synthesised at the "idiophase", the point at which the development of aerial mycelia is initiated. Therefore, the secondary metabolites are associated with differentiation. *Streptomyces coelicolor*, the strain on which the work described in this thesis has been undertaken, is the most thoroughly investigated streptomycete, and produces five secondary metabolites: actinorhodin, undecylprodigiosin and CDA (calcium-dependent antibiotic) are chromosomally-encoded antibiotics; methylenomycin is an antibiotic whose biosynthesis is encoded by genes contained on the plasmid SCP1 [Hopwood &

Wright, 1983]; A-factor is an autoregulatory factor involved in the initiation of secondary metabolism. None of the antibiotics is of industrial significance.

The secondary metabolites from *Streptomyces* often have antibacterial, anti-tumour, anti-parasitic or anti-helminthic activity, or function as immunosuppressants, animal growth promotants, vasodilators, enzyme inhibitors, herbicides, pesticides or insecticides [Demain, 1983]. *Streptomyces* produce over 70% of known antibiotics and, therefore, are industrially-important micro-organisms.

It has been postulated that antibiotics sterilise the soil, precluding the growth of other micro-organisms in the immediate environment, and preventing them from utilising the products of the breakdown of the streptomycete vegetative mycelia, which provide the substrates for aerial mycelia production [Davies *et al.*, 1992]. It is also possible that, with the cessation of vegetative growth, there is a build-up of metabolic intermediates which may be toxic in such high concentrations. Secondary metabolites may represent the end-products of an alternative metabolism of toxic intermediates. Therefore, excretion of secondary metabolites may provide a route for removal of such toxic compounds. They may also have evolved as biochemical modulators or effectors.

1.4 RNA polymerase heterogeneity in *Streptomyces*

Unlike eukaryotes, prokaryotes appear to contain only a single core RNA polymerase enzyme, which is responsible for the transcription of genes to produce the mRNA's used as templates for the synthesis of proteins. However, core RNA polymerase requires the presence of an extra protein, known as a sigma (σ) factor, to provide promoter specificity to the core enzyme. Together, core RNA polymerase and a sigma factor constitute the RNA polymerase holoenzyme [Stragier *et al.*, 1985].

Different genes are transcribed by different RNA polymerase holoenzymes, due to the presence in the holoenzyme of differing sigma factors, which direct the core enzyme to a promoter and instigate transcription from the promoter. Bacteria contain a group of principal sigma factors which direct expression from the housekeeping genes, and various minor sigma factors, which appear to be involved in the transcription of those genes whose products are not expressed throughout the growth cycle, for instance the heat-shock genes and those involved in secondary metabolism.

In certain differentiating bacteria, it has been proposed that regulation of the differentiation processes is effected via the presence of varying sigma factors (or their

relative amounts) at different stages of the life cycle. This theory has been extensively investigated in the Gram-positive, sporulating bacterium *Bacillus subtilis*, in which differentiation is known to be controlled by a cascade of sigma factors, and even by anti-sigma factors which have been implicated in the inhibition of certain sigma factors [Errington, 1993]. Recent studies suggest that a cascade of sigma factors may regulate differentiation in *Streptomyces*.

The existence of multiple sigma factors in *Streptomyces* was first proposed with the discovery that the *veg* and *ctc* genes from *B.subtilis* are transcribed by different holoenzymes in *S.coelicolor* [Westpheling *et al.*, 1985]. Two distinct proteins were discovered to be associated with the core polymerase; one of Mr 35000 which allowed transcription of the *veg* gene (σ^{35}), and one of Mr 49000 (σ^{49}) which stimulated transcription of *ctc*. σ^{35} was proposed as being analogous to the principal (or major) bacterial sigma factors (typified by $E\sigma^{70}$, encoded in *E.coli* by the *rpoD* gene).

Four homologues of σ^{70}, *hrdA-hrdD*, have been cloned from *S.coelicolor* [Tanaka *et al.*, 1988; Buttner *et al.*, 1990], and show significant identity to the family of principal sigma factors [Tanaka *et al.*, 1991]. Disruption of three of these genes, *hrdA*, *hrdC* and *hrdD*, has no apparent effect on growth, differentiation or secondary metabolism in *S.coelicolor*, although *hrdB* disruptants were not obtained, suggesting that such cells are not viable, and that inactivation of σ^{HrdB} is a lethal event. It was demonstrated, by S1 nuclease protection, that *hrdB* and *hrdD* are transcribed when *S.coelicolor* is grown in liquid culture [Buttner *et al.*, 1990]. The *hrdC* gene may also be disrupted without any deleterious effects in *S.lividans* [Takahashi *et al.*, 1988]. Pairwise disruption of *hrdA, hrdC* and *hrdD*, and even simultaneous disruption of all three genes, has no apparent effect on *S.coelicolor* [Buttner & Lewis, 1992], suggesting that they may all be functionally redundant.

Four genes encoding putative sigma factors have been identified in *S.aureofaciens* [Kormanec *et al.*, 1992], all of which contain the *rpoD* box [Tanaka *et al.*, 1988], thus identifying them as putatively encoding principal sigma factors. Three of these bear significant similarities to three of the *hrd* genes of *S.coelicolor*, and were therefore named *hrdA*, *hrdB* and *hrdD*; the other gene was named *hrdE*, although it also bears significant similarity to *hrdB* of *S.coelicolor* (74% identity at the amino acid level). All four deduced protein products show similarity to the principal sigma factors, with σ^{HrdB} showing the greatest identity to σ^{70}. This suggests that σ^{HrdB} is the functional homologue of the principal sigma factors in *S.aureofaciens* as well as in *S.coelicolor*.

The principal sigma factors contain numerous completely conserved amino acids.

However, the sequences of the products of the *hrdA* and *hrdD* genes of *S.aureofaciens* differ at several of these residues, suggesting that their products may have a different function. HrdE only differs at two of these positions, with both changes being conservative substitutions. However, *hrdE* contains two inverted repeats in its coding sequence which may prevent its expression. As with *S.coelicolor*, only HrdB and HrdD were shown to be strongly expressed by Northern blotting when *S.aureofaciens* was grown in minimal medium, with slight expression of HrdA occurring. There are four transcriptional start sites lying upstream of *hrdB*, possibly suggesting a complex pattern of regulation of expression.

Using S1 nuclease mapping, *hrdA*, *hrdB* and *hrdD* all show transcription in *S.aureofaciens* from tandem promoters at various stages of development [Kormanec & Farkasovsky, 1993]. Transcription from the tandem promoters of *hrdB* occurs during all stages of the life cycle, with the tandem promoters of *hrdD* only being active in vegetative mycelia, and those of *hrdA* only during the formation of aerial mycelia. *hrdB* also contains a promoter internal to its coding region. It is possible that temporally-regulated expression of sigma factors is a general method of the control of development, as it is in *B.subtilis*.

The *dagA* gene, encoding an extracellular agarase enzyme, has been cloned in *S.coelicolor* [Kendall & Cullum, 1984]. It is transcribed from four separate promoters [Buttner *et al.*, 1987], at least three of which require distinct sigma factors for their transcription [Buttner *et al.*, 1988], although two of these sigma factors support transcription from the *veg* or *ctc* promoters from *B.subtilis*, and are likely to be those identified previously [Westpheling *et al.*, 1985]. The novel sigma factor was named σ^{28}. Transcription from the promoters *-p1*, *-p2* and *-p4* appears to be developmentally regulated, being significantly enhanced at the onset of formation of aerial mycelia [Servin-González *et al.*, 1994].

Transcription from the *dagA-p4* promoter is stimulated by a 56 kDa sigma factor which also directs transcription of the *B.subtilis veg* gene in *S.coelicolor*. This protein has been identified as the product of the *hrdB* gene [Brown *et al.*, 1992]. It is possible that σ^{35}, which had previously been identified as supporting transcription from the *veg* promoter [Westpheling *et al.*, 1985], is a separate sigma factor, or a proteolytic fragment of σ^{HrdB} which retains some activity.

In both *S.coelicolor* and *S.lividans*, the galactose operon is transcribed from two separate promoters, *gal-p1* which is a promoter induced by galactose but repressed by glucose, and *gal-p2*, a low-level constitutive promoter which lies internal to the operon.

Transcription from the *-p1* and *-p2* promoters is controlled by separate sigma factors, possibly σ^{35} and σ^{28} respectively [Westpheling & Brawner, 1989].

The *whi* genes are involved in sporulation, and hence in differentiation. *whi* mutants do not produce spores, although they do produce aerial mycelia. The *whiG* gene has been cloned [Mendez & Chater, 1987] and found to encode a sigma factor, σ^{WhiG}, which is closely related to a minor motility sigma factor from *B.subtilis* [Chater *et al.*, 1989]. As anticipated, disruption of *whiG* has no effect on vegetative growth and prevents sporulation from aerial mycelia, but overexpression of its product causes the production of spores from substrate mycelium, and a reduction in the synthesis of the secondary metabolites. This latter observation is likely to be due to the lysing of the vegetative mycelium which allows this unusual sporulation event before aerial hyphae are able to form, and therefore prior to synthesis of antibiotics. Thus, the level of active σ^{WhiG} in a cell appears to control the initiation of sporulation. Therefore, it is likely that the production of active σ^{WhiG} is, itself, developmentally-regulated.

1.5 Activators of secondary metabolism

In actinomycetes, these effectors have wide-ranging or even global regulatory effects on the synthesis of secondary metabolites, and possibly on differentiation. They are analogous to the eukaryotic hormones or sexual pheromones, and are able to function at extremely low concentrations. The first such identified compound was A-factor [2-(6'-methylheptanoyl)-3*R*-hydroxymethyl-4-butanolide], which was detected initially in *S.griseus* due to its ability to positively regulate the synthesis of the antibiotic streptomycin and the formation of aerial mycelia. It also has similar activity in *S.coelicolor* and *S.lividans*. A-factor is a member of the γ-butyrolactone family, a group of compounds which are distributed widely throughout the actinomycetes, and which often have similar activities.

A-factor is synthesised at an early stage in the cell cycle but is not active at this time, suggesting the existence of a negative effector [Horinouchi & Beppu, 1990]. It is believed to function by stimulating mRNA synthesis from specific genes. Its biosynthesis is regulated by a protein encoded by *afsA*, which is carried on an unstable extra-chromosomal element, and was originally cloned from *S.griseus*. In *S.coelicolor*, the product of another gene, *afsB*, is required for the transcription of *afsA*. Although the *afsB* gene has not yet been cloned, mutations disabling its activity have been complemented by the *afsR* gene [Horinouchi *et al.*, 1990], the product of which, AfsR, promotes the production of secondary metabolites. Mutations in *afsR*, or

its disruption, reduce the synthesis of the secondary metabolites, but do not affect differentiation and sporulation. AfsR contains two separate consensus ATP binding sites towards its N-terminus, and is activated by phosphorylation, accepting the γ-phosphate of ATP from cell lysates of *S.coelicolor* and *S.lividans* [Hong *et al.*, 1991]. A protein capable of phosphorylating AfsR has also been discovered in *S.lividans*, although this activity could not be found in *S.griseus*. It has been suggested that AfsR is the second protein of a two-component regulatory system, with the first protein being autophosphorylated.

Mutations in *absA* prevent synthesis of all antibiotics in *S.coelicolor*, but have no significant effect on sporulation, differentiation or growth rate [Adamidis *et al.*, 1990]. This locus was shown to be unrelated to any of the *bld* or *afs* loci, and these authors have proposed that mutations in *absA* may deregulate a repressor or produce an inhibitory activity. Unusually, *absA* mutants still retain resistance to methylenomycin, suggesting that not all antibiotic resistance genes are co-ordinately regulated with the respective production genes.

Another two-component regulatory system which stimulates the production of antibiotics and A-factor has been identified in *S.coelicolor* [Ishizuka *et al.*, 1992]. The two genes involved in this system, *afsQ1* and *afsQ2*, are transcriptionally coupled. The product of *afsQ1* shows similarity to the response regulators of prokaryotic two-component systems, and the product of *afsQ2* shows similarity to the protein kinases of such combinations. Thus, the product of *afsQ2*, which is likely to be membrane-bound, is capable of phosphorylating the product of *afsQ1*, which functions as a positive regulator when phosphorylated. Their transcription begins immediately before the production of actinorhodin. Although disruption of this gene pair has no effect on secondary metabolism or differentiation in otherwise wild-type *S.coelicolor*, they are capable of suppression of mutations in *absA*.

The gene *abaA* has also been identified as encoding a positive regulator of secondary metabolism in *S.coelicolor*, inducing the production of actinorhodin, undecylprodigiosin and, to some extent, CDA. However, it has no effect on the synthesis of methylenomycin [Fernández-Moreno *et al.*, 1992]. It has been proposed that the product of *abaA* is involved further down the regulatory network than other global regulators, for example AfsR. An open reading frame ORFA, which may also be involved in the control of secondary metabolism, occurs immediately upstream of *abaA*, and may be transcriptionally coupled to it.

From the work described above, it appears that a number of regulatory cascades, which

may well function in a set progression, are involved in the control of differentiation and the biosynthesis of the secondary metabolites in *Streptomyces.*

1.6 The shikimate pathway

The shikimate pathway is a primary metabolic pathway consisting of seven enzymes (DAHP synthase, DHQ synthase, dehydroquinase, shikimate dehydrogenase, shikimate kinase, EPSP synthase and chorismate synthase) which catalyse the conversion of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) to chorismate (**Figure 1.2**). E4P is synthesised from glucose via the pentose phosphate cycle; PEP is also produced from glucose via the Embden-Meyerhof pathway (Mandelstam *et al.*, 1982). Chorismate is used as a precursor for the synthesis of the aromatic amino acids (phenylalanine, tyrosine and tryptophan), *p*-aminobenzoate (*p*AB: precursor of folic acid), 2,3-dihydroxybenzoate (DHB: precursor of the iron chelator enterochelin), *p*-hydroxybenzoate (precursor of ubiquinone), *o*-succinylbenzoate (precursor of the menaquinones), plastoquinone, chloramphenicol, and the vitamins E and K (**Figure 1.3**). Shikimate is a precursor for the synthesis of the ansamycin and rifamycin antibiotics. DAHP may be an intermediate in the synthesis of 3-hydroxy-5-aminobenzoate, and DHQ in the production of 3-aminobenzoate, both of which are precursors of the ansamycins (which have antibacterial, antiviral and antitumour activities) and the immunosuppressant FK506. The aromatic ring derived from the shikimate pathway may also be incorporated into the coumarin antibiotics, of which there are approximately one thousand currently known. *p*-aminobenzoate and 2,3-dihydroxybenzoate are components in bacterial cell walls, and are precursors for lignin synthesis in higher plants (Bentley, 1990).

The shikimate pathway is present in the bacteria, plants and lower eukaryotes, but is absent from the vertebrates, which must therefore be provided with exogenous aromatic amino acids (although they are able to synthesise tyrosine by hydroxylation of phenylalanine). In plants, the shikimate pathway is localised in, or targeted to, the chloroplast.

1.7 The AROM polypeptide

As early as 1967, a gene cluster encoding a pentafunctional AROM polypeptide (Mr ~200,000) encoding the enzymes catalysing steps two to six of the shikimate pathway (dehydroquinate synthase to EPSP synthase), the *arom* gene cluster, was identified in

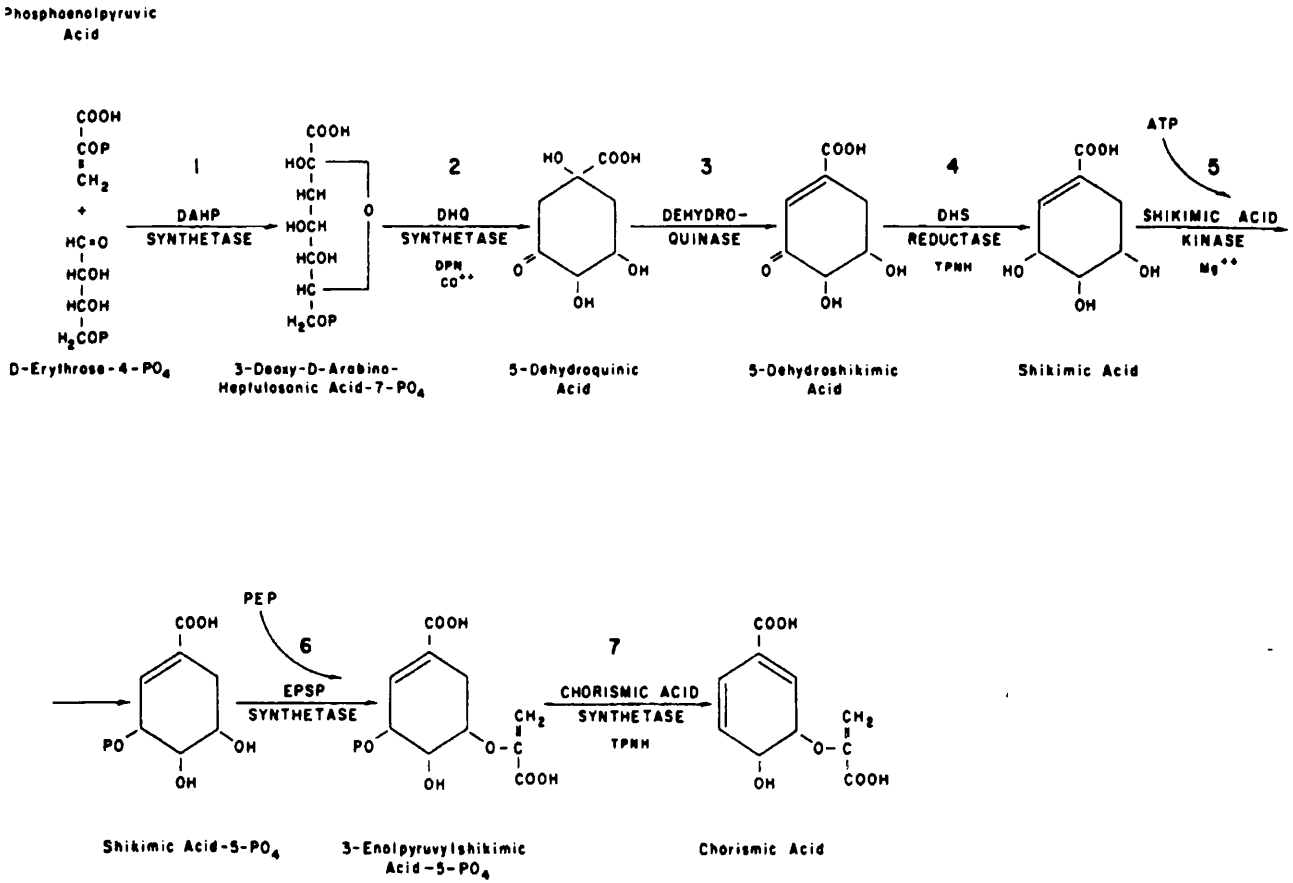


Figure 1.2 The shikimate pathway, taken from Kinghorn & Hawkins, 1982

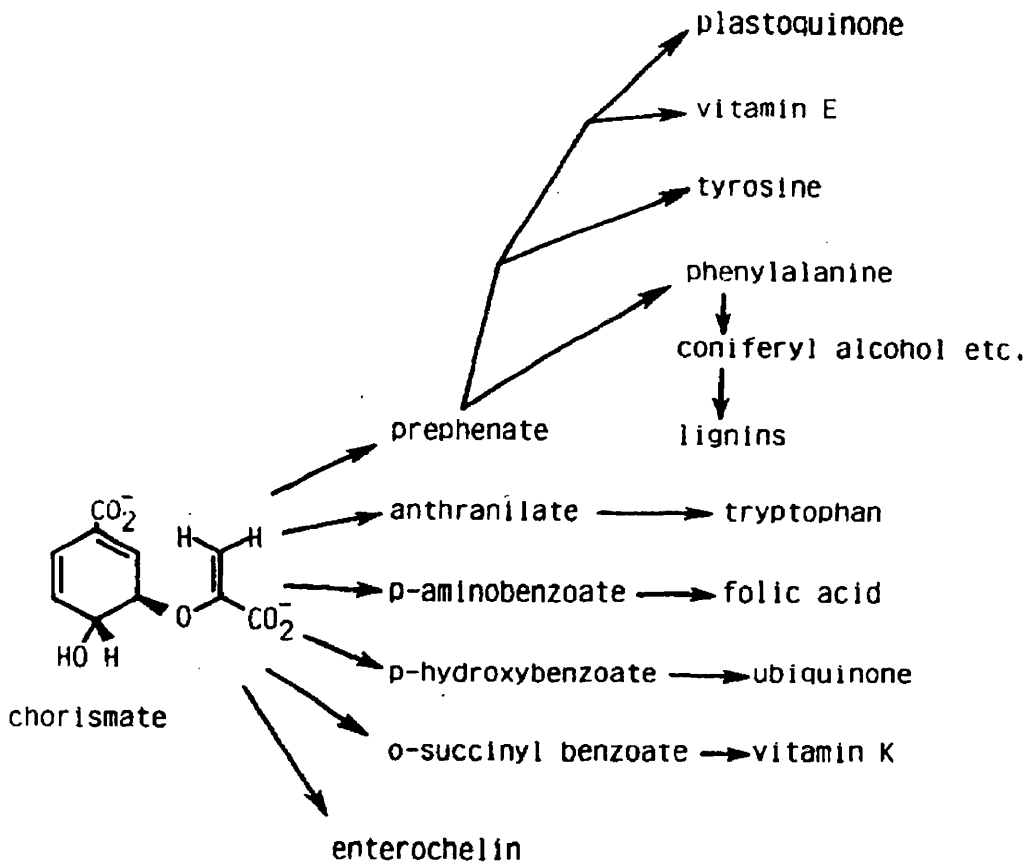


Figure 1.3 Products of chorismate metabolism.

the mould *N.crassa* [Giles *et al.*, 1967] (**Figure 1.4**). The presence of two dehydroquinase activities, one constitutive and occurring as part of the AROM polypeptide, the other inducible and part of a degradative pathway, was reported [Hautala *et al.*, 1975]. It was proposed that the enzyme aggregate may provide a channelling mechanism for the substrates of these two enzymes. The inducible enzyme was purified, and was found to be thermostable at 80°C, but was inhibited by guanidine chloride.

The AROM polypeptide was purified [Gaertner & Cole, 1977] and shown by SDS-PAGE to have an Mr of ~150,000. The native molecular weight was estimated at 300,000, suggesting that the polypeptide functions as a homodimer. This confirmed earlier reports [Case & Giles, 1971] which had also demonstrated that each monomer can dissociate into two distinct sub-aggregates: one containing the dehydroquinate (DHQ) synthase and the EPSP synthase activities; the other containing the dehydroquinase, shikimate kinase and shikimate dehydrogenase domains.

A 4 kb fragment containing part of the *arom* cluster encoding the AROM polypeptide from *N.crassa* was cloned by complementation of an *E.coli aroD* mutant, producing a thermolabile dehydroquinase activity [Catcheside *et al.*, 1985]. This sequence also complements the *qa-2* and *aro-9* dehydroquinase mutations in *N.crassa.*

The AROM polypeptide has most recently been observed as having an Mr of 165,000 [Smith & Coggins, 1983]. As the *arom* cluster of *N.crassa* consists of a single 4·8 kb ORF encoding a protein of estimated Mr ~175,000 [Charles *et al.*, 1986], it appears that the cluster contains only small, or even no introns. Limited proteolysis produces two separable polypeptides of Mr 68,000 and 110,000, the smaller of which retains the dehydroquinase and shikimate dehydrogenase activities.

Ahmed & Giles [1969] used centrifugation in sucrose density gradients to show that six more fungi (*Rhizobus stolonifer, Phycomyces nitens, Absidia glauca, Aspergillus nidulans, Coprinus lagopus* and *Ustilago maydis*) also contained the activities catalysing steps two to six of the shikimate pathway as part of a pentafunctional polypeptide. All six had a thermolabile dehydroquinase activity, but only two showed the thermostable activity. The blue-green algae typically have separable (or at least partially separable) shikimate pathway enzymes, but *Euglena gracilis* contains the pentafunctional AROM polypeptide with an Mr of 249,000 [Berlyn *et al.*, 1970; Patel & Giles, 1979].

Strauss [1979] dissected the AROM-encoding *aro3* gene of *Saccharopolyspora pombe*

into five complementation groups, ordering them by recombination frequencies, and concluded that this *aro3* gene is the functional homologue of the *arom* cluster of *N.crassa*. The *aro3* gene has been cloned and found to complement all tested *aro* mutants of *E.coli* [Nakanishi & Yamamoto, 1984]. Subclones of *aro3* have been isolated which complement both an *aroD* mutation and an *aroD/aroE* double mutation in *E.coli*.

The *aromA* locus of *A.nidulans* has been placed under the control of the powerful *E.coli trc* promoter, and expressed in *E.coli* [Hawkins & Smith, 1991]. The *aromA* gene complements mutations in the *aroA*, *-B*, and-*D* genes, but fails to complement mutations in the *aroE* gene. Hawkins & Smith subcloned the single domains of the *aromA* locus of *A.nidulans* into vectors, allowing separate expression of each of the individual domains of the AROM polypeptide, and showed that the DHQ synthase domain can function independently or as part of a bifunctional protein with EPSP synthase. However, the EPSP synthase domain is only active catalytically when fused to the C-terminus of the DHQ synthase domain, and not when these two polypeptides are merely supplied in *trans*, suggesting that the presence of a DHQ synthase domain covalently attached to its N-terminus is necessary to fold the EPSP synthase domain into an active formation [Moore & Hawkins, 1993]. The dehydroquinase domain functions independently, or as part of a bifunctional polypeptide with either shikimate dehydrogenase or shikimate kinase. It has been suggested that the AROM monomer consists of two distinct domains which can fold and function independently of each other; one consists of the dehydroquinate synthase/EPSP synthase activities, the other of the shikimate kinase/dehydroquinase/shikimate dehydrogenase activities. There is no expression of shikimate kinase from the whole AROM polypeptide or any part of it in *E.coli*. When the biosynthetic dehydroquinase is placed under the control of the *qutE* promoter/terminator, it complements an *A.nidulans qutE* mutation, showing that biosynthetic dehydroquinases are capable of functioning in the quinate pathway.

Integration of twelve copies of the *aromA* gene of *A.nidulans* into an *arom* mutant causes overexpression of the AROM polypeptide [Lamb *et al.*, 1991]. Such overexpressors cannot grow on quinate as sole carbon source or quinate+glycerol, as DHS and DHQ are being fluxed at increased levels into the shikimate pathway, causing the build up of a shikimate pathway intermediate which is toxic at elevated levels. However, when these cells are grown on quinate+glucose, the presence of glucose causes carbon catabolite repression, which reduces the activity of the quinate pathway and allows growth to occur. Elevated levels of quinate prevent full carbon catabolite repression, allowing expression of the quinate pathway and preventing growth. If such overexpressing strains lose many of their copies of the *aromA* gene, effectively

reverting to wild type-levels of AROM, they are again able to grow on quinate as sole carbon source. These authors have also shown that five-fold overexpression of AROM allows a *qutR^c/qutE* mutant to grow on quinate (*qutR^c* denotes a mutation in QutR which leads to constitutive expression of the quinate utilisation genes), suggesting that the AROM protein is leaking DHQ and DHS, which may then be metabolised by the quinate pathway.

The *aromA* gene of *A.nidulans* has been placed under the control of the *qutE* promoter in an *aromA/qutB* double mutant (which is unable to grow in the absence of aromatic amino acids or to grow on quinate as sole carbon source), causing overexpression of the AROM enzyme complex [Moore *et al.*, 1992]. These overexpressing strains showed a significantly reduced growth rate on quinate+glycerol as expected. They produce a quinate-inducible protein of 175 kDa (presumably the AROM protein) whose expression is correlated with the impaired growth on quinate+glycerol.

The *A.nidulans* AROM protein has been fused to the N-terminus of the glutathione S-transferase (GST)-encoding gene under control of the P_{tac} promoter, and has subsequently been overexpressed in *E.coli* [Hawkins *et al.*, 1993]. It has been purified to greater than 98% homogeneity by a single-step affinity purification, and digested with thrombin to remove the GST N-terminus. The dehydroquinase activity is stable in enzyme assays, and the purified AROM polypeptide has a native Mr which suggests that it functions as a monomer in *E.coli*, although it typically functions as a dimer in *A.nidulans*. The dehydroquinase domain of AROM is reduced by borohydride only in the presence of DHQ, confirming its dehydroquinase as a type I enzyme. The K_m of dehydroquinase for DHQ, when the AROM protein is purified from *E.coli*, is ~200 μM (the type I enzyme from *E.coli* has a K_m of 18 μm [Chaudhuri *et al.*, 1986]), suggesting that the dehydroquinase activity of the monomeric AROM protein is inefficient, presumably due to it functioning as a monomer.

The *aro1* gene of *Saccharomyces cerevisiae* encodes a 150 kDa AROM polypeptide which complements missense and nonsense alleles of *S.cerevisiae* and also the *E.coli aroA, aroB, aroD* and *aroE* mutations [Larimer *et al.*, 1983]. The *aro1* gene has been sequenced [Duncan *et al.*, 1987] and shows the same composition and order as the AROM enzymes of *S.pombe* [Nakanishi & Yamamoto, 1984], *A.nidulans* [Hawkins, 1987] and *N.crassa* [Charles *et al.*, 1986]. The AROM polypeptide has also been compared with the corresponding enzymes of *E.coli*, showing that the five monofunctional enzymes of *E.coli* are clearly homologous to separate regions of the multifunctional eukaryotic enzymes. The *E.coli* DHQ synthase, dehydroquinase and EPSP synthase enzymes show extensive similarity to the AROM polypeptide from

A.nidulans, suggesting that the shikimate pathway enzymes have evolved from a common ancestor, and that their genes have fused in eukaryotes to produce a mosaic polypeptide, rather than having become dispersed in bacteria [Hawkins, 1987]. It has also been shown that fragments of this *arom* locus can complement *aroA* and *aroD* mutations of *E.coli.*

The *aroA* function of another fungus, *Pneumocystis carinii*, has been cloned [Banerji *et al.*, 1993] and itself used to clone the complete *arom* locus of this fungus. The deduced protein sequence shows good similarity to the AROM proteins from other fungi and the corresponding shikimate enzymes of *E.coli.*

1.8 Organisation of the shikimate pathway in bacteria and plants

Berlyn & Giles [1969] showed that the enzymes of the shikimate pathway are not aggregated in bacteria (*Escherichia coli, Salmonella typhimurium, Klebsiella pneumoniae, Bacillus subtilis, Pseudomonas aeruginosa* and *Streptomyces coelicolor*), reporting Mr's for the dehydroquinase enzymes as low as 40,000 (*E.coli, S.typhimurium* and *B.subtilis*), with others in excess of 130,000 (*S.coelicolor* and *P.aeruginosa.*). This organisation of separable activities was also identified in *K.pneumoniae* [Berlyn & Giles, 1973].

In plants, there appears to be two copies of the shikimate pathway; one in the plastids (chloroplasts) and the other in the cytosol. Evidence for such a theory has been provided by purification of separate bifunctional dehydroquinase/shikimate dehydrogenase isozymes from *P.sativum* (pea) [Mousdale *et al.*, 1987], EPSP synthase isozymes from *P.hybrida* (petunia) and *L.esculentum* (tomato) [Gasser *et al.*, 1988], and DAHP synthase isozymes from *N.silvestris* [Ganson & Jensen, 1988] and *V.radiata* (mung bean) [Rubin *et al.*, 1982].

In plants, five of the shikimate enzymes exist as separate proteins, but the dehydroquinase and shikimate dehydrogenase activities appear, in all plants studied, to occur in a bifunctional protein, which is active as a monomer with an Mr of ~59,000. The species in which this bifunctional protein has been identified include mosses (*Physcomitrella patens* [Polley, 1978] and *Phaseolus mungo* [Koshiba, 1978]), spinach (*Spinacia oleracea*) and pea (*Pisum sativum* [Mousdale *et al.*, 1987]). This bifunctional activity also appears to exist in certain blue-green algae, including *Chlamydamonas reinhardi* [Berlyn *et al.*, 1970]. It is equivalent to the two *E.coli* enzymes (29 kDa and 32 kDa) and to a proteolytic fragment of the AROM protein

containing only these two activities (63 kDa), and is likely to have evolved by gene fusion.

In the bifunctional enzyme from *Pisum sativum*, the dehydroquinase activity occurs in the N-terminal part of the protein, with the shikimate dehydrogenase domain lying in the C-terminus [Deka *et al.*, 1994]. Thus, this bifunctional enzyme is homologous to, and shows the same organisation as, the smaller of the two polypeptides produced by limited proteolysis of the AROM polypeptide of *A.nidulans* [Smith & Coggins, 1983]. The bifunctional protein may provide some form of channelling, limiting the flux of DHS into the quinate pathway, which is constitutive in plants.

1.9 The enzymes of the shikimate pathway

1.9.1 DAHP synthase

3-deoxy-D-arabino-heptulosonate 7-phosphate synthase [EC 4.1.2.15] (phospho-2-dehydro-3-deoxyheptonate aldolase) is the first enzyme of the shikimate pathway, and as such is suggested to be a likely target for pathway regulation. It catalyses the condensation of D-erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) to form 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP).

The actinomycete *Nocardia mediterranei* uses an early intermediate of the shikimate pathway as a precursor for the synthesis of the antibiotic rifamycin. Its DAHP synthase is inhibited by E4P, which acts as a competitive inhibitor with PEP for the active site [Gygax *et al.*, 1982]. The enzyme, however, shows little or no feedback inhibition by the aromatic amino acids. In the *Streptomyces* species 3022a, DAHP synthase activity is not regulated by any intermediates or end-products of the shikimate pathway, but is inhibited partially by *p*-hydroxybenzoate [Lowe & Westlake, 1971]. It is also inhibited by E4P, but this inhibition is removed by the addition of PEP.

The single form of DAHP synthase from *Streptomyces rimosus* has been purified to 97% homogeneity in this laboratory [Stuart & Hunter, 1993]. It has a subunit Mr of 59,000 and native Mr of 120,000, suggesting that it functions as a dimer. Its activity is apparently inhibited only by tryptophan, which shows cooperative binding to the enzyme. The amino acid sequence of the N-terminus of the DAHP synthase enzyme from *S.rimosus* has been obtained [Walker *et al.*, in press]. An oligonucleotide primer has been designed against this sequence, and the gene encoding this enzyme is currently being cloned and sequenced in this laboratory.

E.coli contains three separate isozymes of DAHP synthase, expression of each being inhibited by one of the aromatic amino acids which acts as an effector for the relevant repressor. The DAHP synthase(tyr) isozyme, encoded by *aroF*, is repressed by TyrR, the tyrosine repressor, which requires either tyrosine or high concentrations (mMolar) of phenylalanine for activity [Brown & Somerville, 1971]. The DAHP synthase(phe) isozyme, encoded by *aroG*, is also repressed by TyrR, but requires both phenylalanine and tyrosine for regulation. The DAHP synthase(trp) isozyme, encoded by *aroH*, is repressed at equivalent levels by the tryptophan repressor, TrpR, which uses tryptophan as its effector [Ogino *et al.*, 1982], and by the tyrosine repressor. The region responsible for regulation of the *aroH* gene contains both a Tyr box (at which TyrR binds), lying 38 bp upstream of the binding site for the Trp repressor, suggesting that the two repressors jointly regulate the expression of this isozyme [Muday *et al.*, 1991]. Therefore, regulation of the DAHP synthase isozymes is by feedback inhibition, with the end-products repressing expression of these isozymes and also inhibiting their activity. Expression of these three isozymes is implicated in controlling the flux through the shikimate pathway.

The plant *Nicotiana silvestris* contains two isozymes of DAHP synthase: one is functional in the cytosol, uses cobalt (Co) as its cofactor, and is inhibited by glyphosate; the other is located in plastids (including the chloroplast), uses manganese (Mn) as its cofactor, and is insensitive to inhibition by glyphosate [Ganson & Jensen, 1988]. Mung bean (*Vigna radiata*) seedlings also contain two DAHP synthase isozymes: the glyphosate-inhibitable Co-dependent enzyme, and the glyphosate-resistant Mn-dependent isozyme [Rubin *et al.*, 1982]. In the presence of glyphosate, the Mn-dependent isozyme is able to synthesise DAHP from E4P and PEP, and this DAHP is then fed into the cytosolic shikimate pathway as well as the plastid-located one. However, as both pathways contain a glyphosate-sensitive EPSP synthase, the shikimate pathway is still blocked, causing the accumulation of shikimate.

DAHP synthase has been purified from *B.subtilis*, and occurs as part of a bifunctional polypeptide which also contains the chorismate mutase activity [Huang *et al.*, 1975]. This polypeptide is found to be associated with a small (approx. 10 kDa) polypeptide conferring shikimate kinase activity. The shikimate kinase enzyme only has activity when it forms part of this trifunctional enzyme complex, and its presence in the complex increases the activity of the other two enzymes by up to 50%.

1.9.2 Dehydroquinate synthase

Dehydroquinate synthase [EC 4.6.1.3] (7-phospho-3-deoxy-D-arabino-heptulosonate

phosphate lyase) catalyses the conversion of DAHP to dehydroquinate (DHQ), and is the first catalysis performed by the AROM pentafunctional polypeptide. In *E.coli*, DHQ synthase is encoded by the gene *aroB*, which has been cloned [Duncan & Coggins, 1984] and sequenced, and is predicted to encode a protein of Mr 39,000 [Millar & Coggins, 1986].

The DHQ synthase domain of the AROM polypeptide from *A.nidulans* has been overexpressed in *E.coli*, complementing an *aroB* mutation [van den Hombergh *et al.*, 1992] providing further evidence that the enzyme activities of AROM are functionally homologous to those of the separable enzymes of bacteria.

DHQ synthase from *B.subtilis* has been purified, and its activity found to be associated with those of the chorismate synthase and flavin reductase enzymes [Hasan & Nester, 1978]. DHQ synthase activity is dependent on the presence of chorismate synthase, but not on the presence of the flavin reductase.

1.9.3 3-Dehydroquinase

Dehydroquinase [EC 4.2.1.10] (3-dehydroquinate dehydratase) catalyses the conversion of dehydroquinate to dehydroshikimate. There are two distinct enzyme forms, type I and type II, which have different mechanisms of action [White *et al.*, 1990]: type I is a biosynthetic enzyme which is active as a dimer, and has a low K_m; type II is a catabolic, degradative enzyme which is active as a dodecamer, and has a considerably higher K_m (650 μM for *S.coelicolor*). The type I enzyme is active in the shikimate pathway of bacteria, plants and lower eukaryotes; the type II enzyme was originally identified as being active in the quinate pathway of lower eukaryotes, although it has also been recently found in the shikimate pathway of certain Gram-positive bacteria and one Gram-negative bacterium (see below).

Polarity mutants of *N.crassa* which were auxotrophic for the aromatic amino acids lack the AROM polypeptide (*arom-2* mutants), and were found to exhibit a high level of dehydroquinase activity when grown on quinate [Giles *et al.*, 1967]. Unlike the constitutive biosynthetic activity which was already known to be associated with the AROM polypeptide, this activity was found to be stable when heated to 71°C, and was induced by the presence of quinate.

When *A.nidulans* is grown on media which do not induce its quinate pathway (and therefore do not induce its type II dehydroquinase), dehydroquinase activity can still be detected. This activity is heat-labile and does not cross-react with antibodies raised

against the *N.crassa* type II enzyme, suggesting that *A.nidulans* has both dehydroquinase enzymes [Kinghorn & Hawkins, 1982]. The type II dehydroquinase gene, *qutE*, is transcribed only when grown on quinate, as it is transcriptionally regulated [Da Silva *et al.*, 1986]. These authors suggested that the levels of the *qutA* gene product are rate-limiting for induction of the quinate pathway.

The sequence encoding the dehydroquinase domain of the *arom* locus of *A.nidulans* was cloned by virtue of its ability to complement an *aroD* mutation in *E.coli* [Kinghorn & Hawkins, 1982]. It has been sequenced [Charles *et al.*, 1985] and its product shows no similarity to that of the *qutE* gene, which encodes the type II enzyme from the same species, suggesting that the two genes evolved separately by convergent evolution, and have not undergone divergent evolution from a common ancestral gene. The dehydroquinase domain has been overexpressed in *E.coli*, and complements an *aroD* mutation [van den Hombergh *et al.*, 1992].

Chaudhuri *et al.* [1986] purified dehydroquinase from *E.coli* to a single band on SDS-PAGE. It had a native Mr of 52,000-58,000 and subunit Mr of 29,000, suggesting that the enzyme functions as a dimer. Its activity is inhibited by $NaBH_4$. The gene encoding this protein, *aroD,* was cloned due to its ability to complement a dehydroquinase mutant [Kinghorn *et al.*, 1981] and found to encode a heat-labile protein. Both the N-terminus of the protein and its gene were sequenced [Duncan *et al.*, 1986; Chaudhuri *et al.*, 1991] and found to have no significant sequence identity to the *qa-2* gene product from *N.crassa* [Berson *et al.*, 1991] which encodes the catabolic dehydroquinase.

The *aroD* gene of *S.typhimurium* was cloned by complementation of an *E.coli aroD* mutation, sequenced, and found to show a high level of sequence similarity to the *E.coli* gene at the DNA level (69%), and at the amino acid level (70%) with the gene product [Servos *et al.*, 1991]. It encodes a protein of estimated subunit size 28 kDa, and codon usage analysis suggests that the *aroD* genes from both *S.typhimurium* and *E.coli* are poorly expressed.

The activity of a type I dehydroquinase is known to be inhibited by chloride anions. As the dehydroquinase activity of the bifunctional protein containing the dehydroquinase and shikimate dehydrogenase domains from pea seedlings (*P.sativum*) is also inhibited by chloride ions, this dehydroquinase activity was identified as a type I enzyme [Mousdale *et al.*, 1987]. However, this bifunctional protein functions as a monomer, unlike typical type I enzymes. It has been purified recently, and has a low K_m for its dehydroquinase activity, which is inactivated by $NaBH_4$ [Deka *et al.*, 1994]. Both of

these results are also indicative of the dehydroquinase portion of this bifunctional protein acting as a type I enzyme. This was confirmed by protein sequencing of the N-terminus and two internal peptides, all of which showed similarity to the type I dehydroquinase enzymes. It also placed the dehydroquinase domain in the N-terminal portion of the protein. The gene encoding the dehydroquinase activity has also been cloned by PCR, and encodes the amino acid sequence anticipated from protein sequencing. Sequencing also confirms that the two residues implicated in the active site of the protein, and conserved among all type I dehydroquinases (Lys-170 and His-143), are also present in this dehydroquinase domain.

Dehydroquinase has been purified to greater than 75% homogeneity from *S.coelicolor*, giving a single band with an estimated Mr of 16,000 on SDS-PAGE [White *et al.*, 1990]. The native Mr is approximately 209,000, so is presumably active as a dodecamer. It is stable at 70°C, has a K_m of 650 μM, and its activity is not inhibited by Cl^- anions. The amino acid sequence of the N-terminus of the protein showed approximately 50% identity to the N-termini of the inducible catabolic dehydroquinases from *N.crassa* and *A.nidulans*. This was the first reported dehydroquinase from the shikimate pathway which resembles the eukaryotic catabolic enzymes, rather than the biosynthetic enzymes. The gene encoding it, *aroQ*, has subsequently been cloned and sequenced in this laboratory [P. White, unpublished results]. *S.coelicolor* was also tested for the presence of the quinate pathway, but is unable to grow on minimal medium with quinate as sole carbon source, and shows no detectable quinate dehydrogenase activity. It will grow on *p*-hydroxybenzoate, which is able to induce the quinate pathway in *K.pneumoniae*, but shows no quinate dehydrogenase activity or elevation of its dehydroquinase activity, suggesting that such growth is due to the presence in *S.coelicolor* of a functional ß-oxoadipate pathway, and not of a quinate pathway. These authors therefore suggested the classification of type I for the biosynthetic enzymes (which occur only in the shikimate pathway), and type II for the catabolic dehydroquinases (found in both the shikimate and quinate pathways).

Garbe *et al.* [1991] cloned the *aroQ* gene from *Mycobacterium tuberculosis* by complementation of an *E.coli aroD* mutation and, due to the similarity of its deduced amino acid sequence to those of the fungal quinate pathway dehydroquinases, classified it as encoding a type II enzyme. They also identified the DHQ synthase gene *aroB* which lies upstream of *aroQ* and overlaps it by 4 bp, suggesting that the two genes are translationally-coupled. Again, this sequence complements its respective *E.coli* mutation. Like *S.coelicolor, M.tuberculosis* is unable to grow on quinate as sole carbon source, so its type II dehydroquinase is likely to function only in the shikimate pathway. When placed under the control of the *groE* promoter, it partially

complements an *A.nidulans aroD* mutation (poor complementation possibly being due to the lack of a functional terminator in *Aspergillus,* or the mutant producing a non-functional protein which could form an inactive heterodimer with the *aroQ* product) [Moore *et al.*, 1992].

The shikimate pathway has also been studied in another actinomycete, *Amycalotopsis methanolica*, which is able to grow on quinate but not shikimate as sole carbon source, although it can metabolise shikimate when grown on either glucose or quinate. This suggests that *A.methanolica* has a functional shikimate pathway, but that it is not induced by shikimate [Euverink *et al.*, 1992]. Apparently only a single dehydroquinase enzyme is present, which is heat-stable and therefore likely to be a type II enzyme, although its activity is not induced by quinate. The dehydroquinase appears to be associated with dehydroquinate synthase, as no single dehydroquinase mutants could be isolated. However, single DHQ synthase mutants do occur, and are able to grow on quinate in the absence of supplementary aromatic amino acids, suggesting that the dehydroquinase enzyme is effective in both biosynthesis of the aromatic amino acids and catabolism of quinate. The dehydroquinase was purified to a single band by SDS-PAGE, and has a subunit Mr of 12,000 and native Mr of 135,000, suggesting that it exists as a dodecamer in its active state. Its activity was not inhibited by $NaBH_4$ and it had a higher K_m than is typical for type I dehydroquinases, supporting its identification as a type II enzyme. Its first 20 N-terminal amino acids have been sequenced and show 45% similarity to the type II enzymes from *N.crassa, A.nidulans* and *S.coelicolor*, and 75% to that of *M.tuberculosis.*

Another gene entitled *aroQ*, and encoding a type II dehydroquinase, has recently been identified in a Gram-negative bacterium, *Actinobacillus pleuropneumoniae* [Lalonde *et al.*, 1994], due to its ability to complement an *aroD* mutation in *E.coli.* Its deduced amino acid sequence shows similarity to the type II enzymes of *N.crassa, A.nidulans* and *M.tuberculosis* (44-62%), and not to that of *E.coli. A.pleuropneumoniae* is able to grow in the absence of tryptophan, suggesting the presence of a pathway leading to the synthesis of chorismate, but is unable to grow on quinate, implying that this type II dehydroquinase also functions solely in the shikimate pathway.

The *qutE* gene of *A.nidulans* weakly complements an *E.coli aroD* mutation [Hawkins *et al.,* 1985], but fully complements such *E.coli* mutations when placed under the control of the *E.coli trc* promoter, showing that a fungal catabolic dehydroquinase is capable of functioning as part of the shikimate pathway [Hawkins & Smith, 1991]. The *N.crassa qa-2* gene, encoding its type II dehydroquinase, has been subcloned under the control of the P_{tac} promoter and expressed in *E.coli*; it complements an *aroD*

mutation, supporting the theory that type II dehydroquinases can function in the shikimate pathway [Hawkins *et al.*, 1993c].

The predicted protein sequences of the type I dehydroquinases from *E.coli* and *A.nidulans*, and of the type II enzymes from *A.nidulans* and *N.crassa,* show conserved sequence motifs which may be involved in the active site or its stabilisation with the substrate/product [Hawkins, 1987].

Kleanthous *et al.* [1992] compared the type I (from *E.coli*) and type II (from *A.nidulans*) enzymes. The type I enzyme consists of ~50% α-helix and 25% ß-sheet; type II is ~75% α-helix. The type II enzyme shows four-fold greater resistance to denaturation by guanidine hydrochloride, and denatures at between 82°C and 95°C, whereas the type I enzyme denatures at 57°C. The type II enzyme has a lower activity at physiological conditions (relatively low flux rates) and has a K_m an order of magnitude higher than that of the type I enzyme. The pH dependence of V_{max} is consistent with the type I having a single ionising group, whereas the complex profile of the type II enzyme suggests that it has several ionising groups. Both enzymes are inactivated by DEPC (diethyl pyrocarbonate) which modifies histidine residues, and both are protected from it by incubation in a substrate/product mixture, implying that the DEPC modifies residues implicated in the active site. Unlike type I, type II is not inactivated by $NaBH_4$, and has no conserved lysine, implying that the type II dehydroquinases do not function via a Schiff's base intermediate. The type I enzymes produce the first double bond in the aromatic ring by *syn* elimination of water, whereas type II enzymes abstract water in *trans* [Harris *et al.*, 1993].

Iodoacetate binds to dehydroquinase, inhibiting the enzyme activity. Radioactively-labelled iodoacetate labels two methionine residues in the *E.coli* dehydroquinase enzyme; one at the N-terminus, the other at the C-terminus [Kleanthous *et al.*, 1990a]. The label is incorporated equally at both sites. The C-terminal methionine (Met-205) is conserved in the type I dehydroquinase enzymes of both *S.cerevisiae* and *A.nidulans*. The N-terminal methionine is not conserved, but residues surrounding it are conserved in both type I and type II dehydroquinases and DHQ synthase enzymes, suggesting that the N-terminal methionine residue (Met-23) is involved in the binding site for DHQ and not the catalytically-active site of the enzyme. The alkylation of the methionine residues by iodoacetate is reversed by incubation with mercaptoethanol, resulting in a 50% reduction in labelling and an increase in enzyme activity from 12% to 50% [Kleanthous *et al.*, 1990b]. The label is only lost from the C-terminal methionine residue, the differential reversibility of alkylation being due presumably to the residues occurring in differing microenvironments.

Incubation of type I dehydroquinase with diethyl pyrocarbonate (DEPC) causes carbethoxylation of six histidine residues, and therefore a loss of enzyme activity, although modification of a single histidine residue appears to be responsible for inactivation [Deka *et al.*, 1992]. Preincubation of the enzyme with DHQ protects against inactivation, suggesting that DEPC affects the active site of the enzyme. Treatment of DEPC-inactivated dehydroquinase with hydroxylamine reverses the carbethoxylation. A modified seven-residue peptide of the *E.coli* dehydroquinase containing two histidines has been isolated, although only one of these residues (His-143) is conserved in type I dehydroquinases.

When a type I enzyme interacts with its substrate (3-DHQ), an imidazole side chain of a histidine residue facilitates abstraction of the *pro-R* proton (hydrogen) from C-2 of the aromatic ring. The essential histidine residue His-143 appears to be a good candidate for this step of the reaction mechanism [Deka *et al.*, 1992]. The hydroxyl group at C-1 is eliminated, producing a carbonium intermediate and allowing the N^6-amino group of a conserved lysine side chain to form a Schiff's base (imine) intermediate with the keto group of dehydroquinate [Kleanthous *et al.*, 1992].

$NaBH_4$ reduces the active-site imine to a stable secondary amine, inactivating the enzyme by trapping dehydroquinate at the active site. This allows labelling of the active site with NaB^3H_4, followed by CNBr cleavage and purification of the radioactively-labelled band. A lysine residue was found to be labelled in both the AROM polypeptide of *N.crassa* and the dehydroquinase of *E.coli* [Chaudhuri *et al.*, 1991]. However, no lysine residue was identified in an appropriate position in the type II enzyme from *S.coelicolor* [P.White, unpublished results]. The imine intermediate has been shown to occur during this reaction using electrospray mass spectrometry [Shneier *et al.*, 1991].

DHQ has been labelled with deuterium at both its C-2 *pro-R* and *pro-S* hydrogens (producing [2-^{2}H]DHQ). Labelled product of the dehydroquinase reaction (DHS) was reduced *in situ* to the more stable intermediate shikimate, with its deuteration being measured by ^{1}H NMR spectroscopy. Type I dehydroquinase removes the *pro-R* hydrogen in a *syn* elimination as has been previously been suggested. However, the type II enzyme eliminates the *pro-S* hydrogen in *trans* [Shneier *et al.*, 1993; Harris *et al.*, 1993]. Thus, the two dehydroquinase isozymes proceed with opposite stereochemistries to achieve the same reaction, evidence of them being distinct proteins which have undergone convergent evolution to produce the same catalytic activity.

Aquaspirillum magnetotacticum, a Gram-negative freshwater bacterium, contains a gene which complements *aroD* mutations in both *E.coli* and *S.typhimurium*. However,

this gene appears not to be homologous to the *E.coli aroD* gene or to the *qa-2* gene which encodes the type II dehydroquinase in *N.crassa* [Berson *et al.*, 1991]. It is therefore possible that this sequence encodes a novel type of dehydroquinase enzyme.

1.9.4 Shikimate dehydrogenase

Shikimate dehydrogenase [EC 1.1.1.25] (shikimate oxidoreductase) converts DHS to shikimate. The enzyme utilised in the shikimate pathway is an $NADP^+$-dependent enzyme, so differs from the quinate/shikimate dehydrogenase enzyme involved in the quinate pathway, which is an NAD^+- or PQQ-dependent enzyme.

In *E.coli*, this enzyme is encoded by the *aroE* gene, which has been subcloned from a lambda bacteriophage library [Anton & Coggins, 1984]. This gene has been sequenced, and placed under the control of the IPTG-inducible P_{tac} promoter [de Boer *et al.*, 1983] for overexpression and purification in *E.coli* [Anton & Coggins, 1988]. It produces a single protein band whose 60 N-terminal amino acids have been sequenced and agree with the amino acid sequence deduced from the nucleotide sequence of the gene. The predicted protein sequence shows similarity to the shikimate dehydrogenase domains from the AROM polypeptides from *A.nidulans* and *S.cerevisiae*, and also to the quinate/shikimate dehydrogenase enzyme from the quinate pathway in *N.crassa*.

Like dehydroquinases, the activity of the shikimate dehydrogenase domain (the C-terminal domain of the dehydroquinase:shikimate dehydrogenase bifunctional protein) from *Pisum sativum* is inactivated by treatment with DEPC [Deka *et al.*, 1994], implying that a histidine residue lies at, or near to, the active site.

1.9.5 Shikimate kinase

Shikimate kinase [EC 2.7.1.71] (ATP:shikimate 3-phosphotransferase) phosphorylates shikimate to produce shikimate 3-phosphate. *E.coli* contains two functional shikimate kinase enzymes, with shikimate kinase II having a K_m two orders of magnitude lower than that of shikimate kinase I. Shikimate kinase II, which functions as a monomer and has an Mr of 17,500, is encoded by *aroL*, which lies upstream of and forms an operon with *aroM*, whose function is unknown [de Feyter *et al.*, 1986]. *aroL* mutants are not auxotrophic, presumably due to the presence of the shikimate kinase I enzyme, but the presence of a functional shikimate kinase II enzyme does result in an increased growth rate in *aroF/aroG* mutants, which facilitated cloning of *aroL* by complementation of *aroF/aroG/aroL* triple mutations [de Feyter & Pittard, 1986].

Like the *aroF* and *aroG* genes (which encode two of the DAHP synthase isozymes of *E.coli*), this operon appears to be regulated by the TyrR repressor, its promoter containing three operator sites [Heatwole & Somerville, 1992]. Tryptophan and tyrosine inhibit the activity of shikimate kinase, but not in a $TyrR^-$ background, suggesting that the TyrR repressor must interact with either tyrosine or tryptophan to regulate the expression of shikimate kinase. The TyrR repressor has been shown to bind tyrosine *in vivo*.

TrpR, the repressor regulated by tryptophan, binds to a 24 bp palindromic sequence, interfering with the ability of RNA polymerase to mediate transcription from promoters which contain this palindrome. TrpR, which regulates transcription of *aroH* (which encodes the third DAHP synthase isozyme), and whose binding sites have also been identified in the *aroL* promoter, has been shown to bind to P_{aroL} in a tryptophan-dependent manner. However, the repression of *aroL* mediated by TrpR only occurs in the presence of TyrR. Repression due to TyrR is increased by the action of TrpR, indicating that the two repressors exhibit a form of synergism. It has been proposed that TrpR is able to bind to its "strong" box, but can only bind to its "weak" box when TyrR is already bound to the promoter. TrpR bound at the weak box is best positioned to inhibit binding of RNA polymerase to the promoter. It is also possible that, if P_{aroL} is a strong promoter, repression by TyrR is required for any repression due to TrpR to be effective.

Shikimate kinase I is encoded by *aroK*, and lies upstream of *aroB* in an operon [Løbner-Olesen & Marinus, 1992]. Its predicted protein sequence shows similarity to a region of shikimate kinase II which includes the ATP-binding site. *aroK* was confirmed as encoding a shikimate kinase by its ability to complement an *aroK/aroL* mutant, producing growth in the presence of shikimate and the absence of the aromatic amino acids. Although the transcription of *aroL* is increased by a knock-out mutation in TyrR, such a mutation has only a minimal effect on transcription of *aroK*, confirming that it is not regulated by TyrR.

1.9.6 EPSP synthase

5-enolpyruvylshikimate 3-phosphate synthase [EC 2.5.1.19] (3-phosphoshikimate 1-carboxyvinyltransferase or 5-O-{1-carboxyvinyl} transferase) catalyses the addition of the enolpyruvyl moiety of phosphoenolpyruvate (PEP) to shikimate 3-phosphate (S3P), producing EPSP. This is the final activity included in the pentafunctional AROM polypeptide.

The broad-spectrum, post-emergence herbicide glyphosate (N-phosphonomethyl-glycine; its monoisopropylamine salt is sold commercially as "Roundup") is known to inhibit the shikimate pathway in plants and bacteria. Although it may inhibit certain DAHP synthase and DHQ synthase activities (for example in the yeast *Candida albicans* [Bode *et al.*, 1984]), glyphosate mainly inhibits the activity of EPSP synthase, and also at much lower concentrations than required to inhibit DAHP synthase or DHQ synthase. Therefore, EPSP synthase has been suggested as the target for glyphosate.

Glyphosate causes accumulation of shikimate, and prevents the conversion of shikimate 3-phosphate to anthranilate in *Klebsiella pneumoniae* [Steinrücken & Amrhein, 1980]. When *K.pneumoniae* was grown in a glyphosate-containing medium, tolerance to glyphosate was correlated with an increase in the specific activity of the EPSP synthase (up to ten-fold increase), although shikimate and S3P were still excreted, suggesting that resistance was achieved by overexpression of the enzyme [Amrhein *et al.*, 1983]. Inhibition of EPSP synthase by glyphosate in *K.pneumoniae* was reversed by addition of the aromatic amino acids. Growth of cultured plant cells from *Corydalis sempervirens* in a glyphosate-containing medium produced a thirty-fold increase in EPSP synthase in glyphosate-tolerant cells which still showed a build-up of shikimate. It has been proposed that shikimate 3-phosphate, which is expected to build up in cells inhibited by glyphosate, is then dephosphorylated to shikimate, which does accumulate.

Glyphosate has been shown to be a competitive inhibitor for the PEP binding site of the EPSP synthase domain of the AROM polypeptide from *N.crassa*, and binds to an EPSP synthase:S3P complex [Boocock & Coggins, 1983]. Glyphosate is a non-competitive inhibitor of EPSP synthase for S3P.

The *E.coli* enzyme has been purified to homogeneity [Lewendon & Coggins, 1983]. It has a subunit Mr of 49,000, and functions as a monomer. The gene encoding EPSP synthase, *aroA*, has been isolated by complementation of an *E.coli aroA* mutation [Duncan & Coggins, 1984]. The protein has been overexpressed and purified, with overexpressing cells found to exhibit resistance to glyphosate at up to seventeen times the level which kills non-overexpressing bacteria [Duncan *et al.*, 1984a]. It appears that overexpression titrates out the glyphosate. The *aroA* gene has also been sequenced, and found to encode a protein with an estimated Mr of 46,112. The N-terminal sequence of the protein shows identity to the predicted protein sequence [Duncan *et al.*, 1984b].

gene, which encodes the enzyme 3-phosphoserine aminotransferase, part of the serine biosynthetic pathway [Duncan & Coggins, 1986]. Serine and chorismate are required in equimolar concentrations for the synthesis of enterochelin, an ionophore which is required for uptake of iron from the environment. Thus, it is possible that having a gene from the serine and chorismate biosynthetic pathways in the same operon allows coordinate expression of these two pathways, and regulation of enterochelin biosynthesis during iron starvation. This operon has also been shown to be under the positive control of cyclic AMP [Lim *et al.*, 1994].

The same two-gene operon has been found in *S.typhimurium* [Hoiseth & Stocker, 1985], *Yersinia enterocolitica* [O'Gaora *et al.*, 1989], *Salmonella gallinarum* (cause of fowl typhoid) [Griffin & Griffin, 1991] and *Klebsiella pneumoniae* [Sost & Amrhein, 1990] with *serC* lying upstream of *aroA* in each operon. The operons in *S.gallinarum* and *E.coli* both contain putative Rho-independent terminators in the intergenic region between *serC* and *aroA*, although their functions are unknown. The operon in *Y.enterocolitica* does not contain such an identifiable terminator sequence. Although this organisation of the operon has not been shown to exist in *S.typhi*, its *aroA* gene has been cloned and sequenced, and its predicted protein sequence shows 97·7% similarity to that of *S.typhimurium* [Chatfield *et al.*, 1990].

The *aroA* gene of *M.tuberculosis* has been cloned by complementation of an *E.coli aroA* mutation, and encodes a 53 kDa protein as shown by Western blots [Garbe *et al.*, 1990]. The *M.tuberculosis aroA* also hybridises to *aroA* from *B.pertussis*, which has been cloned [Maskell *et al.*, 1988], produces a protein of deduced Mr of 46,688, and shows good similarity to the deduced *E.coli* EPSP synthase sequence. The EPSP synthase domain of the gene encoding the AROM polypeptide from *P.carinii* has also been cloned [Banerji *et al.*, 1993]. EPSP synthase has also been purified from seedlings of *Pisum sativum* (pea), has a K_m comparable to that of the *E.coli* enzyme (50 kDa) [Mousdale & Coggins, 1984], and is also inhibited by glyphosate.

The *aroA* gene from *Petunia hybrida* has been cloned from a cDNA library and sequenced, and this clone has been used to isolate the *aroA* gene of *Lycopersicon esculentum* (tomato) from a cDNA library [Gasser *et al.*, 1988]. The *aroA* cDNA of *L.esculentum* has been expressed in *E.coli*, and produces an active protein. Petunia and tomato are both members of the *Solanaceae* family and their EPSP synthases show 93% identity at the amino acid level. Both species also have a second putative EPSP synthase-encoding sequence which hybridises weakly to their isolated cDNA. These sequences have been cloned but may be pseudogenes, or genes from the cytoplasmic shikimate pathway which have diverged from the chloroplast genes.

Unusually, the EPSP synthases of certain bacteria from the genus *Pseudomonas* are not sensitive to inhibition by glyphosate, and such strains do not excrete either shikimate or shikimate-3-phosphate in the presence of glyphosate [Schulz *et al.*, 1985]. EPSP synthase from the cyanobacterium *Anabaena variabilis* also shows reduced inhibition by glyphosate [Powell *et al.*, 1992]. However, the purified enzyme has only a slightly elevated K_m for PEP, unlike the mutant EPSP synthases which have greatly elevated K_m's for PEP. Therefore, the *A.variabilis* enzyme is likely to be more useful than the mutant bacterial enzymes for the production of glyphosate-insensitive transgenic plants, particularly as cyanobacteria are believed to share a common ancestry with chloroplasts from higher plants.

A glycine to alanine substitution (G96A) appears to have no significant effect on the activity of EPSP synthase from a variety of species (Petunia, Soyabean, Maize, Arabidopsis, *E.coli*) but confers a five hundred-fold increase in the tolerance to glyphosate, suggesting that the mutation interferes with binding of glyphosate [Padgette *et al.*, 1991]. This mutation also interferes with binding of PEP to the enzyme, but has no effect on the binding of S3P or EPSP. The naturally-occurring mutant *aroA* gene from *K.pneumoniae* also contains this substitution, and encodes a glyphosate-resistant EPSP synthase [Sost & Amrhein, 1990]. These authors suggest that a glycine at this position is likely to produce a ß-turn in the protein, but an alanine is likely to favour a continuation of the α-helix. Therefore, the substitution may alter the local protein configuration, and prevent glyphosate binding. The proline to serine substitution (P101S) also reduces binding of glyphosate in Petunia. Introducing the G96A mutation into a P101S mutant causes a further decrease in binding of PEP and glyphosate, and by an amount equivalent to that caused by the introduction of this mutation into a wild-type enzyme, suggesting that the two mutations affect different interactions between glyphosate/PEP and EPSP synthase.

A mutant *aroA* gene from *S.typhimurium*, encoding an EPSP synthase containing a proline to serine substitution (P101S), causes resistance to glyphosate, but has no effect on the enzyme kinetics [Stalker *et al.*, 1985]. The mutation lies within a region which is highly conserved between the *aroA* genes of *S.typhimurium* and *E.coli*. The mutant protein is still able to complement an *E.coli aroA* mutation, and also confers resistance to glyphosate in *E.coli* [Stallings *et al.*, 1991]. Expression of this gene from a Ti-based vector in *Agrobacterium tumefaciens*, and its subsequent infection in tobacco plants, confers considerable resistance to glyphosate in the plant, with levels of tolerance to glyphosate which correlate closely with expression levels [Comai *et al.*, 1985].

Treatment of *E.coli* EPSP synthase with DEPC reduces enzyme activity to around 1%, although the enzyme may be protected by preincubation with S3P, PEP or EPSP, particularly when the two substrates (S3P and PEP) are added together [Huynh, 1987]. Inactivation by DEPC is reversible by hydroxylamine, so inactivation must be due to modification of a histidine residue. Four of the eight His residues in EPSP synthase must be modified for complete inactivation, although only one is essential for activity. The modification does not affect binding of substrate or glyphosate. It has been proposed that the essential histidine is located very close to or within the active site and may be involved in catalysis.

1-ethyl-3-(3-dimethylamino-propyl)carbodiimine (EDC) amidates carboxyl groups and modifies four carboxyl groups in EPSP synthase of *E.coli* causing inactivation, although only one is essential for activity [Huynh, 1988]. Modification reduces binding of glyphosate, and may be prevented by preincubation of the enzyme with S3P and glyphosate. It has been shown that the modified residue is Glu-418, which is conserved amongst several EPSP synthases, and possibly lies in the glyphosate-binding site. The negatively-charged carboxyl group of Glu-418 has been proposed as interacting with the positively-charged imino group of glyphosate. This glutamate residue may be involved directly in catalysis or may interact with the essential His-385 residue.

Pyridoxal 5'-phosphate (PALP) modifies lysine residues, and inactivates the EPSP synthase from *E.coli*, with the enzyme being protected from modification by preincubation with S3P, EPSP or S3P+glyphosate [Huynh *et al.*, 1988a]. The modified residue is Lys-22, which may undergo conservative substitution by arginine (which has the same cationic group as lysine) to retain enzyme activity, but replacement with alanine or glycine inactivates the enzyme. The positively-charged amino group of Lys-22 is proposed to be critical in binding of the substrate. The cationic group of Lys-23 of EPSP synthase from *Petunia hybrida* (which is equivalent to Lys-22 from the *E.coli* enzyme) has been shown to be required for binding of the substrate [Huynh *et al.*, 1988b]. These authors propose that this residue is involved in the binding of S3P.

EPSP synthase from *P.hybrida* is inactivated by phenylglyoxal (PGO) and *p*-hydroxy(PGO) which modify three arginine residues, including Arg-28 and Arg-131 [Padgette *et al.*, 1988b]. Incubation with S3P + glyphosate protects the enzyme against inactivation. In *K.pneumoniae*, glyphosate alone appears to protect against inactivation by PGO (although it is possible that either the glyphosate or enzyme used was contaminated with S3P) [Steinrücken & Amrhein, 1984].

DTNB (5,5'-dithiobis-{2-nitrobenzoic acid}) modifies cysteine residues, but only reacts with two of the six cysteines in the EPSP synthase isolated from *E.coli*. In the presence of S3P and glyphosate, only one of these cysteines, Cys-408, is modified, suggesting that this residue lies in or near to the active site of the enzyme [Padgette *et al.*, 1988a]. However, such modification does not inactivate the enzyme, implying that this cysteine is not required for substrate binding or catalysis, and presumably lies close to the ligand binding-site.

It has been shown that, initially, S3P binds to EPSP synthase, forming a complex to which PEP subsequently binds [Anderson *et al.*, 1988], confirming earlier results [Boocock & Coggins, 1983]. PEP does not bind to EPSP synthase in the absence of S3P. The negative charges of both the carboxylate and phosphate moieties of PEP are required for its binding to the enzyme, EPSP synthase showing a high degree of ionic and structural specificity for both PEP and glyphosate [Walker *et al.*, 1991]. It has been proposed that the 5' hydroxyl group of S3P is responsible for a nucleophilic attack on the C-2 of PEP to form a tetrahedral intermediate. This addition-elimination reaction proceeds with the phosphate group of the tetrahedral intermediate becoming protonated, resulting in the production of inorganic phosphate and EPSP. The inorganic phosphate is released from the enzyme first, followed by EPSP.

Anderson *et al.* [1988] have proposed that the binding of glyphosate+S3P at the active site may mimic the binding of the tetrahedral intermediate to the enzyme. However, these authors could find no evidence to support the existence of an enolpyruvyl-EPSP synthase intermediate as proposed by Anton *et al.* [1983]. The existence of the tetrahedral intermediate has been proven by its isolation [Anderson & Sikorski, 1988]; it was shown to contain the shikimate ring from S3P and both the phosphate and enolpyruvyl groups donated by PEP.

An alternative hypothesis was proposed by Anton *et al.* [1983], who showed that purified EPSP synthase of *K.pneumoniae* is able to bind PEP in the absence of S3P, forming an enzyme-enolpyruvyl intermediate. These authors postulated that, in the presence of S3P, a proton is incorporated into the ß-carbon of PEP, allowing release from the enzyme-substrate intermediate of inorganic phosphate, with the protonation of PEP being inhibited by glyphosate. They also proposed that a negatively-charged group from the active site stabilises the carbonium anion of the protonated intermediate, which releases phosphate.

The EPSP synthase from *E.coli* has been crystallised at 3Å resolution [Stallings *et al.*, 1991] and shown to consist of two globular, hemispherical domains with both termini

occurring in the lower domain. Each domain has a three-fold axis of symmetry, producing three subdomains.

As the shikimate pathway is known not to occur in mammals, it has been suggested that the shikimate pathway enzymes may be targets for anti-infective agents. Attenuated mutants of such infectious organisms, caused by mutations in the shikimate pathway enzymes, could also be used as non-virulent live vaccines since, to grow, they would require external *p*-aminobenzoate and dihydroxybenzoate which are not readily available in vertebrate tissue [Griffin & Griffin, 1991]. The *aroA* gene of a virulent *S.typhimurium* strain has been disrupted by Tn*10* transposon mutagenesis, producing auxotrophy [Hoiseth & Stocker, 1981]. Non-revertible mutants, caused by deletion or deletion-inversion events in the transposon, have been isolated and injected into mice, and were found to be non-virulent. Such mice were also found to be resistant to subsequent infection by virulent *S.typhimurium*, suggesting that this *aroA* mutant does function as a live vaccine.

1.9.7 Chorismate synthase

Chorismate synthase (EPSP phosphorylase) [EC 4.6.1.4] is the final enzyme of the shikimate pathway, and does not occur as part of the AROM polypeptide. In *E.coli*, it is encoded by the gene *aroC*, which has been sequenced, overexpressed and purified [White *et al.*, 1988]. It produces a protein of predicted subunit Mr of 40,000, although the native Mr is estimated at 144,000, suggesting that it functions as a tetramer. In *N.crassa*, the subunit Mr of chorismate synthase is predicted to be 50,000, with the native Mr approximately 198,000, so it is also likely to function as a tetramer. Sequences of peptides of the *N.crassa* chorismate synthase show similarity to the predicted amino acid sequence of the *E.coli* enzyme. Both enzymes also require a reduced flavin cofactor for enzyme activity, although the *N.crassa* enzyme needs only flavin, as it also carries a diaphorase activity which allows it to reduce flavin itself. The *aroC* gene of *S.typhi* has also been cloned, and complements an *E.coli aroC* mutation [Charles *et al.*, 1990]. The predicted amino acid sequence shows approximately 95% similarity to that of *E.coli.*

The *aroC* gene has also been cloned from the cyanobacterium *Synechocystis* PCC 6803 and its predicted amino acid sequence shows good similarity to the enzymes from *E.coli*, *S.cerevisiae* and the higher plant *Corydalis sempervirens* [Schmidt *et al.*, 1993]. It has been expressed in *E.coli*, and produces a protein which, in Western blots, cross-reacts with antibodies raised against the chorismate synthase protein from *C.sempervirens*, but not with those raised against the enzyme from *E.coli.*

1.10 The Quinate Pathway

Quinate is present in decaying leaves, comprising approximately 10% by weight of decaying leaf litter, and so acts as a natural nutrient source of fungi and other lower eukaryotes. The quinate pathway catalyses the breakdown of quinate to protocatechuate, which may then enter the ß-ketoadipate pathway, leading to cleavage of the aromatic ring, and subsequently to the production of succinate and acetyl-CoA (**Figure 1.4**). Thus, the quinate pathway is a degradative pathway. Two of the intermediates of this pathway, dehydroquinate (DHQ) and dehydroshikimate (DHS), are also involved in the shikimate pathway, and both pathways have a dehydroquinase enzyme which catalyses the conversion of DHQ to DHS. Although the dehydroquinase enzyme involved in the shikimate pathway may be either a type I or type II enzyme, only the type II enzyme has been found in the quinate pathway.

In *A.nidulans*, which plays a role in the woodland decay cycle, the three structural enzymes in the quinate pathway, quinate dehydrogenase, dehydroquinase and DHS dehydratase, are induced by quinate, with similar induction kinetics [Hawkins *et al.*, 1982]. The dehydroquinase has been purified, its K_m calculated at 500 µM, has a subunit Mr of 18-20,000, and is thermostable at 71°C, typical of a type II enzyme. It is immunoprecipitated by antibodies raised against the *N.crassa* type II dehydroquinase.

Using the *N.crassa qa-2, qa-3* and *qa-4* genes as probes, the three structural genes of the quinate pathway in *A.nidulans* (*qutB, qutC, qutE*) have been isolated [Hawkins *et al.*, 1985]. The *qutE* gene, encoding the type II dehydroquinase, complements *E.coli aroD* mutations weakly. The *N.crassa* type II dehydroquinase-encoding gene, *qa-2*, also complements such *E.coli* mutations.

A clone has been isolated which was shown to contain the *qutA, qutB, qutD, qutE* and *qutG* genes [Hawkins *et al.*, 1988]. The predicted amino acid sequence of the *qutD* gene, encoding a quinate-specific permease [Whittington *et al.*, 1987], shows similarity to that of the Qa-Y protein from *N.crassa*, for which no function had previously been assigned. The QutB protein appears to be analogous to the shikimate dehydrogenase domain of the *A.nidulans* AROM polypeptide and the *N.crassa* Qa-3 protein, and the QutG protein appears to be analogous to the Qa-X protein of *N.crassa.*

When DNA from *A.nidulans* was probed with the *qa-1s* and *qa-4* genes of *N.crassa*, each probe identified two hybridising sequences; a strongly hybridising band and a weakly hybridising one (possibly a pseudogene) [Lamb *et al.*, 1990]. The strongly hybridising sequences complemented their respective *N.crassa* mutations (which the

weakly hybridising sequences did not do), showing that these sequences contain the true functional genes. The predicted protein sequence of QutG shows similarity to myo-inositol monophosphatase from bovine brain, suggesting that QutG may dephosphorylate the activator, QutA, allowing induction of the quinate pathway. These authors produced a physical map for the quinate cluster which is in good agreement with the genetic evidence (**Figure 1.5**).

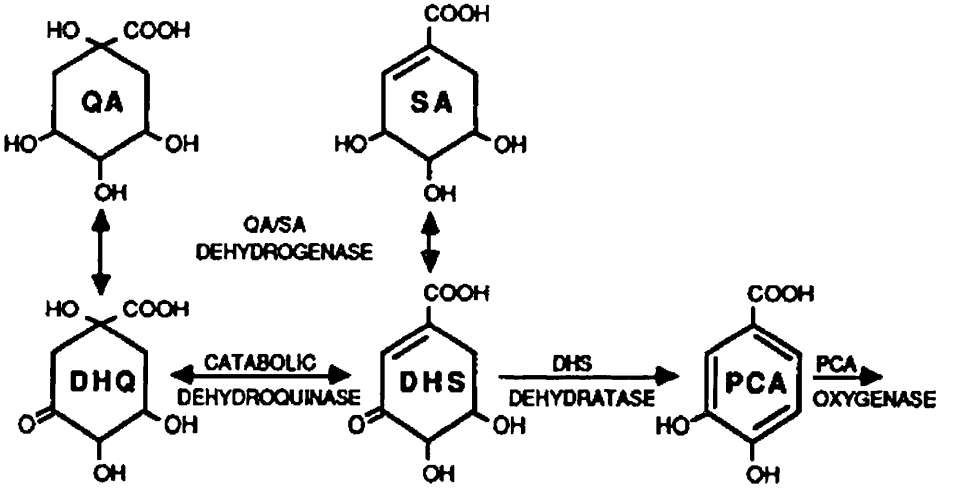


Figure 1.4 The quinate pathway, taken from Geever *et al.*, 1989

N.B. QA denotes quinate
SA denotes shikimate
PCA denotes protocatechuate
DHQ denotes dehydroquinate
DHS denotes dehydroshikimate

The *qutB* gene has been overexpressed in *E.coli* and complements an *aroE* mutation [van den Hombergh *et al.*, 1992], suggesting that the quinate/shikimate dehydrogenase enzyme of the quinate pathway can replace shikimate dehydrogenase from the shikimate pathway, although no quinate/shikimate dehydrogenase activity could be detected in *E.coli*.

The *qutC* gene has been sequenced and shown to be highly similar to the *qa-4* gene of *N.crassa*, which encodes the DHS dehydratase [Lamb *et al.*, 1992a]. Northern blotting has shown that the QutC protein is induced by quinate, and also identified a novel quinate-inducible gene, *qutH*, encoding a 1.45 kb mRNA transcript which is transcribed divergently from *qutC* and lies between the *qutC* and *qutD* genes.

A group of regulatory motifs has been found upstream of each of the structural genes and the permease gene in the quinate pathway, comprising a 22 nt--9 nt--16 nt pattern

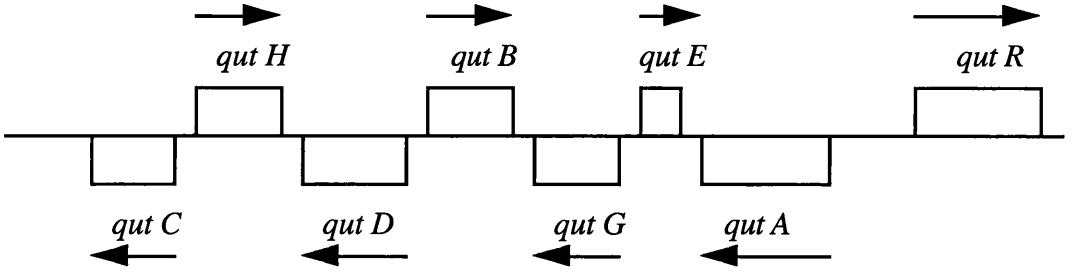


Figure 1.5 Organisation of the quinate pathway of *A.nidulans*

lying between each pair of divergently-transcribed genes, with the *qutA* gene containing a 16 nt variation of the 22 nt motif [Hawkins *et al.*, 1988]. The 22 nt motif has been proposed as a possible target for the DNA-binding finger of the QutA activator protein, and the 16 nt motif is also implicated in the binding of QutA. The 9 nt motif is possibly involved in binding of RNA polymerase II.

qutH lacks the 22 nt--9 nt--16 nt sequence motif organisation which occurs in the upstream sequences of the other *qut* gene pairs. The QutH protein is predicted to have features characteristic of the positively-acting DNA regulatory proteins in eukaryotes, including a "zinc-cluster" DNA binding motif [Vallee *et al.*, 1991]. It shows some similarity to the DNA polymerase from the hepatitis B virus. QutH has been proposed to be a DNA-binding protein involved in controlling the transcription of genes which may be implicated in the metabolism of protocatechuate.

The quinate pathway of *A.nidulans* is regulated by two genes encoding a repressor and an activator, which were identified by mutations [Hawkins *et al.*, 1984]. It has been proposed that the repressor, QutR, is expressed constitutively, and inhibits transcription of the activator-encoding gene, *qutA*, while quinate induces the pathway by inhibiting the activity of the repressor. *qutA* has been cloned and sequenced [Beri *et al.*, 1987] and shows predicted similarity to the Qa-1F regulatory protein of *N.crassa*, which contains a "zinc cluster" as its DNA binding domain [Vallee *et al.*, 1991], and to other eukaryotic regulatory proteins at their putative DNA binding sites. The QutA protein contains a putative helix-turn-helix motif for binding of DNA.

Overexpression of the QutA, -E and -G proteins causes an elevation in the levels of the type II dehydroquinase, but not of any of the other structural proteins of the quinate pathway. The dehydroquinase appears to be regulated normally by the activator when induced by quinate [Beri *et al.*, 1990]. Deleting the *qutG* gene from the overexpressed fragment has no effect on the levels of the structural enzymes, and overexpression of

QutA alone has no effect on their levels, suggesting that the concentration of activator is not the sole limiting factor in expression of the *qut* cluster, possibly due to the cluster containing a limited number of QutA binding sites which may be totally occupied in the wild-type cell.

The current regulatory hypothesis states that the constitutively-expressed QutR repressor of *A.nidulans* phosphorylates the activator, QutA, inactivating it and therefore causing repression of the pathway. In the presence of quinate, DHS or DHQ, the repressor itself is inhibited, preventing its inactivation of QutA. The phosphorylated activator is dephosphorylated by the QutG protein, allowing induction of the quinate pathway. In the absence of any of the inducers of the pathway, the QutR protein has greater activity than the QutG protein, so that the QutA activator protein is phosphorylated, and so is unable to activate the quinate pathway.

The complete quinate (*qa*) cluster of *N.crassa* has been sequenced, and shows considerable similarity to the *qut* cluster of *A.nidulans* [Geever *et al.*, 1989]. The *qa* cluster consists of divergently-transcribed pairs of genes, with the regulatory genes existing as a pair, although the gene order differs from that of the *qut* cluster. In the *qa* cluster, the unpaired *qa-y* gene lies between the regulatory pair of genes and the structural gene pairs. The *qa* cluster appears to contain only seven quinate-inducible genes, one less than the *qut* cluster, and contains no homologue for the *qutH* gene, which is believed to be involved in the regulation of metabolism of protocatechuate.

1.11 Relationship between the shikimate and quinate pathways

The shikimate pathway and quinate pathway are both primary metabolic pathways, and share two intermediates (DHS and DHQ) and the enzyme which interconverts them (3-dehydroquinase). The AROM polypeptide in the shikimate pathway of *N.crassa* was initially proposed as having a channelling function [Hautala *et al.*, 1975] which would prevent the intermediates from the shikimate pathway from being fluxed into the quinate pathway. This may be expected from the relative K_m's of the dehydroquinases from the shikimate pathway and the quinate pathway (the type II dehydroquinases have a considerably higher K_m than their type I homologues [White *et al.*, 1990]), and may provide a function for the fusing of the single enzymes (which occur in the bacteria, where there is no quinate pathway to compete for DHQ and DHS) to produce the pentafunctional AROM polypeptide.

It has, however, been shown that DHS and DHQ leak from the AROM polypeptide of

A.nidulans at a rate comparable to the rate of flux catalysed by the AROM protein, suggesting that, if any channelling occurs, it will only be at a very low level and would be significant only under conditions of very poor availability of nutrients [Lamb *et al.*, 1991]. However, growth under these conditions is quite plausible in the natural environment. *A.nidulans* does not express the quinate pathway enzymes constitutively, indicating that the steady-state levels of DHS and DHQ which leak from the AROM polypeptide are not sufficient for the induction of the enzymes of the quinate pathway.

Supplementation of *A.nidulans* with the aromatic amino acids has no effect on the levels of the type I dehydroquinase or shikimate dehydrogenase activities (i.e. has no effect on the level of the AROM protein) [Lamb *et al.*, 1992b]. Overexpression of QutC (the DHS dehydratase from the quinate pathway) causes an increase in flux through the quinate pathway, and so a drain on the levels of DHS and DHQ available for the shikimate pathway. This produces an increase in the level of shikimate dehydrogenase when grown on MM+glycerol with respect to its level on MM+glycerol supplemented with the aromatic amino acids. This increase is believed to be a response to the limitations imposed on the shikimate pathway. The level of overexpression of DHS dehydratase is correlated with an increase in shikimate dehydrogenase activity (and thus of the other activities performed by the AROM polypeptide). However, if the DHS dehydratase is overexpressed to too high a level, the drain on the resources available to the shikimate pathway is too great, and these cells show extremely poor growth in the absence of aromatic amino acid supplements.

aroB mutants of *A.nidulans* grown on MM+glycerol in the absence of the aromatic amino acids show a two-fold increase in AROM enzyme levels relative to their levels in the presence of these supplements, supporting the hypothesis that a decrease in aromatic amino acid concentrations leads to a starvation response and a subsequent increase in the level of the AROM protein. It is believed that the pentafunctional AROM protein allows only a low-level channelling function which has no effect under steady-state growth conditions, but is significant under physiological conditions of nutrient limitation, ensuring an essential minimal flux through the shikimate pathway.

The Qa-1S repressor from the quinate pathway of *N.crassa* shows similarity to the three C-terminal domains of the AROM polypeptide (dehydroquinase, shikimate dehydrogenase and shikimate kinase activities) and to these enzymes from *E.coli* [Anton *et al.*, 1987]. These authors proposed that the Qa-1S protein has evolved by gene duplication and adaptation from part of the AROM protein, and that the shikimate kinase domain may have evolved into a novel protein kinase domain in the Qa-1S repressor, which catalyses phosphorylation of the Qa-1F activator protein.

A 20-amino acid region of the shikimate kinase enzymes shows approximately 50% similarity to the dehydroquinase and DHS synthase enzymes, and also shows significant similarity to the Qa-1S enzyme of *N.crassa*, with this similarity being conserved across several species [Bugg *et al.*, 1991]. This suggests that this region is involved in substrate binding, possibly binding of the C-4 hydroxyl group which is found in all of the metabolites utilised by these enzymes.

The *qutR* gene of *A.nidulans* has been sequenced, and its predicted protein sequence also shows significant similarity to the three C-terminal domains of the AROM polypeptide of *A.nidulans* (as expected, as it is analogous to the Qa-1S of *N.crassa*) [Hawkins *et al.*, 1992]. The regions round the active sites of the type I dehydroquinase enzymes of *E.coli, S.cerevisiae, A.nidulans* and *N.crassa* show similarity to the repressor proteins, QutR and Qa-1S, of *A.nidulans* and *N.crassa*, although the repressor proteins lack the conserved lysines which, in the dehydroquinase enzymes, have been implicated in catalysis [Chaudhuri *et al.*, 1991]. This suggests that the repressor proteins are able to bind the substrate of the dehydroquinase enzymes (DHQ) but are unable to undertake catalysis on it. The repressor proteins both have the consensus motif for purine nucleotide binding and have kinase activity, supplying further evidence for their ability to phosphorylate the activator proteins.

There is, however, an alternative hypothesis to phosphorylation for the method of action of the QutR repressor. The QutA and Qa-1F activators from *A.nidulans* and *N.crassa* show similarity throughout most of their length to the two N-terminal domains of the AROM polypeptides (DHQ synthase and the N-terminal 80% of EPSP synthase) [Hawkins *et al.*, 1993b], although the activators contain a putative zinc-cluster DNA-binding domain [Vallee *et al.*, 1991] which is absent from the AROM polypeptides. These two domains are able to fold and produce a stable protein, independently of the C-terminal domains of the AROM protein [Moore & Hawkins, 1993]. The QutR and Qa-1S repressors are similar to the remainder of the AROM protein. It has been shown that the AROM protein functions as a dimer, with the monomers contacting each other between the N-terminal (DHQ synthase) and C-terminal (shikimate dehydrogenase) domains [Case & Giles, 1971]. Therefore, it has been proposed that the shikimate dehydrogenase-like domain of QutR lies in close proximity to the DHQ synthase-like domain of QutA. The dehydroquinase-like and/or shikimate kinase-like domain of QutR then occlude(s) residues of QutA which are required for activation, and thus the QutR protein is able to effect repression on the quinate pathway of *A.nidulans* by preventing binding of the activator to the *qut* cluster [Hawkins *et al.*, 1993a].

As has been reported, the activators of the quinate pathway (QutA and Qa-1F) are similar to the N-terminus of the AROM protein, and the repressors of the quinate pathway (QutR and Qa-1S) are similar to the C-terminal domains of the AROM proteins. The AROM proteins of *A.nidulans*, *N.crassa* and *S.cerevisiae* are homologous to each other, and are extremely likely to have evolved from a common ancestor. However, as *S.cerevisiae* does not possess a quinate pathway, it appears that the activators and repressors of the quinate pathway have evolved by splitting the *arom* locus into two portions, and fusion of the activator and repressor genes to produce the *arom* locus is extremely unlikely to have occurred [Hawkins *et al.*, 1993b] (**Figure 1.4**).

1.12 Metabolic Control Analysis

Metabolic Control Analysis (MCA) is a method of studying the flux through intact metabolic systems or pathways, and the concentrations of metabolites within such systems [Kacser & Porteous, 1987]. It is usually performed on systems in a steady state, and can be approached using two methods: studying the individual enzyme reactions in isolation (as typified by enzymology), or studying the whole system *in vivo*.

MCA involves the measurement of three coefficients: the elasticity, concentration control and flux control coefficients.

1.12.1 The elasticity coefficient

This is a measure of the response of an enzyme rate to a change in concentration of a metabolite, and is calculated as the change in reaction velocity relative to the change in metabolite concentration. An enzyme has as many elasticity coefficients as metabolites (substrates, products and effectors) associated with it.

1.12.2 Concentration control coefficient

This is a measure of the response of the concentration of a metabolite to a small change in enzyme concentration. Each metabolite has a concentration control coefficient for each enzyme in the system. The sum of all concentration control coefficients for any metabolite is zero.

1.12.3 Flux control coefficient

This is the response of flux through any one part of the system to a small change in any one enzyme concentration. It is a measure of the importance of an enzyme to the flux through a system; the higher the flux control coefficient, the more control an enzyme exerts on the flux. Branch points in a pathway introduce negative flux control coefficients into a system. The flux summation theorem states that sum of all of the flux control coefficients in a system is unity. The connectivity property states that the ratios of the flux control coefficients of adjacent steps is equal to the ratios of the two elasticity coefficients with respect to the shared substrate.

1.12.4 Application of MCA

Naturally, any external changes in effector or substrate concentrations will affect metabolite (and possibly enzyme) concentrations, and subsequently alter all of the elasticity coefficients to reflect the new steady state of the system, thus producing a new set of flux control and concentration control coefficients, and causing a redistribution of control throughout the system.

In theory, altering the rate of any one reaction in the system will affect the concentrations of metabolites of that particular reaction, the rate of flanking reactions, and ultimately the whole system. Therefore, the fluxes through the various steps are interdependent. However, in a particularly complex system, it is possible that no alteration at any single stage will have a significant effect on the flux through the system; modification of multiple stages, or even all of the stages, in a pathway may be required.

The tryptophan biosynthetic system of *S.cerevisiae* has been studied using MCA [Niederberger *et al.*, 1992]. The five genes of the tryptophan biosynthetic pathway were down-modulated by reducing their copy number from four to zero; the activities of extracted enzymes were copy number-dependent as anticipated, but the effects on tryptophan flux and growth rate were very low. Up-modulation of these five genes by overexpression of any one (10- to 50-fold) also produced only very small (up to 30%) increases in total relative flux. However, overexpression of all five genes in tandem produced an almost nine-fold increase in flux. Addition of the increases produced by singly-overexpressing the enzymes would produce an expected two-fold increase.

Mutation of the *gcd2* gene of *S.cerevisiae* causes constitutive derepression of the genes under general amino acid control, including the shikimate pathway. Mutation of *trp4*

under general amino acid control, including the shikimate pathway. Mutation of *trp4* blocks the tryptophan pathway after anthranilate. Therefore, comparison of the levels of chorismate and anthranilate in *gcd2* and *gcd2/trp4* mutants allows analysis of the flux through the shikimate pathway. Up-modulation of the shikimate pathway caused by the *gcd2* mutation produced a 1·7-fold increase in flux.

1.13 Other amino acid biosynthetic pathways in *Streptomyces*

The biosynthesis of a few other amino acids (arginine, histidine, proline and tryptophan) is also under investigation in *Streptomyces coelicolor*. There is some clustering of the genes involved in the biosynthesis of arginine, with the *argC*, *argJ* and *argB* genes being clustered (although not necessarily coupled transcriptionally) and lying close to *argH*; however, complete clustering does not appear to occur as the *argG* gene is apparently located elsewhere on the chromosome [Hindle *et al.*, 1994]. Similar organisation appears to exist with the genes involved in proline biosynthesis, where the *proA*, *proC* and *proB* genes are clustered but not coupled transcriptionally; those involved in the biosynthesis of tryptophan, with the genes occurring in the "early" (putatively*trpD* and *trpF*) and "late" (*trpCBA*) clusters [Hood *et al.*, 1992]; and those implicated in the biosynthesis of histidine (*hisD, hisC, hisBd, hisH* and *hisA*) being clustered [Limauro *et al.*, 1990]. Therefore, it may be anticipated that there will be some clustering of the genes involved in the shikimate pathway in *S.coelicolor*. Sequencing around the *aroQ* gene is currently being performed to investigate this.

The putative promoter sequence identified immediately upstream of *argC* contains a region showing good similarity to the ArgR repressor binding site, as do sequences in both *E.coli* and *B.subtilis* [Hindle *et al.*, 1994], implying that transcriptional regulation of *argC* (and possibly *argJ* and *argB*) may involve feedback inhibition. Tryptophan biosynthesis is transcriptionally regulated according to growth rate and growth phase in *S.coelicolor* but shows no apparent feedback regulation, and the genes involved in proline biosynthesis are transcribed constitutively [Hood *et al.*, 1992]. These results suggest that feedback regulation is unlikely to be a general method of control for the amino acid biosynthetic pathways in *Streptomyces*.

1.14 Experimental aims

Many of the products and/or intermediates of the primary metabolic pathways of *Streptomyces* are used as precursors for the biosynthesis of secondary metabolites. Therefore, it is of interest to investigate whether an increase in production of these precursors, achieved via up-regulation of the relevant primary metabolic pathways, will lead to an increase in the synthesis of secondary metabolites.

The experiments reported in this thesis have been undertaken as part of an investigation into the effects of up-regulation of the shikimate pathway in *Streptomyces coelicolor.* In order to measure the flux control coefficients of the enzymes in the shikimate pathway, the pathway must be blocked to allow analysis of the build-up of pathway intermediates. Therefore, inactivation of the type II dehydroquinase enzyme (the third enzyme of this pathway, which is encoded by the *aroQ* gene) was attempted by gene disruption.

Gene disruption also allows functional replacement of an enzyme to be effected. Therefore, in anticipation of the successful disruption of *aroQ*, the *aroD* gene encoding the type I dehydroquinase from *E.coli* was expressed in *S.coelicolor.* Even in non-disruptants, expression of an additional enzyme may improve the efficiency of this step of the pathway. Overexpression of the indigenous dehydroquinase enzyme was also attempted, by the introduction of this gene on a multicopy plasmid vector, in order to improve the efficiency of this step.

To allow analysis of the downstream enzymes of the pathway, it is desirable to disrupt the *aroA* gene, which encodes the sixth enzyme in the pathway, EPSP synthase. As the first stage of such a disruption, the cloning of the *aroA* gene was attempted using a variety of methodologies.

CHAPTER 2

Materials and Methods

2.1 Introduction

This chapter contains the general procedures used in the experiments which were the basis of this thesis. The chapter is divided into four main sections for convenience: bacterial strains, vectors and chemicals; microbiological techniques and standard media; general DNA methods; and general protein methods.

2.2 Bacterial strains, vectors and chemicals

2.2.1 Bacterial strains

The bacterial stains used are listed below:

STRAIN	GENOTYPE	REFERENCE/SOURCE
***Escherichia coli* strains**		
DS941	*recF*143, *proA*7, *str*31, *thr*1, *leu*6, *tsx*33, *mtl*2, *his*4, *argE*3, *lacY*$^+$, *lacZ*ΔM15, *lacI*^{q}, *galK*2, *ara*14, *supE*44, *xyl*5.	D. J. Sherratt
TG-1	*supE*, *hsdΔ5*, *thi*, Δ(*lac-proAB*), F' [*traD*36, *proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15].	Gibson (1984)
S17-1	*recA*, *pro*, *hsdR*, *hsdM*$^+$, with integrated RP4.	Mazodier *et al.* (1989)
MB5386	Tn10::*recA*, Tn9::*dam*, *dcm*	MacNeil (1988)
CB51	*dam*	C. Boyd
AB2848	*aroD*	J. R. Coggins
***Streptomyces lividans* strains**		
TK24	*str6*	Hopwood *et al.* (1985)
TK64	*pro2*, *str6*	Hopwood *et al.* (1985)
***Streptomyces coelicolor* strains**		
209	SCP1 NF, SCP2$^-$	
1147	SCP1$^+$, SCP2$^+$	
G216	SCP1$^+$, SCP2$^+$	

Table 2.1 Bacterial strains

2.2.2 Plasmid and bacteriophage vectors

The plasmid pUC18 was obtained from Pharmacia Biotech Ltd. (St. Albans, UK).
Plasmid pIBI24/25 was kindly donated by Dr. M. Anderson, Department of Biochemistry, University of Glasgow.
Plasmid pBluescript^{TM}II (KS+) was obtained from Stratagene (La Jolla, Ca., USA).
Plasmid pMTL23 was kindly donated by Dr. M. Stark, Department of Genetics, University of Glasgow.
pT7Blue vector was obtained from AMS Biotechnology UK Ltd. (Oxford, UK).
Plasmid pGEM7zf(+) was obtained from Promega Corporation (Madison, USA).
M13mp18 and M13mp19 are bacteriophage cloning vectors from which single-stranded DNA may be isolated for DNA sequencing. They were obtained from Pharmacia Biotech Ltd.
The *Streptomyces coelicolor* genomic library [Taylor, 1992] was constructed using the λGEM-11 replacement vector supplied as "*Bam*H I arms" by Promega Corporation.
pKD101 is a pGEM5zf(+)-based vector which has been altered to aid cloning of PCR products (Kovalic *et al.*, 1991).
pIJ486/7 is a general streptomycete cloning vector (pIJ101-based) [Ward *et al.*, 1986].
pUGT1 is a pIJ486-based vector which contains the *ermE* resistance marker, and the *tipA* promoter [Ingham *et al.*, 1995].
All *E.coli* plasmids described above contain a ColE1 replicon.

2.2.3 Chemicals and biochemicals

Ampicillin and ethidium bromide were obtained from Sigma Chemical Co. (Poole, UK).
Bactotryptone, yeast extract and Bactotryptone (agar) were obtained from Difco (Detroit, USA).
ATP, DTT, TEMED (N,N,N',N'- tetramethylethylene diamine), and Tris buffer were obtained from Boehringer Mannheim (Lewes, UK).
DMSO, phenylalanine, polyethylene glycol 8000, and tryptophan were obtained from BDH Chemicals (Poole, UK).
Agarose, IPTG (isopropyl-β-thiogalactoside), phenol (ultrapure), and X-gal (5-bromo-4-chloro-3-indolyl-ß-galactoside) were obtained from BRL (Gibco Ltd., Paisley, UK).
Acrylamide, bisacrylamide and SDS were obtained from FSA Laboratory Supplies (Loughborough, UK).
Oligonucleotides were synthesised on an Applied Biosystems Model 280A DNA synthesiser at the Institute of Genetics, University of Glasgow, using reagents from Cruachem (Science Park, Glasgow).

All other chemicals were of analytical reagent grade and were obtained from one of the following suppliers: BDH Ltd.; Formachem Ltd. (Strathaven, UK); FSA Laboratory Supplies; Koch-Light Ltd. (Haverhill, UK); Sigma Chemical Co.

2.2.4 Enzymes, proteins and kits

A kit for molecular weight determination of proteins was obtained from Sigma Chemical Co.
All restriction enzymes, T4 DNA ligase, and T4 Polynucleotide kinase were obtained from BRL and Promega Corporation.
Sequencing kit used was the Sequenase™ sequencing kit (USB Biochemicals, La Jolla, Ca., USA) supplemented with T7 DNA polymerase (Pharmacia Biotech Ltd.).

2.3 Standard media and microbiological techniques

2.3.1 Media used for growth of *S.coelicolor*

All growth media were sterilised by heating to 120°C for 15 mins in an autoclave. Supplements and buffer solutions were heated to 108°C and $CaCl_2$ to 114°C for 10 mins. Heat-labile solutions, such as antibiotics, were sterilised by filtration through Nalgene 0·22 μm pore membranes (Nalge Co., New York, USA).

2.3.1.1 Complex media

a) Soya Mannitol Agar (SM)
This was used as a general plating medium for *Streptomyces*, particularly for production of spores. It consists of 2% (w/v) mannitol, 2% (w/v) soya bean flour, and 1·6% (w/v) agar, made up in tap water.

b) Yeast extract-Malt extract (YEME)
This was the only liquid medium used for growth of *Streptomyces*. It consists of 0·3% (w/v) Difco yeast extract, 0·5% (w/v) Difco bacto peptone, 0·3% (w/v) Oxoid malt extract, 1% (w/v) glucose, 34% (w/v) sucrose, made up in distilled water. To prevent pelleting, and produce a well-dispersed growth, $MgCl_2$ was added to a final concentration of 5 mM, and glycine to a final concentration of 0·5% (w/v).

2.3.1.2 Regeneration medium

The regeneration medium used for plating out of transformed protoplasts of *S.coelicolor* was R2 medium. This consists of R2 agar and R2 broth added in a 1:1 ratio.

a) R2 agar

This consists of 0·05% (w/v) K_2SO_4, 2·02% (w/v) $MgCl_2.6H_2O$, 0·59% (w/v) $CaCl_2.2H_2O$, 2% (w/v) glucose, 0·6% (w/v) proline, 0·02% (w/v) Casamino acids, 0·4% (v/v) Trace Elements solution, 4·4% (w/v) agar, made up in distilled water.

b) R2 broth

It consists of 1·15% (w/v) MOPS buffer pH7·4, 1% (w/v) Yeast extract, 20·3% (w/v) sucrose, made up in distilled water.

c) Trace Elements solution

This was made up of 0·004% (w/v) $ZnCl_2$, 0·02% (w/v) $FeCl_3.6H_2O$, 0·001% (w/v) $CuCl_2.2H_2O$, 0·001% (w/v) $MnCl_2.4H_2O$, 0·001% (w/v) $Na_2B_4O_7.10H_2O$, 0·001% (w/v) $(NH_4)_6Mo_7O_{24}.4H_2O$ in distilled water.

These three solutions were autoclaved for 15 mins at 121°C. 100 ml R2 agar was melted in a steamer, and added to 100 ml pre-warmed R2 broth. 1 ml of 1% (w/v) KH_2PO_4 was added immediately before pouring the plates.

2.3.2 Media for propagation of *E.coli*

Chemicals of good quality were used in the preparation of the growth media and solutions (AnalaR grade when available). The sources of many of the chemicals varied during the course of this work. The most common suppliers were BDH Chemicals Ltd.; Difco Laboratories (Detroit, Michigan, USA) and Sigma Chemical Co. Ltd.

a) L-broth

1% (w/v) tryptone, 0·5% (w/v) yeast extract, 0·5% (w/v) NaCl, 0·1% (w/v) glucose, 0·002% (w/v) thymine, made up in distilled water and adjusted to pH 7·0 with NaOH.

b) L-agar

As L-broth, but without glucose and with the addition of 1·5% (w/v) bacto-agar.

c) 2xYT medium

1·6% (w/v) bacto-tryptone, 1% (w/v) bacto-yeast extract, 0·5% (w/v) NaCl, made up in distilled water and adjusted to pH 7·0 with NaOH.

d) Top agar

As L-agar, but with only 1% (w/v) bactoagar.

e) Top agarose

As L-Agar, but with 0·65% agarose.

f) Minimal agar

Agar was made to 1·75% with water. To 75 ml of this, 25 ml of D+M salts were added and glucose and thiamine were added to final concentrations of 2 mg/ml and 20 µg/ml, respectively. Other supplements were added if necessary.

g) Davis and Mingoli (D+M) Salts (x4)

2·8% (w/v) K_2HPO_4, 0·8% (w/v) KH_2PO_4, 0·4% (w/v) $(NH_4)_2SO_4$, 0·1% (w/v) trisodium citrate, $MgSO_4.7H_2O$, made up in distilled water.

2.3.3 Growth of *Streptomyces* mycelia in liquid media

Cultures were typically grown in 500 ml conical flasks containing 100 ml of medium at 30°C on an orbital shaker at 200 rpm.

Growth experiments employed YEME complex medium which produced reproducible and rapid growth. Spores from frozen suspensions or suspensions that were freshly-prepared from a frozen slope were used to inoculate the medium. Cells grown on YEME could be harvested after 48-72 hours and stored indefinitely as a cell pellet at -20°C.

2.3.4 Harvesting of mycelia

After growth, mycelia were diluted in an equal volume of distilled dH_2O, recovered from the media by centrifugation (10 mins, 10000 g) and resuspended in dH_20.

2.3.5 Growth of *E.coli*

Liquid cultures of *E. coli* strains from which plasmids were to be isolated were grown in L broth with the appropriate antibiotic selection (typically ampicillin at 100 µg/ml). The volume of broth inoculated depended on the quantity of plasmid required. Routinely, 1·5 ml and 100 ml cultures were used for small and large scale plasmid preparations, respectively (sections **2.4.2.3** and **2.4.2.2** respectively). For the preparation of competent cells, liquid cultures of *E. coli* DS941 were grown in L broth while *E. coli* TG-1 was grown in 2xYT (section **2.3.2**). To maximise aeration of the culture, the volume of the Erlenmeyer flask used was at least five times that of the broth. All cultures were incubated at 37°C in an orbital shaker at ca. 250 rpm.

2.3.6 Production of *S.coelicolor* spores

2.3.6.1 Preparation and storage of suspensions

Concentrated spore suspensions were required for inoculating liquid cultures of *S. coelicolor* or *S.lividans*. The protocol described by Hopwood *et al.* (1985) was followed with minor modifications:
A boiling tube containing a slant of SM agar (produced by pouring ca. 20 ml of molten agar into the tube and allowing it to solidify with the tube held at +5° from the horizontal) was inoculated with 150 µl of a spore or mycelial fragment suspension and incubated at 30°C. After 5-10 days the surface of the culture was covered in a dark grey mass of spores.

The slant was then sealed using parafilm and frozen at -20°C. Spores could be harvested immediately or stored indefinitely at -20°C. The spores were harvested by adding 5 ml of dH_2O to the frozen slant and rubbing the surface of the slant with a 10 ml glass syringe. Contaminating agar or mycelial fragments were removed by a single passage through a cotton wool filter, as described in Hopwood *et al.* (1985). The filtered spore suspension was then either used fresh to inoculate YEME media or frozen at -20°C.

2.3.6.2 Spore counts

Colony forming units were determined by plating suitably diluted spore samples on Soya plates. Counts of the number of colonies were made after incubation at 30°C for 5 days.

2.3.7 Preservation of *E.coli* strains

E.coli strains were stored in glycerol. An 800 µl aliquot of an overnight culture was mixed with an equal volume of 40% (v/v) glycerol, 2% peptone (w/v) and frozen at -70°C. The strains were revived by scraping the surface of the frozen suspension with a toothpick and either inoculating liquid broth or streaking onto agar to isolate a single colony.

2.3.8 Introduction of plasmid DNA into *E.coli*

2.3.8.1 Preparation of competent cells

a) $CaCl_2$ method

An overnight culture of the recipient strain was diluted 1 in 100 into 30 ml L-broth and incubated for 90-120 mins to a density of approximately 10^8/ml cells (OD_{600} 0·4-0·6). The cells were harvested using a centrifuge (12000 g, 4°C for 5 mins) and resuspended in 10 ml of ice-cold 50 mM $CaCl_2$. The cells were pelleted again, resuspended in 1 ml of ice-cold 50 mM $CaCl_2$ and either kept on ice for at least 15 min before use, or stored in 100 µl aliquots following the addition of 200 µl 100% glycerol.

b) Hanahan method

For cells with high transformation efficiencies, the following steps were carried out (Hanahan, 1983). Cells were prepared and harvested as above but were then resuspended in 2.5 ml of ice cold TFB (10 mM MES/KOH pH 6·3, 100 mM RbCl, 45 mM $MnCl_2$, 10 mM $CoCl_2$, 3 mM hexaminecobaltic chloride) and incubated on ice for 15 mins. 100 µl of DMSO was then added and the cells incubated on ice for 5 minutes, followed by the addition of 100 µl of 2·25 M DTT, 40 mM potassium acetate (pH 6·0) and the cells incubated on ice for a further 10 minutes. Finally, 100 µl of DMSO was added, and the cells were stored on ice and used on the day of preparation.

2.3.8.2 Transformation procedure

Transformations were carried out in sterile 1·5 ml microfuge tubes. An aliquot (maximum 10 µl) of ligation mix or plasmid DNA was added to 100 µl aliquots of competent cells and the mixture was incubated on ice for at least 30 minutes. The DNA/cell mix was heat shocked at 42°C for 2 minutes before being placed back on ice. The cells were then plated onto L broth plates containing the appropriate antibiotic/chromogenic substances and incubated overnight at 37°C.

2.3.8.3 Transformation with bacteriophage M13

This procedure is as above, except that, after heat-shock, the transformed cells and 100 µl exponentially-growing (plating) cells were added to 3 ml 0·6% soft agar containing 10 µl IPTG (stock 24 mg/ml) and 50 µl X-gal (stock 20 mg/ml). The mixture was plated onto L-agar plates, which were incubated at 37°C overnight.

2.3.8.4 Selection of pUC-derived recombinant clones

a) Ampicillin

Stock solutions (20 mg/ml made up in water) were added to molten agar (cooled to 55°C) to a final concentration of 50 µg/ml.

b) X-gal (5-bromo-4-chloro-3-indolyl-ß-galactosidase)

This was used in conjunction with IPTG to identify *E. coli* strains containing pUC-based or M13mp18/19 vectors with inserts in their multiple cloning sites. Recombinants containing inserts are generally white while those lacking inserts are blue. X-gal was stored at a concentration of 20 mg/ml in dimethylformamide (DMF) at -20°C while IPTG was stored at a concentration of 24 mg/ml in dH_2O at -20°C. X-gal and IPTG were added to L-agar plates to a final concentration of 20 µg/ml and 50 µg/ml, respectively.

2.3.9 Introduction of plasmid DNA into *Streptomyces*

Plasmids were introduced into protoplasts of *Streptomyces spp.* by genetic transformation using the polyethylene-glycol-mediated protocol described by Hunter (1985).

Reagents

Medium P: 5·73 g N-tris[Hydroxymethyl]methyl-2-aminoethanesulfonic acid (TES), 103 g sucrose, 2·93 g $MgCl_2.7H_2O$, 0·5 g K_2SO_4, 3·68 g $CaCl_2.2H_2O$, 2 ml trace element solution; adjusted to pH 7·4 with NaOH and made up to 1 litre in distilled water.

Lysozyme solution: 10% (w/v) sucrose, 25 mM TES buffer (pH 7·2), 2·5 mM K_2SO_4, 2 ml trace elements (Hopwood *et al.*, 1985), 2·5 mM $MgCl_2$, 2·5 mM $CaCl_2$. KH_2PO_4 (0·005% [w/v]) and lysozyme (0·3 mg/ml) were added immediately prior to use.

PEG solution: 1 g of polyethylene glycol 1540 (supplied by BDH) was melted in a microwave (600 W) for 10 secs on the reheat setting and then mixed with 3 ml of medium P.

2.3.9.1 Preparation of protoplasts

30 ml cultures were grown in the appropriate medium at 30°C. The optimal time to harvest the mycelium in order to recover the most competent protoplasts was species-dependent (i.e. *S.lividans* mycelia were harvested after 65 hrs and *S.coelicolor* after 48

hrs, respectively). The mycelium was pelleted at 12000 g for 10 minutes and washed twice in 10·3% (w/v) sucrose. The pellet was then resuspended in 4 ml of lysozyme solution and incubated at 37°C for 15-30 min. The formation of protoplasts was monitored using a microscope, and the reaction terminated by adding 5 ml of P medium.

The protoplasts were then filtered through cotton wool (Hopwood *et al.*, 1985), pelleted using a centrifuge (12000 g for 10 minutes) and washed twice in P medium. Finally, they were resuspended in 2 ml of medium P, dispensed into 100 µl aliquots and frozen at -70°C.

2.3.9.2 Transformation of protoplasts

The protoplasts were thawed on ice. DNA was added in a volume of less than 10 µl and the mixture incubated on ice for 30 secs. 400 µl PEG solution was added, the solution incubated on ice for a further 1 min and, finally, 800 µl medium P was added. Dilutions of the transformation mix were then made in medium P and plated onto regeneration medium.

2.3.9.3 Regeneration of transformed protoplasts

Selection of transformants, due to plasmid-borne resistance markers, was performed by overlaying the regeneration plates with 1 ml of a 10·3% (w/v) sucrose solution containing a suitable antibiotic after 16-20 hrs of non-selective growth at 30°C. Antibiotics used in this work for plasmid selection in *Streptomyces* were thiostrepton (obtained from E.R. Squibb, New Jersey, USA), gentamicin (Sigma), erythromycin (Sigma), lincomycin (Sigma) and viomycin sulphate (Pfizer). Thiostrepton was dissolved in DMSO to make a 1% (w/v) stock solution. Erythromycin was dissolved in absolute ethanol to produce a 1% (w/v) stock solution. Gentamicin sulphate was dissolved in dH_2O to produce a 1% (w/v) stock solution. Lincomycin was dissolved in dH_2O to produce a 4% (w/v) stock solution. Viomycin was dissolved in dH_2O to produce a 5% (w/v) stock solution. All five of these antibiotic stock solutions were stored at 4°C.

Transformed protoplasts were selected by overlaying with 220 µg thiostrepton, 1·25 mg gentamicin, 4 mg erythromycin, 2 mg lincomycin or 2·5 mg viomycin per plate.

S.lividans and *S.coelicolor* protoplasts were regenerated on R2 agar plates at 30°C.

2.3.10 Conjugation between *E.coli* and *Streptomyces*

A plasmid capable of being maintained in *E.coli* and containing an *oriT* was first transformed into *E.coli* S17-1 [Mazodier *et al.*, 1989]. A transformant was obtained and a 5 ml culture grown in 2xYT.

A slope of *Streptomyces* spores was resuspended in 5 ml of dH_20. The spores were heat shocked at 50°C for 10 mins before being cooled under a cold tap, followed by the addition of 5 ml pre-germination medium containing 0·01 M $CaCl_2$ (Mazodier *et al.*, 1989). This was then placed at 37°C with shaking for 2-3 hrs. The germinating spores were spun down at 12000 g for 10 min and resuspended in 1 ml of dH_2O.

1 ml of the overnight S17-1 culture was spun down at 12000 g for 30 secs and resuspended in 1 ml of fresh 2xYT.

100 µl of germinating spores and 100 µl of S17-1 were plated and spread onto an L-agar plate and left at 30°C for 12 hrs. The plates were washed with L-broth, scraped with a pasteur pipette, and excess liquid taken off with the pasteur pipette before the plates were left to dry. They were then overlaid with the selective drug and nalidixic acid (to kill the S17-1 cells). Transformants grew up in 3-4 days.

2.4 General DNA methods

2.4.1 Commonly used buffers

a) TE buffer (10x)
100 mM Tris-HCl pH 8·0, 10 mM EDTA. Sterilised using an autoclave and stored at room temperature. It was used as a 1x solution for most applications.
b) Phage buffer (1x)
20 mM Tris-HCl pH7·4, 100 mM NaCl, 10 mM $MgSO_4$.
c) TAE Buffer (10x)
4·84% (w/v) Tris, 1·64% (w/v) Na acetate, 0·36% (w/v) $Na_2EDTA.2H_2O$, made up in distilled water, pH adjusted to 8·2 with glacial acetic acid.
d) TBE buffer (10x) pH8·3
10·9% (w/v) Tris, 5·5% (w/v) boric acid, 0·93% (w/v) $Na_2EDTA.2H_2O$ made up in distilled water.
e) Agarose gel loading buffer (10x) pH 7·4
0·5% (w/v) bromophenol blue, 0·05% (w/v) xylene cyanol, 50% (w/v) Ficoll, 1%

(w/v) SDS, 100 mM EDTA.

f) λ/*Hin*d III DNA markers

λ *c*I857 *S7* DNA was obtained from BRL. This DNA was cleaved with the restriction enzyme *Hin*d III and resulting DNA fragments diluted to a final concentration of 27 ng/μl in TE (final concentration 1x) with loading buffer added to 1x concentration. Typically, 10 μl was used on agarose gels as markers for comparing the size and concentration of bands in samples.

Sizes of fragments produced: 23,130 bp; 9,416 bp; 6,557 bp; 4,361 bp; 2,322 bp; 2,027 bp and 564 bp.

2.4.2 Preparation of plasmid DNA

2.4.2.1 Reagents for isolation of plasmid DNA

Protocols based on the alkaline lysis method (Birnboim and Doly, 1979) were used for the isolation of plasmid from small (5 ml) or large (50-200 ml) cultures of *E. coli*. This method, or an alternative alkaline lysis developed by Kieser, was used for isolation of plasmid DNA from *Streptomyces*.

a) Birnboim Doly I (BDI)

50 mM glucose, 25 mM Tris-HCl pH 8·0, 10 mM EDTA. If streptomycete plasmids were being isolated, lysozyme was added immediately before use to a final concentration of 5 mg/ml.

b) Birnboim Doly II (BDII)

0·2 M NaOH, 1% (w/v) SDS which was stored in a plastic container.

c) Birnboim Doly III (BDIII)

5 M KOAc pH 4·8; prepared by mixing equal volumes of 3 M CH_3COOK and 2 M CH_3COOH.

d) DNase-free RNase

Pancreatic RNase (RNase A) was dissolved at a concentration of 10 mg/ml in dH_2O, heated to 100°C for 15 mins and allowed to cool slowly to room temperature. The RNase was then aliquoted and stored at -20°C.

e) TES buffer

25 mM Tris-Cl pH 8·0, 25 mM EDTA, 10% sucrose.

f) Acid phenol/chloroform

50 g phenol was dissolved in 50 ml chloroform and 10 ml dH_2O. Hydroxyquinoline was added to colour the solution.

g) Alkaline SDS

0·3 M NaOH, 2% (w/v) SDS.

2.4.2.2 Large-scale plasmid preparation from *E.coli*

a) Caesium chloride gradient

100 ml cultures of stationary phase cells were harvested using a centrifuge (12000 g, 5 min at 4°C). The pellet was resuspended in 8 ml of Birnboim-Doly I solution and incubated at room temperature for 5 mins. 16 ml of Birnboim-Doly II solution were added and the solution left on ice for 5-10 mins, before 12 ml of cold Birnboim-Doly III solution were added. The suspension was mixed gently and left on ice for 15-30 mins. The cell debris and most of the chromosomal DNA were removed by centrifugation (32000 g, 5 mins at 4°C). The remaining nucleic acid was precipitated by the addition of 0·6 volumes of isopropanol and then harvested by centrifugation (39200 g, 15 mins).

The nucleic acid pellet was washed with 70% (v/v) ethanol. The plasmid DNA was further purified by equilibrium density centrifugation on a caesium chloride/ethidium bromide (CsCl/EtBr) gradient. The nucleic acid pellet was redissolved in 1 ml of dH_2O and 4·5 g of CsCl dissolved in 3·5 ml of dH_2O. The DNA and CsCl solutions were combined with 250 µl of EtBr (10 mg/ml), creating a solution with a density of 1·58 g/ml. The nucleic acid-CsCl solution was spun in a Beckman Ti70 angled rotor at 289,000 g for 16 hours at 20°C. Two bands were visible in the gradients after centrifugation, a lower supercoiled plasmid band and an upper chromosomal and relaxed plasmid DNA band.

The lower band was removed using a 1 ml syringe and the EtBr removed by repeated extractions with water-saturated butanol. After dilution with 3 volumes of dH_2O, 9 volumes of absolute ethanol were added. The precipitate was pelleted by centrifugation (27000 g, 4°C for 30 mins). The resulting plasmid pellet was washed twice with 70% (v/v) ethanol and dried *in vacuo* before being redissolved in 1 ml dH_2O. This procedure yielded very large amounts of pure plasmid DNA (up to 1 mg from *E. coli* cultures) suitable for all *in vitro* manipulations.

b) PEG precipitation

The protocol was the same as that for a), up to the resuspension of the DNA, in dH_2O and CsCl. Instead the DNA was resuspended in 3 ml of dH_2O. 3 ml of ice-cold 5 M LiCl was added and mixed. The precipitated DNA was removed by centrifugation (12000 g, 4°C for 20 mins). The pellet was resuspended in 500 µl 1x TE buffer (containing 10 mg RNase), and incubated at room temperature for 30 mins. 500 µl of 1·6 M NaCl containing 13% PEG-8000 was added, the solution mixed, and centrifuged (12000 g, 4°C for 5 mins). The pellet was washed with 70% EtOH, before

being dried under vacuum and redissolved in dH_2O.

2.4.2.3 Small-scale plasmid preparation from *E.coli*

Routinely, plasmids were isolated from 1·5 ml of *E. coli* cultures. The cells were pelleted by centrifugation in a 1·5 ml microfuge tube (12000 g for 30 secs) and resuspended in 100 µl of BDI, containing lysozyme at a concentration of 1 mg/ml, using a vortex mixer. This was followed by the addition of 200 µl of BDII and repeated inversion of the microfuge tube to thoroughly mix the suspension. Immediately afterwards, 150 µl of pre-chilled BDIII was added to the viscous bacterial lysate, mixed gently on the vortex mixer and placed on ice for 5-10 mins. The cell debris and most of the chromosomal material were harvested by centrifugation (12000 g, 4°C for 10 min) in a microfuge. The supernatant was transferred to a fresh tube and extracted with half volumes of phenol/chloroform and chloroform. The nucleic acid was then precipitated by the addition of 2 volumes of ethanol and allowed to stand at room temperature for 5 mins. The precipitate was harvested by centrifugation in a microfuge (12000 g, 4°C for at least 15 mins). The resulting pellet was rinsed twice with 70% (v/v) ethanol before it was allowed to dry by leaving the tube open on the bench. The nucleic acid was then resuspended in 50 µl dH_2O containing DNase-free RNase (20 µg/ml).

If further purification of the DNA was required (for example for double-stranded sequencing reactions), the pellet was resuspended in 16 µl of 1x TE, to which 4 µl of 4 M NaCl and 20 µl of 13% PEG were added, followed by incubation on ice for 20 mins. Centrifugation for 15 mins at 4°C produced a pellet, which was washed in 70% ethanol, and resuspended in 1x TE.

2.4.2.4 Large-scale plasmid preparation from *Streptomyces*

100 ml *Streptomyces* culture was diluted 1:1 with dH_2O and pelleted by centrifugation (12000 g, 10 mins at room temperature). The pellet was resuspended in 4 ml BDI containing 25 mg/ml lysozyme, and incubated at 37°C for 30 mins. 8 ml BDII was added, followed by 15 mins incubation on ice, and 6 ml BDIII added, followed by 5 mins incubation on ice. Cell debris and most chromosomal DNA was isolated by centrifugation (20000 g, 4°C for 30 mins).

The plasmid DNA contained in the supernatant was precipitated with 0.6 volumes of isopropanol and 15 mins incubation at room temperature. The DNA was pelleted by centrifugation (20000 g for 10 min at room temperature), washed with 70% ethanol,

centrifugation (20000 g for 10 min at room temperature), washed with 70% ethanol, and resuspended in 1x TE buffer.

2.4.2.5 Preparation of plasmid DNA by the "Kieser" method (Hopwood *et al.*, 1985)

Upto 1 ml of *Streptomyces* culture was diluted 1:1 with dH_2O and pelleted by centrifugation. The pellet was washed in TES buffer, resuspended in 500 µl TES buffer (containing 2 mg/ml lysozyme), and incubated at 37°C for up to 30 mins. 250 µl of alkaline SDS was added, the solution homogenised immediately, and incubated at 70°C for 15 mins. After cooling to room temperature, the solution was mixed with 80 µl of acid phenol/chloroform, centrifuged, and the supernatant isolated. This step was repeated using 80 µl of chloroform, and the supernatant added to 60 µl 3 M sodium acetate and 600 µl isopropanol. The solution was mixed and left at room temperature for 10 mins, followed by centrifugation. The pellet was washed in 70% ethanol, and resuspended in 1x TE.

2.4.3 Preparation of "total" DNA from *S.coelicolor*

20 ml *Streptomyces* culture was diluted 1:1 with dH_2O and pelleted by centrifugation (12000g, 4°C for 10 mins), resuspended in 5 ml lysozyme solution [Hopwood *et al.*, 1985] (containing 4 mg/ml lysozyme) and incubated at 37°C for 30 mins, followed by the addition of 2·5 ml 2% SDS and mixing by vortex. Cell debris was removed by extraction with phenol/chloroform and centrifugation (12000g, 4°C for 15 mins). Such extraction was repeated until no interface remained (a minimum of three extractions). DNA was then precipitated by the addition of 0·1 vol. of 3 M Na acetate and 1 vol. isopropanol, with 5 mins incubation at room temperature followed by centrifugation (12000 g, 4°C for 10 mins). The DNA was resuspended in 1x TE buffer (often requiring incubation at 50°C to resuspend the DNA).

2.4.4 Organic solvent extraction

Protein was removed from DNA solutions by phenol/chloroform extraction. An equal volume of TE-saturated phenol was added to samples which were then mixed by vortexing and centrifuged in a microfuge for 1-5 mins. The upper aqueous phase was removed to a fresh microfuge tube and the process repeated this time using phenol/chloroform (1:1 v/v). Finally, traces of phenol were removed by extraction with an equal volume of chloroform in an identical manner. Precipitation with ethanol or isopropanol removed any remaining solvent.

2.4.5 Precipitation of DNA using ethanol or isopropanol

DNA solutions were precipitated by the addition of 1/10 volume of 5 M NaCl and 2 volumes of cold ethanol or an equal volume of isopropanol. After mixing, the DNA was pelleted by centrifugation (27000 g, 4°C, 30 mins for volumes of 7·5-20 ml, or 12000 g, 4°C, 15 mins for small volumes in microfuge tubes). The pellet was washed in 70% (v/v) ethanol and dried briefly in a vacuum desiccator, or in an open tube on the bench.

2.4.6 Spectrophotometric measurement of nucleic acid

Nucleic acid concentrations were determined spectrophotometrically at 260 nm. In a 1 cm path length an absorbency value of 1·0 corresponds to 50 µg/ml for double-stranded DNA , 33 µg/ml for single-stranded DNA and 20 µg/ml for oligonucleotides.

2.4.7 Digestion of DNA with restriction enzymes

Restriction digests were carried out using the BRL restriction enzymes and REact buffers which were provided with each batch of enzyme. There are ten different REact buffers with a range of salt concentrations, each one suitable for a range of enzymes. Alternatively, digests were performed using Promega restriction enzymes and the accompanying restriction buffers. Analytical digests were carried out in a volume of 20 µl at 37°C. Preparative digests were carried out in larger volumes. When DNA was digested with two restriction enzymes, the endonuclease requiring the lower salt buffer was used first. After the recommended duration of digestion, the salt concentration was adjusted and the second enzyme added.

2.4.8 Ligation of DNA fragments

The ligation of DNA fragments was carried out usually at a DNA concentration of 6 mg/ml. The molar ratio of insert fragment to vector was 3:1, when the vector could not ligate to itself (for example when using a vector that has been dephosphorylated or has been cut with two non-complementary enzymes). A molar ratio of 10:1 was used when the ends of the vector could ligate to each other. Ligations were performed usually in 10 µl of 1x ligation buffer provided by BRL, containing 1 U of T4 ligase per µg of DNA. The reactions were incubated for 4 hours at room temperature or overnight at 16°C.

2.4.9 T-tailing of linearised, blunt-ended vectors

After the linearised, blunt-ended vector DNA had been purified from an agarose gel, it was denatured by heating to 90°C for 5 mins, followed by three-fold dilution of the solution in 1x Taq-buffer. dTTP was added to 2 mM, followed by the addition of 2U of Taq polymerase and 2 hours incubation at 70°C. The T-tailed vector was again run on an agarose gel and purified, to remove the enzyme and any unincorporated dTTP.

2.4.10 Removal of the 5' phosphate from linearised DNA

10x CIP Buffer: 200 mM Tris-HCl pH 8·0, 10 mM $MgCl_2$, 10 mM $ZnCl_2$ and 0·5 mg/ml Bovine Serum Albumin.

Procedure: Calf Intestinal Alkaline phosphatase (CIP) was used to remove the 5' phosphate from DNA. Around 5 pmoles of 5'-terminal phosphorylated DNA with 5' protruding ends (approximately 7 μg of a 5 kb molecule) were incubated in 1x CIP buffer, containing 0·1 U of CIP at 37°C for 30 mins. The reaction was terminated by heating to 65°C in 1x gel loading buffer for 10 mins. The 5'-terminal dephosphorylated DNA was recovered from an agarose gel after electrophoresis.

2.4.11 Agarose gel electrophoresis

DNA was visualised on horizontal neutral agarose gels. Although 0·8% (w/v) gels were most commonly used, 1-2% (w/v) gels were occasionally used to separate fragments of <1·5 kb. Gels were routinely prepared and run in TBE buffer. However, TAE buffer was used when DNA fragments were to be isolated from the gels (see section **2.4.13**). λ-Hind III markers were used on all gels as size markers and for quantification of the amount of DNA by comparing the intensity of bands to those of the samples (**2.4.1**f).

a) Mini gels

BRL model H6 gel kits were used for the rapid analysis of DNA after digestion with restriction enzymes or precipitation steps. 0·16 g agarose was added to 20 ml of 1x TBE (or TAE), boiled, and then cooled to 60°C. EtBr was added to 200 ng/ml and the molten agarose poured into a 7·6 cm x 5·1 cm gel caster with an 8 well slot former (4·1 x 0·8 mm wells). After the gel had set, the slot former was removed and the gel placed in the tank with 500 ml of 1x TBE (or TAE).

separated by electrophoresis for 30-60 mins with an applied voltage of 2-10 V/cm. 1/10 volume loading buffer (**2.4.1**e) was added to the DNA samples before they were loaded onto the gel. The separated DNA molecules were visualised on a 302 nm UV transilluminator.

b) Large gels

200 ml gels were also used to ensure good separation of DNA fragments for accurate sizing and/or Southern analysis. They were made by pouring 200 ml of molten agar, containing 200 µg EtBr, into a 16·5 x 23 cm gel former with a 14 space slot former. The gels were run overnight at 20 V in 1x TAE or TBE buffer. DNA samples were mixed with 1/10 volume of 10x loading buffer (**2.4.1**e) before loading onto the gel.

2.4.12 Photography of agarose gels

Gels stained with ethidium bromide were viewed on a 302 nm UV transilluminator and photographed using Polaroid type 67 land film or using a Pentax 35 mm SLR loaded with Ilford HP5 film. Both cameras were fitted with Kodak Wratten filters (No. 23A). Alternatively, pictures were obtained using a Mitsubishi video copy processor attached to a UVP video camera.

2.4.13 Recovery of DNA from agarose

a) using SPIN-X tubes

SPIN-X tubes were obtained from Costar UK Ltd. These tubes contain a cellulose acetate membrane, which allows the passage of buffer and DNA, but stops agarose from passing through.

The DNA band was excised from the gel and placed in the upper chamber of the SPIN-X tube. This was then placed at -70°C for 5 minutes, thawed by placing at 37°C for 5 minutes, and spun in a centrifuge at 12000 rpm for 5 mins. The filtrate contained the DNA in a state which could be used immediately.

b) using Qiaex-suspension

A Qiaex gel extraction kit was obtained from Qiagen, Germany. The gel slice containing the excised DNA fragment was mixed with 3 vols. QX1 solubilisation buffer and Qiaex-suspension (10 µl per 5 µg DNA). The gel slice was incubated at 50°C for 10 mins with frequent mixing to solubilise the gel and allow the DNA to absorb to the Qiaex-suspension. The suspension was pelleted, washed twice with 500 µl QX2 wash buffer, and twice with 500 µl QX3 wash buffer. The suspension was

pelleted and air-dried for 10-15 mins, followed by resuspension in 1x TE buffer, 5 mins incubation at room temperature and pelleted again. The DNA was contained in the supernatant, in a condition ready for any subsequent manipulations.

2.4.14 Techniques with *E.coli* bacteriophage λ

2.4.14.1 Preparation of plating bacteria for infection with bacteriophage λ

50 ml of L-broth supplemented with 0·2% (w/v) maltose, 20 mM-$MgSO_4$ was inoculated with a single colony of the appropriate *E.coli* strain (*e.g.* NM621), and grown overnight at 37°C on an orbital shaker. The cells were pelleted by centrifugation (12000 g, 4°C, 5 mins) and resuspended in 0·5 volumes of sterile, ice cold 10 mM $MgSO_4$. The cell suspension was diluted if the OD_{600} was greater than 2 ($1{\cdot}6 \times 10^9$ cells/ml). The cells were stored at 4°C and remained viable for at least 3 weeks.

2.4.14.2 Infection of bacteriophage λ, plating and titre

Serial 10-fold dilutions of λ phage stock (or packaged λ DNA) were prepared in phage buffer (**2.4.1**b). Bacteriophage λ infection was achieved by adding 100 µl aliquots of each dilution to 100 µl (1.5×10^8 cells) of a suspension of plating bacteria. The samples were incubated at 37°C for 20 minutes. 3 ml of top agarose at a temperature of 45°C was added and the mixture was poured onto plates containing bottom agar. The plates were left to stand for 5 minutes at room temperature to allow the top agarose to harden and then incubated at 37°C overnight. The plaques were counted and the titre determined for each dilution assayed.

2.4.14.3 Isolation of bacteriophage particles from a plaque

The plaque of interest was stabbed out of the plate using the narrow end of a sterile glass Pasteur pipette to form a plug of agar. The plug was left in 1 ml of phage buffer containing 70 µl of DMSO (to kill any cells) for two hours at room temperature (or overnight at 4°C) to allow bacteriophage particles to diffuse out of the agar. An average plaque yielded 10^6-10^7 infectious bacteriophage particles, which could be stored indefinitely at -70°C in phage buffer/DMSO without loss of viability.

2.4.15 Techniques for handling *E.coli* bacteriophage M13

2.4.15.1 Transfection and plating of M13

Cloning in M13 vectors is essentially the same as cloning in plasmids. M13 produces a double stranded, replicative form (RF) which is isolated and treated just like an ordinary plasmid such as pUC18/19. After introduction of foreign DNA into such vectors by ligation, they can be introduced into a suitable *E.coli* host which produces pili, such as TG-1.

E.coli TG-1 was made competent by one of the methods above (**2.3.8.1**) and the transformation protocol followed up to and including the heat-shock stage. After this step, 200 µl of a fresh exponential TG-1 culture were added to the transfected cells, followed by 10 µl of IPTG (24 mg/ml) and 50 µl of X-gal (20 mg/ml). The cells were then mixed and added to 2·5 ml of molten water-agar (0·6% w/v, pre-cooled to 45°C), which were poured onto dried L-agar plates. Plaques containing recombinant phage appeared white on the agar, whereas non-recombinant phage appeared blue.

2.4.15.2 Preparation of single-stranded DNA from M13

The single-stranded M13 templates were prepared as described in the "M13 Cloning/Dideoxy sequencing Instruction Manual" published by Bethesda Research Laboratories.

a) Minipreparations

A single M13 plaque was used to infect 1·5 ml of 2xYT broth containing 15 µl of an overnight culture of *E.coli* TG-1. This culture was grown at 37°C for 5-6 hrs with vigorous shaking, then transferred to a microfuge tube and harvested by centrifugation at room temperature for 5 mins. The supernatant, containing the phage particles, was recovered and respun. The remaining supernatant was mixed with 200 µl of a solution of 20% (w/v) PEG (8000), 2·5 M NaCl and left to stand at room temperature for 15 mins to precipitate the phage particles. These were recovered by centrifugation at room temperature in a microfuge for 15 mins.

The supernatant was discarded, the pellet respun and all traces of supernatant removed. The pellet was then resuspended in 100 µl dH_2O and extracted twice with phenol/chloroform and twice with chloroform. The DNA was then precipitated from the aqueous phase with sodium acetate and ethanol, and recovered by centrifugation in a microfuge. The ssDNA was then washed with 70% (v/v) ethanol, dried *in vacuo*

b) Maxipreparations

The same overall procedure was followed as for the "minipreparations", except that all the volumes were scaled up 20-fold. The cells from a 30 ml culture were spun out (14000 g for 2 mins), the supernatant recovered and respun as before. The supernatant (20 ml) that remained was then precipitated with 5 ml of 20% (w/v) PEG, 2·5 M NaCl for 10 mins at room temperature and the phage harvested by centrifugation (14000 g at 20°C for 15 mins). The phage pellet was resuspended in 1 ml of dH_2O, then reprecipitated and processed as for the minipreparations with the volumes scaled up accordingly.

2.4.16 Labelling of DNA with γ-^{32}P or Digoxygenin

2.4.16.1 Labelling of the 5' termini of oligonucleotides with γ-^{32}P using bacteriophage T4 polynucleotide kinase

Bacteriophage T4 polynucleotide kinase catalyses the transfer of the γ-phosphate group from ATP to a free hydroxyl group on the 5' terminus of DNA. In a total reaction volume of 10 µl the mixture contained:
8 pmoles of purified oligonucleotide, kinase buffer (50 mM Tris-HCl pH 8·0, 10 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine), 8 pmoles (γ-^{32}P) ATP and 10 units of T4 polynucleotide kinase. The reaction was carried out at 37°C for 30 minutes by which time it had gone to completion. Unincorporated label was removed by gel filtration (see below) .

2.4.16.2 Removal of unincorporated radionucleotide

Unincorporated label from end-labelling reactions was removed by gel filtration chromatography on a 20x1 cm Sephadex G-50 column. Sephadex G-50 was hydrated in 1x TE and poured columns equilibrated in 1x TE. The reaction mixture was mixed with an equal volume of Blue Dextran dye in 1x TE and loaded directly on top of the column. The radioactivity was monitored as it passed down the column and as it approached the bottom, fractions were collected manually. Labelled oligonucleotide was co-eluted first with the Blue Dextran dye, followed by a trough of radioactivity, and then a second peak corresponding to the unincorporated label. The fractions containing the largest number of incorporated counts were pooled.

2.4.16.3 Labelling of DNA using the "random priming" technique

The method of "random primed" DNA labelling is based on the hybridisation of a

mixture of all possible hexanucleotides to the DNA to be labelled. The complementary strand is synthesised from the 3'-hydroxyl termini of the random hexanucleotide primer, with Klenow enzyme (DNA polymerase I). A random priming kit from Boehringer Mannheim, UK. was used for this purpose. Protocols were followed as per the manufacturer's instructions.

2.4.16.4 Labelling of DNA with digoxygenin-dUTP (DIG-dUTP) using the "random priming" technique

The method of "random primed" DIG DNA is essentially the same as for labelling with ^{32}P (**2.4.16.3**), except that the labelling reaction was carried out overnight. A random priming kit from Boehringer Mannheim, UK., was used for this purpose. Protocols were followed as per the manufacturer's instructions.

2.4.17 Southern blotting

After electrophoresis and photography, the resolved DNA fragments were transferred under alkali conditions to nylon membrane (adapted from Southern, 1975), as described in "Blotting and hybridisation protocols for Hybond-N^{TM} membranes" (published by Amersham International plc), or by a modification of this protocol, "dry blotting".

2.4.17.1 Reagents

a) Denaturing solution
1·5 M NaCl, 0·5 M NaOH.
b) Alkali transfer buffer
1·5 M NaCl, 0·25 M NaOH.
c) 20x SSC
3 M NaCl, 0·3 M tri-sodium citrate, pH 7·0.

2.4.17.2 Procedure

The gel was rinsed in distilled water, placed in enough denaturing solution to immerse it completely and left for 30 mins to chemically denature the DNA contained within the agarose. The gel was then equilibrated for 10-15 mins in alkaline transfer buffer. The DNA was transferred to the nylon membrane in transfer buffer, by capillary action (disposable nappies proved a particularly useful absorbent material for driving the transfer process).

After blotting for a least 4 hours (but usually overnight), the membrane was washed briefly in 2x SSC to remove any adhering agarose. The DNA was fixed to the membrane either by UV crosslinking using a Statagene "Stratalinker", or by baking at 80°C for 2 hours.

2.4.17.3 Dry blotting

The procedure was similar to that above, except that the gel was first washed twice in 0·25 M HCl for 15 mins to depurinate the DNA, and was washed briefly in dH_2O before denaturation as above. No equilibration in alkali transfer buffer was performed, although the 3MM Whatman paper placed on top of the membrane was soaked in transfer buffer. Transfer took only 1-2 hrs, followed by fixing as above.

2.4.18 Hybridisation of oligonucleotides to filter-bound nucleic acid

2.4.18.1 Prehybridisation

Nylon filters were not pre-wetted, but were placed directly into a hybridisation tube containing prehybridisation solution (6x SSC, 0·05% (w/v) sodium pyrophosphate, 200 µg/ml heparin, 0·05% (w/v) SDS); 200 µls of the prehybridisation solution was used per cm^2 of filter surface area. The tube was placed in a hybridisation oven and the filter prehybridised for at least 4 hours at 50-68°C, depending on individual prehybridisations.

Alternatively, filters were prehybridised in QuickHyb solution (Stratagene), containing sonicated Salmon sperm DNA as blocking reagent, for 1 hr at 50-68°C.

2.4.18.2 Hybridisation and washing

After prehybridisation, the prehybridisation solution was replaced by hybridisation solution. The hybridisation solution was the same as the prehybridisation solution, except that it contained 0·5% (w/v) SDS and the salt concentration was varied according to the conditions required (see results of individual hybridisations).

50 µls of the solution was used per cm^2 of filter surface area. 10 pmoles of labelled oligonucleotide (**2.4.16**) was added to the bag before it was sealed. Hybridisation conditions were as described for each individual hybridisation.

Hybridised filters were washed twice in large volumes of buffer at ionic strengths

appropriate to the experimental conditions. The temperature of this buffer was also varied experimentally. After washing, the filters were left damp, wrapped in Saran wrap and exposed to Fuji RX film using intensifying screens at either -70°C (for radiolabelled probes) or room temperature (for DIG-labelled probes). Films were developed by a Kodak X-OMAT processor.

2.4.19 Screening of plasmid clones by colony hybridisation

Recombinant pIBI-based clones were screened using a modification of the method described in the Hybond-N protocol manual. Nylon filters were placed on duplicate agar plates containing the selective antibiotic. Bacterial colonies were crossed onto a master plate (containing antibiotic) and then onto the nylon filters (onto which alignment marks had been made). The plates were inverted and grown overnight at 37°C. The filters were removed and placed colony side up on a pad of absorbent filter paper soaked in denaturation solution (1·5 M NaCl, 0·5 M NaOH) and left for 7 mins.

The filters were then transferred, colony side up, to a pad of filter paper soaked in neutralising solution (1·5 M NaCl, 0·5 M Tris-HCl pH 7·2, 1 mM EDTA), and left for 3 mins. This step was repeated with a fresh pad soaked in the same solution. The filters were then washed briefly in 2x SSC, transferred to dry filter paper and allowed to dry in air, colony side up. Finally, the filters were baked at 80°C for 2 hrs, or UV-crosslinked. The filter was then hybridised with a nucleic acid probe (**2.4.18**).

2.4.20 Screening of a bacteriophage λ library

2.4.20.1 Primary screening

Cells from a prepared bacterial suspension were infected (**2.4.14.1**) with phage from the bacteriophage λ library at a multiplicity of 10^4 pfu/10^8 cells. $2x10^3$ pfu were plated onto 10x10 cm petri dishes using 8 ml of 0·6% (w/v) top agarose in L broth. The plates were incubated overnight at 37°C.

Up to six impressions could be taken from one plate onto nylon membranes, provided that sufficient time was allowed for fresh phage to diffuse to the top agarose surface (Sambrook *et al.*, 1989). Alignment marks were made on the plate and on the filter using a syringe needle. Filters were removed and treated as described for colony filters (**2.4.19**).

2.4.20.2 Secondary screening

Single plaques were isolated as described in **2.4.14.3**. A lawn of bacteria (10^8 cells) were plated onto a 9 mm circular petri dish along with a suitable dilution of phage particles removed from the agar plug. The plates were then grown up overnight at 37°C. An impression was taken of the plate onto duplicate nylon filters which were treated as in (**2.4.19**a).

2.4.21 Hybridisation, and detection of DIG-labelled DNA hybrids

Prehybridisation and hybridisation of filters with DIG-labelled random primed DNA was carried out as described in the Digoxygenin protocols manual from Boehringer Mannheim, UK.

DIG-bound DNA was detected using AMPPD as a chemiluminescent substrate for alkaline phosphatase. Again, all protocols were followed as per the manufacturer's instructions.

2.4.22 Stripping filters

Filters were boiled for approximately 30 mins in 0·1% (w/v) SDS solution, and the efficiency of stripping detected by autoradiography overnight. If the autoradiogram was clear when developed, the filter was suitable for reprobing.

2.4.23 DNA sequencing techniques

2.4.23.1 Single-stranded sequencing

Dideoxy sequencing (Sanger *et al.,* 1977) was carried out on single-stranded M13 templates using either a Sequenase™ kit (supplied by United States Biochemical Corporation) or a T7 sequencing kit (supplied by Promega). All sequencing strategies used deoxy-7-deazaguanosine triphosphate (dc^7GTP) as a replacement for dGTP, to reduce sequence compressions [Mizosawa *et al.*, 1986]. Extension and labelling reactions were performed as suggested by the manufacturers. However, termination was carried out at 42°C instead of the normal 37°C. ^{35}S-dATP was used for labelling.

2.4.23.2 Double-stranded sequencing

Plasmid DNA prepared by alkaline lysis of 1·5 ml *E.coli* culture, followed by

precipitation with PEG (**2.4.23**), was resuspended in 5 µl TE, to which 1 pmol of sequencing primer and 1 µl of 1 M NaOH were added, followed by 30 mins incubation at 37°C. 1 µl of 1 M HCl and 2 µl of 5x sequencing buffer were then added, followed by a further 10 mins incubation at 37°C. A further 1 µl of 100 mM DTT, 2 µl of 1x labelling mix, 0·5 µl of ^{35}S-dATP and 2 µl of diluted T7 polymerase (1·5 U/µl) were added, and the labelling mixture incubated at room temperature for 5 mins.

To each of four tubes, 2·5 µl of one of the four ddNTP's was added. The tubes were pre-warmed for 2 mins at 37°C, followed by the addition of 3·5 µl of the labelling mixture. The reactions were incubated at 37°C for a further 5 mins, followed by the addition of 4 µl of stop solution. The reactions were heated at 70°C for 3 mins before 2·5 µl of each was loaded onto a sequencing gel.

2.4.23.3 Autoradiography of DNA sequencing gels

Autoradiography was performed in metal cassettes (medical chest X-ray type) using Kodak X-ray film. All autoradiography of sequencing gels was performed at room temperature, as autoradiography at -70°C produces bands that are diffuse and therefore more difficult to read. The X-ray films were developed using a Kodak X-OMAT automatic processor, Model ME-I.

2.4.24 Denaturing polyacrylamide gel electrophoresis for DNA sequencing

A BRL sequencing unit (Model S2) was used for high voltage polyacrylamide gel electrophoresis.

2.4.24.1 Preparation of polyacrylamide gels

6% (w/v) denaturing polyacrylamide gels were used for sequencing. The gels were prepared from the following stock solutions:

40% (w/v) acrylamide stock	9 ml
urea	30 g
10x TBE	6 ml
dH_2O	21 ml

The urea was dissolved by heating the mix to 37°C and then cooled to room

temperature. The sequencing gel solution could be stored at 4°C for several weeks without loss of resolution. Before pouring the gel, 300 µl of freshly prepared 10% (w/v) ammonium persulphate and 50 µl of TEMED were added to 50 ml of the stock solution.

2.4.24.2 Preparation of glass plates, and pouring of the gel

The plates (40 cm x 33 cm) were cleaned thoroughly with water and alcohol, and assembled using spacers (0.4 mm thick) along the vertical sides and 3MM Whatman paper along the bottom of the gel. The entire assembly was held in place by clamps. The gel solution was poured from a beaker down one edge of the plates while tilting the plates at an angle of approximately 30°. The plates were then laid at an angle of 5° and the sharks tooth combs inserted in an inverted fashion. The gel polymerised usually within 30 mins at room temperature.

2.4.24.3 Electrophoresis of sequencing gels

The gel was pre-electrophoresed for 30 mins at a constant power of 60 W. Prior to loading, the samples containing sequencing loading buffer were heated to 95°C for 5 mins, placed on ice and loaded onto the gel. If two cycles were performed, the first set of samples was run until the darker blue dye front reached the bottom of the gel, when the second set of samples was loaded. When the dark blue dye front had run approximately half-way down the gel, 61·52 g sodium acetate was added to the 500 ml of 1x TBE in the lower buffer tank, and electrophoresis continued until the darker blue dye front reached the bottom of the gel. This allowed upto 400 nucleotides to be read.

2.4.25 The polymerase chain reaction

DNA was amplified from a number of sources, but all following the same general procedure, detailed below:

Reaction cocktail:
- Unamplified DNA (~100 ng genomic DNA, ~5 ng plasmid)
- PCR buffer, containing $MgCl_2$ at concentrations of 0·5-3 mM)
- DMSO (10% v/v)
- dNTPs (1·25 mM)
- Primers (100 pmol each)

Protocol:
- Denature DNA (94°C for 5 mins)
- + 1U Taq polymerase

Denature DNA (94°C for 1 min)
Anneal primers (45-65°C, depending on the annealing temperature of the primers used; typically for 30 secs)
Extension of the DNA (72°C for 30 secs to 2 mins, depending on the length of DNA being amplified)
Repeat above procedure 20-30 times

Denature DNA (95°C for 4 mins)
Anneal primers (as above, except lengthen this step to typically 1-2 mins)
Extension of the DNA (72°C for 4 mins)

Typically, a tenth of the PCR reaction was run on a 1% agarose gel to determine if amplification of the DNA had occurred. DNA could be extracted and purified as described in section **2.4.13**.

2.5 General protein methods

2.5.1 Preparation of crude protein extracts

Cell pellets were resuspended in ice-cold extraction buffer (100 mM potassium phosphate buffer pH7·0 [6·15 ml 1M K_2HPO_4, 3·85 ml 1M KH_2PO_4 in 100 ml], 5 mM EDTA, 1 mM benzamidine, 0·4 mM DTT; benzamidine and DTT added immediately before lysis) and lysed by passage through an automatic French Pressure cell, under 750 psi pressure. The suspension was centrifuged at 10,000 g for 15 mins at 4°C, and the supernatant isolated.

2.5.2 Assay for activity of 3-dehydroquinase

The activity of the dehydroquinase enzymes in the lysed cells was assayed by measuring the change in OD_{234} caused by the addition of DHQ to the cell supernatant (DHQ and DHS absorb at a wavelength of 234 nm).
The assay solution consisted of: 50 mM Tris-HCl, pH 8·0
0·5 mM DHQ [White *et al.*, 1990]
supernatant from the cell lysate (typically 1-5 µl)

The supernatant was added to the solution of Tris-HCl in dH_2O in a 1 ml glass cuvette, which was inverted approximately three times to mix the contents, and the absorbency

at 234 nm set at zero. The substrate (dehydroquinate) was added, the cuvette inverted again, and the absorbency measured on a spectrophotometer at 234 nm for 30 secs. The rate of change in the absorbency was measured at least three times per sample, and an average calculated.

Since the type I dehydroquinase enzyme is thermolabile, whilst the type II enzyme is thermostable, supernatants were heated at 70°C for 5 mins to denature the typeI enzyme. *E.coli* supernatants were centrifuged (12,000 g for 15 mins) to remove denatured protein from the supernatant.

To analyse the inhibitory effect of chloride anions on the activity of the dehydroquinases, 100 mM potassium phosphate (pH 7·0) [Chaudhuri *et al.*, 1986] was used as the assay buffer in the absence or presence of various concentrations of KCl. The absorbency was again set to zero prior to the addition of substrate to the cuvette, and the reaction measured at 234 nm.

2.5.3 Estimation of protein concentration

This was determined by the method of Bradford (1976) using bovine serum albumin (BSA) as a standard. 950 µl Bradford's reagent was added to the BSA, and the volume made up to 1 ml by the addition of dH_2O, and the OD_{595} measured approximately 30 mins after the reaction was set up. A standard curve of OD_{595} versus µg BSA was plotted, and protein concentrations of the cell lysates estimated from this curve.

2.5.4 Denaturing polyacrylamide gel electrophoresis of proteins

Electrophoresis of the supernatant from the cell lysate was performed in the presence of 0·1% (w/v) SDS by the method of Laemmli (1970) in a 5% (w/v) stacking gel and a 12% (w/v) running gel. A ratio of 30 : 0·8 acrylamide : bis-acrylamide was used in the gel, and polymerisation was induced by the addition of 0·1% (w/v) fresh ammonium persulphate and 0·1% (v/v) TEMED.

100 µl 2x protein loading buffer was added to an equal volume of cell lysate, and the mixture boiled for 10 mins, followed by centrifugation (12000 g for 5 mins at 4°C). Extracted supernatant was loaded onto the gel, and the gel electrophoresed at a constant current of 20 mA. A molecular weight ladder, supplied by Sigma Chemical Co., was also run on the gel to allow estimation of subunit Mr.

The gel was stained after electrophoresis with Coomassie blue (0·1% (w/v) Coomassie brilliant blue G250 in dH_2O) for 10 mins, and was destained in 10% (v/v) acetic acid, 10% (v/v) methanol.

2.5.5 Analysing levels of pathway intermediates

Approximately 100 µg of protein from a crude extract, or 20 µl of a cell culture supernatant diluted 1:1 with dH_2O, were run down an Aminex HPX-87H Organic Acids Analysis column (300 x 7·8 mm I.D.; 9 µm particle diameter) on an HPLC in 5 mM H_2SO_4 at a flow rate of 1 ml per minute for 45 mins to measure the levels of five of the intermediates (DHQ, DHS, shikimate, S3P and EPSP) and the product (chorismate) from the shikimate pathway. The absorbency of the column eluate was measured at 215 nm on a spectrophotometer. The retention times for these six compounds had already been estimated under identical conditions [Mousdale & Coggins, 1985]. 100 pmoles of DHQ and shikimate were run separately on the column to allow correlation of the reported times with the estimated times derived from the chromatograms. The concentrations of intermediates and product were measured in mAbs units.

CHAPTER 3

Attempted disruption of the *aroQ* gene of *S.coelicolor*

3.1 Introduction

One of the ultimate aims of the shikimate project in this laboratory is to increase the rate of flux through the shikimate pathway, in order to provide higher levels of the intermediates and products of this pathway, which may then be utilised in the synthesis of primary and secondary metabolites. In order to do this, we wished to study the metabolic flux through this pathway, to allow identification of the enzymes with the greatest flux control coefficients, and therefore the enzymes which exert the greatest level of control over the rate of flux through the whole pathway (see section **1.12**). These enzymes provide the most obvious targets for upregulation, in an attempt to ultimately increase the flux through the pathway.

Dehydroquinase, which converts dehydroquinate to dehydroshikimate, catalyses the third step of the shikimate pathway. In *S.coelicolor*, the dehydroquinase enzyme is a type II enzyme, which has a high K_m (650 μM), is thermostable, and functions as a dodecamer [White *et al.*, 1990]. The *aroQ* gene, which encodes this type II enzyme in *S.coelicolor*, has previously been cloned and sequenced in this laboratory [P. White, unpublished results].

The high K_m of the *S.coelicolor* dehydroquinase identifies it as being relatively poor at sequestering its substrate (DHQ), which may produce an increase in the pool size of DHQ in the cell, and suggesting that it may have a high flux control coefficient. Therefore, under conditions of upregulation of the pathway, which may occur immediately prior to, or even during, differentiation and secondary metabolism, the dehydroquinase enzyme may impose a limitation on the rate of flux through the pathway.

One of the aims of this project was to replace the dehydroquinase gene with the *aroQ* gene of *E.coli* [Kinghorn *et al.*, 1981], which encodes a type I dehydroquinase enzyme [Chaudhuri *et al.*, 1986] whose K_m is approximately one twentieth that of the enzyme from *Streptomyces* [White *et al.*, 1990]. Thus, the type I enzyme is much more effective at sequestering its substrate than the type II enzyme, and may therefore have a lower flux control coefficient. Therefore, functional replacement of the indigenous type II dehydroquinase with the type I enzyme of *E.coli* may produce an increase in flux through the shikimate pathway.

3.1.1 Transformation of *S.coelicolor* with DNA derived from *E.coli*

Numerous replicative vectors carrying a variety of different antibiotic resistance

markers, multiple cloning sites and expression systems have been created for use in streptomycetes. However, it is still desirable to perform certain subcloning manipulations in *E.coli*, and to then transform the constructs into a particular *Streptomyces* strain. Transfer of plasmids from *E.coli* to *Streptomyces* also allows the use of both bifunctional and suicide (integrative) vectors, which have numerous applications in streptomycete research. However, many prokaryotes have developed methods for distinguishing their own DNA from foreign DNA, particularly through the use of chemical modification of the host DNA. They have also developed defence mechanisms which cause the degradation of DNA which has been recognised as foreign, preventing its replication and the expression of foreign genes. The streptomycetes in particular have effective modification and defence systems.

The *E.coli dam* gene encodes a Dam methylase which modifies the sequence GATC, converting the adenine base to N^6-methyladenine. Dam methylation is involved in the repair of DNA mismatches, and also in the modulation of gene expression and DNA replication [Palmer & Marinus, 1994]. The *dcm* gene of *E.coli* encodes a Dcm methylase which modifies the sequence CC(A/T)GG, converting the internal cytosine residue to 5-methylcytosine.

A number of *Streptomyces* strains contain methyl-dependent restriction systems: *S.coelicolor* has been found to contain a restriction system sensitive to Dam methylation [MacNeil, 1988] and digests any DNA containing such a methylation; *S.rimosus* restricts Dcm-methylated DNA; *S.avermitilis* restricts DNA modified by either Dam or Dcm methylases. Therefore, *S.coelicolor*, *S.rimosus* and *S.avermitilis* cannot be transformed with DNA from the majority of *E.coli* strains. *S.lividans*, however, does not methylate its DNA in such a way, so that its DNA may be used to transform protoplasts of other *Streptomyces* species. Also, *S.lividans* does not restrict DNA which has been modified by either Dam or Dcm methylases, and may therefore be transformed with DNA from all *E.coli* strains.

During the experiments presented in this thesis, two *E.coli* strains, MB5386 and CB51, were used to provide DNA which was not Dam-methylated for transformation into *S.coelicolor*. MB5386 contains a transposon, Tn*9*, disrupting the *dam* gene, and a mutant copy of the *dcm* gene, preventing expression of either the Dam or Dcm methylases (MacNeill, 1988). The Tn*9* transposon may be selected for by growth on chloramphenicol-containing medium, as the transposon contains the *cat* gene, which encodes a chloramphenicol acetyl transferase enzyme, and therefore mediates resistance to chloramphenicol. CB51 contains a mutant *dam* gene, so does not methylate the adenine residue in the GATC sequence. Therefore, DNA isolated from either of these

strains is able to avoid the restriction barrier of *S.coelicolor.*

In order to perform a functional replacement of the host dehydroquinase enzyme, it was necessary to first disrupt the *aroQ* gene of *S.coelicolor*, which encodes the host's type II dehydroquinase enzyme. Gene disruption would also allow analysis of the first two steps of the pathway, catalysed by DAHP synthase and DHQ synthase, in isolation, and enable analysis of whether either of these genes is a suitable target for replacement, or whether any of the other upregulation methodologies, such as overexpression, may be advantageous in increasing the flux through the pathway.

3.1.2 Gene disruption in *S.coelicolor*

Two different gene disruption strategies were available; integration of an internal fragment of the gene, and functional replacement with a mutant gene copy. Both of these methods are discussed below:

3.1.2.1 Disruption via integration of an internal fragment

In this method, an internal fragment of a cloned gene is inserted into a non-replicative (suicide) vector, the plasmid is transformed into the strain targeted for disruption, and the internal fragment integrates by Campbell-type homologous recombination between the cloned insert and its chromosomally-located parental gene copy; otherwise the plasmid is lost as it is unable to replicate in the chosen host. This recombination event produces two gene copies; one with an N-terminal deletion, the other with a C-terminal deletion. The vector sequence lies between these two terminally-deleted gene copies. If the cloned internal fragment has either sufficiently large deletions, or deletions of essential fragments, at each end of the gene, then neither terminally-deleted copy of the gene will encode a functional protein, and disruption will have been effected successfully. The existence on the vector of an antibiotic-resistance marker which is functional in the transformed strain enables selection for such a disruption event, as the vector is unable to replicate stably in the host, so resistant transformants must contain a copy of the plasmid integrated into the chromosome. Legitimate recombination should only occur between the chromosomally-located parental gene copy and the internal gene fragment on the plasmid, as no other regions of the suicide vector are likely to bear any significant homology to the genome of the host.

Suitable vectors for the above disruption strategy include ColE1-based vectors, naturally-occurring temperature-sensitive host vectors, and *attP*-deleted ØC31 vectors:

ColE1-based vectors, including the *E.coli* pUC-based vectors, are unable to replicate in streptomycetes, and so can only survive in a *Streptomyces* host by integration into the chromosome. Such vectors are suitable for use as suicide vectors in *Streptomyces*. It has been demonstrated that suicide vectors are able to transform into protoplasts of *S.coelicolor* and to integrate into the chromosome, but at a much lower frequency (circa $2{\cdot}5x10^4$ reduction)than replicative double-stranded plasmid DNA [Smith, 1992].

Temperature-sensitive vectors will replicate stably in their host cells at permissive temperatures. However, they are unable to replicate at elevated, non-permissive temperatures. Therefore, under such conditions, they are only able to survive in the host by integration into the chromosome. Thus, temperature-sensitive plasmids are suitable for use as both expression and suicide vectors. This strategy was utilised for disruption of the *pat* gene, encoding resistance to phosphinothricyl-alanyl-alanine (PTT), of *S.viridochromogenes* using the naturally-occurring temperature-sensitive plasmid pSG5 from *S.ghanaensis* [Muth *et al.*, 1989]. However, no naturally-occurring temperature-sensitive plasmids have been isolated from *S.coelicolor*.

ØC31 is a temperate *Streptomyces* bacteriophage which can undergo a lytic or lysogenic infection of its host. During a lytic infection it does not integrate into the chromosome, remaining as an extra-chromosomal DNA molecule, and lyses its host to release progeny phage. However, when ØC31 enters its lysogenic life-cycle, it usually integrates site-specifically into the chromosome at the *attB* site due to homologous recombination between the *attB* site and the *attP* site of the phage, with the two sites exhibiting almost complete identity. Deletion of the *attP* site from the phage prevents this site-specific recombination event from occurring, so that lysogens are only produced through homologous recombination between the insert and its chromosomally-located parental gene copy. Disruption using this strategy has already been used successfully in *S.coelicolor*, for example in the disruption of the *hrd* genes, which encode putative sigma factors [Buttner *et al.*, 1990].

Disruption via integration of an internal gene fragment requires as little as 620 bp of homologous DNA for the integration event to occur, as shown by the disruption of *hrdD* using a ØC31 lysogen [Buttner *et al.*, 1990]. Unfortunately, the coding sequence of the *aroQ* gene of *S.coelicolor* is only 639 bp in length. In order to produce non-functional, terminally-deleted proteins by this method, sufficient sequence from each end of the gene must be deleted. This would considerably reduce the size of the coding region in the *S.coelicolor aroQ* gene, possibly to a size which is unable to undergo recombination with the chromosomal copy. Also, integration of ØC31 is performed by an integrase protein which it encodes, and which increases the frequency of

recombination.

It is also possible that, in *S.coelicolor*, *aroQ* forms part of a bifunctional or multifunctional operon. Disruption by this method produces a gene deleted at its 3' end, with the integrated ØC31 or plasmid sequence and the 5'-deleted gene lying downstream of it. Such an integration event is likely to prevent transcription through the 5'-deleted *aroQ* gene copy, due to the length of the intervening sequence, and therefore is also likely to prevent transcription into any downstream gene which may be contained within this operon. It is also possible that the 5'-deleted *aroQ* gene copy and any downstream genes will be brought under the control of ØC31- or plasmid-encoded regulatory sequences, causing abnormal expression of any gene(s) downstream of *aroQ*. Disruption via gene replacement is less likely to suffer from either of these problems. For these reasons, this strategy for disruption of the *aroQ* gene of *S.coelicolor* was not attempted. The sequences upstream and downstream of *aroQ* are currently being sequenced in this laboratory, possibly followed by transcriptional analysis.

3.1.2.2 Disruption via gene replacement

Through this method, the functional chromosomally-located gene is replaced by a mutant plasmid-borne gene copy. The plasmid used is often a non-replicating suicide vector, or may be an unstable host vector. The mutation may be due to an internal deletion, a substitution (often performed by PCR mutagenesis), or an insertion [for example, Buttner & Lewis, 1992]. This strategy requires two recombination events to occur between the two gene copies, one on either side of the mutation. It allows the mutation to be effectively exchanged from the gene copy carried on the plasmid into the chromosomal copy, with the functional gene copy being moved onto the plasmid (**Figure 3.1**). The first recombination event causes integration of the plasmid into the chromosome; the second event liberates the plasmid, carrying either the mutant copy (therefore regenerating the original, non-mutated host) or the functional gene copy (therefore causing the production of a disrupted mutant).

If a suicide vector is used to carry the mutated gene copy, the plasmid will be lost from the cells after the second recombination event due to its inability to replicate in the host. If an unstable vector is used, it may survive in the cells for a few generations, but is likely to be lost quickly from the population if grown under non-selective conditions.

Cells into which the plasmid has integrated by a single recombination event are likely to eventually lose the plasmid via a second homologous recombination event when grown

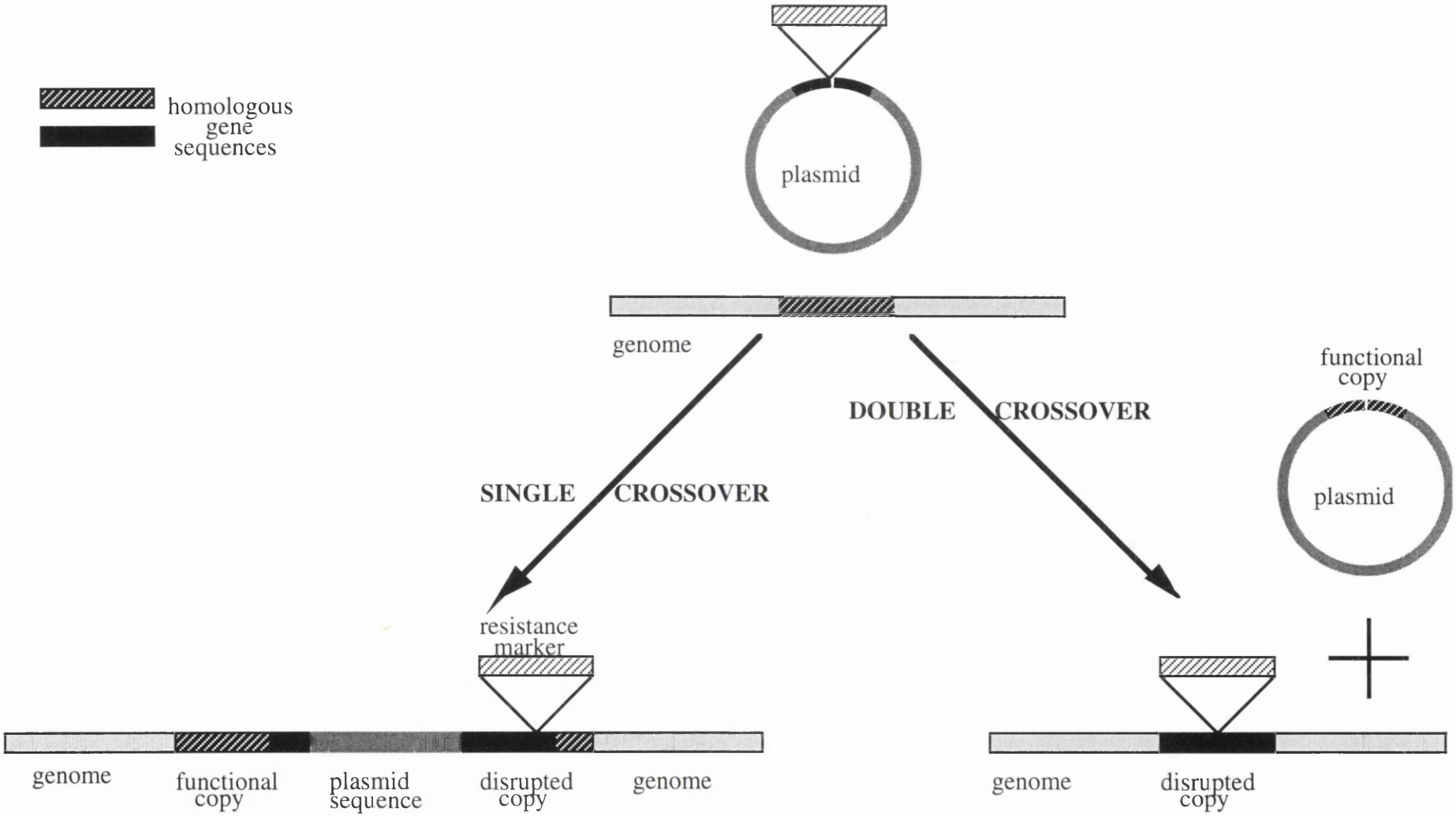


Figure 3.1 Strategy for disruption via gene replacement

under conditions which do not select for maintenance of the integrated copy. However, it is also possible to promote the secondary recombination event in certain *Streptomyces* strains through protoplasting (which ensures that only a single copy of the genome is present in each cell, and therefore increases the ease of identification of the desired cell) and regeneration on non-selective media (which releases the selection pressure, and allows a secondary recombination event to occur). It has been demonstrated that the required secondary recombination event occurs in approximately 3 to 4% of colonies after a single round of sporulation on non-selective solid media in *S.coelicolor* [Smith, 1992]. This strategy was used for disruption of the gene encoding the fifth step of the bialaphos biosynthetic pathway in *S.hygroscopicus* [Anzai *et al.*, 1988].

Insertion mutagenesis is performed by the stable insertion of a transposon, or by the subcloning of an antibiotic-resistance gene, into the coding region of the plasmid-borne gene which has been targeted for disruption. The resistance gene should be subcloned into a restriction site as close as possible to the centre of the insert fragment carried on the vector, in order to allow as much of the coding sequence on either side of the resistance marker to function as a template for homologous recombination. This method also carries the disadvantage that it is likely to interfere with the transcription of any genes lying downstream of, and transcriptionally coupled to, the disrupted gene.

3.1.3 Transformation of *S.coelicolor*

There are currently four methods available for transformation of *Streptomyces*:

3.1.3.1 Transformation of protoplasts with double-stranded plasmid DNA

This basically follows the PEG-mediated method of Hopwood *et al.* [1985]. Non-methylated plasmid DNA is transformed into protoplasts, which are plated onto R2 regeneration medium and incubated at 30°C. Selection is achieved by overlaying the plates after 16-20 hrs with 1 ml of 10·3% sucrose solution containing the relevant antibiotic. Transformants typically appear after around 5 days, and sporulate after 7 to 10 days. *S.lividans* protoplasts are able to be transformed with DNA isolated from either *E.coli* or other *Streptomyces* strains. *S.coelicolor* protoplasts can be transformed with plasmid DNA derived from *S.lividans*, but may only be transformed with DNA from *E.coli* which has not been Dam-methylated. Transformation of protoplasts was the method most used during the work reported in this thesis.

Although it is possible to transform linearised plasmids into certain bacterial species

[for example in *E.coli*; Winans *et al.*, 1985] which may promote insertion of this plasmid DNA into the chromosome via homologous recombination, many species contain exonuclease enzymes which degrade such unprotected linear DNA (but not the giant linear plasmids which occur naturally in streptomycetes and whose ends are capped with protein, protecting them from such degradation [Chen *et al.*, 1993]). This transformation methodology, however, often requires mutations in the host genes encoding these exonucleases (*recB* and *recC* genes in *E.coli*) in the recipient strain to prevent degradation of the linear DNA. Unfortunately, such mutations are not available in *Streptomyces* strains.

An alternative way of preventing degradation of linear DNA molecules to mutations in exonuclease genes is the capping of the ends of the linear DNA, using one of two possible strategies: either with oligonucleotides which are able to ligate to both strands of each end of the plasmid, forming a looped end (analogous to a hairpin structure); or to cap the ends of the linearised plasmid with proteins, as occurs with naturally-occurring linear plasmids and linear chromosomes. However, neither of these systems have been reported in *Streptomyces* species. Therefore, transformation of protoplasts with linearised plasmid DNA is not currently considered as a viable experimental procedure in *Streptomyces*.

3.1.3.2 Transformation of protoplasts with single-stranded plasmid DNA

Single-stranded plasmid DNA can be obtained from vectors containing a bacteriophage f1 origin of replication by single-stranded rescue using a helper bacteriophage, for example M13K07, which has been mutated to promote an elevated level of single-stranded plasmid DNA isolated relative to phage DNA. Hillemann *et al.* [1991] have developed a method for gene disruption in *S.viridochromogenes* by transformation with single-stranded suicide vectors. These authors showed that transformation by single-stranded DNA increased integration by approximately 10- to 100-fold relative to transformation with double-stranded DNA from the same plasmid, with both systems showing approximately equal transformation efficiencies. They were able to utilise homologous fragments as small as 200 bp in length.

Single-stranded DNA which is transformed into *S.coelicolor* will undergo replication to produce a double-stranded plasmid. The daughter strand will not be methylated as it is synthesised by the replication machinery of *S.coelicolor*. If the single-stranded DNA used for transformation is from an *E.coli* strain containing a Dam methylase, it will be methylated at the adenine residue of the sequence GATC. The plasmid, therefore, will

be hemi-methylated. However, *S.coelicolor* has been shown to restrict hemi-methylated plasmid DNA [C. Smith, personal communication] as well as fully methylated DNA. Although single-stranded plasmid DNA derived from a methylating host will produce transformants of *S.coelicolor*, it does so at a greatly reduced frequency (circa $2x10^5$ reduction) compared to replicating double-stranded plasmid DNA [Smith, 1992].

The *S.coelicolor* restriction-methylation barrier can be evaded by transformation with DNA isolated from a non-methylating *E.coli* strain (for example MB5386 or CB51). However, none of these non-methylating *E.coli* strains currently contain the F plasmid, which is required for isolation of single-stranded plasmid DNA. Therefore, all single-stranded DNA transformed into *S.coelicolor* will form a hemi-methylated double-stranded plasmid which will be restricted to produce a linear molecule, and will subsequently be degraded by exonucleases. Thus, although this method of transformation was briefly attempted as part of this project prior to the discovery that hemi-methylated DNA is degraded in *S.coelicolor*, a method for transformation of *S.coelicolor* by single-stranded DNA is currently unavailable.

3.1.3.3 Transconjugation of spores

Transconjugation of plasmids between various *Streptomyces* strains has long been known to occur [Hopwood *et al.*, 1969]. However, transfer of plasmids by conjugation has also been shown to occur from *E.coli* to a variety of streptomycetes using a bifunctional plasmid [Mazodier *et al.*, 1989]. This method of transfer requires the presence on the plasmid of a *mob* (mobilisation) region, which contains an *oriT* (origin of transfer), and a *tra* gene supplied *in trans*. The *tra* gene from the RP4 plasmid (an IncP plasmid) is required for mobilisation of the plasmid, and is often integrated into the chromosome, whilst *oriT* is the region of the plasmid at which the initial incision is made to produce the linearised DNA molecule which is passed across the "conjugation tube".

Mazodier *et al.* isolated plasmid DNA from exconjugants of *S.lividans* when matings were performed on a solid medium, but conjugation did not appear to occur when the donor (*E.coli*) and recipient (*S.lividans*) were grown together in liquid medium. Also, no transfer to mycelia was observed, and the efficiency of transfer to spores was increased 5- to 10-fold by pregermination of the spores.

It has also been demonstrated that an IncQ plasmid (RSF1010) is able to transfer from *E.coli* to both *S.lividans* and *Mycobacterium smegmatis* (another Gram-positive

bacterium which is very closely related to the streptomycetes), and to self-replicate stably in both species [Gormley & Davies, 1991]. IncQ plasmids appear to be relatively independent of the host bacterium's replication machinery, supplying the enzymes necessary for the initiation and propagation of their own replication. They require only DNA polymerase III from an *E.coli* host, and are able to stably replicate in *S.lividans*. For mobilisation, they also require the *tra* gene to be supplied in *trans* in the donor strain. Again, *S.lividans* exconjugants contained plasmids which were restricted with *Bcl* I, and therefore were not methylated.

3.1.3.4 Electroporation of mycelia

Streptomyces rimosus mycelia exhibiting well-dispersed growth in a complete liquid medium may be harvested, treated with lysozyme, and transformed with up to 2 μg of plasmid DNA if subjected to a brief, high-voltage electric discharge [Pigac & Schrempf, 1995]. The electrical pulse results in the formation of transient pores in the membrane, through which uptake of DNA can occur. Transformants are plated, after growth in non-selective liquid medium to allow expression, onto antibiotic-containing medium to select for transformants. These authors claim that this method has advantages relative to transformation of protoplasts, including a 10-fold increase in transformation efficiency, increased reproducibility, easier preparation of transformable cells, and growth of transformants within 72 hours.

This transformation methodology is currently being applied to a range of *Streptomyces* strains. However, in this laboratory, suitably-dispersed growth of mycelia cannot currently be achieved with *S.coelicolor*, so that we are unable to obtain enough electrocompetent mycelia to achieve a sufficient transformation efficiency. Therefore, this method of transformation has not been utilised in the experiments presented in this thesis. Based on all of the knowledge described above, the strategy for gene disruption chosen in the work described below was insertion mutagenesis with a gene encoding antibiotic resistance, using an *E.coli* (suicide) plasmid as the transmission vector.

3.2 Results

3.2.1 Attempted disruption of *aroQ* with *ermE*

3.2.1.1 Disruption of a plasmid-borne *aroQ* gene with *ermE*

The *aroQ* gene of *S.coelicolor*, which had previously been cloned in this laboratory [P.

White, unpublished results], encodes a type II dehydroquinase enzyme. It is contained within an approximately 3 kb *Hin*d III/*Sst* I fragment, which was previously subcloned into the vector pIBI25 to generate pPW63. p113 is a pIBI24-based vector, created by replacement of the *Ssp* I fragment of the vector with the *tsr* gene [H. Hacker, personal communication].

The *tsr* gene, which was cloned from *S.azureus*, functions as a selectable marker in *S.coelicolor* by inducing resistance to thiostrepton by modification of ribosomes [Thompson *et al.*, 1986]. It encodes a rRNA methylase which methylates an adenosine residue in the ribosomes, converting it to 2'-O-methyladenosine, and preventing the binding of thiostrepton to the ribosomes [Thompson *et al.*, 1982c].

The *ermE* gene was originally cloned from *Streptomyces erythraeus* (subsequently reclassified as *Saccharopolyspora erythraea*) [Thompson *et al.*, 1982a]. It encodes a rRNA methylase which modifies adenine residues in the ribosomes, converting them to N^6,N^6-dimethyladenine [Thompson *et al.*, 1982c]. This modification confers resistance to both the macrolide antibiotic erythromycin, and to lincomycin.

The 3 kb *Hin*d III/*Sst* I fragment from *S.coelicolor*, which contains the *aroQ* gene, was isolated from pPW63 and subcloned into p113, creating plasmid pPM1 (**Figure 3.2**). The *ermE* gene of plasmid pIJ4026 was excised on a 1·8 kb *Bgl* II fragment. It was used to disrupt the *aroQ* gene of *S.coelicolor* by subcloning this fragment into the unique *Bgl* II site of pPM1, which occurs inside the coding region of *aroQ*, generating the plasmid pPM3.

pPM3 contains a unique *Pst* I site which lies in a region of the multiple cloning site transferred from pIJ4026 during the subcloning of *ermE*, and lies upstream of the *ermE* gene on the *Bgl* II fragment. To allow this plasmid to be transferred from *E.coli* strains containing a *tra* gene by conjugation, a plasmid *mob* region containing an *oriT* was subcloned into the unique *Pst* I site of pPM3, creating plasmid pPM4 (**Figure 3.2**).

The above subcloning events were performed in *E.coli* DS941. However, this strain fully methylates plasmid DNA contained within it, and so plasmid DNA derived from it cannot be used for transformation of protoplasts of *S.coelicolor*. Therefore, plasmid pPM3 was transformed into competent *E.coli* MB5386 cells, and selected for by growth on L agar supplemented with ampicillin. Ampicillin-resistant transformants of MB5386 were streaked onto L agar supplemented with chloramphenicol to ensure that these transformants were true *dam*$^-$ cells. After the efficacy of this mutation had been ascertained, non-methylated pPM3 was isolated for transformation of protoplasts of

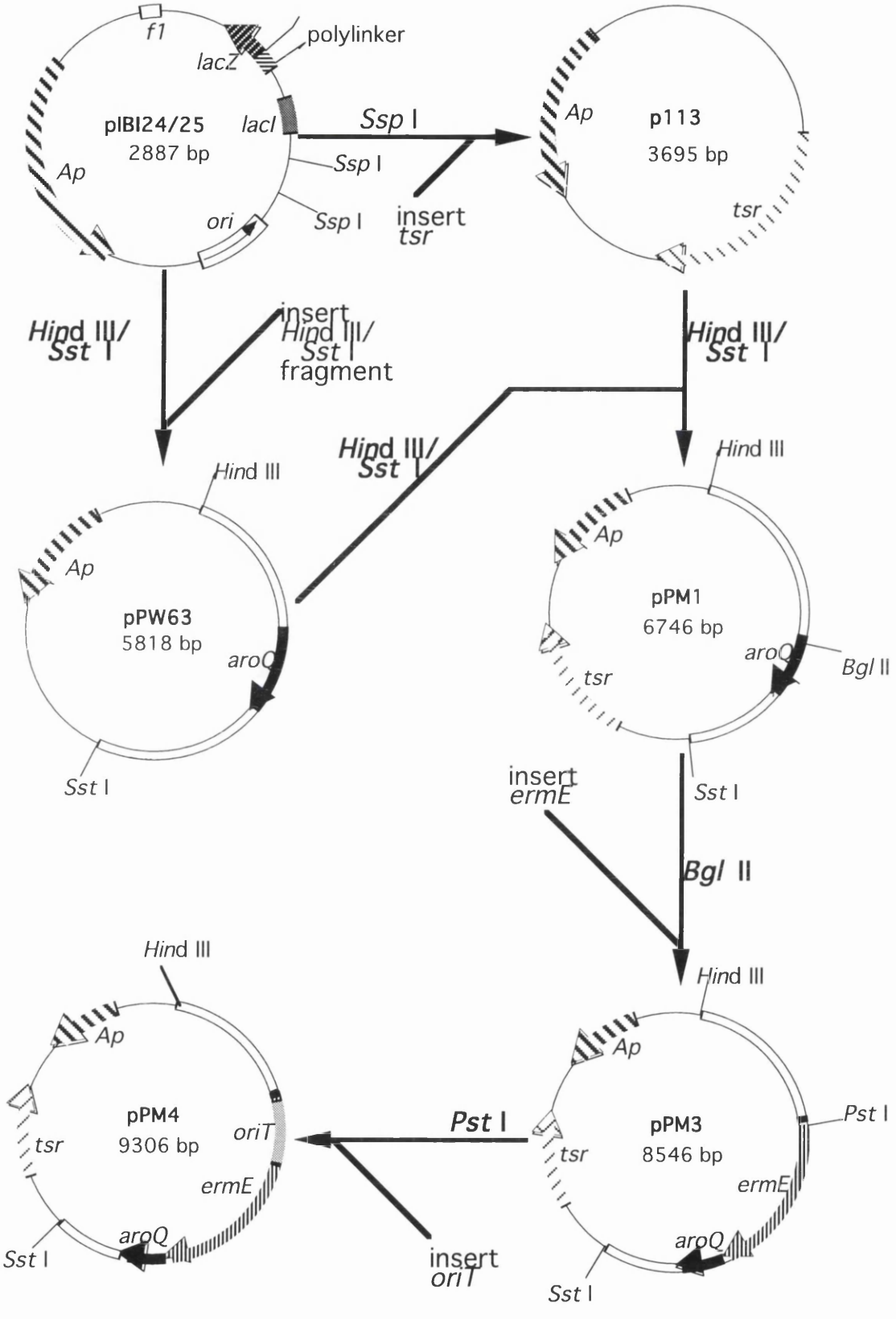


Figure 3.2 Disruption of a plasmid-borne *aroQ* gene with *ermE*

N.B. Restriction sites in bold denote digestions performed during subcloning

S.coelicolor.

3.2.1.2 Selection for *ermE*

Spores of *S.coelicolor* 209 were streaked onto Soya agar plates supplemented with lincomycin at concentrations of 200, 400, 600, 800 μg/ml and 1 mg/ml. They were also streaked onto Soya plates supplemented with erythromycin at a concentration of 200 μg/ml. Although no spores grew on the plates in the presence of erythromycin, they grew on all of the lincomycin-containing plates. Therefore, erythromycin was preferred for selection of the presence of *ermE* in *Streptomyces* during the experiments described below.

Spores of *S.coelicolor* 1147 were plated onto Soya agar plates supplemented with between 1 and 5 mg erythromycin per plate (40-200 μg/ml). Colonies grew on plates containing 1 mg and 2 mg erythromycin only. Therefore, transformed protoplasts of *S.coelicolor* were overlaid with 4 mg erythromycin per plate, or plated onto agar containing 200 μg/ml (as suggested for *S.lividans* [Hopwood *et al.*, 1985]).

3.2.1.3 Transformation of protoplasts of *S.coelicolor* 1147 with pPM3

pPM3 plasmid DNA was isolated from the non-methylating *E.coli* strain MB5386 [MacNeil, 1988], and transformed into high-efficiency protoplasts of *S.coelicolor* 1147 with or without digestion with *Pst* I endonuclease. Transformation with a linear DNA molecule is known as a "hit and run" transformation. The transformation mixtures were plated onto R2 regeneration plates, incubated at 30°C, and overlaid either with erythromycin to select for both single- and double-recombinants, or with thiostrepton to select for single-crossovers. No colonies grew up on the plates which were overlaid with thiostrepton. However, two colonies from each of the *Pst* I-restricted and unrestricted transformations grew on erythromycin plates. Transformants were streaked again onto erythromycin-containing and thiostrepton-containing Soya agar plates to test the efficacy of selection of protoplasts, and again no growth was observed on thiostrepton; three of the colonies (including only one of those transformed with digested plasmid) grew up on the erythromycin-containing plates. The resultant colonies were patched onto both erythromycin-containing and thiostrepton-containing Soya agar plates. Two colonies were isolated which were resistant to erythromycin, but not to thiostrepton, indicative of a double-crossover. To again test the efficacy of selection, these two isolates were re-streaked onto erythromycin- and thiostrepton-containing plates and, as anticipated, grew only on erythromycin-containing plates.

To check for any auxotrophic requirements of such putative disruptants, the two erythromycin-resistant isolates and an untransformed 1147 colony were streaked onto minimal medium with and without a supplement of the aromatic amino acids. All three strains grew on the supplemented medium, but only 1147 grew on medium lacking the aromatic amino acids (phenylalanine, tryptophan and tyrosine). This suggested that *S.coelicolor* strains lacking a functional dehydroquinase enzyme require an exogenous supply of the aromatic amino acids for growth, as the shikimate pathway is unable to produce chorismate, thus preventing the synthesis of these amino acids which are required for growth.

Chromosomal DNA was prepared from the two putative dehydroquinase-disrupted strains and untransformed 1147, approximately 1·5 μg of each was digested with either *Hin*d III or *Sst* I restriction endonucleases, and was run out on a 0·8% TBE agarose gel. This DNA was Southern-blotted onto nylon membrane, and the membrane was probed with a radiolabelled *ermE*-containing probe (the 1·7 kb *Kpn* I fragment of pIJ4026). However, no DNA band on the membrane lit up, suggesting that the *ermE* probe did not hybridise to any bands on the membrane, and therefore that neither isolate contains *ermE* (and hence a disrupted dehydroquinase gene), or that the labelling *ermE* probe or the probing of the filter was not successful. Unfortunately, no positive control had been included on the filter to test the efficiency of the probing.

It was concluded from this result that the erythromycin-resistant isolates arose by spontaneous mutation to resistance. This theory is supported by the fact that no thiostrepton-resistant colonies were obtained at any point during this experiment. Also, although two erythromycin-resistant colonies were produced by the "hit and run" transformation, it has since been demonstrated that exonuclease enzymes of *S.coelicolor* digest linear DNA, and so it is extremely unlikely that the linearised plasmid DNA would have been able to integrate into the chromosome.

Following the dismissal of these two colonies as containing an *aroQ* disruption, they were not investigated further.

3.2.1.4 Transconjugation of spores of *S.coelicolor* 1147 with pPM4

Plasmid pPM4 was transformed into *E.coli* strain S17-1, which contains a chromosomally-located *tra* gene, and is therefore capable of mobilisation of plasmids (including pPM4) carrying a *mob* region, which contains the origin of transfer, *oriT*.

Approximately 10^8 *S.coelicolor* 1147 spores were plated onto Soya agar plates along

with approximately $2x10^8$ S17-1/pPM4 cells, and incubated at 30°C. Plates were overlaid with erythromycin (which selects for transconjugants) and nalidixic acid (which kills *E.coli*, but to which *Streptomyces* are naturally insensitive). This produced an almost confluent lawn of putative transconjugants, which were replica-plated onto plates supplemented with either erythromycin or thiostrepton. Again, growth on erythromycin was nearly confluent, but considerably fewer colonies grew on thiostrepton-containing plates.

44 colonies from the thiostrepton-containing plates were streaked onto non-selective plates to allow a secondary recombination event to occur, and then streaked back onto plates containing either thiostrepton or erythromycin. All 44 colonies grew in the presence of both antibiotics. This round of non-selective growth followed by selective growth was repeated twice, to produce a single colony which was erythromycin-resistant, thiostrepton-sensitive (disruptant #1). After a further round of non-selective growth, another colony which had apparently undergone the required secondary recombination event was isolated (disruptant #2).

The two putative disruptants, a putative pPM4 integrant (erythromycin- and thiostrepton-resistant), and untransformed 1147 were grown up in YEME liquid medium supplemented with $MgCl_2$ (to 5 mM), glycine (to 0·5% w/v) and, for the putative disruptants, erythromycin. Total DNA was isolated from all four strains.

Approximately 5 µg total DNA from each of the four strains (untransformed 1147, putative single-crossover, and two putative mutants) was digested with either *Pst* I or *Sst* I, and electrophoresed on a 0·8% TAE agarose gel (**Figure 3.4**). The DNA was transferred to nylon membrane by Southern blotting, and the membrane probed successively with labelled *aroQ* (from *S.coelicolor*) and *ermE* genes.

When the filter was probed with the radiolabelled *aroQ* gene from *S.coelicolor* (the 3 kb *Hin*d III/*Sst* I fragment of pPW63), it was anticipated that the gDNA isolated from an untransformed colony would exhibit a single band to which the probe would hybridise; the gDNA isolated from the putative single crossover isolate would contain two bands of unknown size (totalling approximately 21·5 kb) when digested with *Pst* I, and three bands (totalling approximately 14·5 kb) to which the *aroQ* probe would hybridise when digested with *Sst* I; and the genomic DNA (gDNA) isolated from the putative disruptants would show two bands (totalling approximately 13·7 kb; this probe would not hybridise to the *oriT* fragment, which has *Pst* I ends) when the DNA was digested with *Pst* I, and two bands (of approximately 1·3 kb and 6·1 kb) when cut with *Sst* I

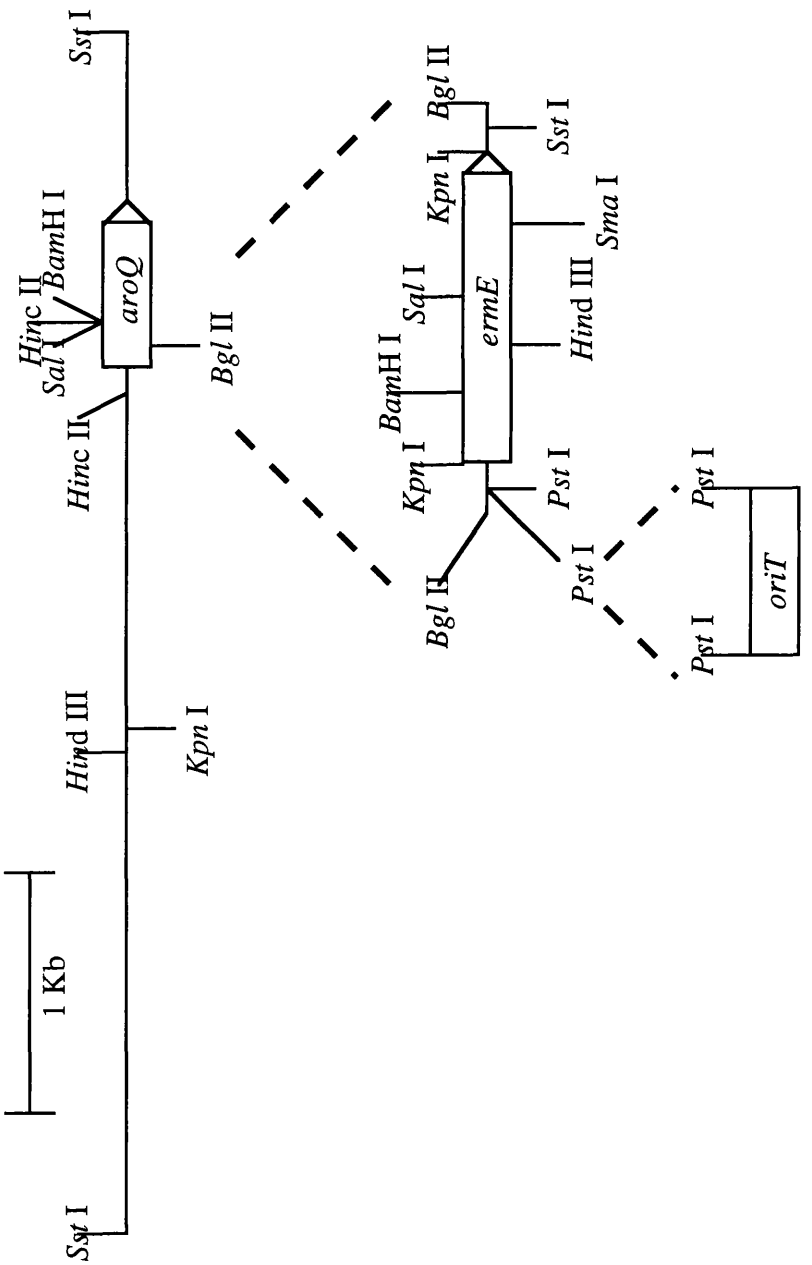


Figure 3.3 Predicted restriction map of *aroQ* disrupted by insertion of *ermE* and *oriT*

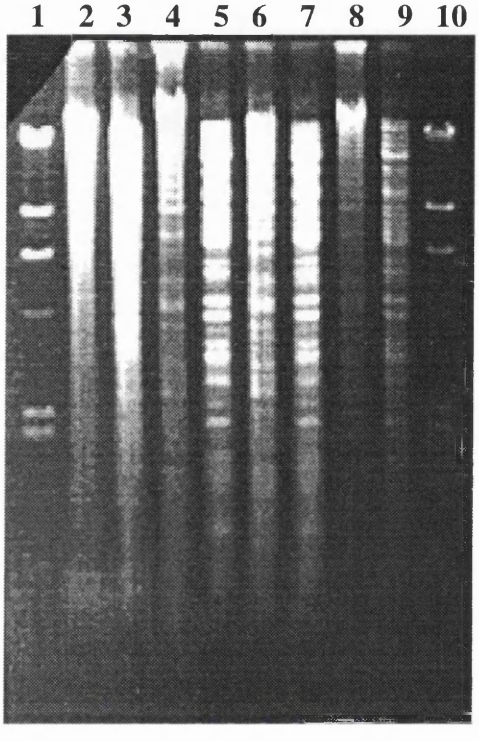


Figure 3.4 gDNA isolated from untransformed 1147, two putative disruptants and a putative 1147/pPM4 transformant

lanes 1 and 10: λ/*Hin*d III markers

lanes 2 and 3: gDNA from 1147 digested with *Pst* I and *Sst* I respectively

lanes 4 and 5: gDNA from disruptant #1 cut with *Pst* I and *Sst* I respectively

lanes 6 and 7: gDNA from disruptant #2 cut with *Pst* I and *Sst* I respectively

lanes 8 and 9: gDNA from 1147/pPM4 cut with *Pst* I and *Sst* I respectively

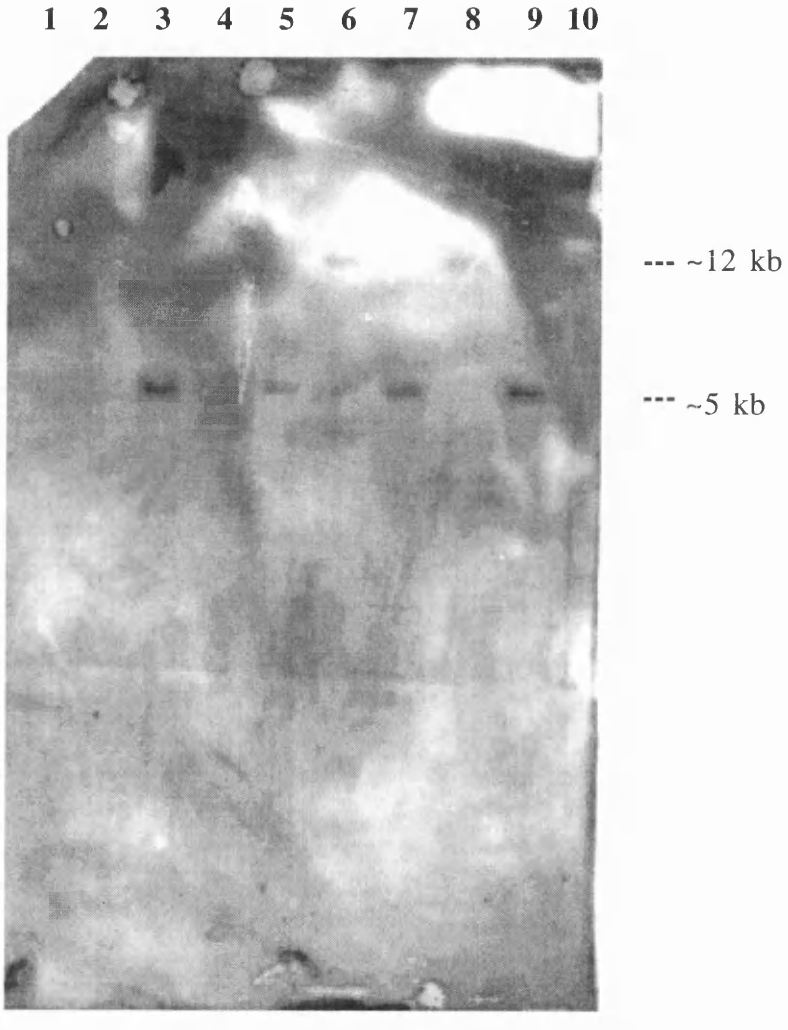


Figure 3.5 Autoradiogram of membrane carrying gDNA from 1147 isolates probed with a fragment containing the *aroQ* gene from *S.coelicolor*

(lanes as in **Figure 3.4**)

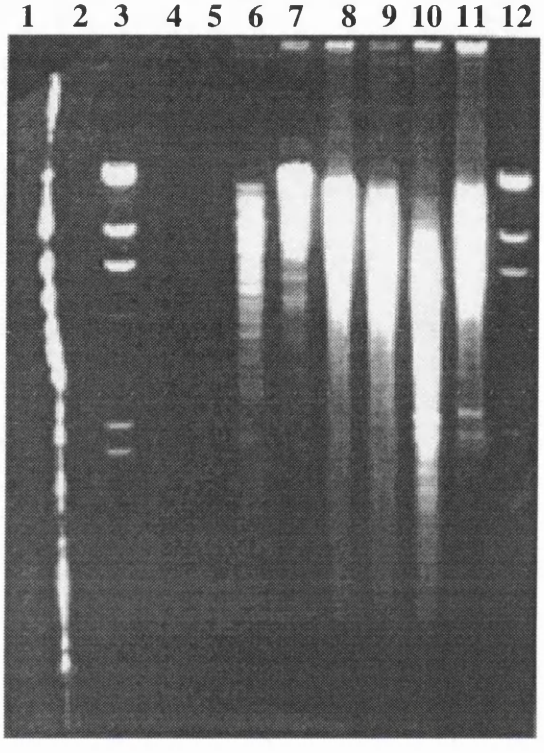


Figure 3.6 gDNA from two putative disruptants of 1147 and one single-crossover isolate

lanes 3 and 12: λ/*Hin*d III markers

lanes 1 and 4: 10 ng and 2 ng (respectively) of the *tsr* gene

lanes 2 and 5: 10 ng and 2 ng (respectively) of the *ermE* gene

lanes 6 and 7: gDNA from disruptant #1 cut with *Bcl* I and *Bgl* II respectively

lanes 8 and 9: gDNA from disruptant #2 cut with *Bcl* I and *Bgl* II respectively

lanes 10 and 11: gDNA from 1147/pPM4 cut with *Bcl* I and *Bgl* II respectively

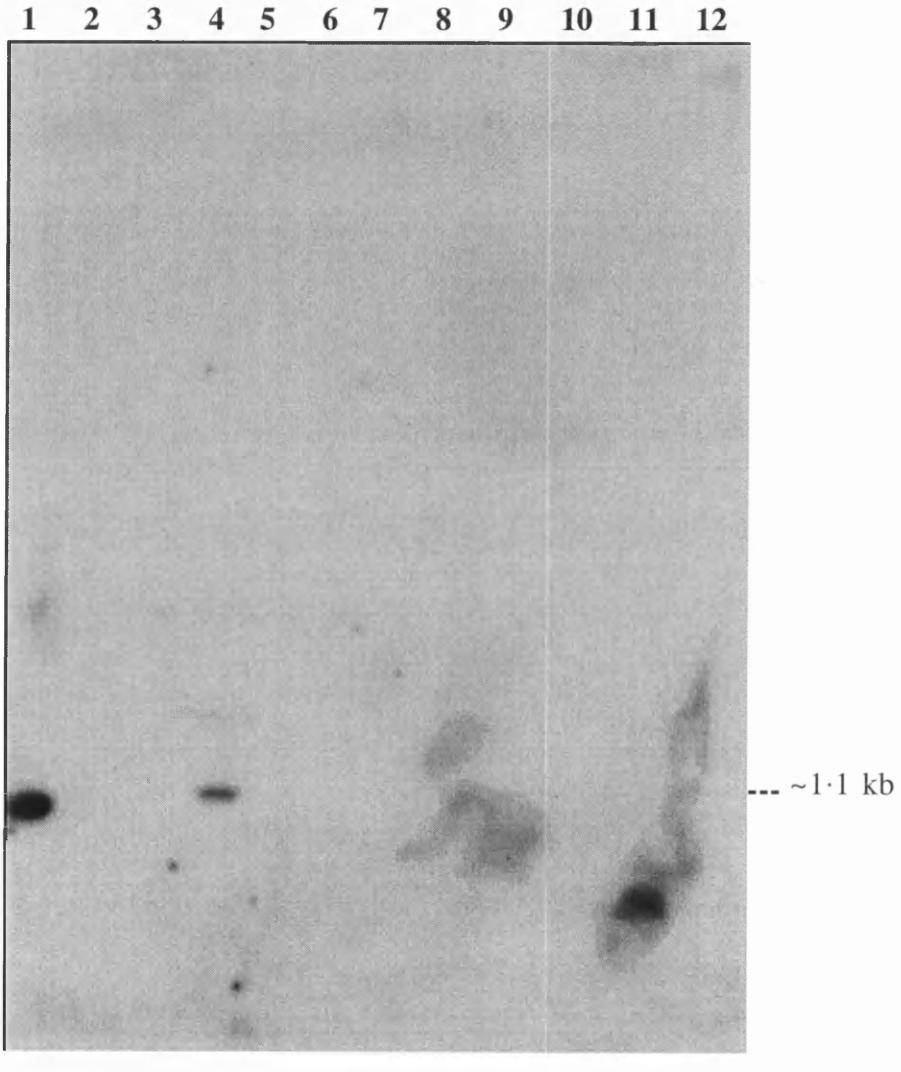


Figure 3.7 Autoradiogram of membrane carrying gDNA from 1147 isolates probed with a fragment containing the *tsr* gene

(lanes as in **Figure 3.6**)

(the "blob" in lane 11 is believed to be an artefact)

[see **Figure 3.3**]. However, all four isolates, when digested with either *Pst* I or *Sst* I, exhibited only the same single band on the autoradiogram (**Figure 3.5**).

Approximately 5 μg of gDNA isolated from the two putative disruptants and the putative 1147/pPM4 transformant were digested with either *Bcl* I or *Bgl* II, and run out on a 0·8% TAE gel. The *Streptomycete* genome is estimated at 6-9 megabases [Hopwood & Kieser, 1990]; the *tsr* gene is 1·1 kb in length, and would constitute around 0·015% of the genome in an integrant. Therefore, around 1 ng of *tsr* DNA (0.015% of 5 μg) would represent a single copy of the *tsr* gene in the gDNA digests. Approximately 10 ng and 2 ng of the unlabelled *tsr* and *ermE* probes were also run on the gel as positive controls (**Figure 3.6**). The DNA was subsequently transferred by Southern blotting to a nylon membrane. When the membrane was probed with the labelled *tsr* gene, it was anticipated that the probe would hybridise to the *tsr* control and to the gDNA isolated from the putative single-crossover. However, the probe apparently only hybridised to the *tsr* control, even though the membrane was exposed to X-ray film for up to 24 hrs at -70°C (lanes 1 and 4 in **Figure 3.7**). When the filter was probed with the *ermE* probe, it was anticipated that it would hybridise to the *ermE* control, the gDNA isolated from the putative single-crossover isolate and the gDNA isolated from the two putative double-crossovers. However, only the *ermE* control lit up on the autoradiogram. These results suggest that the two putative disruptants and the putative single crossover are actually wild-type *S.coelicolor* 1147. These results provide further evidence that *ermE* is a poor resistance marker in *S.coelicolor* 1147.

3.2.1.5 Transformation of protoplasts of *S.coelicolor* G216 with pPM3

The *S.coelicolor* strain G216 has the same genotype as that of strain 1147, but protoplasts isolated from it have a higher efficiency of transformation. Therefore, it was chosen to substitute for 1147 for disruption of the *aroQ* gene.

Protoplasts of G216 were transformed with pPM3 isolated from the non-methylating *E.coli* strain MB5386, the transformation mixture plated on R2 supplemented with the aromatic amino acids, and overlaid with 2 mg lincomycin. The excessive number of transformants (approximately 500 per plate) suggested that lincomycin was not an effective selecting agent in G216. Growth of non-transformed G216 spores on medium containing lincomycin supports this theory.

102 of the colonies which had grown on R2 overlaid with lincomycin were streaked onto Soya plates without antibiotic supplements, or Soya plates supplemented with either erythromycin or thiostrepton. Only four colonies grew on the erythromycin-

containing plates, and two colonies on the thiostrepton-containing plates.

The colonies from the non-selective plates were re-streaked onto plates containing either erythromycin or thiostrepton. Only 84 of these isolates showed any growth on erythromycin-containing plates, giving rise to 243 colonies, and just a single isolate growth on the plates containing thiostrepton. 207 erythromycin-resistant colonies were again patched onto plates containing either erythromycin or thiostrepton, with 162 growing on the plates containing erythromycin and 10 on the thiostrepton-containing medium.

19 erythromycin-resistant, thiostrepton-sensitive colonies and one double-resistant colony (G216/pPM3) were grown in YEME liquid medium supplemented with $MgCl_2$, glycine and erythromycin. Untransformed G216 was grown under similar conditions, but without the medium being supplemented with antibiotic. Total DNA was isolated from all twenty-one of these isolates. 10 μg of gDNA from each isolate was digested with *Sst* I and run out on a 0·8% TAE gel, along with 5 ng of each of the *ermE* and *tsr* genes (**Figure 3.8**). The DNA was transferred to a nylon membrane by Southern blotting, and the membrane probed with a 600 bp digoxygenin-labelled *aroQ* fragment. It was anticipated that a single band of approximately 5 kb in length would light up in gDNA isolated from untransformed G216; gDNA from a single-crossover would contain three bands (totalling approximately 13·7 kb in length); the double-crossover would show two bands, approximately 1·3 kb and 5·4 kb in size [see **Figure 3.3**]. However, all isolates showed the same single *aroQ* band, approximately 5 kb in length, as the untransformed G216 (**Figure 3.9**). Some of the isolates (particularly in lane **10** of **Figure 3.9**) appear to show two bands to which the *aroQ* probe hybridises. However, as this pattern of banding does not comply with any of the anticipated patterns, it is assumed that the higher of the two bands was an artefact, and was possibly present due to only partial digestion of some of the gDNA in these lanes.

The filter was stripped to remove the labelled *aroQ* probe, and probed with the *tsr* gene(the *Bam*H I/*Bgl* II fragment of pLus801) which had been labelled with digoxygenin. It was anticipated that the *tsr* control and the putative single-crossover would show a single band. However, only the *tsr* control lit up (**Figure 3.10**).

A further 10 μg of the total DNA from each isolate was digested with *SstI*, run out on a 0·8% TAE gel, along with 1 ng of the *ermE* gene (the 1·7 kb *Kpn* I fragment) as a control (**Figure 3.11**), and Southern blotted onto nylon membrane. The membrane was probed with a digoxygenin-labelled *ermE* fragment (the 1·7 kb *Kpn* I fragment of pIJ4026).

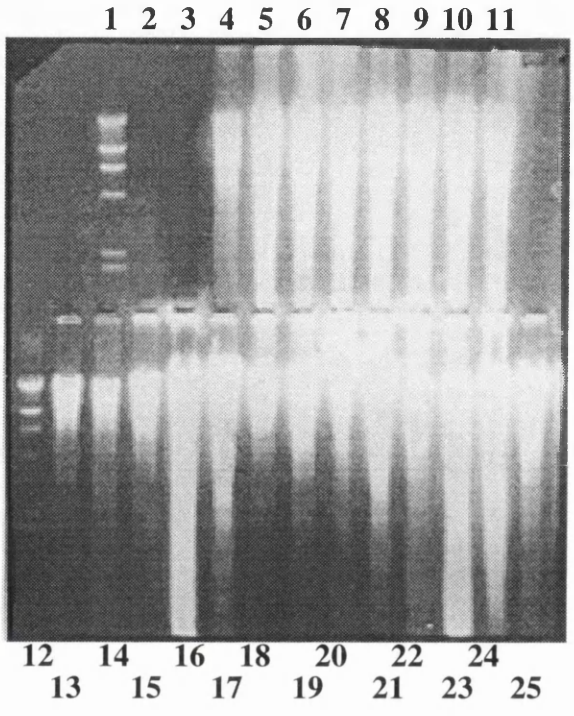


Figure 3.8 gDNA from G216 isolates

lanes 1 and 12: λ/*Hin*d III markers

lane 2: 5 ng *ermE* (1·7 kb *Kpn* I fragment of pIJ4026)

lane 3: 5 ng *tsr* (1·1 kb *Bam*H I/*Bgl* II fragment of pLus801)

lanes 4-11 and 13-23: gDNA from putative disruptants, cut with *Sst* I

lane 24: gDNA from putative G216/pPM3 isolate, cut with *Sst* I

lane 25: gDNA from untransformed G216 isolate, cut with *Sst* I

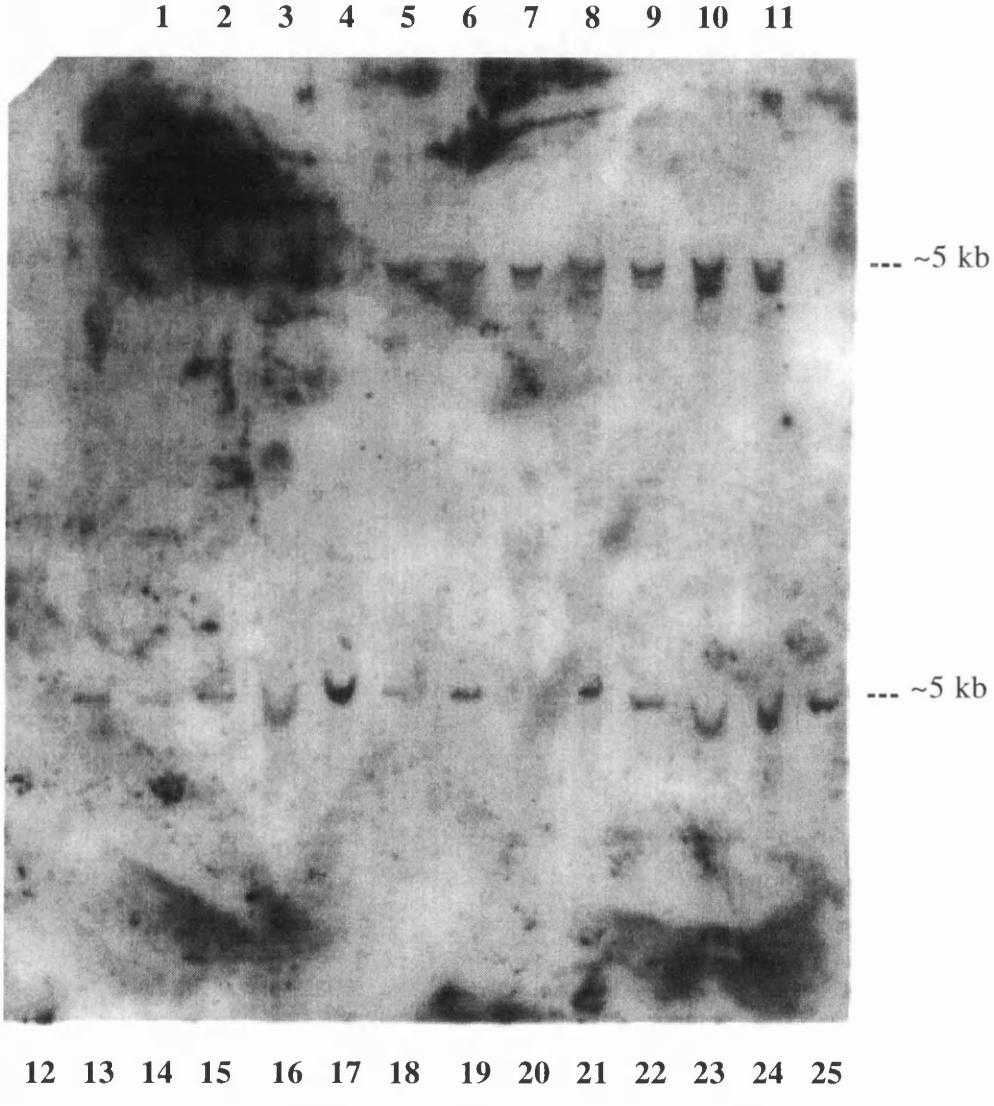


Figure 3.9 Autoradiogram of membrane carrying gDNA from G216 isolates probed with a fragment containing the *aroQ* gene of *S.coelicolor*

(lanes as in **Figure 3.8**)

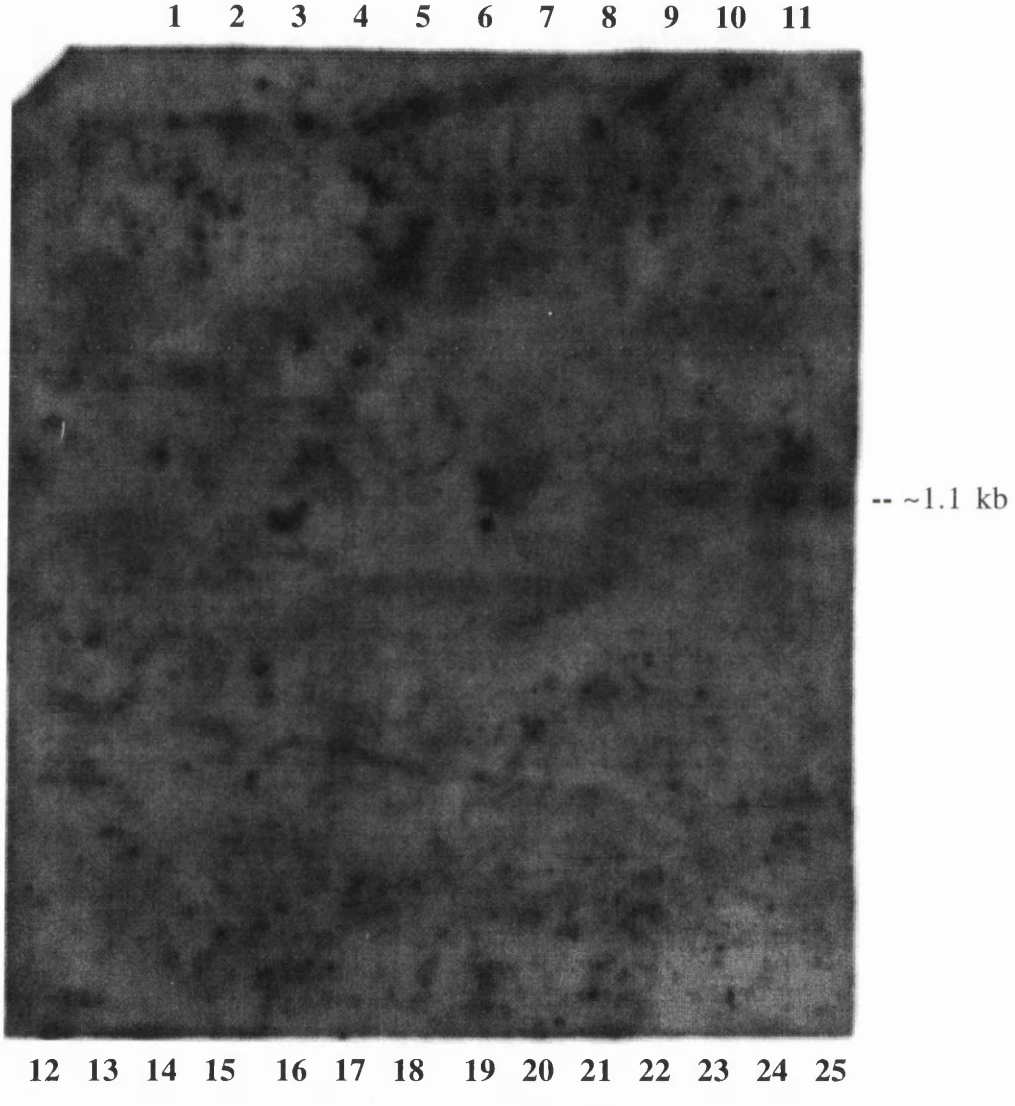


Figure 3.10 Autoradiogram of membrane carrying gDNA from G216 isolates probed with the *tsr* gene

(lanes as in **Figure 3.8**)

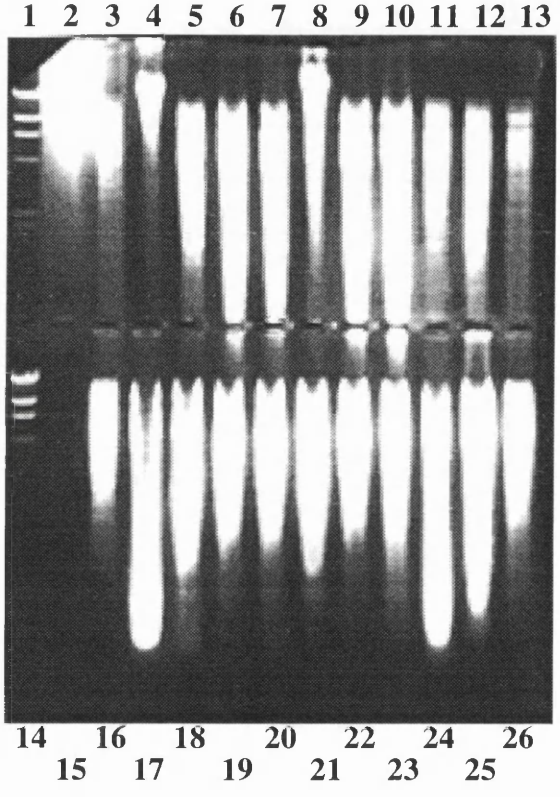


Figure 3.11 gDNA from G216 isolates

lanes 1 and 14: λ/*Hin*d III markers

lanes 2 and 15: ~1 ng of *ermE*

lanes 3-12 and 16-24: gDNA from putative disruptants, cut with *Sst* I

lanes 13 and 26: gDNA from untransformed G216, cut with *Sst* I

lane 25: gDNA from putative G216/pPM3 isolate, cut with *Sst* I

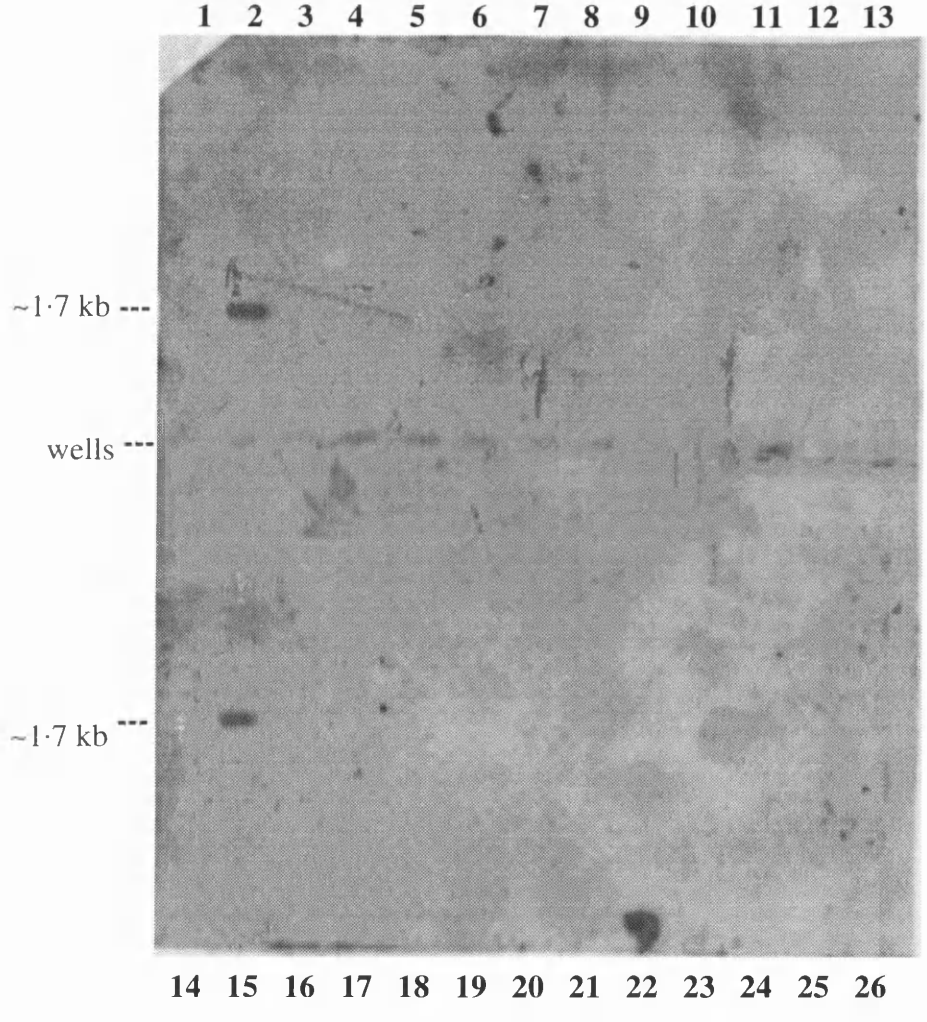


Figure 3.12 Autoradiogram of membrane carrying gDNA from G216 isolates probed with the *ermE* gene

(lanes as in **Figure 3.11**)

It was anticipated that the *ermE* control (1·7 kb in size) and the putative single- and double-crossovers (both approximately 5·4 kb in length) would show a single band to which the *ermE* probe hybridises. However, only the *ermE* control lit up on the autoradiogram (**Figure 3.12**). These results imply that the putative double-crossovers and the putative single-crossover were actually wild-type for *aroQ* and provides more evidence that *ermE* is not a reliable resistance marker in *S.coelicolor.*

3.2.2 Attempted disruption of *aroQ* with *grmA*

The aminoglycoside group of antibiotics, which includes streptomycin, neomycin, kanamycin and gentamicin, are antibacterial agents which inhibit protein synthesis by causing premature chain termination and misreading during translation, and block the initiation of replication of DNA. They also cause the incorporation of aberrant proteins into the cytoplasmic membrane, increasing its permeability which subsequently results in an increased uptake of the aminoglycosides.

The gentamicin-resistance gene had previously been cloned from the gentamicin producer *Micromonospora purpurea* and found to induce resistance to both gentamicin and kanamycin in *M.purpurea*, and also in *S.lividans* [Thompson *et al.*, 1985]. In *S.lividans* strains containing the gentamicin/kanamycin resistance determinant, resistance to these aminoglycosides is due to methylation of the 16S ribosomes [Skeggs *et al.*, 1987]. This gentamicin/kanamycin resistance determinant has been subcloned into the multicopy streptomycete plasmid pIJ702, and has been shown to encode a rRNA methylase protein which methylates a guanosine nucleoside of the 16S ribosomes, producing 7-methyl guanosine. This clone has subsequently been sequenced and named *grmA* (gentamicin resistance methyltransferase) [Kelemen *et al.*, 1991].

An alternative method of providing resistance to gentamicin exists in Gram-negative prokaryotes. The transposon Tn*1696* was originally isolated from *Pseudomonas aeruginosa* due to its ability to induce transfer of plasmid R1033, into which it had inserted, between Gram-negative species [Rubens *et al.*, 1979]. The transposon encodes resistance to gentamicin, streptomycin, chloramphenicol, sulfamethoxazole and mercuric chloride. It contains a gene, *aacC1*, which confers resistance to gentamicin in *E.coli*, and its product functions by acetylation of gentamicin. This gene has also been shown to induce gentamicin resistance in certain *Streptomyces* strains, including *S.lividans* [Muth *et al.*, 1989]. The bifunctionality of this gene would allow selection of a plasmid-borne gene disruption event in *E.coli* before transformation into

Streptomyces.

3.2.2.1 Disruption of a plasmid-borne *aroQ* gene with *grmA*

grmA is contained within an approximately 1·5 kb *Sph* I/*Bam*H I fragment of the pLST14 insert. This fragment has previously been subcloned, via the vector pIBI24 along with a polylinker fragment from vector pLus801 creating p503, into the vector pGEM7 to generate plasmid p504. *grmA* was subcloned into pIBI24 to create pPM2 using the *Xba* I and *Pst* I restriction endonucleases. The *Bam*H I fragment of this plasmid contains the resistance gene, and was used to disrupt the plasmid-borne dehydroquinase-encoding gene (*aroQ*) of *Streptomyces coelicolor* by insertion into the internal *Bam*H I site of this gene, creating plasmid pPM19 (**Figure 3.13**).

Plasmid pPM19 contains the origin of replication (*ori*) from *E.coli*, and all of the above subcloning procedures were performed in *E.coli*. However, this plasmid does not contain the *Streptomyces ori*, so will not replicate in a *Streptomyces* strain, and therefore should function as a suicide vector in *S.coelicolor*.

As a control, the *actI* region from *S.coelicolor* was also disrupted by the *grmA* gene. The *actI* gene cluster had previously been subcloned into the pUC18 vector and subsequently disrupted by insertion of the *ermE* gene, encoding erythromycin resistance, into its coding region, creating plasmid pIJ5604. The *ermE* gene was removed from pIJ5604, and was directly replaced with the *Kpn* I fragment of p504, disrupting the plasmid-borne *actI* cluster with the gentamicin/kanamycin resistance marker, and creating plasmid pPM20. Like pPM19, pPM20 carries only the *E.coli* origin of replication, and also should function as a suicide plasmid in *S.coelicolor*.

3.2.2.2 Selection for *grmA*

The minimum inhibitory concentration (MIC) of gentamicin for protoplasts of *S.coelicolor* G216 was calculated by the growth of such protoplasts on R2 medium overlaid with varying amounts of gentamicin. Protoplasts grew on plates overlaid with 10 μg/ml, but not on plates overlaid with 25 μg/ml or any greater amounts. The concentration of drug chosen for addition to plates was twice the lowest concentration on which untransformed protoplasts were unable to grow. Therefore, a concentration of gentamicin of 50 μg/ml agar was used for all subsequent *S.coelicolor* transformations. This figure was also the concentration chosen for selection in *S.lividans* [Skeggs *et al.*, 1987].

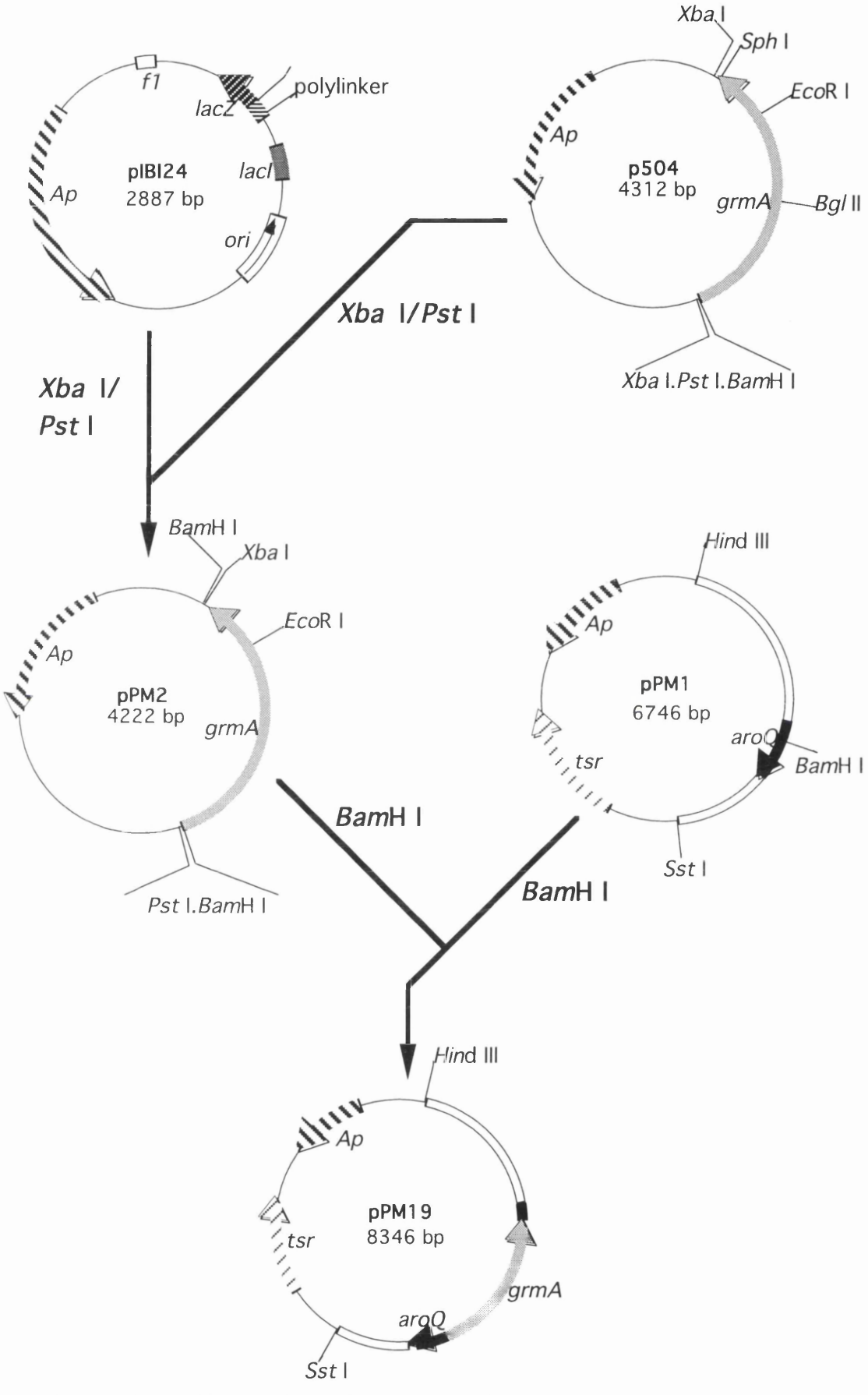


Figure 3.13 Disruption of a plasmid-borne *aroQ* gene with *grmA*

In order to allow transformation of protoplasts of *S.coelicolor* G216, the plasmids pPM19 and pPM20 were transformed into the *dam*⁻ *E.coli* strains MB5386 and CB51, to circumvent the restriction barrier of *S.coelicolor*.

Transformants were selected by growth on L agar supplemented with ampicillin, and were streaked, along with DS941 as a negative control, onto L agar supplemented with gentamicin at concentrations ranging from 5 μg/ml to 20 μg/ml. Although all three strains grew on L agar without supplement, none of them grew on the gentamicin-containing agar, suggesting that (unlike the *aacC1* gene product) the *grmA* gene product is not expressed in *E.coli*, that its product does not methylate the 16S ribosomal component of *E.coli*, or that such methylation does not induce resistance to gentamicin in *E.coli*.

The ampicillin-resistant isolates of MB5386 were also streaked onto L agar supplemented with both chloramphenicol and ampicillin. Chloramphenicol selects for the transposon which is used for the insertion mutagenesis of the *dam* gene of MB5386. Therefore, these two antibiotics, when used together, select for non-methylated pPM19 and pPM20 plasmids, which may be subsequently used for transformation of protoplasts of *S.coelicolor*. The *dam*⁻ mutation in CB51 is a more stable, naturally-occurring mutation, and is non-selectable.

To test the efficacy of selection with gentamicin in *S.coelicolor* G216, protoplasts of G216 were transformed with pLST14, plated onto R2, and overlaid with thiostrepton (pLST14 is a derivative of pIJ702, and so contains a *tsr* resistance marker). Transformants were patched onto Soya supplemented with gentamicin or thiostrepton. All transformants grew on both plates.

3.2.2.3 Transformation of protoplasts of *S.coelicolor* G216 with pPM19 and pPM20

Protoplasts of *S.coelicolor* G216 were transformed with approximately 5 μg of non-methylated pPM19 or pPM20 plasmid DNA (isolated from either MB5386 or CB51 *E.coli* transformants). Transformants were plated on R2 medium, supplemented with the aromatic amino acids for pPM19 transformations, and selected by overlaying with gentamicin after 16-20 hours incubation at 30°C.

These protoplasts were also transformed with 100 ng of pIJ487 plasmid DNA, plated on R2 medium and selected by overlaying with thiostrepton, to test the transformation competency of the protoplasts. The transformation efficiency of these protoplasts was

calculated as approximately $3x10^5/\mu g$.

110 colonies from the G216(MB5386)/pPM19 plates, 20 colonies from the G216(CB51)/pPM19 plates, 13 from G216(MB5386)/pPM20 plates and one from the G216(CB51)/pPM20 plates were picked, and patched onto Soya agar supplemented with either gentamicin or no drug. One G216(CB51)/pPM20 colony [C20] was identified which grew on Soya supplemented with gentamicin. Three G216(MB5386)/pPM19 colonies [M19-1, M19-2, M19-4] grew on Soya supplemented with gentamicin but not with thiostrepton, as did one of the G216(CB51)/pPM19 colonies [C19]. As no colonies grew on Soya supplemented with thiostrepton (an anticipated result in the control transformations since pPM20 does not contain the *tsr* resistance marker), these cells showed the resistance profiles of double-crossovers. However, since no single-crossovers (which it is anticipated are more likely to occur) were obtained, it was anticipated that these five isolates may have undergone a spontaneous mutation to produce resistance to gentamicin. To test their resistance profiles, the cells were again streaked onto Soya supplemented with either gentamicin or thiostrepton. All five isolates again showed resistance to gentamicin only.

These five isolates were grown in YEME liquid medium containing $MgCl_2$, glycine and gentamicin, and were grown to stationary phase at 30°C. Untransformed G216 mycelia was grown in YEME supplemented with $MgCl_2$ and glycine. Total DNA was prepared from each culture, and approximately 3 μg of DNA from each culture was digested with either *Sst* I or *Sal* I and run out on a 0·8% TAE gel. 5 ng of pLST14 digested with *Xho* I (as a control for the *grmA* gene) and pIJ487 digested with *Xba* I (as a control for the *tsr* gene) were also run on the gel (**Figure 3.15**).

The DNA on the gel was transferred to nylon membrane by Southern blotting, and the membrane probed with radiolabelled *aroQ* from *S.coelicolor* . Genomic DNA isolated from the parental G216 strain would be expected to show two bands of unknown size when digested with *Sal* I, and a single 5 kb band when digested with *Sst* I. It was anticipated that the gDNA isolated from disruptants (double-crossovers) would show two bands (one the same as one of the bands from parental gDNA, the other 1·7 kb larger than its parental copy) on an autoradiogram when digested with *Sal* I, and a single band (approximately 5 kb in size) when digested with *Sst* I. gDNA from an integrant (a single-crossover) would be expected to show three bands of unknown size when digested with *Sal* I, and two bands (approximately 4·5 kb and 7 kb in size) when digested with *Sst* I [see **Figure 3.14**]. However, when the DNA was digested with

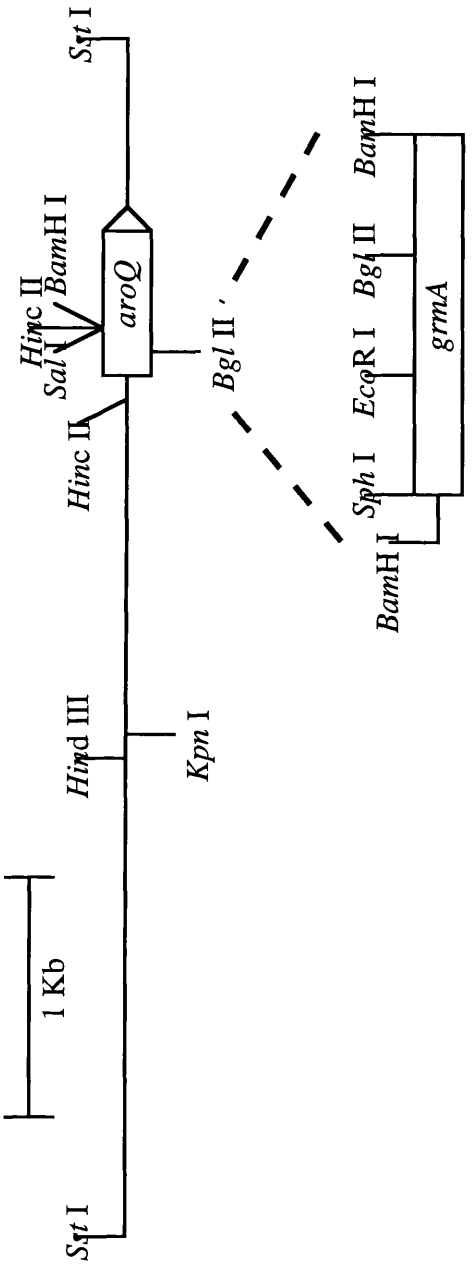


Figure 3.14 Restriction map of the *aroQ* gene disrupted by insertion of *grmA*

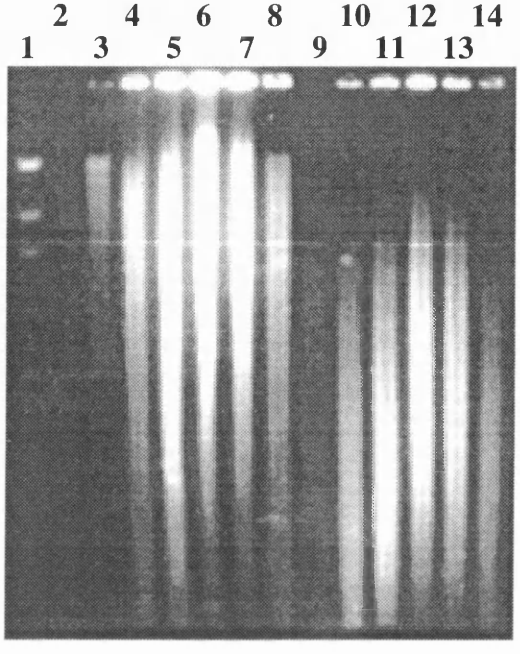


Figure 3.15 gDNA from G216 isolates

lane 1: λ/*Hin*d III

lane 2: 5 ng pLST14 cut with *Xho* I (~8·8 kb)
and 3 ng pIJ487 cut with *Xba* I (~6·2 kb)

lanes 3, 9: gDNA from untransformed G216 cut with *Sst* I or *Sal* I respectively

lanes 4, 10: gDNA from M19-1 cut with *Sst* I or *Sal* I respectively

lanes 5, 11: gDNA from M19-2 cut with *Sst* I or *Sal* I respectively

lanes 6, 12: gDNA from M19-4 cut with *Sst* I or *Sal* I respectively

lanes 7, 13: gDNA from C19 cut with *Sst* I or *Sal* I respectively

lanes 8, 14: gDNA from C20 cut with *Sst* I or *Sal* I respectively

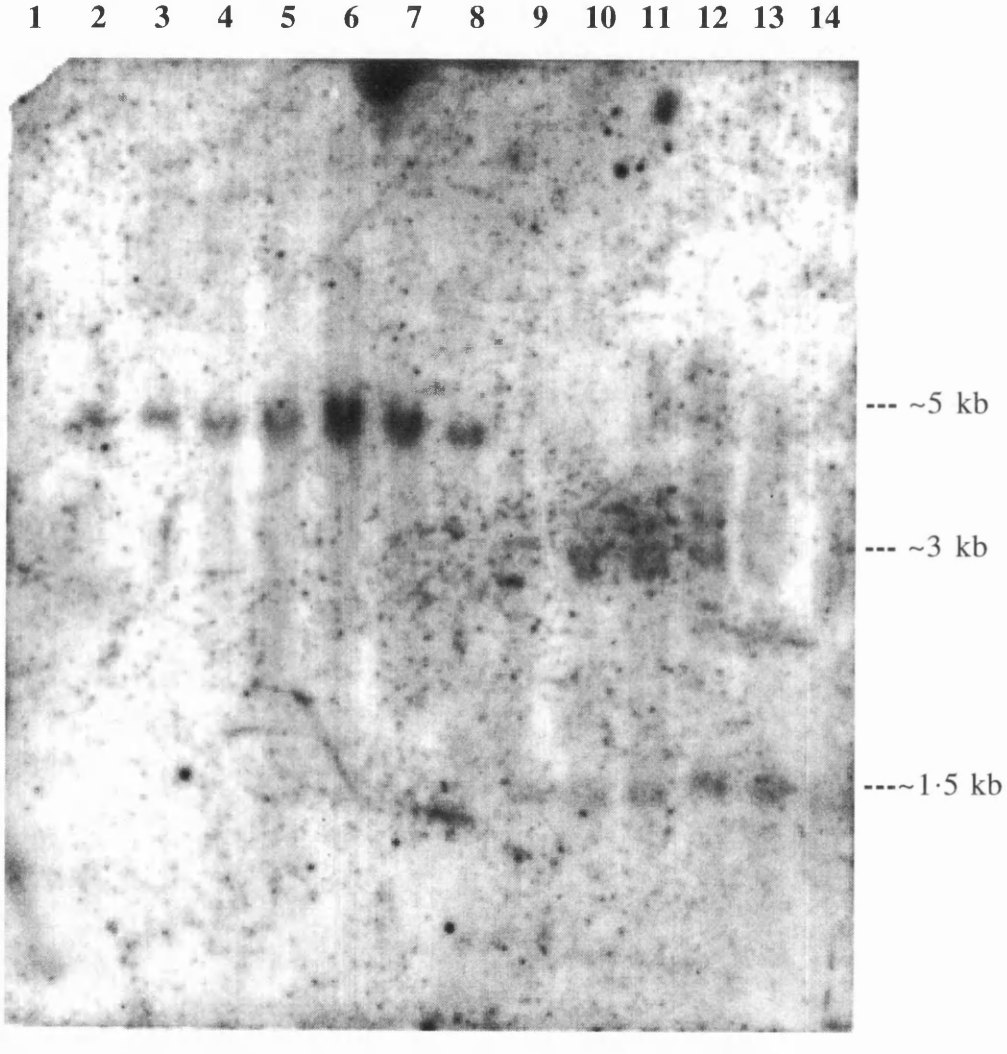


Figure 3.16 Autoradiogram of membrane carrying gDNA from G216 isolates probed with a fragment containing the *aroQ* gene from *S.coelicolor*

(lanes as in **Figure 3.15**)

was digested with *Sal* I, all six lanes exhibited the same two bands, approximately 3 kb and 1·5 kb in size, suggesting that the five gentamicin-resistant isolates were actually wild-type for *aroQ* (**Figure 3.16**).

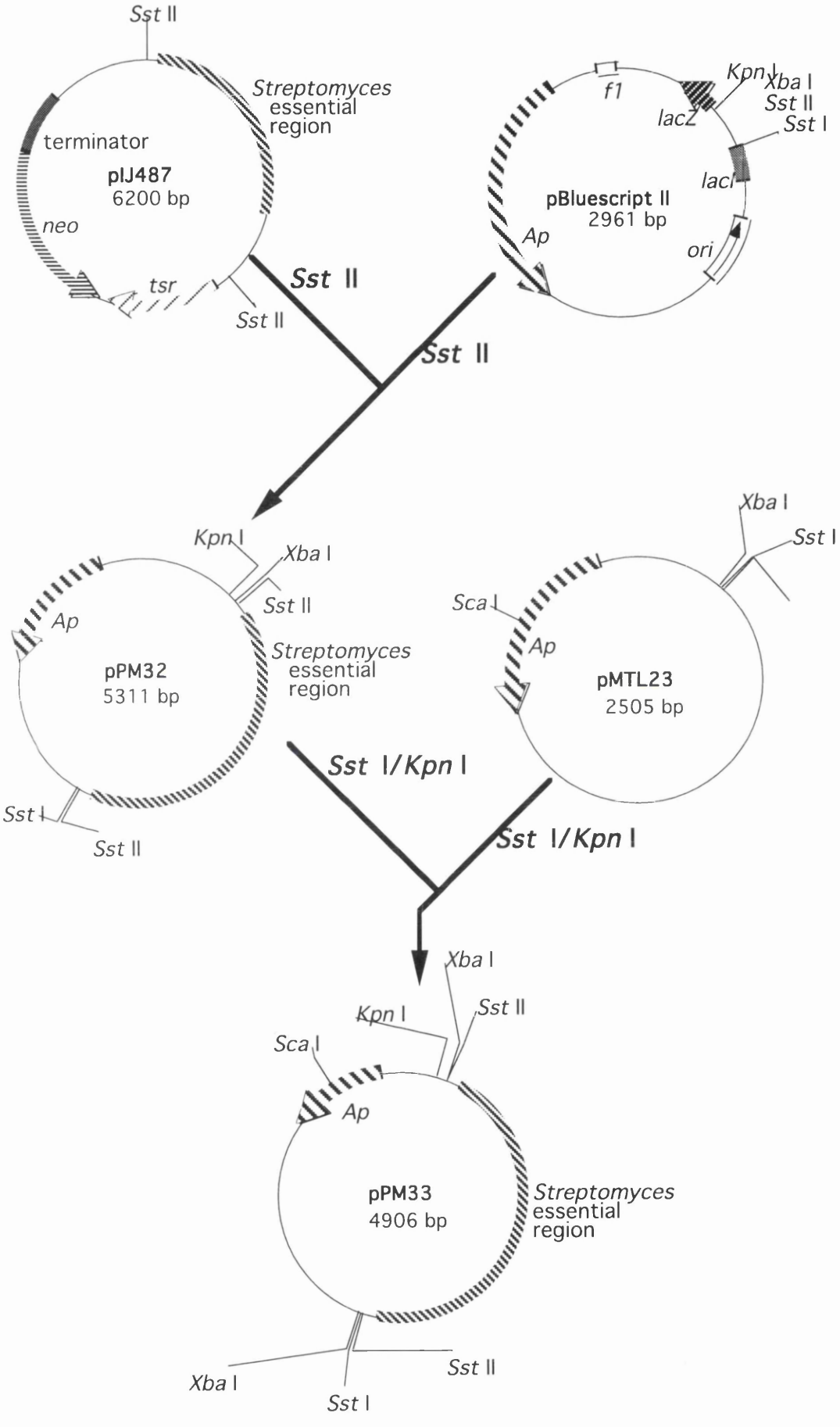
3.2.2.4 Attempted insertion of a *Streptomyces ori* into pPM19

An alternative method for disruption of the chromosomal copy of *aroQ* would be to transform protoplasts of G216 with a plasmid which replicates stably in *S.coelicolor*, and contains a disrupted copy of the *aroQ* gene. Transformants could be grown on a non-selective medium, which would allow plasmid loss. It is possible that a copy of a plasmid could become integrated into the chromosome due to homologous recombination between the chromosomal and plasmid-borne copies of *aroQ*, particularly when selection for maintenance of the plasmid in the cell is removed.

Therefore, insertion of a *Streptomyces* plasmid origin of replication (*ori*) into pPM19 was desired. This would prevent the plasmid from functioning as a suicide vector, and hopefully increase the likelihood of recombination between the plasmid and chromosome. After growth (up to three rounds) on a non-selective medium, cells would be patched onto selective media to screen for those which had not lost one or even both of the resistance genes on the plasmid. Colonies which had lost neither resistance would undergo further rounds of growth on a non-selective medium, followed by growth on selective media; those which had lost resistance to thiostrepton would be investigated for the possibility of having undergone a double-crossover event.

To convert plasmid pPM19 into a vector which was able to replicate stably in *S.coelicolor*, the streptomycete plasmid origin of replication, contained in the *Streptomyces* essential region, was isolated from pIJ487 by digestion with *Sst* II, and subcloned into pBluescript II, generating pPM32 (**Figure 3.17**). The origin was isolated from pPM32 using *Kpn* I and *Sst* I, and subcloned into pMTL23, creating pPM33. Both of the subcloning procedures described above were performed in the *E.coli* strain DS941.

The streptomycete plasmid origin was isolated from pPM33 (isolated from MB5386, as restriction by *Xba* I is partially-sensitive to Dam methylation) by digestion with *Xba* I and *Sca* I (*Sca* I cuts the vector DNA into two fragments, each having *Xba* I/*Sca* I ends, preventing either of these fragments from ligating to the *Xba* I-restricted pPM19 DNA), and was ligated to pPM19 (isolated from MB5386) which had been digested


Sst II
Streptomyces essential region
terminator
pIJ487
6200 bp
neo
tsr
Sst II
Sst II
Sst II
f1
Kpn I
Xba I
Sst II
Sst I
lacZ
lacI
pBluescript II
2961 bp
Ap
ori
Kpn I
Xba I
Sst II
Ap
pPM32
5311 bp
Streptomyces essential region
Sst I
Sst II
Xba I
Sst I
Sca I
Ap
pMTL23
2505 bp
Sst I/Kpn I
Sst I/Kpn I
Xba I
Kpn I
Sst II
Sca I
Ap
pPM33
4906 bp
Streptomyces essential region
Sst II
Xba I
Sst I

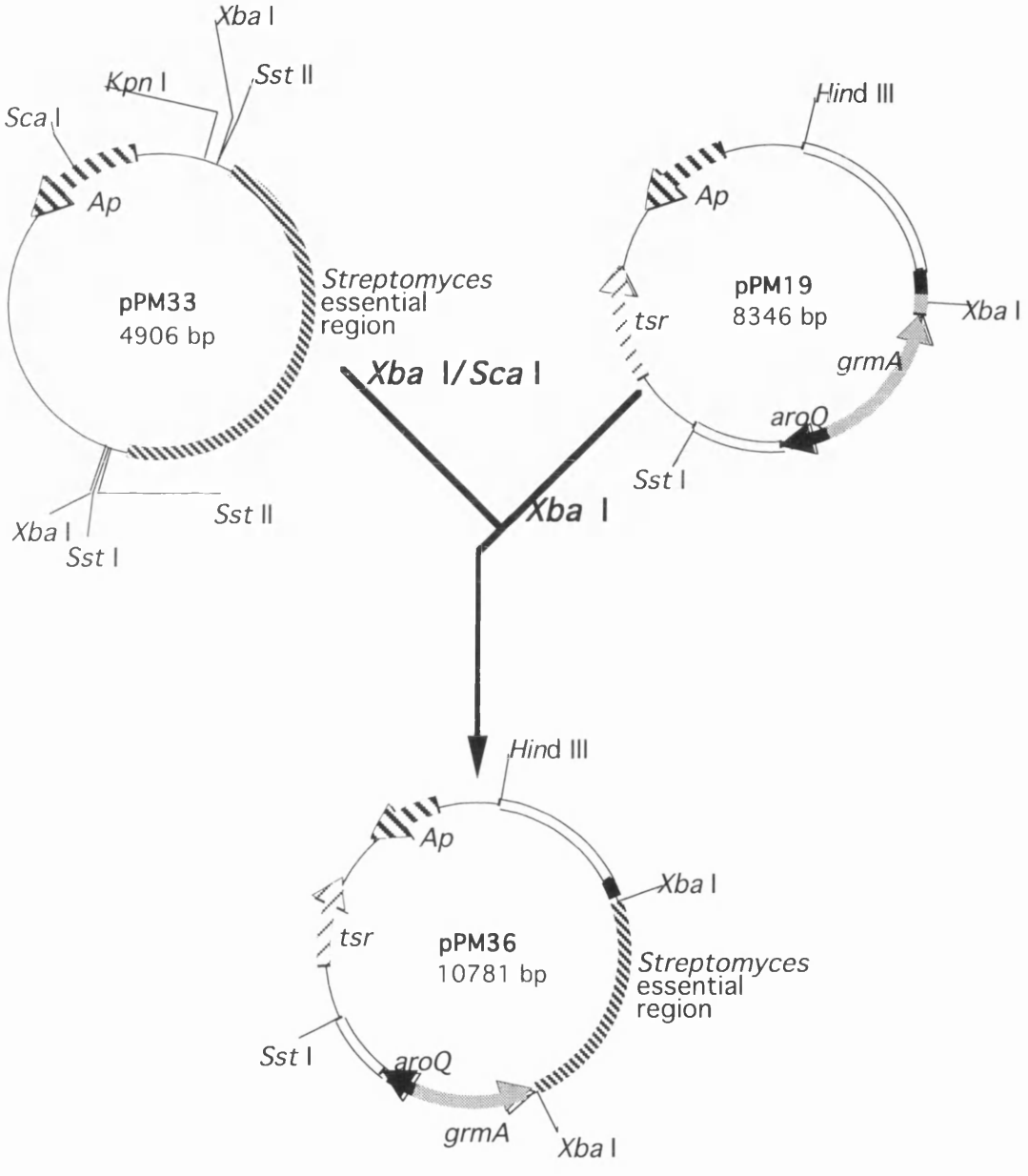


Figure 3.17 Insertion of a *Streptomyces ori* into pPM19

with *Xba* I and had also been dephosphorylated, to create pPM36. It was thought to be possible to transform the ligation mix into *E.coli*, select with ampicillin, and isolate plasmid DNA from transformants to analyse for insertion of the streptomycete *ori*. Plasmid DNA showing the correct restriction pattern could then be transformed into a non-methylating strain of *E.coli*, and subsequently be transformed into protoplasts of *S.coelicolor*. However, no transformants of *E.coli* containing pPM36 were generated.

Alternatively, as both plasmids pPM19 and pPM33 were isolated from a non-methylating strain of *E.coli* prior to digestion and ligation to generate pPM36, the ligation mixture could be transformed directly into protoplasts of *S.coelicolor*. Any true transformants would contain either a stably-replicating plasmid (pPM36) or an integrated plasmid (either pPM36, or pPM19 produced by religation). However, this method of disruption was unsuccessful since no transformants were obtained when protoplasts of G216 were transformed with the pPM36 ligation mixture.

A third method of transformation of *S.coelicolor* with pPM36 was also attempted, involving transformation of the pPM36 ligation mixture into a non-methylating strain of *E.coli* (MB5386), dilution of the transformation mixture with 1 ml of 2xYT medium supplemented with ampicillin, and growing the culture directly at 37°C until the cells reached stationary phase. Plasmid DNA was subsequently isolated from this culture, and transformed directly into protoplasts of *S.coelicolor* G216, which were plated onto R2 medium and overlaid with thiostrepton. However, again no transformants were obtained using this method. Thus, none of the three methods attempted for subcloning of the streptomycete plasmid *ori* into pPM19 were successful.

3.3 Discussion

These results suggest that *S.coelicolor* has a high rate of spontaneous mutation to produce resistance to erythromycin and possibly also to gentamicin (although when tested, no non-transformed protoplasts of G216 grew when overlaid with a concentration of gentamicin higher than 10 μg/ml (see section **3.2.2.2**). Therefore, these antibiotics appear not to be good selecting agents in *S.coelicolor* under the conditions used in this work. Resistance to thiostrepton also appears to occur spontaneously in *S.coelicolor*. However, as this antibiotic has long been used as a reliable selecting agent in *S.coelicolor*, it is possible that spontaneous mutations induced by growth in the presence of erythromycin or gentamicin induce multiple resistance to antibiotics. This situation is most likely to occur through either mutations in membrane proteins which prevent the uptake of antibiotics, or mutations which

increase the rates of efflux of the antibiotics. However, it is also possible that the dual spontaneous resistance to erythromycin and thiostrepton could arise by spontaneous changes in the ribosomes of resistant cells, as resistance to both of these antibiotics may be brought about by methylation of adenosine residues in the ribosomes (see **3.2.1.1**).

Although transformants showing resistance to antibiotics may arise through spontaneous mutations, it might also be anticipated that some cells exhibiting the correct resistance profile for a disruptant would have occurred due to double recombination events, causing insertion of the resistance gene into the chromosomal copy of *aroQ*. The lack of such disruptants may be due to the existence of a functional dehydroquinase being essential to the survival of the cell. Although mutations in the *aroD* gene of *E.coli* have arisen, which are auxotrophic for the three aromatic amino acids (or the end-product of the shikimate pathway, chorismate), no *aroD* disruptants have been produced in *E.coli*, despite the efforts of three separate research groups. This implies that a null mutation of the gene encoding dehydroquinase is lethal to the cell. This, in turn, would suggest that the *aroD* mutants which have been identified contain a leaky mutation which allows a low but sufficient level of flux through the pathway to produce some other essential product of the pathway (see **Figure 1.3** for examples), but not sufficient flux for the synthesis of the required amounts of the aromatic amino acids. However, a null mutation would prevent any flux through the pathway, and thus impede the synthesis of this essential product (or at least a product which has not been supplied in any of the growth media which have been used when trying to isolate null mutations of the *aroD* or *aroQ* genes from *E.coli* and *S.coelicolor*, respectively).

CHAPTER 4

Overexpression of the 3-dehydroquinase from *S.coelicolor* in *S.lividans*

4.1 Introduction

As the type II 3-dehydroquinase enzyme from *S.coelicolor* has a high K_m, it is possible that it may also have a high flux control coefficient, and exert a high level of control over flux through the shikimate pathway, particularly when the rate of flux is high (for example during secondary metabolism). Therefore, it is desirable to overexpress the dehydroquinase enzyme in an attempt to upregulate the flux.

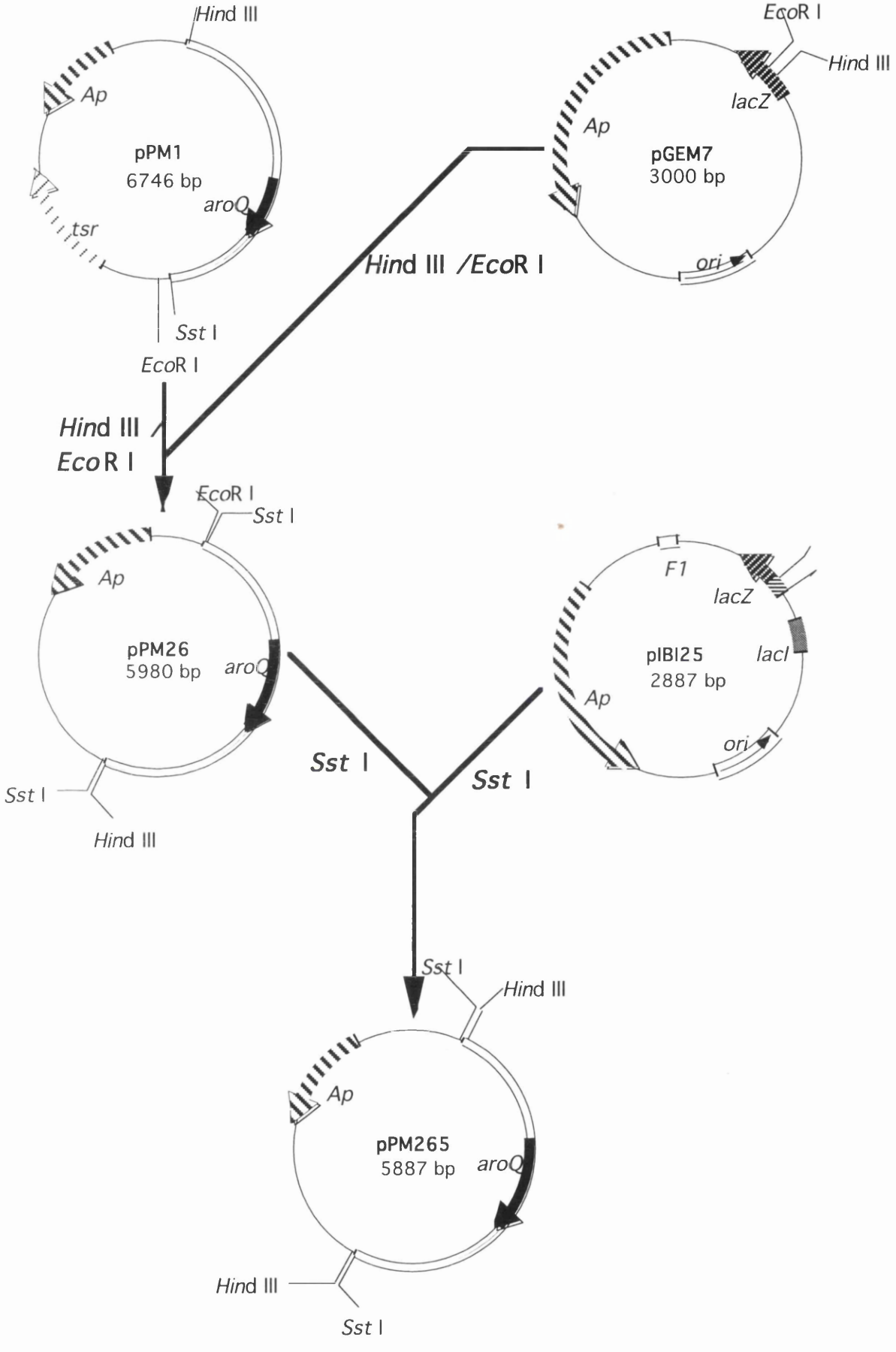
It was intended to achieve overexpression by subcloning of the *aroQ* gene, which encodes the type II dehydroquinase of *S.coelicolor*, into high copy-number *Streptomyces* vectors. pIJ487 was chosen as the most suitable vector for overexpression. It is a pIJ101-based vector which contains an *E.coli* transcription terminator upstream of the multiple cloning site, and a promoterless aminoglycoside phosphotransferase gene, *neo*, downstream of the cloning site [Ward *et al.*, 1986]. Therefore, this plasmid is a suitable vector for the overexpression of a gene from its own promoter, and also for use as a promoter-probe vector.

The plasmid pUGT1 is a pIJ487-based vector into which the thiostrepton-inducible *tipA* promoter has previously been subcloned [Ingham *et al.*, 1995]. This promoter allows transcription towards the *neo* gene, and lies on the far side of the multiple cloning site relative to the *neo* gene. Therefore, a promoterless gene may be expressed from pUGT1, its expression being regulated by the presence of thiostrepton in the medium.

4.2 Results

4.2.1 Construction of plasmids facilitating overexpression

The 3 kb *Hin*d III/*Eco*R I fragment containing the *aroQ* gene from *S.coelicolor* [P.White, unpublished results] was isolated from pPW63, subcloned into the pGEM7 vector to generate pPM26 (**Figure 4.1**), transformed into *E.coli* DS941 competent cells, and plated onto L agar supplemented with ampicillin. The *aroQ*-containing insert was isolated from pPM26 using the *Sst* I restriction enzyme, and was subcloned into pIBI25 which had been linearised by *Sst* I and dephosphorylated, creating plasmid pPM265. The insert was re-isolated from pPM265 by digestion with *Hin*d III, and was ligated into pIJ487 which had been digested with *Hin*d III and dephosphorylated, creating plasmid pPM27. Alternatively, the *aroQ*-containing fragment was isolated from pPM26 with the *Sst* I restriction enzyme, and ligated into pUGT1 which had been


Hind III
Ap
pPM1
6746 bp
aroQ
tsr
Sst I
EcoR I
EcoR I
Hind III
lacZ
Ap
pGEM7
3000 bp
ori
Hind III /EcoR I
Hind III
EcoR I
EcoR I
Sst I
Ap
pPM26
5980 bp
aroQ
Sst I
Hind III
F1
lacZ
pIBI25
2887 bp
lacI
Ap
ori
Sst I
Sst I
Sst I
Hind III
Ap
pPM265
5887 bp
aroQ
Hind III
Sst I

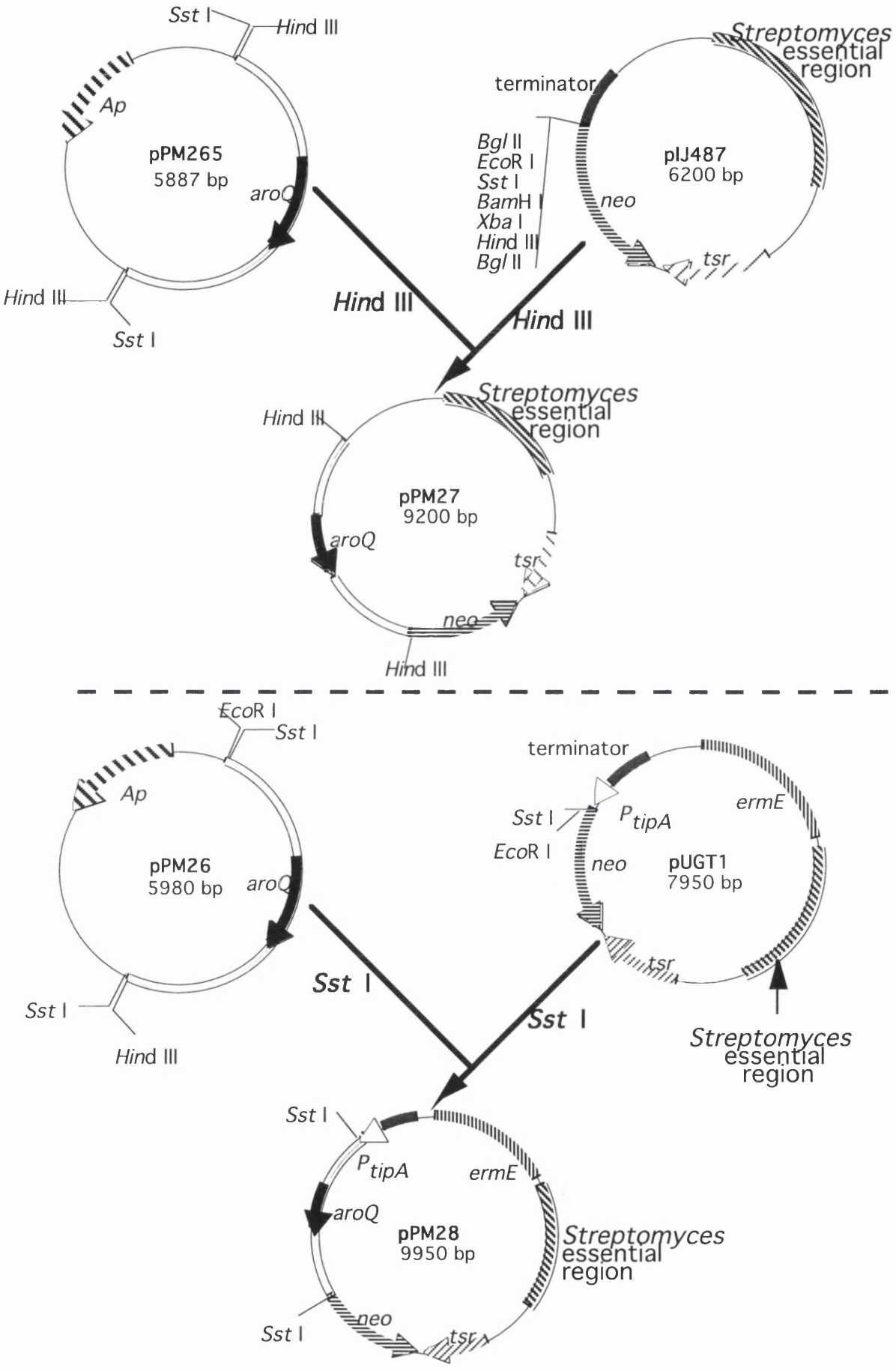


Figure 4.1 Construction of plasmids enabling overexpression of dehydroquinase

linearised with *Sst* I and dephosphorylated, creating plasmid pPM28.

4.2.2 Transformation of protoplasts of *S.lividans*

Since both ligation mixes contained DNA which had been isolated from a dam-methylating *E.coli* strain DS941, and any resultant plasmids would therefore be degraded if transformed directly into protoplasts of *S.coelicolor*, the resultant plasmids were initially transformed into protoplasts of *S.lividans* strains TK54 or TK64. The transformation mixtures were plated onto R2 regeneration medium supplemented with thiostrepton to select for transformants. A single thiostrepton-resistant transformant was obtained when protoplasts of TK54 were transformed with each of the pPM27 and pPM28 ligation mixes, and a single putative pPM27-containing colony was obtained from transformation of protoplasts of TK64.

4.2.3 Investigation of overexpression of dehydroquinase

In order to ascertain whether any of the transformants contained plasmids which were not merely religations, the three colonies were grown in liquid medium supplemented with thiostrepton, and isolation of plasmid DNA was attempted from each. Although alkaline lysis using the methods of both Birnboim & Doly (**2.4.2.4**) and Kieser (**2.4.2.5**) were attempted, plasmid DNA could not be obtained from any of these three cultures.

An alternative method to investigate whether any of these three isolates was overexpressing the type II dehydroquinase enzyme by running the total cellular protein extracts on a denaturing polyacrylamide gel was subsequently attempted (**2.5.4**). Although neither the putative TK64/pPM27 isolate nor the putative TK54/pPM28 isolate exhibited overexpression of a protein of the appropriate size on the gel, the putative TK54/pPM27 isolate did show an overexpressed protein of approximately 15-18 kDa in size (the type II dehydroquinase enzyme from *S.coelicolor* was estimated at 16 kDa [White *et al.*, 1990]). This isolate was, therefore, confirmed as TK54/pPM27.

To ensure that this protein was due to the presence of an insert in pIJ487, protein extracts from TK54/pPM27, untransformed TK54 and TK24/pIJ487 were run out on a gel by SDS-PAGE (**Figure 4.2**). The overexpressed protein which had been identified previously was only evident in the extract from TK54/pPM27. Therefore, it was assumed that this isolate was indeed overexpressing the type II dehydroquinase from *S.lividans*.

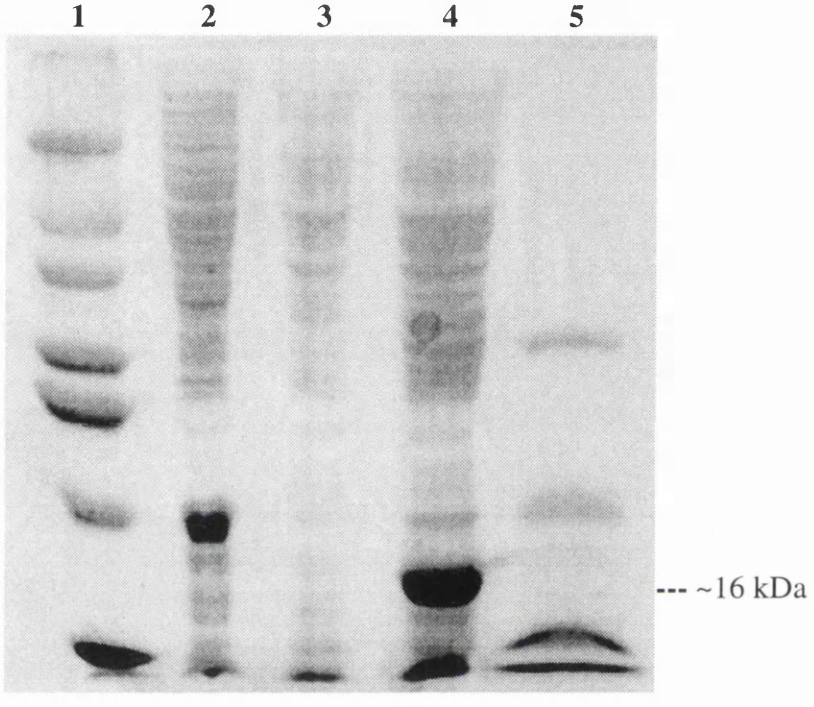


Figure 4.2 Crude protein extract from TK54/pPM27 exhibiting overexpression of the dehydroquinase from *S.coelicolor*

lane 1: VII-L protein markers (14·2 / 20·1 / 24 / 29 / 36 / 45 / 66 kDa)
lane 2: approx. 50 µg crude protein extract from TK24/pIJ487
lane 3: approx. 50 µg crude protein extract from TK54
lane 4: approx. 50 µg crude protein extract from TK54/pPM27
lane 5: low molecular weight protein markers

4.2.4 Assaying for dehydroquinase activity

TK54 and TK54/pPM27 were grown in 200 ml YEME supplemented with glycine and magnesium chloride (and also with thiostrepton in the case of the overexpressing strain). Crude protein extracts were obtained from these cultures (**2.5.1**), which were assayed for the presence of dehydroquinase, with each assay being repeated a minimum of three times and the averages calculated (**2.5.2**). Total protein concentrations from each of the crude extracts were calculated using the Bradford's assay.

Strain	Activity (Units / ml)	Protein conc (mg / ml)	Specific activity (Units / mg)	Standard Deviation (σ_{n-1})
TK54	1·9	7·3	0·26	0·14
TK54 / pPM27	5·0	2·6	1·92	0·27

Table 4.1 Activities of dehydroquinase and total protein concentrations in the crude extracts

N.B. One unit of enzyme activity is defined as the amount of enzyme required to catalyse the conversion of 1μmol of substrate or the synthesis of 1μmol of product in one minute.

As can be seen from **Table 4.1**, the overexpressing strain (TK54/pPM27) exhibits an elevated level of dehydroquinase activity compared to the untransformed strain of *S.lividans* (TK54). The specific activities for the two strains were calculated by dividing the activities derived from the assays by the protein concentrations estimated by the Bradford's assay. The TK54/pPM27 strain shows a 7·38-fold increase in activity relative to the untransformed TK54 strain. Therefore, it may be concluded that the plasmid pPM27 contained in this strain is indeed causing overexpression of an active form of the type II dehydroquinase enzyme from *S.coelicolor*.

4.3 Discussion

The type II dehydroquinase enzyme from *S.coelicolor* has been overexpressed in *S.lividans*, and produced greater than a 7-fold increase in dehydroquinase activity in this strain relative to its untransformed parent strain. It was intended that this plasmid would subsequently be transformed into *S.coelicolor*, but this experiment has yet to be

attempted due to the difficulty in isolating plasmid DNA from the overexpressing strain. This experiment would facilitate analysis of the flux through the shikimate pathway to identify whether an increase in the activity of the dehydroquinase enzyme would produce any consummate increase in metabolic flux through this pathway.

pPM27, which facilitates the overexpression of dehydroquinase, is a pIJ487-based plasmid. pIJ487 is a multicopy *Streptomyces* vector which lacks a promoter immediately upstream of the multiple cloning site, and therefore should not cause any vector-driven transcription of genes in a fragment which has been inserted into the multiple cloning site. Thus, any expression from a gene (or genes) contained on such an insert must be due to the existence on the insert of a promoter which lies in the correct orientation with respect to the gene, and is functional in the particular strain in which the construct is analysed. Therefore, expression of the type II dehydroquinase from *S.coelicolor* in TK54/pPM27 must be due to the presence of a promoter upstream of, and in the correct orientation to, the *aroQ* gene on the inserted fragment. This promoter is presumably the natural promoter for the *aroQ* gene in *S.coelicolor.*

As there is apparently only a single protein which is overexpressed in TK54/pPM27 (relative to the untransformed TK54 strain), it is assumed that the *aroQ* gene of *S.coelicolor* does not exist as part of an operon which is wholly, or at least partially, contained on this 3 kb fragment, and therefore that this promoter lies immediately upstream of the *aroQ* gene on this fragment. It is feasible that another protein lies downstream of, and is transcriptionally coupled to, the *aroQ* gene, but that only its 5' end lies on this fragment, as there are approximately only 800 bp downstream of the *aroQ* gene on this fragment. This could mean that only the N-terminus of this downstream protein is actually expressed from the insert in pPM27, and might not be seen on the protein gel.

The regions upstream and downstream of *aroQ* are currently being sequenced in this laboratory to ascertain the existence of other genes, and in an attempt to identify any potential promoter-like sequences.

CHAPTER 5

Attempted cloning of the *aroA* gene

5.1 Introduction

The *aroA* gene of *S.coelicolor* encodes the enzyme EPSP synthase, which catalyses the conversion of shikimate-3-phosphate to EPSP. This enzyme has been widely studied as it is the target of glyphosate, which has herbicidal activity. Therefore, the *aroA* gene is a target for mutation to produce a glyphosate-resistant enzyme in crop plants. *aroA* has also been disrupted in virulent bacteria for use as live vaccines [for example, Hoiseth & Stocker, 1981].

Cloning of *aroA* would allow gene disruption to be performed, which would cause a blockage late in the pathway (EPSP synthase catalyses the fifth step of six in the shikimate pathway), and allow analysis of flux through the majority of the pathway. Chorismate synthase is a notoriously difficult enzyme to assay for, and so was not chosen. EPSP synthase was chosen for investigation as it is the enzyme immediately preceding chorismate synthase.

Ideally, purification of EPSP synthase from *S.coelicolor*, followed by N-terminal protein sequencing, would have enabled the design of oligonucleotides against this protein sequence. Such oligonucleotides would have then been used for PCR cloning or genomic cloning of *aroA* from *S.coelicolor*. However, EPSP synthase has not been purified from *Streptomyces*, and so alternative approaches were utilised.

5.2 Results

5.2.1 Design of oligonucleotides primer 1 and primer 2

At the beginning of this investigation, deduced amino acid sequences were available of the EPSP synthase enzymes from nine species of prokaryotes, lower eukaryotes and plants. When these amino acid sequences were lined up, they showed extremely good similarity across a number of highly conserved regions (**Figure 5.1**). Two of these regions were used for the design of oligonucleotide primers. Using a codon preference table compiled from 67 streptomycete coding regions (**Figure 5.2**) [Taylor, 1992], the most likely DNA sequences which could encode these predicted amino acid sequences were derived, and were used as templates for the design of oligonucleotides (**Figure 5.3**).

Primer 1 was designed against a region of sequence (NAGTAMRP, residues **176-183** in **Figure 5.1**) which is extremely well conserved in the nine sequences available, and is

```
                1                                                  50
A.nidulans      KCLAAYGLPT SLKDARIRKL TAGKHCSVDQ LMFNMALDKK NDGPKKKIVL
S.cerevisiae    KILVAYGLPV SPDEKWFKEL TLHKKTPLDI LLKKMSIDKK NEGSKKKVVI
E.coli          .......... .......... .......... .......... ..........
S.typhimurium   .......... .......... .......... .......... ..........
P.hybrida       .....MAQIN NMAQGIQTLN PNSNFHKPQV PKSSSFLVFG SKKLKNSANS
P.sativum       .......... .......... .......... .......... ..........
L.esculentum    .MAQISSMAQ GIQTLSLNSS NLSKTQKGPL VSNSLFFGSK KLTQISAKSL
A.thaliana      .MAQVSRICN GVQNPSLISN LSKSSQRKSP LSVSLKTQQH PRAYPISSSW
B.pertussis     .......... .......... .......... .......... ..........

                51                                                 100
A.nidulans      LSAIGTPYET RASVVANEDI RVVLAPSIEV HPGVA..HSS NVICAPPGSK
S.cerevisiae    LESIGKCYGD SAQFVSDEDL RFILTDETLV YPFKDIPADQ QKVVIPPGSK
E.coli          .......... .......... ........ME SLTLQPIARV DGTINLPGSK
S.typhimurium   .......... .......... ........ME SLTLQPIARV DGAINLPGSK
P.hybrida       MLVLKKDSIF MQKFCSFRIS ASVATAQKPS EIVLQPIKEI SGTVKLPGSK
P.sativum       .......... .......... ....KPSTAP EIVLEPISEI SGTITLPGSK
L.esculentum    GVFKKDSVLR VVRKSSFRIS ASVATAEKPH EIVLXPIKDI SGTVKLPGSK
A.thaliana      GLKKSGMTLI GSELRPLKVM SSVSTAEKAS EIVLQPIREI SGLIKLPGSK
B.pertussis     .......... .......... .....MSGLA YLDLPAARLA RGEVALPGSK

                101                                                150
A.nidulans      SISNRALVLA ALGSGTCRIK NLLHSDDTEV MLNALERLGA ATFSWEEEGE
S.cerevisiae    SISNRALILA ALGEGQCKIK NLLHSDDTKH MLTAVHELKG ATISWEDNGE
E.coli          TVSNRALLLA ALAHGKTVLT NLLDSDDVRH MLNALTALGV SYTLSADRTR
S.typhimurium   SVSNRALLLA ALPCGKTALT NLLDSDDVRH MLNALSALGI NYTLSADRTR
P.hybrida       SLSNRILLLA ALSEGTTVVD NLLSSDDIHY MLGALKTLGL HVEEDSANQR
P.sativum       SLSNRILLLA ALSEGTTVVE NLLDSEDIHY MLEALKTLGL RVEDDKTTQR
L.esculentum    SLSNRILLLA ALSEGRTVVD NLLSSDDIHY MLGALKTLGL HVEDDNENQR
A.thaliana      SLSNRILLLA ALSEGTTVVD NLLNSDDINY MLDALKRLGL NVETDSENNR
B.pertussis     SISNRVLLLA ALAEGSTEIT GLLDSDDTRV MLAALRQLGV SV......GE

                151                                                200
A.nidulans      VLVVNGKGGNL QASSSP.... .LYLGNAGTA SRFLT.TVAT LANSSTVDS
S.cerevisiae    TVVVEGHGGST LSACADP... .LYLGNAGTA SRFLTSLAAL VNSTSSQKY
E.coli          CEII.GNGGPL HAEGAL.... ELFLGNAGTA MRPL.AA.AL CLGSND...
S.typhimurium   CDIT.GNGGAL RAPGAL.... ELFLGNAGTA MRPL.AA.AL CLGQNE...
P.hybrida       AVVE.GCGGLF PVGKESKEEI QLFLGNAGTA MRPLTAAVTV AGGNSR...
P.sativum       AVVE.GSGGLF PTGRESKDEV NLFLGNAGTA MRPLTAALVA AGGNTR...
L.esculentum    AIVE.GCGGQF PVGKKSEEEI QLFLGNAGTA MRPLTAAVTV AGGHSR...
A.thaliana      AVVE.GCGGIF PASIDSKSDI ELYLGNAGTA MRPLTAAVTA AGGNAS...
B.pertussis     VAD..GCVTIE GVARFPTEQA ELFLGNAGTA FRPLTAALAL MGGDYRLS.

                201                                                250
A.nidulans      SVLTGNNRMK QRPIGDLVDA LTANVLPLNT SKGRASLPLK IAASGGFAGG
S.cerevisiae    IVLTGNARMQ QRPIAPLVDS LRANGTKIEY LNNEGSLPIK VYTDSVFKGG
E.coli          IVLTGEPRMK ERPIGHLVDA LRLGGAKITY LEQENYPPLR ..LQGGFTGG
S.typhimurium   IVLTGEPRMK ERPIGHLVDS LRQGGANIDY LEQENYPPLR ..LRGGFTGG
P.hybrida       YVLDGVPRMR ERPISDLVDG LKQLGAEVDC FLGTKCPPVR IVSKGGLPGG
P.sativum       YILDGVPRMR ERPIGDLVSG LKQLGADVDC FLGTNCPPVR IIGKGGLPG.
L.esculentum    YVLDGVPRMR ERPIGDLVDG LKQLGAEVDC SLGTNCPPVR IVSKGGLPGG
A.thaliana      YVLDGVPRMR ERPIGDLVVG LKQLGADVEC TLGTNCPPVR VNANGGLPGG
B.pertussis     ...GVPRMHE RPIGDLVDAL RQFGAGIEYL GQAGYPPLRI GGGSIRVDGP
```

```
                251                                                 300
A.nidulans      NINLAAKVSS QYVSSLLMCA PYAKEPVTLR LVGGKPISQP YIDMTTAMMR
S.cerevisiae    RIELAATVSS QYVSSILMCA PYAEEPVTLA LVGGKPISKL YVDMTIKMME
E.coli          NVDVDGSVSS QFLTALLMTA PLAPEDTVIR IKGDLVSKPY IDITLNLM.K
S.typhimurium   DIEVDGSVSS QFLTALLMTA PLAPKDTIIR VKGELVSKPY IDITLNLM.K
P.hybrida       KVKLSGSISS QYLTALLMAA PLALGDVEIE IIDKLISVPY VEMTLKLMER
P.sativum       KVKLSGSISS QYLTALLMAA PLALGDVEIE IIDKLISVPY VEMTLKLMER
L.esculentum    KVKLSGSISS QYLTALLMAA PLALGDVEIE IIDKLISVPY VEMTLKLMER
A.thaliana      KVKLSGSISS QYLTALLMSA PLALGDVEIE IVDKLISVPY VEMTLKLMER
B.pertussis     VRVEGSVSSQ FLTALLMAAP VLARRSGQDI TIEVVGELIS KPYIEITLNL

                301                                                 350
A.nidulans      SFGIDVQKST TEEHTYHIPQ G..RYVNPAE YVIESDASCA TYPLAVAAVT
S.cerevisiae    KFGINVETST TEPYTYYIPK G..HYINPSE YVIESDASSA TYPLAFAAMT
E.coli          TFGVEIENQH YQQFVV...K GGQSYQSPGT YLVEGDASSA SYFLAAAAIK
S.typhimurium   TFGVEIANHH YQQFVV...K GGQQYHSPGR YLVEGDASSA SYFLAAGAIK
P.hybrida       FGISVEHSSS WDRFFV...R GGQKYKSPGK AFVEGDASSA SYFLAGAAVT
P.sativum       FGVSVEHSDN WDRFLV...H GGQKYKSPGN AFVEGDASSA SYFLAGAAVT
L.esculentum    FGVFVEHSSG WDRFLV...K GGQKYKSPGK AFVEGDASSA SYFLAGAAVT
A.thaliana      FGVSVEHSDS WDRFFV...K GGQKYKSPGN AYVEGDASSA SYFLAGAAIT
B.pertussis     MARFGVSVRR DGWRAFTIAR DAVYRGPGRM AIEGDASTAS YFLALGAIGG

                351                                                 400
A.nidulans      GTTCTVPNIG SASLQGDARF AVE....VLR PMGCTVEQTE TSTTVTGPSD
S.cerevisiae    GTTVTVPNIG FESLQGDARF ARD....VLK PMGCKITQTA TSTTVSGPPV
E.coli          GGTVKVTGIG RNSMQGDIRF AD.....VLE KMGATICWGD DYISCT....
S.typhimurium   GGTVKVTGIG RKSMQGDIRF AD.....VLE KMGATITWGD DFIACT....
P.hybrida       GGTITVEGCG TNSLQGDVKF AE.....VLE KMGAEVTWTE NSVTVKGPPR
P.sativum       GGTITVIGCG TSSLQGDVKF AE.....VLE KMGAKVTWTE NSVTVTGPPR
L.esculentum    GGTVTVEGCG TSSLQGDVKF AE.....VLE KMGAEVTWTE NSVTVKGPPR
A.thaliana      GETVTVEGCG TTSLQGDVKF AE.....VLE KMGCKVSWTE NSVTVTGPPR
B.pertussis     GP.VRVTGVG EDSIQGDV.. AF....AATL AAMGADVRYG PGWIETRGVR

                401                                                 450
A.nidulans      GILRATSKRG Y..GTNDRCV PRCFRTGSHR PMEKSQTTPP VSSGIANQRV
S.cerevisiae    GTLKPLKHVD MEPMTDAFLT ACVVAAISHD SDPNSANTTT IE.GIANQRV
E.coli          ........RG ELNAIDMDMN HIPDAAMTIA TAALFAKGTT RLRNIYNWRV
S.typhimurium   ........RG ELHAIDMDMN HIPDAAMTIA TTALFAKGTT TLRNIYNWRV
P.hybrida       ....SSSGRK HLRAIDVNMN KMPDVAMTLA VVALYADGPT AIRDVASWRV
P.sativum       ....DSSGRK VLQGIDVNMN KMPDVAMTLA VVALFANGPT AIRDVASWRV
L.esculentum    ....NSSGMK HLRAIDVNMN KMPDVAMTLA VVALFADGPT TIRDVASWRV
A.thaliana      ....DAFGMR HLRAIDVNMN KMPDVAMTLA VVALFADGPT TIRDVASWRV
B.pertussis     ....VAEGGR .LKAFDADFN LIPDAAMTAA TLALYADGPC RLRNIGSWRV

                451                                                 500
A.nidulans      KECNFIKAMK DELAKFGVIC REHDDGLE.. ....IDGIDR SNLRQPVGGV
S.cerevisiae    KECNFILAMA TELAKFGVKT TELPDGIQVH GLNSIKDLKV PSDSSGPVGV
E.coli          KETDRLFAMA TELRKVGAEV EEGHDY.... ....IRITPP EKLNF..AEI
S.typhimurium   KETDRLFAMA TELRKVGAEV EEGHDY.... ....IRITPP AKLQH..ADI
P.hybrida       KETERMIAIC TELRKLGATV EEGPDY.... ....CIITPP EKLNV..TDI
P.sativum       KETERMIAIC TELRKLGATV EEGPDY.... ....CVITPP EKLNV..TSI
L.esculentum    KETERMIAIC TELRKLGATV VEGSDY.... ....CIITPP EKLNV..TEI
A.thaliana      KETERMIAIC TELRKLGATV EEGSDY.... ....CVITPP KKVKT..AEI
B.pertussis     KETDRIHAMH TELEKLGAGV QSGADW.... ....LEVAPP EPGGWD.AHI
```

```
                501                                                  550
   A.nidulans   FCYDDHRVAF SFSV..... ...LSLVTPQ PTLILEKECV GKTWPGWWDT
  S.cerevisiae  CTYDDHRVAM SFSLLAGMVN SQNERDEVAN PVRILERHCT GKTWPGWWDV
       E.coli   ATYNDHRMAM CFS..... ...LVALSDT PVTILDPKCT AKTFPDYFEQ
 S.typhimurium  GTYNDHRMAM CFS..... ...LVALSDT PVTILDPKCT AKTFPDYFEQ
     P.hybrida  DTYDDHRMAM AFS..... ...LAACADV PVTINDPGCT RKTFPNYFDV
     P.sativum  DTYDDHRMAM AFS..... ...LAACGDV PVTIKDPGCT RKTFPDYFQV
  L.esculentum  DTYDDHRMAM AFS..... ...LAACADV PVTIKNPGCT RKTFPDYFEV
    A.thaliana  DTYDDHRMAM AFS..... ...LAACADV PITINDSGCT RKTFPDYFQV
    B.pertussis GTWDDHRMAM CFL..... ...LAAFPAA V.RILDPGCV SKTFPDYFDV

                551   559
   A.nidulans   LRQLFKVKL
  S.cerevisiae  LH
       E.coli   LARISQAA..
 S.typhimurium  LARMSTPA..
     P.hybrida  LQQYSKH...
     P.sativum  LERFTKH...
  L.esculentum  LQKYSKH...
    A.thaliana  LERITKH...
    B.pertussis YAGLLAARD.
```

Figure 5.1 Line-up of the deduced amino acid sequences of the EPSP synthases from nine species of bacteria, plants and lower eukaryotes

completely conserved in six of these sequences. It contains a proline residue (conserved in seven of the sequences), substitution of which with serine induces an increase in tolerance to glyphosate [Stalker *et al.*, 1985; Sost & Amrhein, 1990; Stallings *et al.*, 1991]. This consensus region also contains a glycine residue (Gly-96 in *E.coli*) which is conserved in all nine sequences, and which has also been implicated in the binding of both PEP and glyphosate to the enzyme, since substitution of this amino acid induces an increase in tolerance to glyphosate of up to 500-fold [Sost & Amrhein, 1990; Padgette *et al.*, 1991]. Indeed, the region against which primer 1 was designed is included in a highly conserved consensus region (^{90}LxLGNAGTAxRxL^{102}, where x denotes a non-conserved residue) identified in an alignment of the EPSP synthase sequences from Gram-negative bacteria, plants and lower eukaryotes [Padgette *et al.*, 1991].

Primer 2 was designed against the consensus region CTAKTFPD (residues **539-546** in **Figure 5.1**), which is fairly well conserved amongst the nine enzyme sequences, and is completely conserved in two of the three bacterial sequences. This region contains a cysteine which has been suggested as lying proximal to the binding-site for S3P [Padgette *et al.*, 1988a].

These oligonucleotides were intended to be used as primers for PCR cloning of the *aroA* gene. Therefore, primer 1 (the N-terminal oligonucleotide) was designed to anneal to the coding strand of the DNA, and primer 2 (the C-terminal oligonucleotide) designed against the reverse of the coding strand, so that it anneals to the non-coding strand (see **Figure 5.4**).

The alanine residues contained in the region against which primer 1 was designed may be encoded by any one of four nucleotide triplets, which differ only in the nucleotide in the third position. According to the codon preference table, the two triplets encoding alanine which are most commonly used in *Streptomyces* genes, GCG and GCC, encode approximately 92% of the alanine residues in the proteins analysed to date. Therefore, (GC) degeneracies were incorporated into the oligonucleotide to represent the third nucleotide of these alanine codons (**Figure 5.3**). A similar situation occurred with the triplets encoding the glycine and arginine residues, where (GC) degeneracies were also used to represent the third nucleotides of these codons.

At the third residue of the region against which primer 2 was designed, there are four different amino acids in the nine sequences (glycine in the two eukaryote sequences, alanine in two of the bacterial sequences with serine in the other one, and arginine in the four plant sequences). Alanine was included in the predicted protein sequence at

	G		A		T		C		
G	Gly	19	Glu	81	Val	39	Ala	33	G
	Gly	10	Glu	19	Val	2	Ala	4	A
	Gly	8	Asp	4	Val	4	Ala	3	T
	Gly	64	Asp	96	Val	56	Ala	59	C
A	Arg	6	Lys	95	Met	100	Thr	30	G
	Arg	1	Lys	5	Ile	3	Thr	3	A
	Ser	3	Asn	4	Ile	5	Thr	2	T
	Ser	28	Asn	96	Ile	92	Thr	65	C
T	Trp	100	END	14	Leu	3	Ser	26	G
	END	83	END	3	Leu	0	Ser	3	A
	Cys	13	Tyr	5	Phe	1	Ser	1	T
	Cys	87	Tyr	95	Phe	99	Ser	39	C
C	Arg	36	Gln	93	Leu	55	Pro	52	G
	Arg	5	Gln	7	Leu	0	Pro	2	A
	Arg	7	His	6	Leu	2	Pro	3	T
	Arg	46	His	94	Leu	39	Pro	43	C

Figure 5.2 Codon preference table for 67 genes from *Streptomyces*

N.B. a) The first nucleotide of each codon is provided in the left-hand column; the second nucleotide is shown in the top row; the third nucleotide is shown in the right-hand column

b) The figures show the relative use (expressed as a percentage) of each nucleotide triplet in encoding each amino acid in the 67 genes analysed

Data taken from Taylor, 1992

Primer 1:

Predicted protein sequence:

N ^{176}-Asn Ala Gly Thr Ala Met Arg Pro-^{183} C

Derived DNA sequence (non-coding strand):

5' -AAC GCn GGn ACn GCn ATG CGn CCn- 3'

Primer sequence:

5'	-AAC	GC(GC)	GG(GC)	AC(GC)	GC(GC)	ATG	CG(GC)	C- 3'
	96%	92%	83%	95%	92%	100%	82%	100%

Primer 2:

Predicted protein sequence:

N ^{539}-Cys Thr Ala Lys Thr Phe Pro Asp-^{546} C

Derived DNA sequence (non-coding strand) :

5' -TGC ACn GCn AAG ACn TTC CCn GAC- 3'

Derived DNA sequence (coding strand) :

3' -ACG TGn CGn TTC TGn AAG GGn CTG- 5'

Reverse of derived DNA sequence (coding strand) :

5' -GTC nGG GAA nGT CTT nGC nGT GCA- 3'

Primer sequence :

5'	-GTC	(GC)GG	GAA	(GC)GT	CTT	(GC)GC	(GC)GT	GC- 3'
	96%	95%	99%	95%	95%	82%	95%	87%

Figure 5.3 Design of primer 1 and primer 2 for cloning of *aroA*.

a) (GC) denotes degeneracies, where 50% of the oligonucleotides synthesised have a G at this position, the other 50% a C

b) Figures presented under the primer sequences are the percentage use of each codon for the amino acids which they encode in the 67 streptomycete coding regions used in the codon preference table [Taylor, 1992] (**Figure 5.2**).

c) n denotes that more than one of the four bases could occur in this position

d) Residue numbers in small block font refer to the amino acid residues in **Figure 5.1**

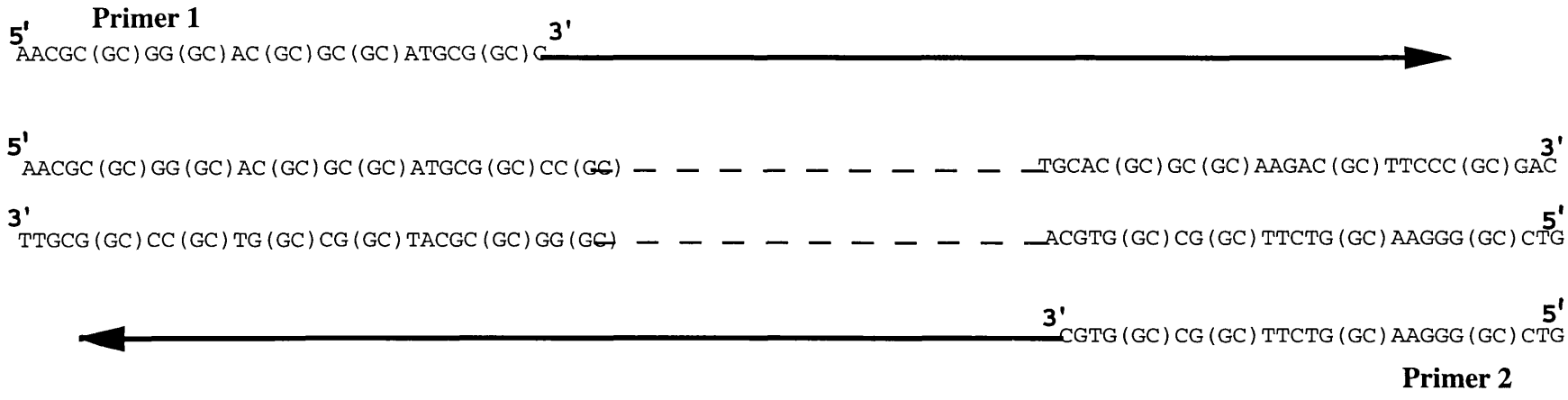


Figure 5.4 PCR strategy for cloning of *aroA*

this position as this amino acid is conserved in two of the three bacterial sequences, and it was anticipated that the EPSP sequence from *S.coelicolor* would show greater similarity to other bacterial sequences than to the enzymes from the eukaryotes and plants. In the DNA sequence of primer 2, (GC) degeneracies were used to encode the nucleotides in the third positions of the triplets encoding the threonine, alanine and proline residues.

The melting temperatures (T_m) of the oligonucleotides were calculated using the formula:

$$T_m = 59{\cdot}9 + 0{\cdot}5(G+C)\% - 600/L$$

where L is the length of the primer (in numbers of nucleotides).

The primers were designed to produce T_m's as close as possible to each other, allowing PCR to be performed as close as possible to the optimum T_m of each oligonucleotide. Each degeneracy in a primer sequence doubles the number of oligonucleotides synthesised for that sequence; an optimal primer sequence will contain no degeneracies, so that only a single oligonucleotide is synthesised for this sequence.

Primer 1 was 22 nt in length, contained four degeneracies (producing 16 individual oligonucleotides), was 77% G+C rich, and had a calculated T_m of approximately 71°C.

Primer 2 was 23 nt in length, also contained four degeneracies, was 70% G+C rich, and had a calculated T_m of approximately 69°C.

5.2.2 Attempted PCR-cloning of *aroA* with primer 1 and primer 2

The Polymerase Chain Reaction (PCR) allows two primers, which anneal to opposite strands of a denatured DNA duplex, to prime DNA replication from the duplex, which acts as a template for binding of the primers (**Figure 5.4**). The first round of replication produces two partial duplexes, each consisting of a single strand of template DNA and a newly-synthesised strand with one of the primers at the 5' end. For each duplex thus produced, a second round of denaturation, followed by primer binding and DNA replication, will produce a duplex identical to the template for the second round of replication, and a duplex consisting of one strand which was synthesised during the first round of replication (with one of the primers at its 5' end) and its complement, which was synthesised during the second round of replication, and has the other primer at its 5' end. Further rounds of replication will cause further

amplification of this novel duplex (the PCR product), with one of the primers at each 5' end.

Typically, PCR using Taq polymerase produces a product with a single A base overhanging at the 3' end of the product. This base may be removed for ligation to a blunt-ended, linearised vector; or the PCR product may be ligated into a linearised vector which has a single overhanging T base at both 3' ends. All of the ligations of PCR products described in this thesis used the latter method of ligation.

The PCRs described in this section were performed using *S.coelicolor* genomic DNA (gDNA) as the initial template, and primer 1 and primer 2, either singly (for single primer controls) or in tandem. The PCRs were performed according to section **2.4.25**, using an annealing temperature of 64°C, and a 60 second extension. The 10x Taq buffer used contained no Mg^{2+}, so $MgCl_2$ was added to supplement the PCR reaction cocktail at concentrations of Mg^{2+} ranging from 0·5 mM to 3 mM. Each PCR reaction was covered with a single drop of mineral oil to prevent evaporation during the reaction. After the initial denaturation step, 1 U Taq polymerase was added to each reaction to catalyse polymerisation. It was anticipated that these primers would direct the synthesis of a PCR product of between 950 and 1100 bp in length.

PCR of 1147 gDNA using either 2 or 2·5 mM Mg^{2+} produced two PCR products of approximately 0·5 to 1 kb in size. These two products were ligated into the vector pDK101 which had been digested with the restriction endonuclease *Xcm* I (this digestion produces a single base-pair stagger at the restriction site, with an overhanging T at the 3' end of the DNA, to which the single overhanging A at the 3' end of the PCR product can ligate [Kovalic *et al.*, 1991]), transformed into competent *E.coli* TG-1 cells, and the transformation mixture was plated onto L-agar plates supplemented with ampicillin to select for transformants, and with X-gal and IPTG to screen for transformants carrying plasmids which contain inserts (i.e. not merely religated plasmid). Of the putative transformants, only one contained the smaller PCR product (approximately 800 bp in length), and none of the transformants contained the larger product. The plasmid containing the insert was named pTA+800.

pTA+800 was sequenced by the double-stranded sequencing protocol using T7 polymerase and T7 primer. 54 nt of insert sequence were derived (**Figure 5.5**). The first 22 nucleotides showed complete identity to the primer 1 sequence. The deduced amino acid sequence from this insert lying downstream of the region against which primer 1 was designed showed no similarity to that of any of the deduced protein sequences in **Figure 5.1**. Therefore, this PCR product was rejected as containing part

of the *aroA* gene.

```
1  AACGCGGGGACGGCGATGCGGCGCnCGGCCAGGCGGCCCnnCAGTTCCCTCACC  54

   N  A  G  T  A  M  R  R  ?  A  R  R  P  ?  S  S  L  T
   T  R  G  R  R  C  G  A  R  P  G  G  P  ?  V  P  S
   R  G  D  G  D  A  A  ?  G  Q  A  A  ?  Q  F  P  H
```

Figure 5.5 Partial DNA and deduced amino acid sequence of pTA+800

N.B. Primer 1 sequence is denoted in bold type

5.2.3 Design of oligonucleotide primer 4

Primer 4 was designed against a region (WRVKET(D/E)R) of deduced amino acid sequences for the EPSP synthase enzymes shown in **Figure 5.1** (residues **448-455**). Again, this region is highly conserved in all nine deduced amino acid sequences. Primer 4 was designed against the deduced DNA sequence of the coding strand, allowing it to anneal to the non-coding strand of the *aroA* gene and to allow DNA to be synthesised towards the 5' end of the gene (and therefore towards primer 1), thus providing an alternative to primer 2 as the C-terminal oligonucleotide. Therefore, primer 2 and primer 4 function as nested primers, permitting primer 4 to anneal towards the downstream end of the *aroA* PCR product formed using primer 1 and primer 2.

5.2.4 Attempted PCR-cloning of *aroA* with primer 1 and primer 4

In an attempt to prevent annealing of the primers at incorrect sites internal to the PCR product, and thus producing spurious smaller PCR products, "touchdown" PCR [Don *et al.*, 1991] was performed with primer 1 and primer 4, either singly (single primer controls) or in tandem. In this method, the annealing temperature is gradually decreased during the PCR reaction, biasing the reaction towards longer (and hopefully correct) PCR products. The touchdown PCRs described below were performed using 100 ng of gDNA isolated from mycelia of *S.coelicolor* 1147 as template, and a 5 mins extension time. It was anticipated that these primers would direct the synthesis of a PCR product of 750-850 bp.

Touchdown PCRs were performed by decreasing the annealing temperature by 1°C every second cycle. The target temperatures (the temperature at which the primers anneal to the denatured DNA in the 30^{th} cycle) were varied, in order to discover the

Primer 4:

Predicted protein sequence:

N ^{448}-Trp Arg Val Lys Glu Thr Asp/Glu Arg-^{455} **C**

Derived DNA sequence (non-coding strand):

5' -TGG CGn GTn AAG GAG ACn GAn CGn- **3'**

Derived DNA sequence (coding strand):

3' -ACC GCn CAn TTC CTC TGn CTn GCn- **5'**

Reverse of derived DNA sequence (coding strand):

5' -nCG nTC nGT CTC CTT nAC nCG CCA- **3'**

Primer sequence:

5'	-(GC)CG	(GC)TC	(GC)GT	CTC	CTT	(GC)AC	(GC)CG	CC-	**3'**
	82%	88·5%	95%	81%	95%	95%	82%	100%	

Figure 5.6 Design of primer 4 for cloning of the *aroA* gene

optimal temperature for this reaction, from 65°C to 68°C. Also, in order to circumvent primer-primer annealing or primer self-annealing, the primers were added together along with the dH_2O required for the PCR, boiled for 10 mins to denature any such partial (or even complete) duplexes, and immediately placed on ice. The rest of the PCR cocktail (apart from the Taq polymerase, which is added after the initial denaturation step of the reaction) is made up on ice. However, no PCR products which were not single-primer products were identified.

5.2.5 Attempted PCR-cloning of *aroA* with primer 1, primer 2 and primer 4

PCRs were then performed with primer 1 and primer 2, using an annealing temperature of 60°C, and 0·5 to 3 mM Mg^{2+}. After completion of the reactions, the PCR reactions were run out on agarose gels, and the products purified from the gels. These PCR products were then used as templates for secondary PCR reactions performed at identical Mg^{2+} concentrations. However, these secondary reactions were performed at an annealing temperature of 66°C, and primer 2 (which was used as the C-terminal oligonucleotide in the primary PCRs) was replaced by primer 4 in these reactions.

The products from these six secondary PCRs were electrophoresed on 0·8% TAE gels; a single band was isolated from the 0·5 mM Mg^{2+} reaction, and three bands isolated from the 2 mM Mg^{2+} reaction. However, all of these products were considerably smaller than anticipated, ranging up to only 500 bp for the largest of the three products from the 2 mM Mg^{2+} reaction. All four PCR products were purified, and were subsequently ligated into pIBI25 vector which had been digested with *Hin*c II (which produces blunt-ended restriction fragments) and T-tailed (section **2.4.9**). The resultant plasmids were transformed into competent *E.coli* DS941 cells, and the mixture plated onto L-agar supplemented with ampicillin, X-gal and IPTG. Only the largest fragment was cloned by this method, generating plasmid p41.

The insert was isolated from p41 by digestion with the restriction endonucleases *Hin*d III and *Eco*R I, and subcloned into suitably-restricted M13mp18 and M13mp19, creating m41-18 and m41-19. Single-stranded DNA was prepared from each of these two vectors for sequencing with the universal primer.

The sequence derived from each of the two bacteriophages began with M13 polylinker sequence, followed by pIBI25 sequence upto the triplet GTC, the 5' half of the *Hin*c II restriction site. This was immediately followed by the T nucleotide ligated onto the linearised pIBI25 vector by T-tailing, and subsequently by the insert.

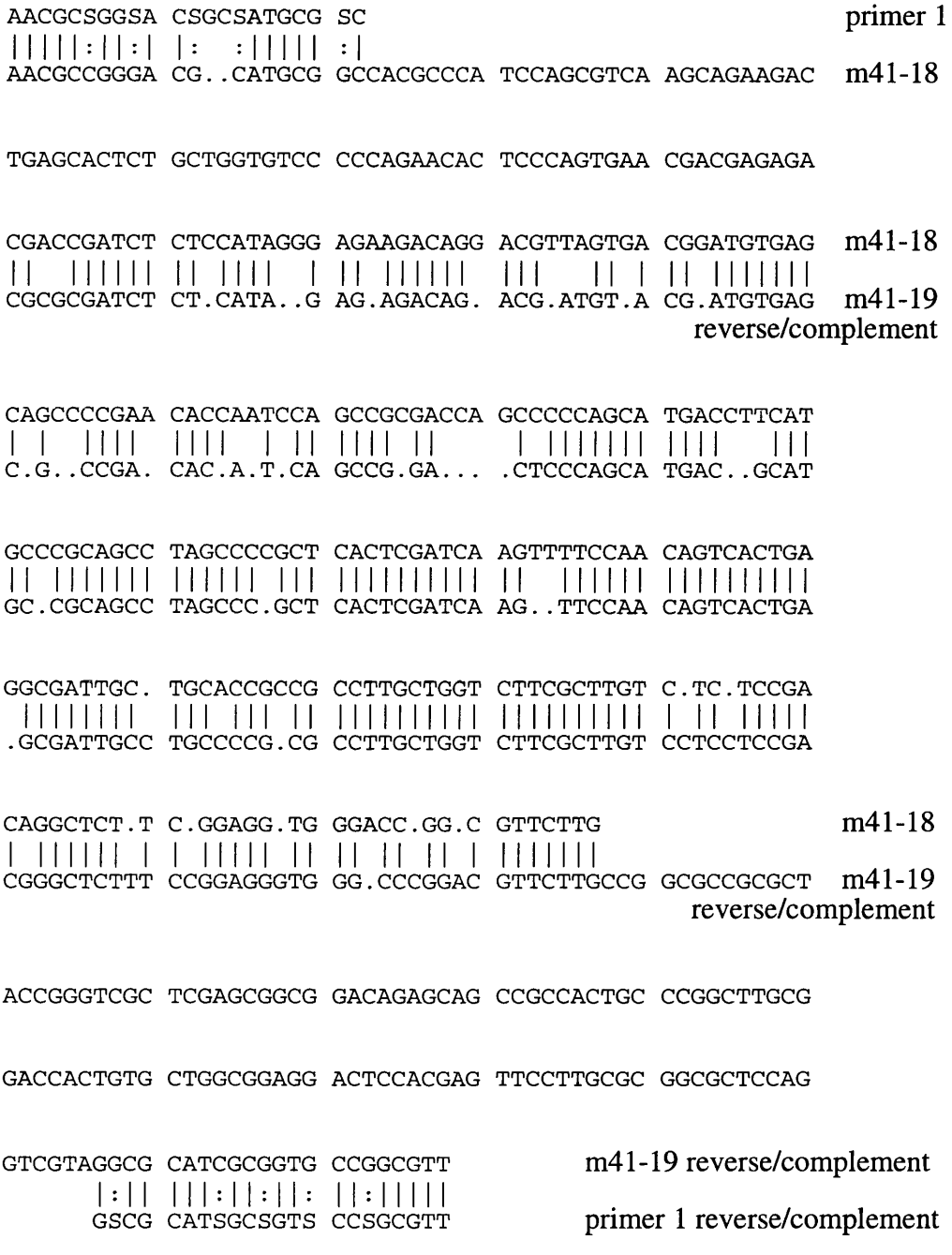


Figure 5.7 Partial nucleotide sequence of the insert from p41

The sequences derived from both m41-18 and m41-19 contained a region at their 5' end which bore considerable similarity to the primer 1 sequence. Therefore, it appeared that the insert contained in p41 has primer 1 at each end, and was generated by PCR using only a single primer.

The two sequences were translated to produce the amino acid sequence they encode [**Figure 5.8**]. The nucleotide sequence derived from m41-18, when lined up with the primer 1 sequence, contained an apparent 2 nt gap. This may be due to a sequencing error, as the PCR product must begin with a primer at each 5' end. Therefore, the adapted m41-18 sequence, beginning with the primer 1 sequence, was also translated. These derived amino acid sequences were used to search for similarity to any of the protein sequences lodged on the protein databases. However, no significant matches were produced with any of these sequences.

Amino acid sequence derived from m41-18 nucleotide sequence:

```
NAGTHAATPI QRQAED*ALC WCPPEHSQ*T TRDDRSLHRE KTGR**RM*A
APNTNPAATS PQHDLHARSL APLTRSSFPT VTEAIAAPPP CWSSLVSPTG
VFQQSLRRLL HRRLAGLRLS LRQALRRWDR RS
```

Amino acid sequence derived from the adapted m41-18 nucleotide sequence:

```
NAGTAMRPRP SSVKQKTEHS AGVPQNTPSE RRETTDLSIG RRQDVSDGCE
QPRTPIQPRP APSMTFMPAA *PRSLDQVFQ QSLRRLLHRR LAGLRLSLRQ
EFSNSH*GDC CTAALLVFAC LSDRLFGGGT GVL
```

Amino acid sequence derived from m41-19 nucleotide sequence:

```
NAGTAMRLRP GAPRKELVES SASTVVRKPG SGGCSVRRSS DPVRAAPART
SGPTLRKEPV GGGQAKTSKA RGRQSLSDCW NLIE*AG*AA ACVMLGVRLM
CRRSHRTSSV SL*EIA
```

Figure 5.8 Amino acid sequences derived from the p41 insert

N.B. * denotes a STOP codon

5.2.6 Genomic probing with primer 1

Since cloning of *aroA* by PCR was not successful, it was decided to attempt the cloning of *aroA* directly from the genome by the screening of a genomic sub-library with the PCR primers already available. The library was produced by cloning a portion of the genome into a plasmid vector. The size of fragments required for cloning was determined by screening digests of genomic DNA isolated from *S.coelicolor* 1147 with the three primers described above. An oligonucleotide (or fragment of single-stranded DNA) which hybridises to the gene sequence should light up a small number of bands in the genomic digest which will contain all, or at least part of, the gene being sought. If a fragment contains only part of the gene, it may be used as a single-stranded probe for screening gDNA digested with another restriction enzyme.

Genomic DNA was digested to completion with a number of different restriction enzymes, singularly or in pairs, and the fragments produced separated by agarose gel electrophoresis. The DNA was transferred by Southern blotting to nylon membrane, and was fixed to the filter by baking at 80°C for 2 hours. The filter was prehybridised (see section **2.4.18**) at 50°C, and probed with primer 1.

Primer 1 was radio-labelled using bacteriophage T4 polynucleotide kinase enzyme, which attaches the γ-phosphate from a molecule of ATP to the hydroxyl group at the 5' end of the oligonucleotide. Various hybridisation and washing conditions were used, with differing concentrations of salt and at various temperatures, with suitable conditions being identified as:

hybridise overnight at 50°C

wash twice for 20 minutes each at 58°C in 1x SSC,
0·1% SDS,
0·1% sodium pyrophosphate.

Hybridising fragments of a suitable size for cloning and further examination were identified in *Bam*H I/*Eco*R I and *Pst* I/*Sst* I double-digests (**Figure 5.9**).

5.2.7 Construction of genomic sub-libraries

Approximately 15 μg of *S.coelicolor* G-94 genomic DNA was digested to completion in pairwise fashion with either *Bam*H I and *Eco*R I, or *Pst* I and *Sst* I restriction endonucleases, and the fragments separated by electrophoresis through a 0·8% TAE agarose gel. Fragments of approximate size 4 to 4·5 kb were cut out of the gDNA which had been digested with *Bam*H I and *Eco*R I; 3·5 to 4 kb fragments were cut out

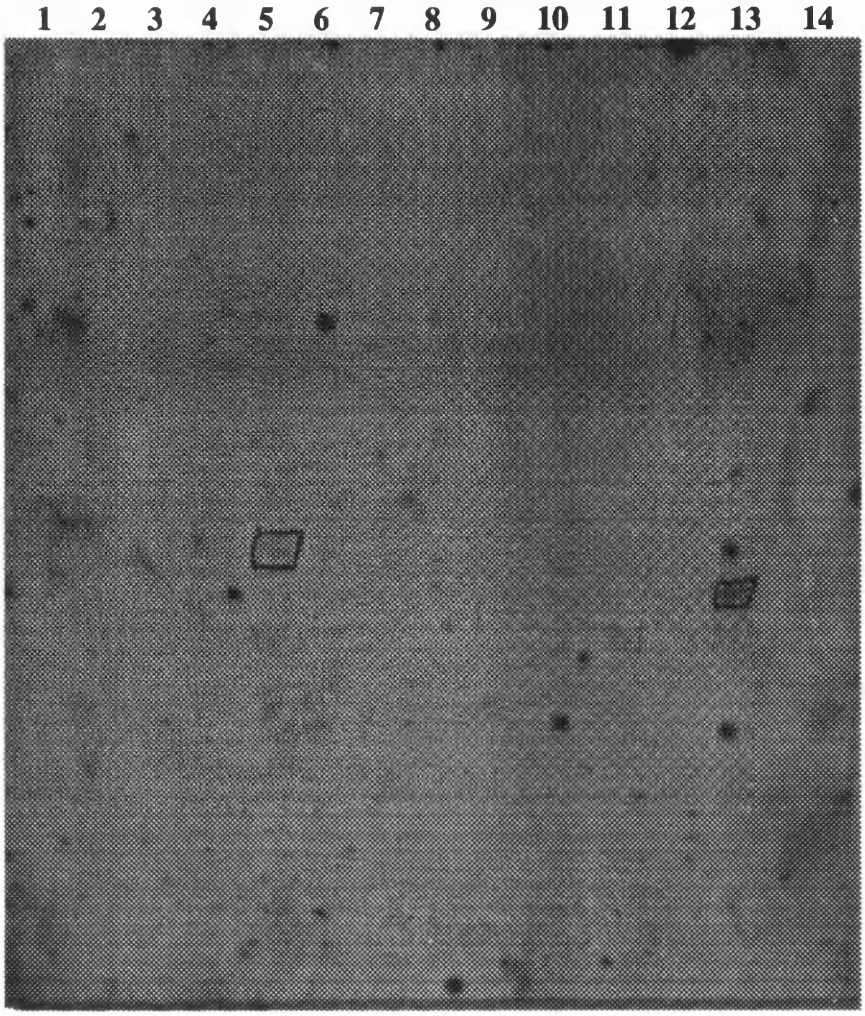


Figure 5.9 Probing with radio-labelled primer 1 of a filter carrying gDNA from 1147 digested with a variety of restriction enzymes

lanes 1&14: λ/*Hin*d III markers
lane 2: 1147 gDNA cut with *Bam*H I
lane 3: 1147 gDNA cut with *Eco*R I
lane 4: 1147 gDNA cut with *Sal* I
lane 5: 1147 gDNA cut with *Bam*H I/*Eco*R I (boxed region ~4 to 4·5 Kb)
lane 6: 1147 gDNA cut with *Bam*H I/*Sal* I
lane 7: 1147 gDNA cut with *Eco*R I/*Sal* I
lane 8: 1147 gDNA cut with *Hin*d III
lane 9: 1147 gDNA cut with *Pst* I
lane 10: 1147 gDNA cut with *Sst* I
lane 11: 1147 gDNA cut with *Hin*d III/*Pst* I
lane 12: 1147 gDNA cut with *Hin*d III/*Sst* I
lane 13: 1147 gDNA cut with *Pst* I/*Sst* I (boxed region ~3·5 to 4 Kb)

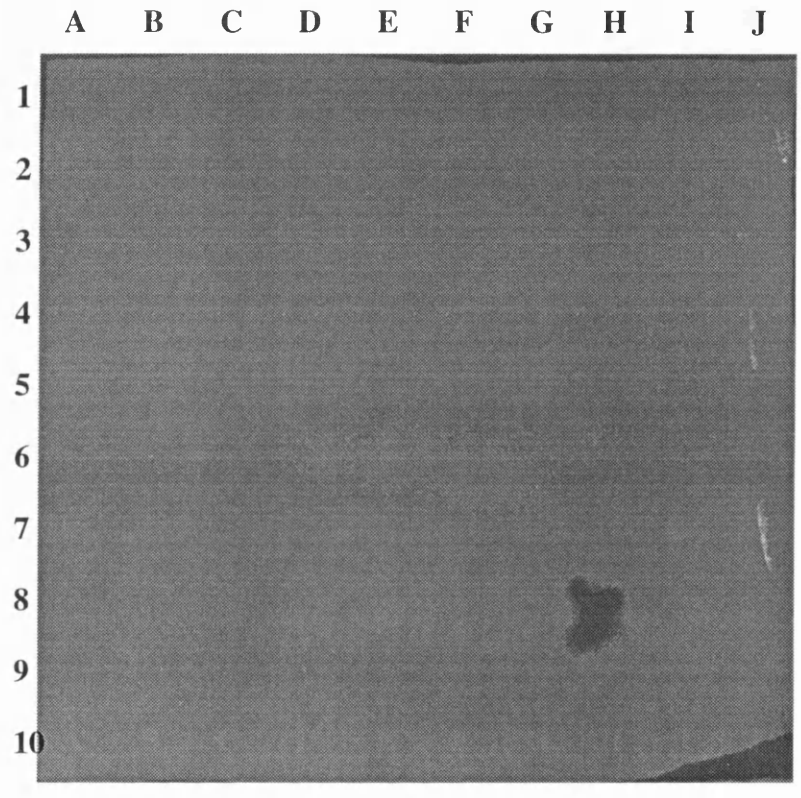


Figure 5.10 Probing of filter carrying putative pPM7-containing colonies with radio-labelled primer 1

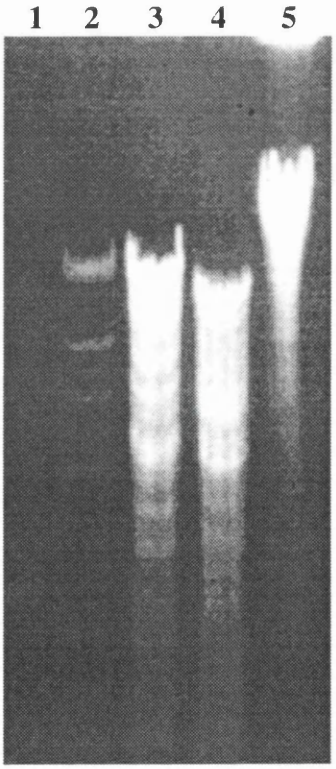
of the gDNA which had been digested with *Pst* I and *Sst* I. Both sets of fragments were eluted from the agarose, and ligated into suitably-restricted pIBI25 plasmid vector, producing genomic sub-libraries. The ligation mixes were transformed into competent *E.coli* DS941 cells, transformants with inserts being selected for by growth on L agar supplemented with ampicillin, and with X-gal and IPTG. 200 white (putative insert-containing) colonies from each transformation were crossed onto L agar supplemented with ampicillin for maintenance of stocks, and onto nylon membranes placed on top of L agar supplemented with ampicillin. This procedure allowed screening of the sub-libraries with primer 1 to identify transformants containing the fragments to which the primer hybridised during the probing of the genomic blot. The clones constituting the sub-libraries which had been plated onto nylon membrane had their DNA transferred to the membrane on which they had grown, and the filters probed with radio-labelled primer 1 by colony hybridisation (**2.4.19**). A single clone (colony **H8** in **Figure 5.10**) containing a *Bam*H I/*Eco*R I fragment to which primer 1 hybridised was identified from the filters. The plasmid contained in the hybridising clone was named pPM7. No clones containing hybridising *Pst* I/*Sst* I fragments were obtained, despite repeated screening.

To investigate why the *Bam*H I/*Eco*R I fragment contained in pPM7 was of an estimated length of 2·3 kb despite the DNA which was excised from the gel being approximately 4 to 4·5 kb in length, 5 μg of 1147 gDNA was digested with *Bam*H I and/or *Eco*R I was run on a 0·8% TAE gel, along with 5 ng of pPM7 DNA which had been digested with *Bam*H I and *Eco*R I to separate the insert and vector bands (**Figure 5.11a**). The DNA was transferred to nylon membrane by Southern blotting, the membrane probed with radio-labelled insert from pPM7 and subsequently washed at the conditions previously identified. From the autoradiogram (**Figure 5.11b**) it may be seen that an approximately 2·3 kb band lights up in the lane containing digested pPM7 DNA as anticipated. However, in the lane containing 1147 gDNA digested with both enzymes, a single band of approximately 3 kb in length lit up. As these two bands of DNA are likely to be the same fragment, it appears that the genomic DNA is running higher on the gel than anticipated from the λ/*Hin*d III markers. This result suggests that the insert in pPM7 is likely to be the fragment identified in the initial genomic DNA hybridisation, but that the gDNA runs aberrantly, possibly due to the high concentration of gDNA run on the gel.

5.2.8 Restriction mapping of pPM7

Plasmid pPM7 DNA was isolated from the hybridising clone, and was shown to contain an *Eco*R I/*Bam*H I genomic insert of approximately 2·3 kb in length. This

a) Digests of 1147 gDNA and pPM7



b) Autoradiogram

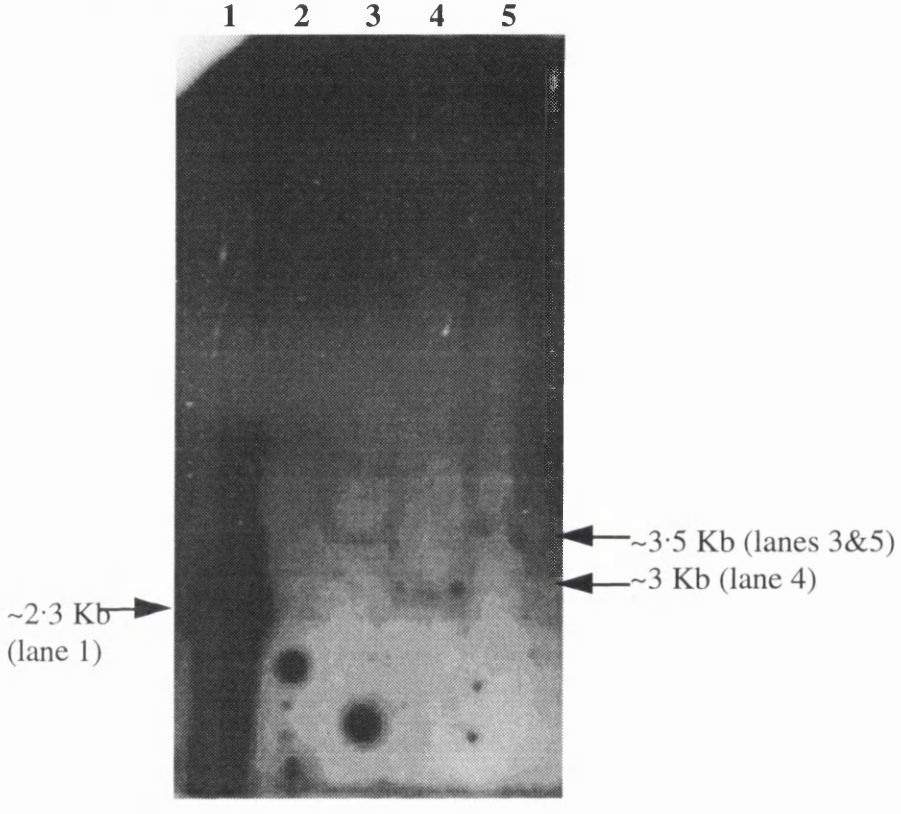


Figure 5.11 Digests of plasmid and genomic DNA probed with radiolabelled primer 1

lane 1: ~5 ng pPM7 digested with *Bam*H I/*Eco*R I

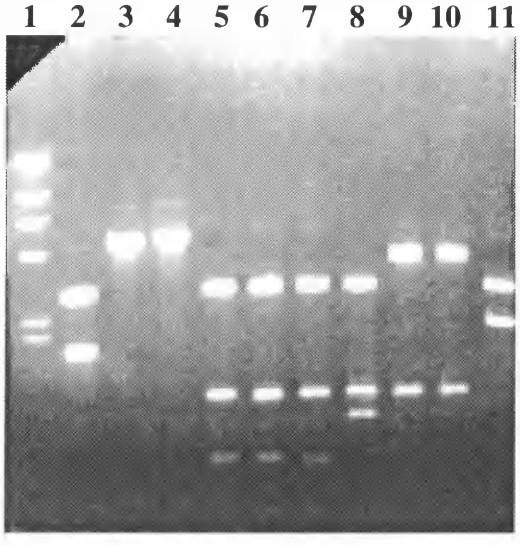
lane 2: λ/*Hin*d III markers

lane 3: 1147 gDNA digested with *Bam*H I

lane 4: 1147 gDNA digested with *Bam*H I/*Eco*R I

lane 5: 1147 gDNA digested with *Eco*R I

a) **Restriction digests of pPM7**



b) **Autoradiogram**

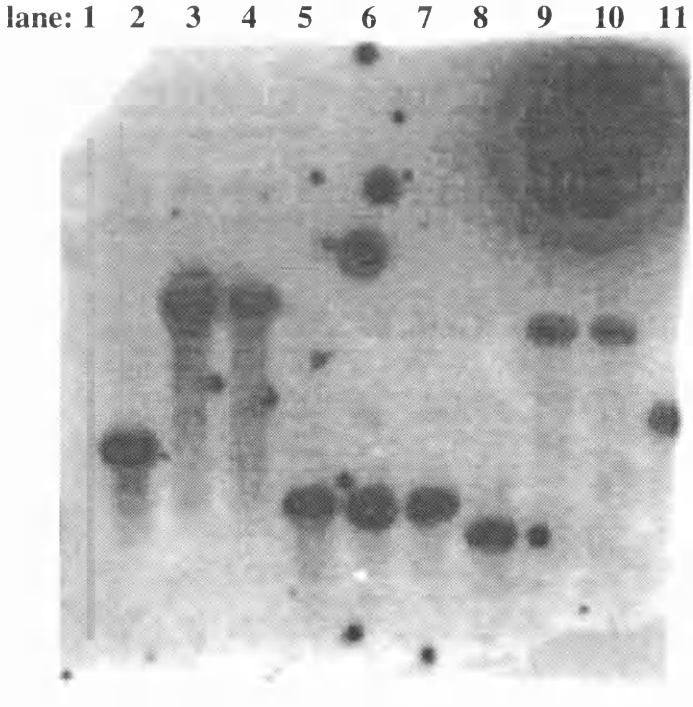


Figure 5.12 Probing of restriction digests of pPM7 with radiolabelled primer 1

lane 1: λ/*Hin*d III
lane 2: pPM7/*Eco*R I/*Pst* I
lane 3: pPM7/*Pst* I
lane 4: pPM7/*Bam*H I/*Pst* I
lane 5: pPM7/*Eco*R I/*Xho* I
lane 6: pPM7/*Xho* I
lane 7: pPM7/*Bam*H I/*Xho* I
lane 8: pPM7/*Eco*R I/*Sal* I
lane 9: pPM7/*Sal* I
lane 10: pPM7/*Bam*H I/*Sal* I
lane 11: pPM7/*Eco*R I/*Bam*H I

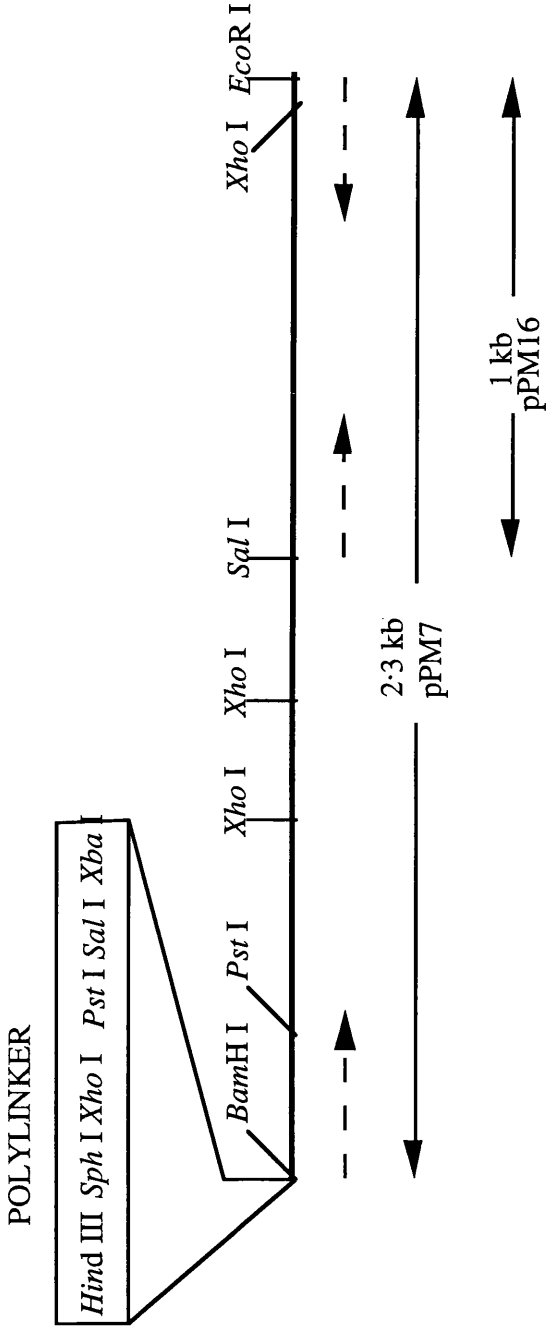


Figure 5.13 Restriction mapping and sequencing strategy of plasmid pPM7

fragment was mapped with a variety of restriction endonucleases. Of those enzymes used, only *Xho* I, *Sal* I and *Pst* I were found to cut in this 2·3 kb fragment; there were three *Xho* I sites, along with one site for each of the other two endonucleases in the fragment.

The plasmid pPM7 was digested with these three endonucleases, either singly or in tandem with the endonucleases which cut at the ends of the insert (*Eco*R I or *Bam*H I). The digests were then run out on a 0·8% TAE agarose gel, and the DNA transferred to nylon membrane by Southern blotting. The membrane was probed with radio-labelled primer 1. The smallest fragment which lit up on the autoradiogram was an approximately 1 kb *Eco*R I/*Sal* I fragment (**Figure 5.12**). From the restriction patterns produced by single and double digests using these three endonucleases, a restriction map was produced (**Figure 5.13**).

5.2.9 Further subcloning, partial sequencing and analysis of pPM7

The 2·3 kb insert was isolated from pPM7 by digestion with the *Bam*H I and *Eco*R I endonucleases, and subcloned into the bacteriophage vectors M13mp18 and M13mp19, creating mPM7-18 and mPM7-19 respectively, for single-stranded sequencing using the M13 Universal primer. The M13 bacteriophage vectors allow the isolation of single stranded DNA, when transformed into an *E.coli* strain carrying an F', for use in single-stranded sequencing reactions.

The 1 kb *Eco*R I/*Sal* I fragment identified above (**Figure 5.12**) as containing the binding-site for primer 1 was isolated from pPM7 using the *Eco*R I and *Sal* I endonucleases, and subcloned into suitably-restricted pIBI25, creating plasmid pPM16, and also into suitably-restricted M13mp18, creating mPM16-18. This insert was not subcloned into M13mp19, as analysis of this clone would produce the same sequence as that of mPM7-19.

5.2.9.1 Sequencing of mPM7-18

Single-stranded DNA was isolated from mPM7-18 (see **2.4.15.2**), and sequenced using Universal primer. 348 nt of insert sequence were derived from mPM7-18, beginning with the *Bam*H I restriction site (GGATCC) (**Figure 5.14**). This sequence confirms the position of the *Pst* I restriction site (CTGCAG; nucleotides **338-343** in **Figure 5.14**), previously identified by restriction mapping (see **Figure 5.13**), which lies in the 2·3 kb *Bam*H I/*Eco*R I insert.

This sequence was analysed using the BIBBPROG (Bibb *et al.*, 1984). This programme analyses the average G+C percentage at each position to identify an open reading frame. The codons in genes from *Streptomyces* typically contain a relatively intermediate level (around 70%) of G or C nucleotides in the first position, a relatively low level (around 50%) in the second position, and a high level (90% or greater) in the third position.

```
  1   GGATCCTCTG  ACTCCTCACG  AGCAGAGGCG  CGAGCCCCGG  CTATCACGAT
 51   GTCCTCAGCA  CACCAGGTCC  GGCGACACGC  GTCGTCCCTC  GTCGATGCGG
101   ACATCGTCGT  CACCGCAACG  TGCCTGACAG  CGCCACGTCC  TGCGGACCGC
151   CCGACTCCGT  CGCCGACTGG  CGGCCGCGGA  CCCTGCGGGC  CAAACCGACG
201   AGGACCGAGC  GCAACGCCTC  GGCCCGACGG  CTGGATGGGA  CGAACATCTT
251   GCTGCTTGAG  CCGCGCGGCG  CGCTTGGTGC  TTGGTGATGA  ACAGACGCAG
301   CCGGCTCCAG  GCGTCAGGCG  CGCAGGTCGT  CGGGTSTCTG  CAGGCCGG
```

Figure 5.14 Partial nucleotide sequence of mPM7-18

The BIBBPROG analysis identified an open reading frame in this DNA sequence, which contains neither an in-frame start (ATG or GTG) or stop (TGA, TAG or TAA) codon. The DNA sequence was then translated to produce the putative amino acid sequence. This polypeptide sequence was used to search the NIH protein databases using the TBLASTN programme from the BLAST suite, to investigate whether it shows any significant similarity to any of the proteins contained in these databases. However, this putative polypeptide sequence showed no significant similarity to any of the deduced amino acid sequences on the database. This suggests that the putative open reading frame included in this sequence is likely to form part of a gene whose cloning and sequencing had not been reported previously, or whose gene product is not similar to any protein yet sequenced.

5.2.9.2 Sequencing of mPM7-19

Single-stranded DNA was isolated from mPM7-19, and sequenced using the Universal primer. 392 nt of insert sequence were derived from this clone, beginning with the *Eco*R I restriction site (GAATTC) (**Figure 5.15**). This sequence confirms the position of the *Xho* I restriction site (CTCGAG; nucleotides **36-41** in **Figure 5.15**), as estimated from the restriction map (**Figure 5.13**).

Sequence analysis using the BIBBPROG suggests that this DNA contains a putative open reading frame, although no in-frame start or stop codons were identified. The DNA sequence was translated to produce an amino acid sequence, which was used to search the NIH nucleic acid databases using TBLASTN. This putative polypeptide sequence also exhibited no significant similarity to any of the proteins contained on the databases.

```
  1   GAATTCGATC  TCGTGGTCGG  CGCGGACGGG  ATGCGCTCGA  GCGTGCGCCG
 51   GATGGTGTTC  GGTCCGGACG  AGGACTACCT  GGCGAACTGG  AAGGCGATGA
101   TCTGCGCGTT  CCAGATGAAG  GAGCAGGTTC  CCACCTACGA  GGCGAGCGAC
151   AGCATCATCG  TCTCGCGCCC  CAAGCGTGCG  ATGTGGGTGT  TCGGACTCGC
201   CGATCACGCG  CCCTGCGTGC  TGACCTACCG  CACCGACGAC  ATTCCGGACC
251   GGTTCGCCGG  GCCGTCGATC  GAGCAGGTGC  GCGCGGCCTT  TTCCGGGTTA
301   GAGGACGACC  CCGTGGTGCT  ACACTCCTGG  ACTCTCTGGT  AGCAGGCCCC
351   TGACCACGTG  TTCGATTACG  TACCATCAGG  TGAAGATGGA  CC
```

Figure 5.15 Partial nucleotide sequence of mPM7-19

5.2.9.3 Sequencing of mPM16-18

Single-stranded DNA was isolated from mPM16-18, and sequenced using the Universal primer. 313 nt of insert sequence were derived from this clone, beginning with the *Sal* I restriction site (GTCGAC) (**Figure 5.16**).

```
  1   GTCGACCGCG  GCCTGTCGGC  CTCAGGGAAC  AGCCCGACGA  AGTAGCCCCC
 51   GGTGCGCCGC  TCGGGCGCCG  TTCGACGATC  ACCGGCGTCC  AACCGGCCCG
101   GCGCASCGAT  CGCCGCGGAC  ATCCCGCGAT  CCCGAGCCCC  ACCACCAACG
151   CCCTCTTCTG  CATGGCCGTC  ACCGCGTCCT  GCCGGAGGGG  ACAGCGGCCG
201   CAGTCGGAGC  CGCAGCCGCG  TGAACTCGCG  CAGCGCGTGG  ACGACCTGCG
251   GACAGATCGA  GAACGCCGAA  ACCAGCACCT  GACGCGCTGA  CGTACGACGT
301   GTTGTTTGAT  CGT
```

Figure 5.16 Partial nucleotide sequence of mPM16-18

This nucleotide sequence contains a putative open reading frame, as suggested by the BIBBPROG, and so was translated in-frame to produce the amino acid sequence which is, putatively, encoded by this DNA sequence. This sequence lacks both a start and stop codon. The putative polypeptide sequence was used to search the NIH nucleic acid databases using TBLASTN. However, no significant matches were identified by this search. Therefore, this deduced polypeptide does not have similarity to any sequences already contained in the databases.

```
     3'                                                  5'
150  GCCCTCTTCTGCATGGCCGTCACCGCGTCCTGCCGGAGGGG  190
               :  ||||:||:||:  |:||||
     .......GSCGCATSGCSGTSCCSGCGTT............
```

Figure 5.17 Putative binding-site for primer 1 in the insert of pPM7

The nucleotide sequence derived from mPM16-18, and the reverse/complement of the primer 1 sequence, running 5' to 3', are shown above to demonstrate the similarity between primer 1 and the derived sequence. In the diagram above, primer 1 would be intended to anneal from nt **178** towards nt **150**, and prime DNA synthesis 5' to 3' as the sequence is written.

This nucleotide sequence contains the putative binding-site for primer 1. This putative binding-site contained 18 matches with the primer, including the five degeneracies in the primer (see **Figure 5.17**). However, two of the three nucleotides at the 3' end of the oligonucleotide did not match with the nucleotides in the derived binding site sequence. As probing of genomic DNA with primer 1 only lit up a single band, it is assumed that primer 1 only has a single binding-site in the genome. Therefore, the mismatches at the 3' end of the primer are likely to prevent this oligonucleotide from initiating any DNA replication in PCR.

5.2.10 Line-up of the new deduced amino acid sequences of EPSP synthases

Since the attempted cloning of the gene encoding EPSP synthase (as described above) was begun, a number of genes which encode EPSP synthase enzymes have been cloned and sequenced from Gram-positive and Gram-negative bacteria, lower eukaryotes and plants.

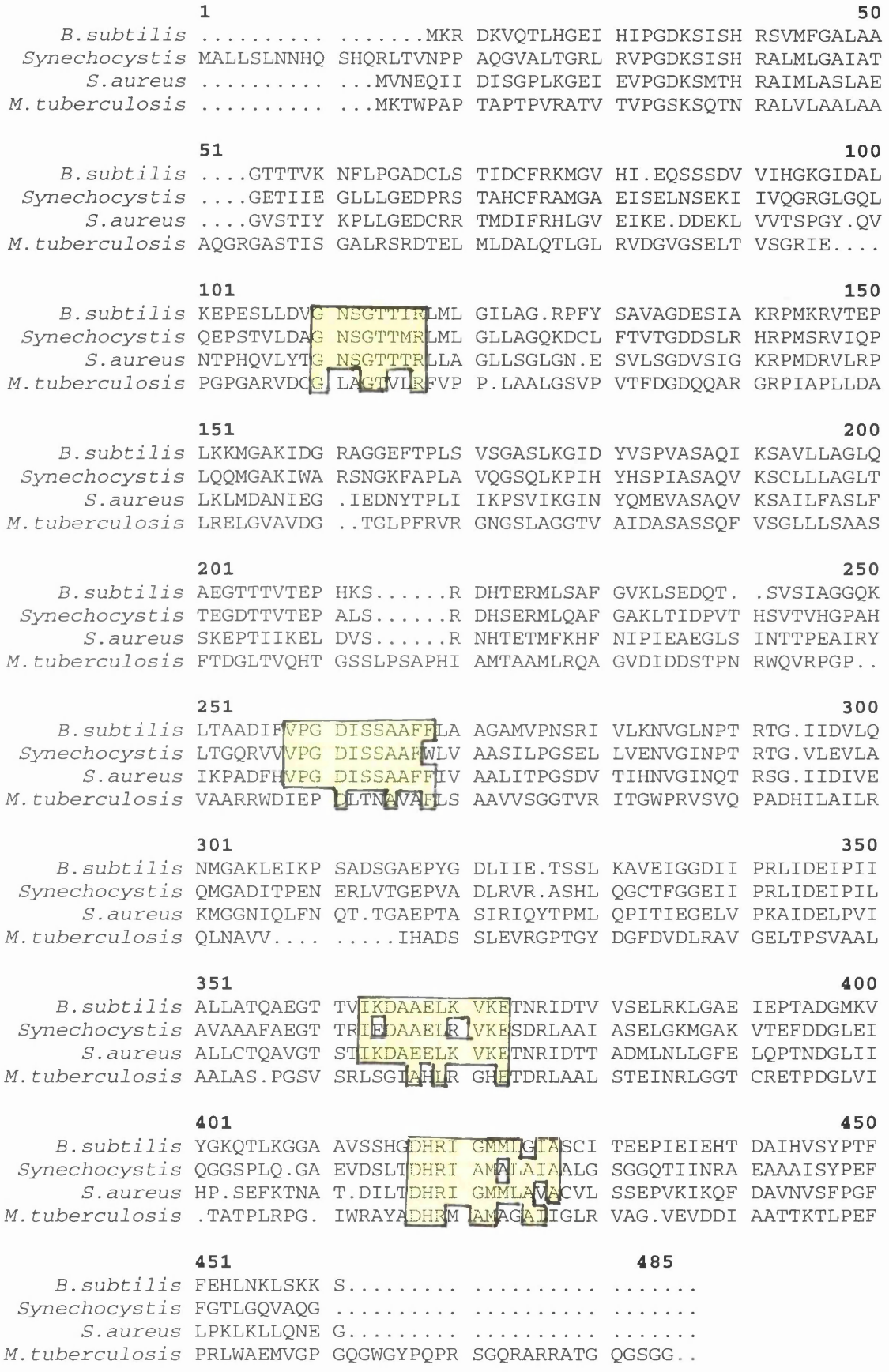


Figure 5.18 Line-up of the protein sequences of the EPSP synthase enzymes from the Gram-positive bacteria and a cyanobacterium

```
                1                                                   50
          Tobacco ..........  ..........  ..........  ..........  ..........
         Tobacco2 SIFIGSKKIT  QNSAKSLWVS  KEDSVLRVAK  SPFRISASVV  TAQKPNEIVL
          B.napus PFSVSLKTHQ  PRASSWGLKK  SGTMLNGSVI  RPVKVTASVS  TSEKASEIVL
          S.typhi ..........  ..........  ..........  ..........  ....MESLTL
      S.gallinarum ..........  ..........  ..........  ..........  ....MESLTL
      K.pneumoniae ..........  ..........  ..........  ..........  ....MESLTL
         Y.entero. ..........  ..........  ..........  ..........  ...MLESLTL
      H.influenzae ..........  ..........  ..........  ..........  ....MEKITL
       P.multocida ..........  ..........  ..........  ..........  MIKDATAITL
     A.salmonicida ..........  ..........  ..........  ..........  .....NSLRL
   P.pseudomallei ..........  ..........  ..........  ..........  MAFQWPRFPL
       P.carinii MSIDKKNDSN  NKKIVLLSAI  GKTYEKKASS  VSDDDIRTIL  SQNILLYGIP
       D.nodosus ..........  ..........  ..........  ..........  ..MMTNIWHT

                51                                                  100
          tobacco ..........  ..........  ..........  ..........  ..........
         tobacco2 QPIKDISGTV  KLPGSKSLSN  RILLLAALSK  GRTVVDNLLS  SDDIHYMLGA
          B.napus QPIREISGLI  KLPGSKSLSN  RILLLAALSE  GTTVVDNLLN  SDDINYMLDA
          S.typhi QPIARVDGAI  NLPGSKSVSN  RALLLAALAC  GKTVLTNLLD  SDDVRHMLNA
      S.gallinarum QPIARVDGAI  NLPGSKSVSN  RALLLAALAC  GKTVLTNLLD  SDDVRHMLNA
      K.pneumoniae QPIARVDGTV  NLPGSKSVSN  RALLLAALAR  GTTVLTNLLD  SDDVRHMLNA
         Y.entero. HPIALINGTV  NLPGSKSVSN  RALLLAALAE  GTTQLNNLLD  SDDIRHMLNA
      H.influenzae APISAVEGTI  NLPGSKSLSN  RALLLAALAK  GTTKVTNLLD  SDDIRHMLNA
       P.multocida NPISYIEGEV  RLPGSKSLSN  RALLLSALAK  GKTTLTNLLD  SDDVRHMLNA
     A.salmonicida EPISRVAGEV  NLPGSKSVSN  RALLLAALAR  GTTRLTNLLD  SDDIRHMLAA
   P.pseudomallei QPWRHVTGHL  RLPGDKSISN  RSLLLGALAE  GVTEVTGLLD  SDDARAMLNA
       P.carinii LNAFQKHTTI  TLPGSKSISN  RALILASLSN  GICYLKNFLH  SDDTYYMLSA
       D.nodosus APVSALSGEI  TICGDKSMSH  RALLLAALAE  GQTEIRGFLA  CADCLATRQA

                101                                                 150
          tobacco ..........  ..........  ..........  ..........  .........L
         tobacco2 LKTLGLHVED  DNENQRAIVE  GCGGQFPVGK  KSEEEIQLFL  GNAGTAMRPL
          B.napus LKKLGLNVER  DSVNNRAVVE  GCGGIFPASL  DSKSDIELYL  GNAGTAMRPL
          S.typhi LSALGINYTL  SADRTRCDIT  GNGGPLRASG  TLE....LFL  GNAGTAMRPL
      S.gallinarum LSALGINYTL  SADRTRCDIT  GNGGPLRAPG  ALE....LFL  GNAGTAMRPL
      K.pneumoniae LSALGVHYVL  SSDRTRCEVT  GTGGPLQAGS  ALE....LFL  GNAGTAMRPL
         Y.entero. LQALGVKYRL  SADRTRCEVD  GLGGKLVAEQ  PLE....LFL  GNAGTAMRPL
      H.influenzae LKALGVRYQL  SDDKTICEIE  GLGGAFNIQD  NLS....LFL  GNAGTAMRPL
       P.multocida LKELGVTYQL  SEDKSVCEIE  GLGRAFEWQS  GLA....LFL  GNAGTAMRPL
     A.salmonicida LTQLGVKYKL  SADKTECTVH  GLGRSFAVSA  PVN....LFL  GNAGTAMRPL
   P.pseudomallei LRDLGV.VIE  GPHQGRCTVH  GVGLHGLKAP  PGP....LFL  GNAGTAMRPL
       P.carinii LEKLNAAEFK  WEQDGDVLVV  KGKSGYLENP  QME....LYL  GNSGTTARFL
       D.nodosus LRALG.VDIQ  REKEIVTIRG  VGFLGLQPPK  APL.....NM  GNSGTSMRLL

                151                                                 200
          tobacco TAAVA....V  AGGNSRYVLD  GVPRMRERPI  GDLVDGLKQL  GAEVDCFLGT
         tobacco2 TAAVT....V  AGGHSRYVLD  GVPRMRERPI  GDLVDGLKQL  GAEVDCFLGT
          B.napus TAAVT....A  AGGNASYVLD  GVPRMRERPI  GDLVVGLKQL  GADVECTLGT
          S.typhi AAALC.....  .LGQNEIVLT  GEPRMKERPI  GHLVDSLRQG  GANIDYLEQE
      S.gallinarum AAALC.....  .LGQNEIVLT  GEPRMKERPI  GHLVDSLRQG  GANIDYLEQE
      K.pneumoniae AAALC.....  .LGSNDIVLT  GEPRMKERPI  GHLVDALRQG  GAQIDYLEQE
         Y.entero. AAALC.....  .LGKNDIVLT  GEPRMKERPI  GHLVDALRQG  GAQIDYLEQE
      H.influenzae TAALC...LK  GNHEVEIILT  GEPRMKERPI  LHLVDALRQA  GADIRYLENE
       P.multocida TAALCLSTPN  REGKNEIVLT  GEPRMKERPI  QHLVDALCQA  GAEIQYLEQE
     A.salmonicida CAALC.....  .LGSGEYMLG  GEPRMEERPI  GHLVDCLALK  GAHIQYLKKD
   P.pseudomallei SAA......L  ALQPFDTTLT  GDPRMSERPI  NRLVDALREM  GAVIEYLAQE
       P.carinii TSICTLVQPN  SRENHLILTG  SNRMKQRPIG  PLVDALKNNG  CCIEYLELEN
       D.nodosus AGILAAQRFE  SVLCGDESLE  KRPMQRIITP  LVQMGAKIVS  HSNFTAPLHI
```

```
                201                                                 250
      tobacco KCPPVRIVSK GGLPGGKVKL SGSISSQYLT ALLMAAPLAL GDVEIEIIDK
     tobacco2 NCPPVRIVSK GGLPGGKVKL SGSISSQYLT ALLMAAPLAL GDVEIEIIDK
      B.napus NCPPVRVNAN GGLPGGKVKL SGSISSQYLT ALLMAAPLAL GDVEIEIIDK
      S.typhi NYPPLR.LRG .GFIGGDIEV DGSVSSQFLT ALLMTAPLAP ED.TIIRVKG
  S.gallinarum NYPPLR.LRG .GFIGGDIEV DGSVSSQFLT ALLMTAPLAP KD.TIIRVKG
  K.pneumoniae NYPPLR.LRG .GFTGGDVEV DGSVSSQFLT ALLMASPLAP QD.TVIAIKG
     Y.entero. NY.RRC.IAG .GFRGGKLTV DGSVSSQFLT ALLMTAPLAE QD.TEIQIQG
  H.influenzae GYPPLA.IRN KGIKGGKVKI DGSISSQFLT ALLMSAPLAE ND.TEIEIIG
   P.multocida GYPPIA.IRN TGLKGGRIQI DGSVSSQFLT ALLMAAPMAE AD.TEIEIIG
 A.salmonicida GYPPLV.VDA KGLWGGDVHV DGSVSSQFLT AFLMAAPAMA PVIPRIHIKG
P.pseudomallei GYPPLTIRGG GSVSSQFLTA LLM...... .........T APMASAQIKSG
      P.carinii CLPLLIKPKE IGLYGGNINL SATVSSQYVS SILMCSPYAK TQVTLSLIGG
      D.nodosus SGRPLTGIDY ALPLPSAQLK SCLILAGLLA DGTTRLHTCG ISRDHTERML

                251                                                 300
      tobacco LISVLYVEMT LKLMERFGIS VEHSSSWDRF VVRGGQKYKS PGK...AYVE
     tobacco2 LISVPYVEMT LKLMERFGVS VEHTSSWDKF LVRGGQKYKS PGK...AYVE
      B.napus LISVPYVEMT LKLMERFGVS AEHSDSWDRF FVKGGQKYKS PGN...AYVE
      S.typhi ELVSKPYIDI TLNLMKTFGV EIANHHYQQF VVKGGQQYHS PGR...YLVE
  S.gallinarum ELVSKPYIDI TLNLMKTFGV EIANHHYQQF VVKGGQQYHS PGR...YLVE
  K.pneumoniae ELVSRPYIDI TLHLMKTFGV EVENQAYQRF IVRGNQQYQS PGD...YLVE
     Y.entero. ELVSKPYIDI TLHLMKAFGV DVVHENYQIF HIKGGQTYRS PGI...YLVE
  H.influenzae ELVSKPYIDI TLAMMRDFGV KVENHHYQKF QVKGNQSYIS PNK...YLVE
   P.multocida ELVSKPYIDI TLKMMQTFGV EVENQAYQRF LVKGHQQYQS PHR...FLVE
 A.salmonicida ELVSKPYIDI TLHIMNSSGV VIEHDNYKLF YIKGNQSIVS PGD...FLVE
P.pseudomallei LLLSKPYIDI TLNVM.PFGV PTR.DHTERI FAVSAIRYPS PAV...LRLE
      P.carinii KPISQPYIDM TISMMSSFGI KVTRSHSKEN TYYIPKGCYT CPSEYIIEGD
      D.nodosus PLFGGALEIK KEQIIVTGGQ KLHGCVLDIV .......... ..........

                301                                                 350
      tobacco GDASSASYFL AGAAVTGGTV TVEGCGTSSL QGDVKFAEVL EQMGAEVTWT
     tobacco2 GDASSASYFL AGAAVTGGTV TVEGCGTSSL QGDVKFAEVL EKMGAEVTWT
      B.napus GDASSASYFL AGAAITGETV TVEGCGTTSL QGDVKFAEVL EKMGCKVSWT
      S.typhi GDASSASYFL AAGGIKGGTV KVTGIGGKSM QGDIRFADVL HKMGATITWG
  S.gallinarum GDASSASYFL AAGAIKGGTV KVTGIGRKSM QGDIRFADVL EKMGATITWG
  K.pneumoniae GDASSASYFL AAGAIKGGTV KVTGIGRNSV QGDIRFADVL EKMGATVTWG
     Y.entero. GDASSASYFL AAAAIKGGTV RVTGIGKQSV QGDTKFADVL EKMGAKISWG
  H.influenzae GDASSASYFL AAGAIKG.KV KVTGIGKNSI QGDRLFADVL EKMGAKITWG
   P.multocida GDASSASYFL AAAAIKG.KV KVTGVGKNSI QGDRLFADVL EKMGAHITWG
 A.salmonicida GDASSASYFL AAGAIKG.KV RVTGIGKHSI GDIHFADVLE RMGARITWGD
P.pseudomallei GDATSASYFL AAAGIKG..V PVTGIGRHSM QGDSWFPRAL RRMGAR.SCG
      P.carinii ATSATYPLAI AAITGGSCTI SNVGSASLQG DSKFSEYILK PMGCEVVQSP
      D.nodosus GDLSAAAFFM VAALIAPRAE VVIRNVGINP TRAAIITLLQ KMGGRIELHH

                351                                                 400
      tobacco ENSVTVKGPP RNSSAMKHLR AIDVNMNKMP ...DVAMTLA VVALFADGPT
     tobacco2 ENSVTVKGPP RNSSGMKHLR AVDVNMNKMP ...DVAMTLA VVALFADGPT
      B.napus ENSVTVTGPS RDAFGMRHLR AVDVNMNKMP ...DVAMTLA VVALFADGPT
      S.typhi DDFIACT... .....RGELH AIDMDMNHIP ...DAAMTIA TTALFAKGTT
  S.gallinarum DDFIACT... .....RGELH AIDMDMNHIP ...DAAMTIA TTALFAKGTT
  K.pneumoniae EDYIACT... .....RGELN AIDMDMNHIP ...DAAMTIA TAALFARGTT
     Y.entero. DDYIECS... .....RGELQ GIDMDMNHIP ...DAAMTIA TTALFADGPT
  H.influenzae EDFIQAE... .....HAELN GIDMDMNHIP ...DAAMTIA TTALFSNGET
   P.multocida DDFIQVE... .....KGNLK GIDMDMNHIP ...DAAMTIA TTALFAEGET
 A.salmonicida D FIEAEQG. .......PLH GVDMDMNHIP ...DVGHDHS GQSHCLPRVP
P.pseudomallei SSMIVCPRGE .......LRA AVRSDSNSIP ...DAAMTLA TRS...AGAR
      P.carinii TTTYIKGPPK GKLKSLGSIN MESMTDTFLT ...AAVLASV AYEESKPYVT
      D.nodosus QRFWGAEPVA DIVVYHSKLR GITVAPEWIA NAIDELPIFF IAAACAEGTT
```

```
                401                                                  450
      tobacco AIRDVASWRV KETEFMIAIC TELRKLGATV EEGPDYCI.. ITPPEKLNVT
     tobacco2 AIRDVASWRV KETEFMIAIC TELRKLGATV VEGSDYCI.. ITPPEKLNVT
      B.napus TIRDVASWRV KETEFMIAIC TELRKLGATV EEGSDYCV.. ITPPAKVKPA
      S.typhi TLRNIYNWRV KETDFLFAMA TELRKVGAEV EEGHDYIR.. ITPPAKLQHA
  S.gallinarum TLRNIYNWRV KETDFLFAMA TELRKVGAEV EEGHDYIR.. ITPPAKLQHA
  K.pneumoniae TLRNIYNWRV KETDFLFAMA TELRKVGAEV EEGEDYIR.. ITPPLTLQFA
     Y.entero. VIRNIYNWRV KETDFLSAMA TELRKVGAEV EEGQDYIR.. VVPPAQLIAA
  H.influenzae VIRNIYNWRV KETDFLTAMA TELRKVGAEV EEGEDFIRIQ PLALNQFKHA
   P.multocida VIRNIYNWRV KETDFLTAMA TELRKVGAEV EEGEDFIRIQ PLNLAQFQHA
 A.salmonicida PHSQHLQLAV RDDRCTPCTH GHRRAQAGVS EEGTTFITRD AADPAQ..AR
P.pseudomallei WAATANHIRV AGEGDGSAVC NVHGAGGGWR ASGSRCWSSW LPSARKVVLR
      P.carinii KITGISNQRI KECNRINAMV CELKKFGIEA GELPDGIYVK ALNTSNLPYS
      D.nodosus FVGNLSELRV KESDRLAAMA QNLQTLGVAC DVGADFIHIY GRSDRQFLPA

                451                                                  500
      tobacco EIDT..YDDH RMAMAFSLAA C.ADVPVTIN DPGCTRKTFP NYFDVLQQYS
     tobacco2 EIDT..YDDH RMAMAFSLAA C.ADVPVTIK DPGCTRKTFP NYFDVLQQYS
      B.napus EIDT..YDDH RMAMAFSLAA C.ADVPVTIK DPGCTRKTFP DYFQVLESIT
      S.typhi DIGT..YNDH RMAMCFSLVA L.SDTPVTIL DPKCTAKTFP DYFEQLARMS
  S.gallinarum DIGT..YNDH RMAMCFSLVA L.SDTPVTIL DPKCTAKTFP DYFEQLARMS
  K.pneumoniae EIGT..YNDH RMAMCFSLVA L.SDTPVTIL DPKCTAKTFP DYFGQLARIS
     Y.entero. EIGT..YNDH RMAMCFSLVA L.SDTPVTIL DPKCTAKTFP DYFEQLARLS
  H.influenzae NIET..YNDH RMAMCFSLIA L.SNTPVTIL DPKCTAKTFP TFFNEFEKIC
   P.multocida ELNI...HDH RMAMCFALIA L.SKTSVTIL DPSCTAKTFP TFLILFTLNT
 A.salmonicida RDRH..LQRS RIAMCFSLVA L.SDIAVTIN DPGCTSKTFP DYFDKLASVS
P.pseudomallei CAVPKRFPDG NVLLAASAWR TGCETSWIPA APTRRRIVIE GGAIGS....
      P.carinii VEGINCYNDH RIAMSFSVLA CISSKPTTIL DKACVNKTWP YWWDILNSTF
      D.nodosus RVNS..FGDH RIAMSLAVAG V.RAAGELLI DDGAVAAVSM PQFRDFAAAI

                501         514
      tobacco KH........ ......
     tobacco2 KH........ ......
      B.napus KH........ ......
      S.typhi TPA....... ......
  S.gallinarum TPA....... ......
  K.pneumoniae TLA....... ......
     Y.entero. QA........ ......
  H.influenzae LKN....... ......
   P.multocida REVA...... ......
 A.salmonicida QAV....... ......
P.pseudomallei .......... ......
      P.carinii K
      D.nodosus GMNVGEKDAK NCHD..
```

Figure 5.19 Line-up of the remaining deduced amino acid sequences of the EPSP synthase enzymes from plants, eukaryotes and Gram-negative bacteria

These sequences were lined up, using the PILE-UP programme, to examine the similarity between them. Unexpectedly, the deduced EPSP synthase sequences from the three species of Gram-positive bacteria show similarity to each other, but not to the Gram-negative bacteria, plant or lower eukaryote enzymes (**Figure 5.18**). The protein sequence of the enzyme from the cyanobacterium (*Synechocystis*) also shows similarity to the enzymes from the three Gram-positive bacteria. However, the sequence of the *M.tuberculosis* enzyme only exhibits limited similarity to any of the other three enzymes in **Figure 5.18**, and shows even lower similarity to any of the other protein sequences available.

The three plant, one eukaryotic and nine Gram-negative bacterial EPSP synthase enzymes (i.e. those not shown in **Figure 5.1**), with the possible exception of the enzyme from *D.nodosus* , show significant similarity to each other, and to the nine sequences previously examined (**Figure 5.19**). As in **Figure 5.1**, the region against which primer 1 was designed (amino acid sequence NAGTAMRP; positions **142-148** in **Figure 5.19**) is conserved very highly, with complete conservation in ten of the twelve sequences containing this region. The region against which primer 2 was designed is also conserved fairly well (sequence CTAKTFPD; amino acids **484-491** in **Figure 5.19**), with complete conservation in four of these sequences. However, no conservation was apparent with the sequence from the enzyme of *D.nodosus*. The region against which primer 4 was designed (WRVKET(D/E)R; residues **408-415** in **Figure 5.19**) is completely conserved in nine of these thirteen sequences, poorly conserved in two, and not apparent in the other two sequences. The high levels of similarity in these three regions amongst most of the thirteen sequences in **Figure 5.19** lends support to the choice of these regions for the design of primers 1, 2 and 4.

The EPSP sequences from the Gram-positive bacteria and the cyanobacterium only show limited similarity to the regions against which primer 1 (residues **111-117** in **Figure 5.18**) and primer 4 (residues **370-376**) were designed, and no apparent similarity to the region against which primer 2 was designed. As *S.coelicolor* is a Gram-positive bacterium, it was anticipated that the protein sequence of its EPSP synthase enzyme is most likely to show similarity to the sequences from the other Gram-positive bacteria.

5.2.11 Design of oligonucleotides primer 57 and primer 60

Since the protein sequences of the EPSP synthases from the Gram-positive bacteria show only poor similarity to the other EPSP synthase sequences available, it was assumed that the sequence of the enzyme from *S.coelicolor* is also likely to show only

poor similarity to the enzymes from the Gram-negative bacteria, plants and eukaryotes. Therefore, it was considered that primers designed against regions of consensus from the Gram-positive bacteria and the cyanobacterium would be more likely to be successful in the cloning of the *aroA* gene of *S.coelicolor*.

Regions of consensus were identified in the sequences of the enzymes from the Gram-positive bacteria and the cyanobacterium, and primers were designed as before to allow a region of the *aroA* gene to be duplicated by PCR. Two regions of consensus were identified: GNSGTT(I/M/T)R (residues **110-117**) and DHRIGMML (residues **417-424** in **Figure 5.18**).

Primer 60 was designed against the sequence GNSGTT(I/M/T)R, and acts as the N-terminal oligonucleotide (**Figure 5.20**). It was designed against the deduced DNA sequence of the coding strand, and was intended to anneal to the non-coding strand of *aroA* and to prime DNA synthesis towards the 3' end of the gene. Although this region is only poorly conserved in the enzyme from *M.tuberculosis*, seven out of the eight residues are completely conserved in the other three sequences in **Figure 5.18**. The N-terminal glycine residue in this region is the Gly-96 residue which has been identified as being involved in the binding of glyphosate to EPSP synthase [Sost & Amrhein, 1990; Padgette *et al.*, 1991]. However, the proline residue (Pro-101) which has also been implicated in binding of glyphosate [Stalker *et al.*, 1985] is not contained in any of the four amino acid sequences in **Figure 5.18**.

This oligonucleotide contains inosine nucleotides, which may base-pair with A, C, G or T nucleotides. Therefore, the use of inosines in a primer increases the likelihood of the primer binding to denatured (single-stranded) DNA. However, inosines also reduce the specificity of an oligonucleotide, due to their non-specific base-pairing.

The region of consensus against which primer 60 was designed contains a serine residue which may be encoded by any one of six nucleotide triplets. In the *Streptomyces* genes which have been analysed to date, three of these six codons each account for over 25% of the serine codons used (AGC: 28%; TCG: 26%; TCC: 39%). However, these three triplets show no conservation of nucleotides at any position. Therefore, it is extremely difficult to design a an oligonucleotide which will anneal only to a deduced DNA sequence encoding a serine for use in *Streptomyces*, as each position of this triplet requires the incorporation of either a degeneracy or an inosine. In the design of this oligonucleotide, it was decided to incorporate three inosine nucleotides for the triplet encoding the serine residue.

Primer 60:

Predicted protein sequence :

N	^{110}-Gly	Asn	Ser	Gly	Thr	Thr	Ile/Met/Thr	Arg-^{117}	**C**

Deduced DNA sequence :

5'	-GGn	AAC	nnn	GGn	ACn	ACn	Ann	CGn-	**3'**

Primer :

5'	-GGI	AAC	III	GGI	AC(GC)	AC(GC)	AII	CG-	**3'**
	100%	96%	100%	100%	95%	95%	100%	93%	

Primer 57:

Predicted protein sequence:

N	^{417}-Asp	His	Arg	Ile	Gly	Met	Met	Leu-^{424}	**C**

Deduced DNA sequence (non-coding strand) :

5'	-GAC	CAC	CGn	ATC	GGn	ATG	ATG	CTn-	**3'**

Deduced DNA sequence (coding strand) :

3'	-CTG	GTG	GCn	TAG	CCn	TAC	TAC	GAn-	**5'**

Reverse of deduced DNA sequence (coding strand) :

5'	-nAG	CAT	CAT	nCC	GAT	nCG	GTG	GTC-	**3'**

Primer:

5'	-AG	CAT	CAT	ICC	GAT	ICG	GTG	GTC-	**3'**
	100%	100%	100%	100%	92%	93%	94%	95%	

Figure 5.20 Design of primer 57 and primer 60 for cloning of *aroA*

N.B. a) An I in the DNA denotes an inosine nucleotide, which may pair with all four of the other nucleotides, and therefore increases the likelihood of the primer annealing to DNA

b) Numbers in small block point represent the positions of amino acids in **Figure 5.18**

c) n denotes a position which could contain more than one of the four nucleotides

The region of consensus against which primer 60 was designed also includes a non-conserved residue (residue **116** in **Figure 5.18**). At this position, either an isoleucine (Ile), methionine (Met) or threonine (Thr) occurred in the three sequences which exhibit significant conservation. In the *Streptomyces* genes analysed to date, isoleucine is typically (92%) encoded by ATC; methionine is only encoded by ATG; threonine is typically encoded by ACG (30%) or ACC (65%). Only the nucleotide occurring at the first position is conserved amongst these four codons; either inosines or degeneracies ((AT) in the second position, (GC) in the third) must be included at the other positions. In this oligonucleotide, it was decided to utilise inosines to solve this problem.

In order to increase the possibility of this oligonucleotide annealing to the anticipated sequence in the *Streptomyces* genome, inosines were also included at the third position of the two triplets encoding glycine residues. Although it is possible to include a (GC) degeneracy in this position, this would only account for 83% of the triplets which encode glycine, whereas the use of an inosine at this position would account for all of the triplets which encode glycine.

Melting temperatures for oligonucleotides containing inosine residues were calculated by ignoring any inosines. Therefore, primer 60 had an estimated T_m of 56·8°C, and was approximately 69% G+C rich.

Primer 57 was designed against the sequence DHRIGMML, and acts as the C-terminal oligonucleotide (**Figure 5.20**). It was designed against the deduced DNA sequence of the non-coding strand, and was intended to anneal to the non-coding strand of *aroA*, and to prime DNA synthesis towards the 5' of the gene (and therefore towards primer 60). The region against which it was designed is well conserved amongst three of the sequences, and is completely conserved in two of these sequences. However, it is only poorly conserved in the sequence from *M.tuberculosis* (4 out of 8 residues).

As with primer 60, an inosine residue was included in this oligonucleotide at the third position of the triplet encoding the glycine residue in the consensus against which this oligonucleotide was designed. An inosine was also used at the third position of the triplet encoding the arginine contained in this consensus sequence. Use of a (G/C) degeneracy at this position would account for only an estimated 82% of the triplets which encode arginine, whereas use of an inosine accounts for approximately 93% of these codons. It also contains the Gly-418 residue which has been implicated in the binding of glyphosate to EPSP synthase [Huynh *et al.*, 1988]. This oligonucleotide had an estimated T_m of 59·9°C, and was 57% G+C rich.

5.2.12 Attempted PCR-cloning of *aroA* with primer 57 and primer 60

The PCRs described in this section were undertaken using primer 57 and primer 60 in an attempt to replicate an internal fragment of the *aroA* gene, using gDNA isolated from *S.coelicolor* G216, *S.lividans* TK24 or *S.rimosus* 4018 mycelia as the template. PCRs were performed according to section **2.3.24**, and the generation of a product of 850 to 950 bp in length was anticipated.

Initially, the PCRs were performed at an annealing temperature of 48°C, with a Mg^{2+} concentration of 1 to 2 mM. Although no obvious products were identified when the reaction using *S.rimosus* gDNA as template was run out on an agarose gel, products of approximately the anticipated size were identified from *S.coelicolor* gDNA and *S.lividans* gDNA (at 1·5 and 2 mM Mg^{2+}). These products were still identifiable when the annealing temperature was raised to 52°C, but were also identifiable from reactions in which primer 57 only was used. Therefore, these products were being produced by a single primer alone, and so were not further investigated.

5.2.13 Design of oligonucleotides: primer A, primer B and primer C

These three oligonucleotide primers were also designed against regions of consensus in the amino acid sequences of the EPSP synthase enzymes from the Gram-positive bacteria and the cyanobacterium. They were intended for use not only as PCR primers, but also for use in genomic cloning of *aroA*, and were therefore longer than the oligonucleotides described above.

Primer A was designed against the deduced DNA sequence of the coding strand, to anneal to the non-coding strand, and to prime synthesis towards the 5' end of the gene in PCR. It was designed against the amino acid consensus sequence DHRI(G/A)MMLAIA (residues **416-427** in **Figure 5.18**), which is well conserved amongst the four sequences available. The glycine residue was previously identified as the Gly-418 residue which is implicated in the binding of glyphosate to EPSP synthase [Huynh *et al.*, 1988].

The region against which this oligonucleotide was designed contains two positions at which the amino acids are the same in the sequences from *B.subtilis* and *S.aureus* (Gly and Met at positions **421** and **423** respectively), but are different to the two amino acids in the sequences from *Synechocystis* and *M.tuberculosis* (Ala at both positions). At position **421**, the triplets encoding both amino acids have a G at the first position, so a G was used in this position in the oligonucleotide. All of the

triplets encoding glycine have a G at the second position, which is replaced in all of the triplets encoding alanine by a C. Therefore, a (GC) degeneracy was introduced at the second position of primer A (**Figure 5.21**). A (GC) degeneracy was also introduced at the third position of this codon, which represents an estimated 83% of the triplets encoding glycine, and approximately 92% of those encoding alanine. Use of an inosine at this position would have represented all of the triplets encoding both amino acids and may, in retrospect, have been the preferable option in the design of this oligonucleotide. At position **423**, only the methionine codon was represented, as the codons for these two amino acids have different nucleotides in the first two positions, so that three degeneracies or inosines would be required to represent the codons for both amino acids, which would significantly reduce the specificity of the oligonucleotide. Methionine was chosen as the amino acid represented at this position of primer A as it occurred in the sequences from *B.subtilis* and *S.aureus*, which are the two Gram-positive bacteria whose amino acid sequences show the greatest similarity throughout the entire coding sequence. Primer A had an estimated T_m of 75°C, and was approximately 66·7% G+C rich.

Primer B and primer C were designed as nested primers, against the coding strand, to initiate DNA synthesis towards the 3' end of the gene during PCR.

Primer B was designed against the consensus amino acid sequence IKDA(A/E)ELKVKE (residues **363-373** in **Figure 5.18**), which is very well conserved in three of the sequences available, and completely conserved in two of them. However, this sequence is not apparent in the enzyme sequence from *M.tuberculosis*.

At the fifth residue of this consensus sequence, three of the sequences have an alanine residue, whilst the other sequence, that of the enzyme from *S.aureus*, has a glutamine residue. It was decided to incorporate both of these possible residues into the oligonucleotide. All four of the triplets encoding alanine have the nucleotides G and C in the first two positions respectively, with the two glutamate codons containing G and A at these positions. Therefore, an (AC) degeneracy was used in the second position of this triplet (**Figure 5.21**). A (GC) degeneracy in the third position of these codons represents an estimated 93% of the alanine codons and 81% of the glutamate codons in *Streptomyces* genes, and was therefore incorporated into the oligonucleotide at this position. However, use of an inosine nucleotide at this position would represent all six of these possible codons, and may have been a preferable option. Primer B had an estimated T_m of between 69·3 and 70·8°C (depending on whether an A or C is used at the 14th residue of this oligonucleotide) and was between 56·2 and 59·4% G+C rich.

Primer A:

Predicted protein sequence :

N ^{417}-Asp His Arg Ile Gly/Ala Met Met Leu Ala Ile Ala-^{427} **C**

Deduced DNA sequence (non-coding strand) :

5' -GAC CAC CGn ATC G(GC)n ATG ATG CTn GCn ATC GCn- **3'**

Deduced DNA sequence (coding strand) :

3' -CTG GTG GCn TAG C(GC)n TAC TAC GAn CGn TAG CGn- **5'**

Deduced DNA sequence (coding strand) :

5' -nGC GAT nGC nAG CAT CAT n(GC)C GAT nCG GTG GTC- **3'**

Primer sequence:

5' -(GC)GC GAT (GC)GC(GC)AG CAT CAT (GC)(GC)C GAT(GC)CG GTG GTC- **3'**

92% 92% 92% 94% 100% 100% 88% 92% 82% 94% 96%

Primer B:

Predicted protein sequence :

N -Ile Lys Asp Ala Ala/Glu Glu Leu Lys Val Lys Glu- **C**

Deduced DNA sequence (non-coding strand) :

5' -ATC AAG GAC GCn G(AC)n GAG CTn AAG GTn AAG GAG- **3'**

Primer sequence:

5' -ATC AAG GAC GC(GC) G(AC)(GC) GAG CT(GC) AAG GT(GC) AAG GA- **3'**

92% 95% 96% 92% 87% 81% 94% 95% 94% 95% 100%

Primer C:

Predicted protein sequence :

N -Val Pro Gly Asp Ile Ser Ser Ala Ala Phe Phe-**C**

Deduced DNA sequence (non-coding strand) :

5' -GTn CCn GGn GAC ATC nnn nnn GCn GCn TTC TTC-**3'**

Primer sequence:

5' -GT(GC) CC(GC)GG(GC) GACATC I(GC)II(GC)(GC) GC(GC)GC(GC) TTC TTC-**3'**

95% 94% 83% 96% 92% 100% 93% 92% 92% 99% 99%

Figure 5.21 Design of primer A, primer B and primer C for cloning of *aroA*

Primer C was designed against the consensus amino acid sequence VPGDISSAAFF (residues **258-268** in **Figure 5.18**), which is also extremely well conserved in three of the sequences available, and is completely conserved in two of them. However, like the sequence against which primer B was designed, this sequence is apparently not conserved in the sequence of the *M.tuberculosis* enzyme. The region against which primer C was designed contains two serine residues, with the codons representing them both containing an inosine at the first position and a (GC) degeneracy at the second position. However, it was decided to use an inosine as the third nucleotide of the triplet encoding the first of the two serines, and a (GC) degeneracy in the third position of the triplet encoding the second serine residue. Primer C had an estimated Tm of 76·6°C, and was approximately 73% G+C rich.

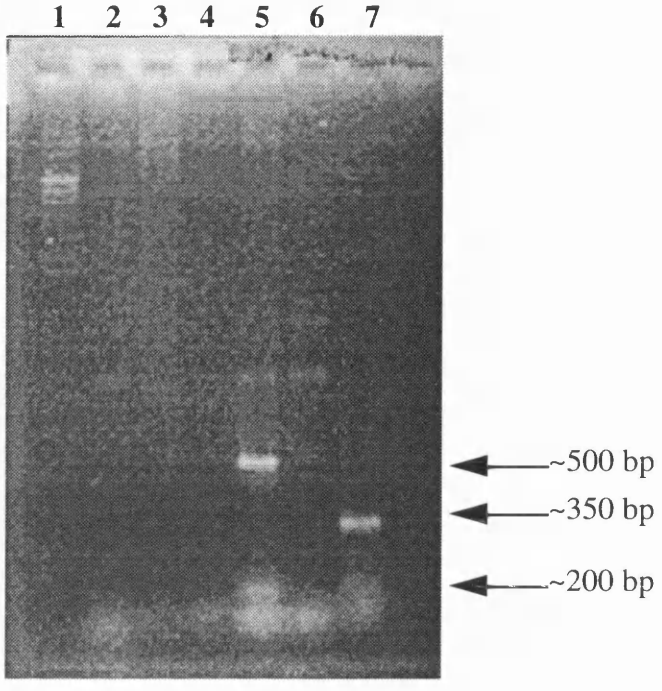
5.2.14 Attempted PCR-cloning of *aroA* with primer A, primer B and primer C

Primer A, primer B and primer C were used in an attempt to amplify an internal fragment of the *aroA* gene from *S.coelicolor*. The template used in these PCRs was *S.coelicolor* G216 gDNA, and the reactions were performed according to section **2.3.24**. It was anticipated that PCR using primer A and primer C would give a product of approximately 500 to 550 bp in length, whilst PCR using primer A and primer B would give a product of 180 to 200 bp.

PCRs were initially performed at an annealing temperature of 56°C and at 1·5 mM Mg^{2+}, with single-primer PCRs included as controls (**Figure 5.22**). Although primer A and primer B each created a product in single-primer controls, these products were significantly in excess of 500 bp, and so were discounted. The PCR using both primer A and primer B generated two products of estimated sizes of 500 and 200 bp. However, the PCR with primer B and primer C, which was intended merely as a negative control, also created a product, approximately 350 bp in length. This was an unexpected result, as primer B and primer C were both designed to anneal to the same strand. Since neither primer alone gave a product of this size, it appears that one of these two primers is annealing to the opposite strand of the genomic DNA. PCR with primer A and primer C did not generate any suitable products.

These PCRs were re-performed at $MgCl_2$ concentrations ranging from 0·5 to 3·5 mM Mg^{2+} (with increments of 0·5 mM Mg^{2+}) (**Figure 5.23**). Although no obvious products were generated with any of the primers at 0·5 or 1 mM Mg^{2+}, primer B and primer C together generated the previously-identified 350 bp product at 1·5 to 3·5 mM Mg^{2+}. Primer A and primer B together generated the 500 and 200 bp products at 1·5 to 2·5 mM Mg^{2+}, with the 500 bp product being the major product, and generated

a) 10 µl of each 50 µl sample



b) 20 µl of each 50 µl sample

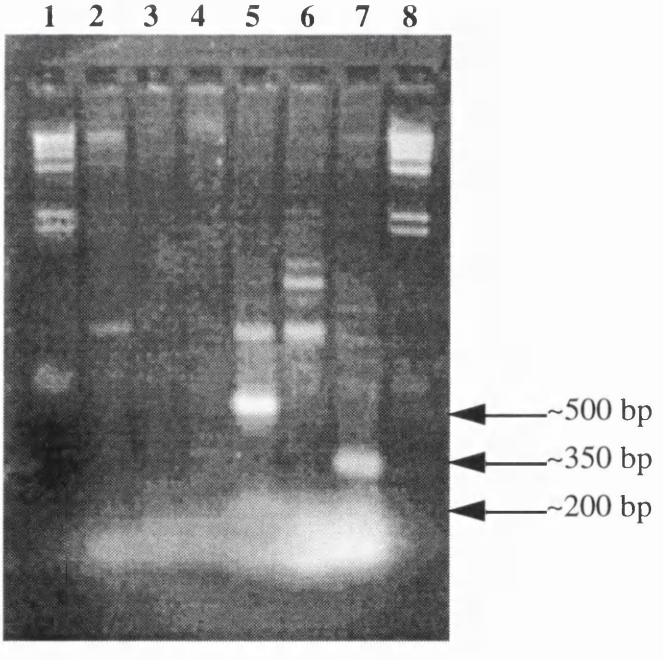


Figure 5.22 PCRs on gDNA from G216 at 56°C and 1·5 mM Mg^{2+}

lanes 1 and 8: λ/*Hin*d III markers [lane 1 only in **a**)]

lane 2: primer A alone	lane 5: primer A and primer B
lane 3: primer B alone	lane 6: primer A and primer C
lane 4: primer C alone	lane 7: primer B and primer C

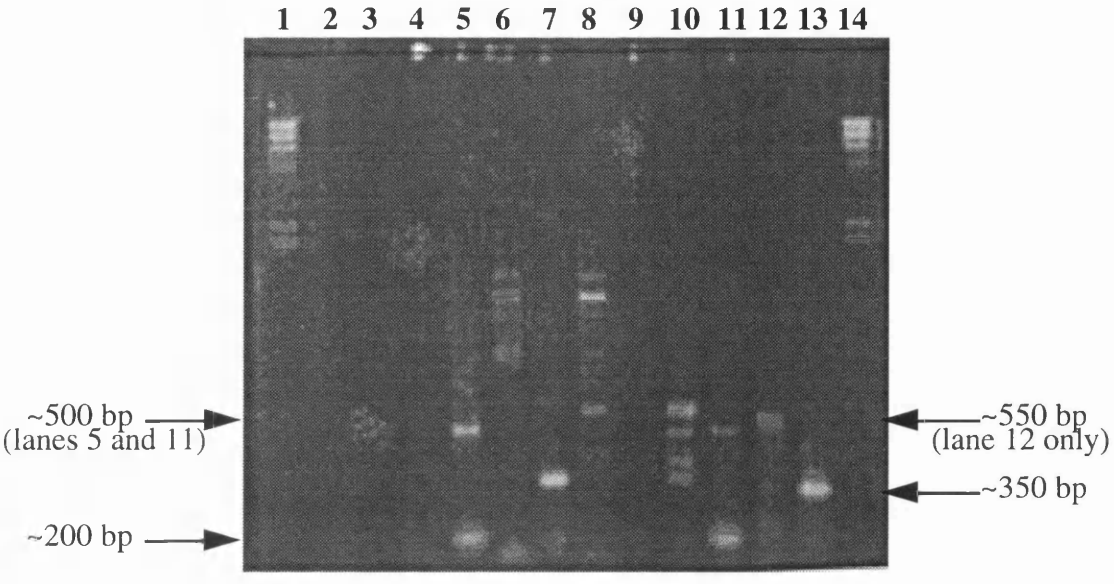


Figure 5.23 PCRs on gDNA from G216 at 56°C and 2-2·5 mM Mg^{2+}

lanes 1 and 14: λ/*Hin*d III markers

lanes 2-7: 2 mM Mg^{2+}	lanes 8-13: 2·5 mM Mg^{2+}
lanes 2 and 8: primer A alone	lanes 5 and 11: primer A and primer B
lanes 3 and 9: primer B alone	lanes 6 and 12: primer A and primer C
lanes 4 and 10: primer C alone	lanes 7 and 13: primer B and primer C

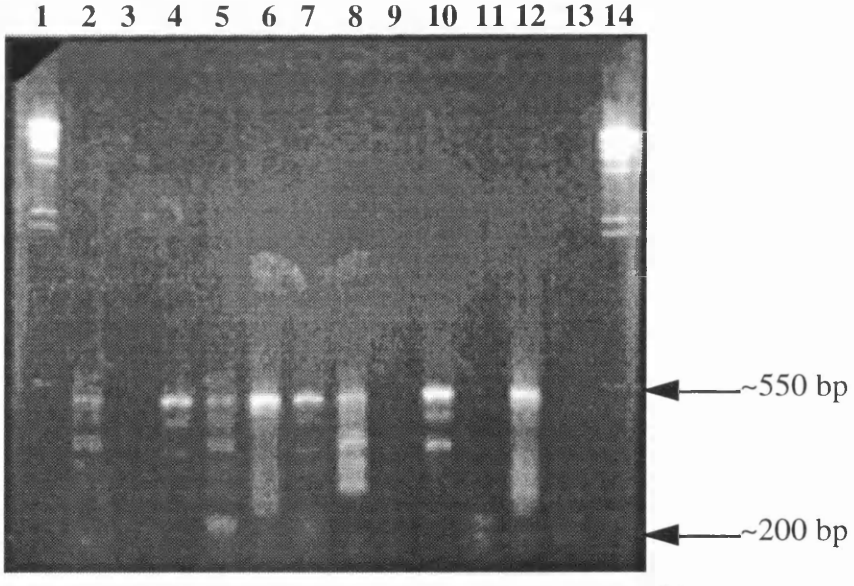


Figure 5.24 PCRs on 550 bp product at 56°C and 2·5 mM Mg^{2+}

lanes 1 and 14: λ/*Hin*d III markers

lanes 2-7: 10 pg 550 bp template	lanes 8-13: 100 pg 550 bp template
lanes 2 and 8: primer A alone	lanes 5 and 11: primer A and primer B
lanes 3 and 9: primer B alone	lanes 6 and 12: primer A and primer C
lanes 4 and 10: primer C alone	lanes 7 and 13: primer B and primer C

the 200 bp product alone at 3 and 3·5 mM Mg^{2+}. Primer A and primer C generated a single product with an estimated size of 550 bp at 2·5 to 3·5 mM Mg^{2+} which was not synthesised in any of the single primer reactions.

The 550 bp product generated by PCR with primer A and primer C, using gDNA from G216 as the template, was purified from a TAE gel, and used as the template for a further series of PCRs at 2·5 mM Mg^{2+} (**Figure 5.24**). Again, all three primers were used, either singly or in pairwise combination, in these reactions. It was anticipated that PCR using primer A and primer C in combination would regenerate the 550 bp product, and primer A and primer B together would cause the synthesis of the 200 bp product.

The concentrations of template used were approximately 1 ng, 100 pg or 10 pg per reaction, and the annealing temperature was maintained at 56°C. At all three template concentrations, primer A and primer C in isolation regenerated the 550 bp product, which was therefore visible in all pairwise reactions. However, primer A and primer B also generated the 200 bp product at a higher level than the 550 bp product with all three concentrations of template. The same results were obtained at an annealing temperature of 57°C.

In an attempt to sequence the two PCR products generated by primer A and primer B (200 bp) and primer A and primer C (550 bp), they were both purified from a TAE gel and ligated into pIBI25, which had previously been digested with *Hin*c II (which produces blunt-ended ends) and subsequently T-tailed. The ligation mixture was transformed into competent *E.coli* DS941 cells, and the transformation mixture plated onto L agar supplemented with ampicillin. However, none of the resultant transformants contained inserts.

Cloning of the two purified PCR products was then attempted via ligation into the commercially-available pT7Blue T-vector (linearised by digestion with the *Eco*R V restriction enzyme, which also produces blunt ends, and then T-tailed). The ligation mix was transformed into competent *E.coli* DS941 cells, and the transformation mix plated onto L agar supplemented with ampicillin, X-gal and IPTG. However, the resultant white colonies again failed to contain any inserts.

Direct sequencing of the two PCR products was then attempted. Two PCR sequencing methodologies were utilised: cycle sequencing, and asymmetric PCR. In cycle sequencing, PCRs containing only one of the primers, one of the four ddNTPs

and ^{32}P-dATP are performed. Therefore, the products are radiolabelled, and DNA polymerisation is likely to be terminated by incorporation of a dideoxynucleotide into the newly-synthesised strand. As only one of the two PCR primers is used in the reaction, only one of the two strands is used as a template for polymerisation, so the reaction only yields sequence of its complementary strand.

Cycle sequencing, using the SequiTherm Cycle Sequencing Kit, was undertaken with primer A and primer B to generate sequence from both strands of each PCR template. The reactions underwent 30 cycles, each consisting of a denaturation step (95°C for 30 secs), an annealing step (50°C for 30 secs) and an extension step (70°C for 1 min). Stop solution (containing electrophoresis dyes) was added after the reactions had ended, the products were denatured at 70°C, and the PCRs products separated by PAGE. However, no sequence could be read from the PCRs.

In asymmetric PCR, reactions are performed under normal conditions, but the concentration of one of the primers is limiting in the PCR (1 pmol, compared with 50 pmol of the other primer). The reaction therefore synthesises an excess of one strand, which may then be sequenced according to the single-stranded sequencing protocol.

Asymmetric PCR was performed using both of the PCR products which had previously been identified (550 bp and 200 bp) as templates. The primer which was limiting in the asymmetric PCR was used as the sequencing primer, which had previously been radiolabelled with ^{32}P-dATP. Again, no readable sequence was obtained from either template with either sequencing primer (primer A and primer C for the 550 bp product, and primer A and primer B for the 200 bp product).

The 550 bp and 200 bp PCR products were again purified from a 1·6% TAE gel, and ligated (to give p550 and p200 respectively) into pDK101 which had previously been digested with *Xcm* I and gel-purified. The ligation mixtures were transformed into TG-1 cells, and the mixture plated onto L agar containing ampicillin, IPTG and X-gal. A single putative p200-containing colony and eleven putative p550-containing transformants were selected for further investigation.

These colonies were streaked to single colonies, one (per transformant) of which was resuspended in 100 µl of Doly I solution, and plasmid DNA isolated by the standard Birnboim-Doly procedure. The resultant pellets were resuspended in dH_2O and used as templates for PCR reactions using primers A and B for the putative p200 plasmid DNA, and primers A and C for the putative p550 plasmid. 2·5 mM $MgCl_2$ was used in all PCRs.

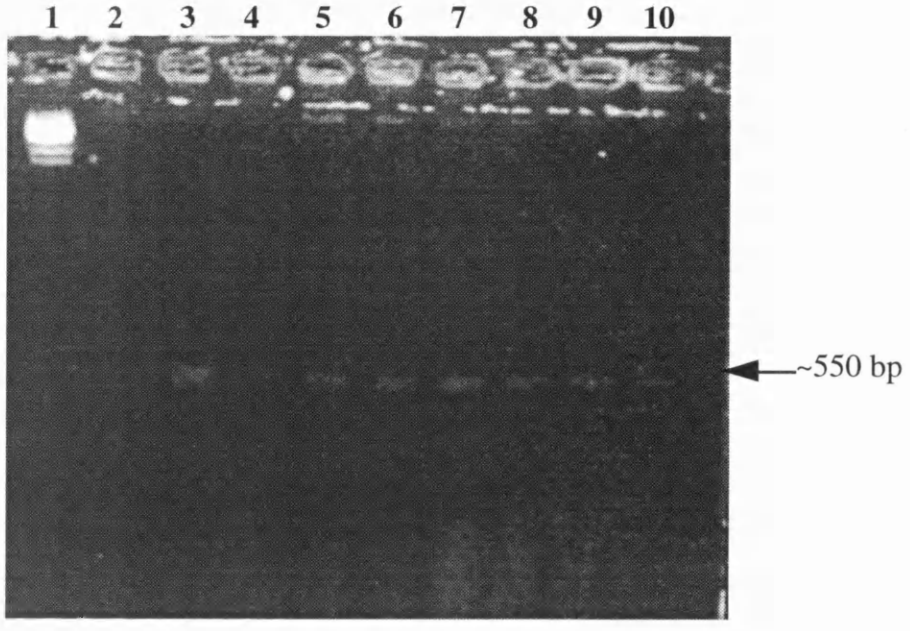


Figure 5.25 PCR using primers A and C on putative p550 plasmids

lane 1: λ/*Hin*d III markers
lanes 2-10: PCRs on putative p550 plasmids

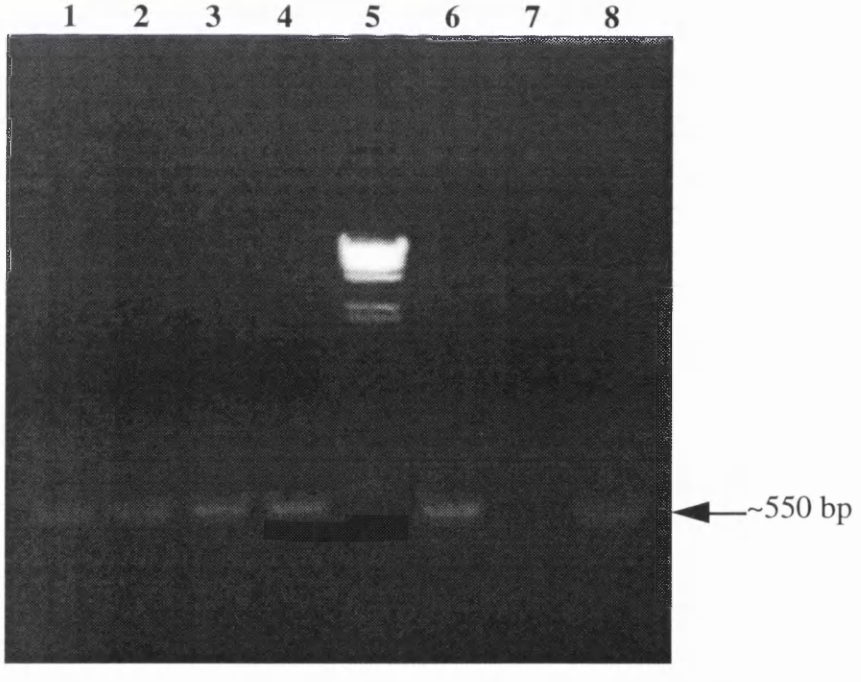


Figure 5.26 PCR using primers A and C on putative m550-18 and m550-19 single-stranded DNA

lanes 1-4: PCR on putative m550-18 templates
lane 5: λ/*Hin*d III markers
lanes 6-8: PCR on putative m550-19 templates

The products were run out on a 2% TAE gel (**Figure 5.25**). Although no product was observed with the putative p200, eight of the putative p550 plasmids gave a single product of the anticipated size.

The insert was isolated from one of these p550 plasmids by digestion with *Sst* I and *Sph* I, ligated into the M13mp18 and M13mp19 vectors (to give m550-18 and m550-19 respectively), and the mixtures transformed into TG-1 cells. Four white plaques were selected from each transformation, grown up as in the standard M13 protocol, and the supernatants used as templates for PCR reactions with primers A and C at 2·5 mM Mg^{2+}. The products were run out on a 2% TAE gel (**Figure 5.26**). Four of the putative m550-18 transformants and two of three putative m550-19 transformants gave a single product of the anticipated size. Single stranded DNA was therefore prepared from one each of the m550-18 and m550-19 transformants. The single-stranded DNA was sequenced using the -40 primer. However, the sequence derived from m550-18 consisted of M13mp18 polylinker sequence upto the *Sph* I restriction site, followed by pDK101 polylinker sequence upto the *Sst* I restriction site, and then more M13mp18 polylinker sequence. The sequence derived from m550-19 was the reverse-complement of this sequence. Therefore, it appears that the plasmid p550 contains no insert DNA, and is actually pDK101. The m550-18 and m550-19 clones constructs merely contain the *Sst* I-*Sph* I polylinker sequence. This suggests that when pDK101 was digested with the *Xcm* I enzyme, the enzyme only cut at one of its restriction sites in the polylinker, creating only a linear plasmid which had a single overhanging T at one end, but a single overhanging A at the other. Thus, a PCR product (which contains overhanging A's at both ends) would not be able to ligate to this linearised plasmid DNA, which must have merely undergone a religation event.

It also appeared that the approximately 550 bp PCR product generated with primers A and C using p550, m550-18 and m550-19 as templates was produced by one (or both) of the primers annealing to the vector DNA sequences outside of the polylinkers. To test this theory, PCR was performed on approximately 5 ng of pDK101 and 5 ng of M13mp18 using the same conditions as when genomic 1147 DNA (or p550, m550-18 or m550-19) was used as the template. When the PCR reactions were run out on a 2% TAE gel, both reactions were found to contain a single product of approximately 550 bp in length (**Figure 5.27**).

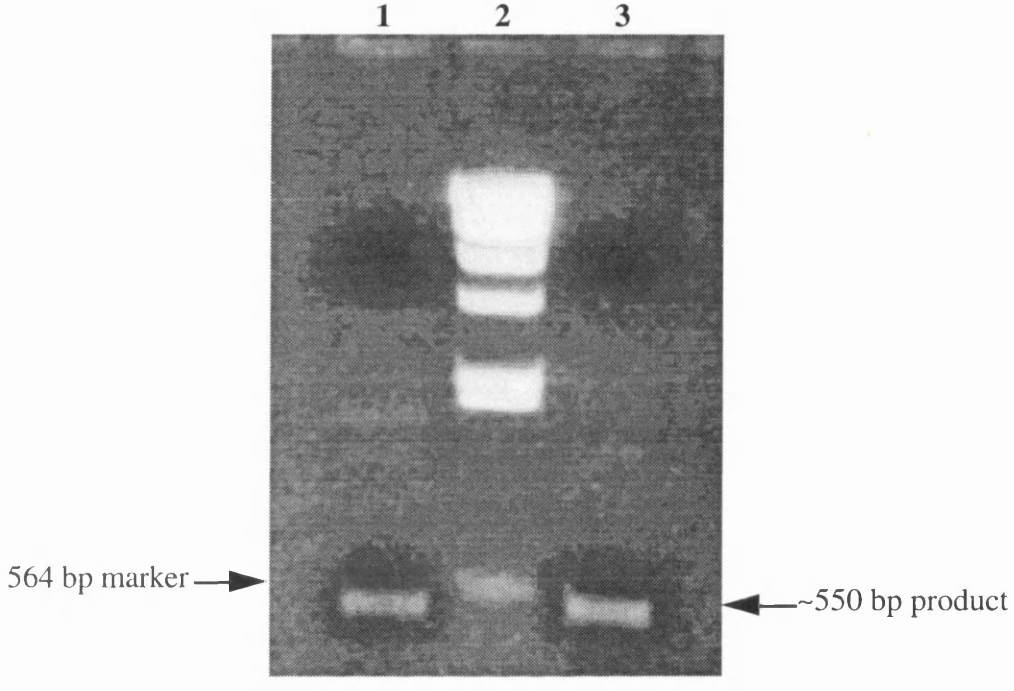


Figure 5.27 PCR using primers A and C on pDK101 and M13mp18

lane 1: PCR on ~5 ng pDK101

lane 2: λ/*Hin*d III markers

lane 3: PCR on ~5 ng M13mp18

5.2.15 Attempted cloning of *aroA* from a λ library

Radiolabelled primer C was used to probe a fully-representative λ-Gem11 genomic library of *S.coelicolor*, which comprised approximately 9000 plaques (Taylor, 1992). The filters were hybridised at 56°C in Quickhyb hybridisation buffer, and washed twice at 56°C in 4x SSC, 0·1% SDS, 0·1% sodium pyrophosphate. From this primary screen, five λ plaques were identified which lit up on duplicate filters. These plaques were picked and replated in *E.coli*, and the sub-library blotted onto nylon membranes to allow a secondary screening to be undertaken.

These filters were also probed with radiolabelled primer C, this secondary screen identifying 10 clones (produced from only two of the initial five clones) which lit up on autoradiograms. Again, these plaques were picked from the secondary plates and replated, and blotted onto nylon membranes. However, when these filters were probed with radiolabelled primer C, no λ plaques lit up on autoradiograms.

5.3 Discussion

Ideally, the EPSP synthase enzyme (encoded by *aroA*) of *S.coelicolor* would have been purified and N-terminal sequencing undertaken, allowing nucleotide primers to be designed against the deduced DNA sequence of the coding region of *aroA*. These oligonucleotides would then have been used to clone the *aroA* gene, either by PCR using genomic DNA isolated from *S.coelicolor* as template, or by probing genomic digests or λ-libraries. When the work described in this thesis was begun, it was anticipated that the EPSP synthase would be purified by collaborators with this research group from the Biochemistry Department of the University of Glasgow. However, such purification was not attempted. Therefore, alternative methods of cloning the *aroA* gene were undertaken.

Oligonucleotides were designed against DNA sequence derived from regions of protein sequence conserved across a number of EPSP synthase sequences. Initially, the only sequences available were from plants, Gram-negative bacteria and lower eukaryotes. Although PCR products of approximately the correct size were identified, two of which were cloned and sequenced, attempts at cloning a fragment of the *aroA* gene using the primers designed against these sequences were not successful.

One of the oligonucleotides designed against these amino acid sequences was subsequently used in an attempt to clone the *aroA* gene from digests of genomic DNA. Two genomic DNA bands of a suitable size for cloning were identified, one of which was cloned. However, the DNA sequence of part of this fragment encodes an amino acid sequence which shows no similarity to the protein sequences from any other EPSP synthases.

While such attempts at cloning of the *aroA* gene were being undertaken, the cloning of the genes encoding the EPSP synthase enzymes from a number of other species, including plants, lower eukaryotes and Gram-negative bacteria, was reported. The deduced amino acid sequences from three Gram-positive bacteria and one cyanobacterium were also reported. Although the amino acid sequences from all of the other species exhibited significant levels of similarity to each other, the enzymes from the three Gram-positive bacteria and the cyanobacterium showed only limited similarity to the other enzymes, but significant similarity to each other (although the deduced protein sequence from the enzyme of *Mycobacterium tuberculosis* showed poor similarity to any of the other deduced protein sequences).

These four deduced amino acid sequences were lined up, and showed regions of consensus sequence, particularly across the enzymes from *B.subtilis*, *S.aureus* (both Gram-positive bacteria) and *Synechocystis* (the cyanobacterium). Since it was anticipated that the enzyme from *S.coelicolor* would show greatest similarity to those of other Gram-positive bacteria, oligonucleotides were designed against these regions of consensus, and again used for cloning of the *aroA* gene from *S.coelicolor*. However, these attempts to clone *aroA* were also unsuccessful.

CHAPTER 6

Expression of the 3-dehydroquinase from *E.coli* in *S.coelicolor*

6.1 Introduction

6.1.1 Difference in K_m between the type I and type II dehydroquinase enzymes

If the shikimate pathway is not functioning at its maximum capacity in *S.coelicolor*, then increasing the flux through the pathway may be facilitated by deregulation of the pathway, upregulation of each step of the pathway, or even by upregulation of the steps in the pathway which have high flux control coefficients, and are therefore involved in the control of the rate of flux through the pathway (see section **1.12.3**). Methods of producing upregulation of a catalytic step include overexpression of the enzyme, or the introduction of a more efficient homologue of the enzyme, either as a replacement for the indigenous enzyme or in addition to it.

The type II dehydroquinase enzymes have higher K_m's than those of the type I enzymes; the *S.coelicolor* enzyme has been reported as having a K_m of 650 μM, which is approximately twenty times that of the *E.coli* type I enzyme [White *et al.*, 1990]. Due to the differing K_m's of the two dehydroquinase enzymes, the type I enzyme is more capable of sequestering its substrate than its type II counterpart. It was therefore proposed that the *E.coli* type I enzyme may perform more efficiently in the shikimate pathway of *S.coelicolor* than the host type II enzyme, particularly during conditions of low substrate concentration which may occur during the life-cycle and lead to the initiation of differentiation, or during periods of high levels of flux through the pathway, for example during secondary metabolism. Therefore the working hypothesis for the experiments described in this chapter was that expression of the *E.coli* type I dehydroquinase in *S.coelicolor* might upregulate this particular step, and possibly increase the overall flux through the pathway. As the work for this thesis was undertaken as part of a project aimed at increasing the rates of synthesis of the intermediates and products of the shikimate pathway, which may then be used as substrates for the biosynthesis of secondary metabolites, increasing the flux through the shikimate pathway was one of the major aims of this work.

Expression of the *E.coli aroD* gene in a strain of *S.coelicolor* containing a disrupted *aroQ* gene would demonstrate whether the *E.coli aroD* gene is able to complement this mutation, and therefore whether a type I dehydroquinase is able to replace the type II enzyme in *S.coelicolor*. Complementation would also provide further evidence that the disrupted gene is involved in the conversion of DHQ to DHS, and is either a structural gene (likely to encode the dehydroquinase enzyme) or a regulatory gene acting on the expression of dehydroquinase.

Since this research was begun, the *aroQ* gene of *S.coelicolor* has been expressed in *E.coli*, under the control of the inducible T7 promoter [G. Young, unpublished results]. This construct is able to complement the *E.coli* strain AB2848, which contains an *aroD* mutation. Also, when expression from the T7 promoter in this construct is induced in *E.coli*, a protein of the correct size for the *S.coelicolor* type II dehydroquinase was overexpressed. This protein was subsequently purified and characterised biochemically, and shown to be the *S.coelicolor* type II dehydroquinase enzyme.

6.1.2 The *tac* promoter

The *tac* promoter (P_{tac}) is a hybrid promoter created from two *E.coli* promoters, the *trp* and *lac* UV5 promoters [de Boer *et al.*, 1983]. It consists of the -35 region from the strong *trp* promoter, and the -10 region from the *lac* UV5 promoter. In *E.coli*, it is repressed by the *lac* repressor, and derepressed by IPTG. It is stronger than both parental promoters.

It has been demonstrated that *Streptomyces* have an RNA polymerase holoenzyme which is capable of transcribing certain promoters from *E.coli* [Jaurin & Cohen, 1984]. Indeed, it was shown that P_{tac} was expressed in *S.lividans* [Horinouchi & Beppu, 1985]. It has been suggested that P_{tac} shows a high level of similarity to the "consensus" promoter sequences of both *E.coli* and *Streptomyces*, and so its ability to function in *S.lividans* was not unexpected [Alves, 1989].

P_{tac} is repressed by the *lac* repressor (encoded by the *lacI* gene) and derepressed by IPTG in both *E.coli* and *S.lividans*. However, it has also been suggested that there is no naturally-occurring repressor capable of causing significant repression in *S.lividans* [Alves, 1989]. Therefore, although the addition of IPTG to an *E.coli* culture for transcription from P_{tac} is necessary, the same is not true for *S.lividans*. It was anticipated that the same situation exists in *S.coelicolor* as in *S.lividans*, and so no IPTG induction of the P_{tac}-*aroD* construct was attempted.

6.2 Results

The *aroD* gene of *E.coli*, encoding its type I 3-dehydroquinase, has been cloned [Kinghorn *et al.*, 1981] and sequenced [Duncan *et al.*, 1986; Chaudhuri *et al.*, 1991]. It was previously subcloned into a pKK223/3 vector, bringing it under the control of the *E.coli tac* promoter (P_{tac}) to facilitate overexpression.

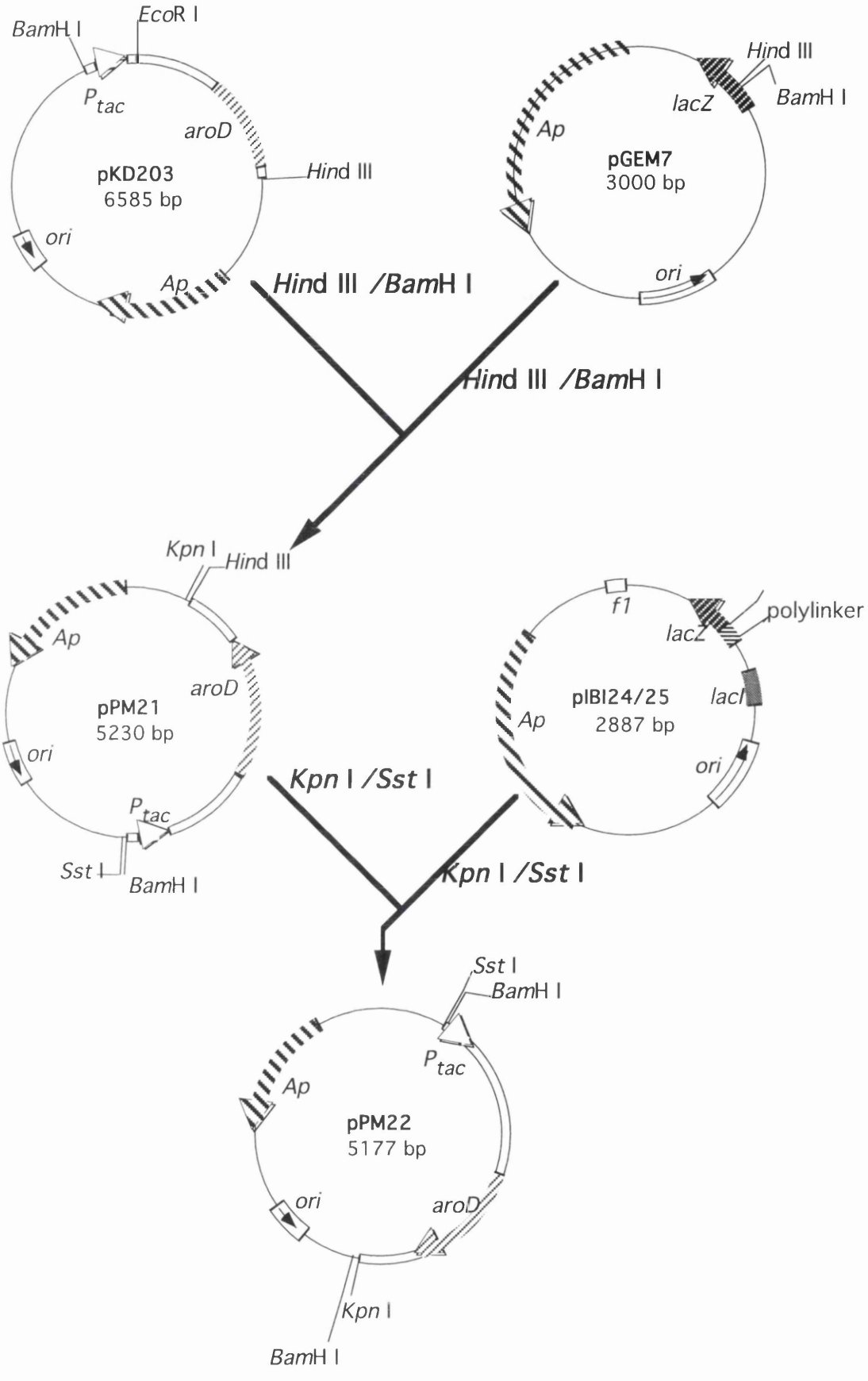
The pWHM3 vector is a bifunctional *E.coli-Streptomyces* shuttle vector, containing a plasmid origin of replication (*ori*) from both species [Vara *et al.*, 1989]. It also contains genes encoding resistance to ampicillin and thiostrepton, providing selectable resistance markers in *E.coli* and *Streptomyces* respectively. The *Bcl* I fragment containing the thiostrepton resistance marker was removed previously, the plasmid religated, and the *vph* gene subcloned into the *Sph* I site, producing the bifunctional shuttle vector pKL69 [K.Linton, personal communication]. This *vph* gene, originally cloned from *Streptomyces vinaceus* [Thompson *et al.*, 1982b], encodes the viomycin phosphotransferase enzyme which provides resistance to the antibiotic viomycin by phosphorylating the antibiotic, producing phosphoviomycin which has no antibiotic activity [Thompson *et al.*, 1982c]. The *vph* gene, like *tsr*, functions as a selectable marker in *Streptomyces*.

6.2.1 Construction of a bifunctional vector allowing expression of the 3-dehydroquinase from *E.coli*

In order to express the P_{tac}-*aroD* gene construct in *S.coelicolor*, the construct underwent two subcloning steps to produce the restriction sites to allow it to be subcloned into the shuttle vector pKL69, ultimately producing plasmid pPM23 (**Figure 6.1**). This plasmid was isolated from the *E.coli* TG-1 strain in which it had been constructed, and transformed into the *E.coli aroD* strain, AB2848. Two transformants, along with two AB2848 colonies and a TG-1 colony were streaked sequentially onto minimal agar, minimal agar supplemented with the aromatic amino acids (phenylalanine, tyrosine and tryptophan), and L agar. The plates were incubated at 37°C for up to 2 days. Although non-transformed AB2848 did not grow on minimal agar due to its aromatic amino acid auxotrophy (one of the colonies grew very poorly on minimal agar, presumably due to the residual aromatic amino acids contained in the cells when streaked onto the minimal plate), cells transformed with pPM23 did grow, demonstrating that this construct was able to complement the *aroD* mutation in AB2848, and therefore that the type I dehydroquinase was expressed from this construct from the P_{tac} promoter in *E.coli* [**Figure 6.2**]. TG-1 was used as a positive control, and grew on minimal agar as anticipated. All five colonies grew on both minimal agar supplemented with the aromatic amino acids and on L agar.

6.2.2 Expression of the 3-dehydroquinase from *E.coli* in *S.coelicolor*

For transformation of protoplasts of *S.coelicolor* with pPM23, two methods of evading the restriction barrier of *S.coelicolor* were attempted: passage of plasmid from *E.coli*


BamH I
EcoR I
P tac
aroD
pKD203
6585 bp
Hind III
ori
Ap
Hind III
BamH I
lacZ
Ap
pGEM7
3000 bp
ori
Hind III /BamH I
Hind III /BamH I
Kpn I
Hind III
Ap
aroD
pPM21
5230 bp
ori
P tac
Sst I
BamH I
f1
lacZ
polylinker
pIBI24/25
2887 bp
lacI
Ap
ori
Kpn I /Sst I
Kpn I /Sst I
Sst I
BamH I
P tac
Ap
pPM22
5177 bp
ori
aroD
Kpn I
BamH I

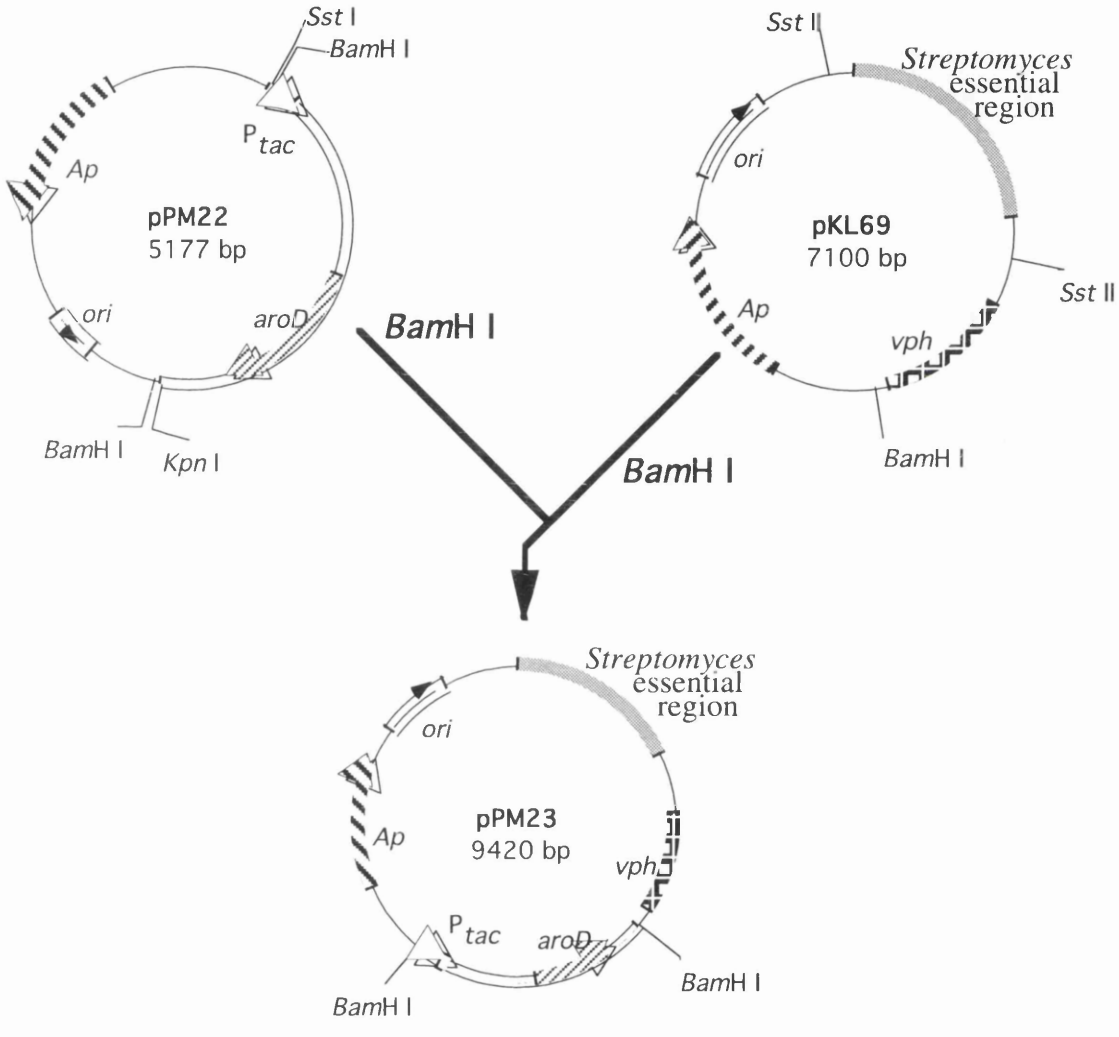


Figure 6.1 Construction of a bifunctional plasmid expressing the 3-dehydroquinase from *E.coli*

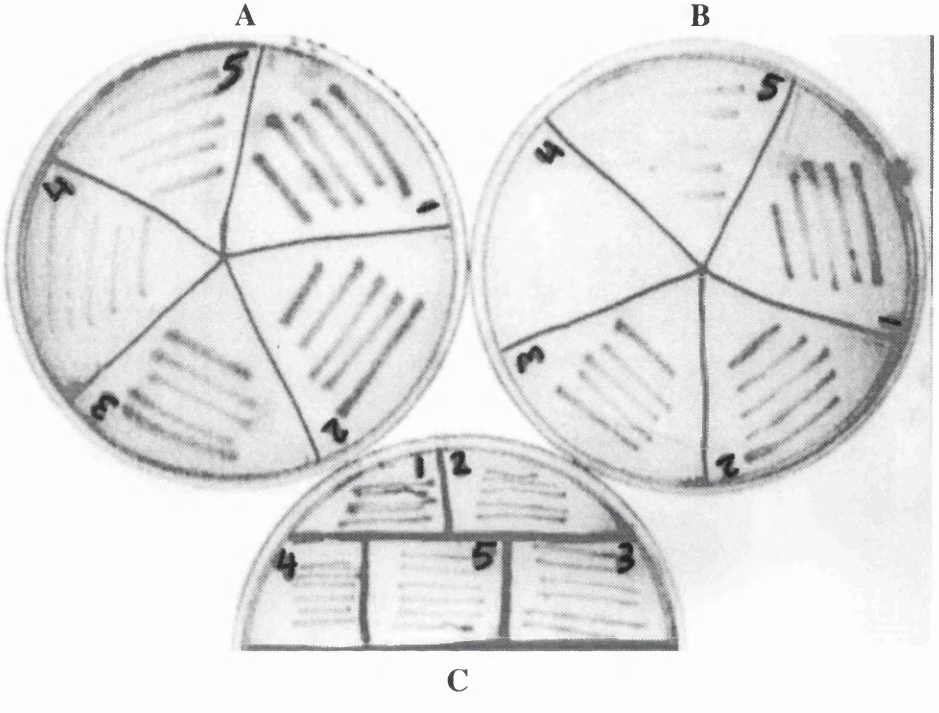


Figure 6.2 Complementation of an *aroD* mutation in *E.coli* with pPM23

Plate **A**: Minimal medium + phenylalanine, tyrosine and tryptophan

Plate **B**: Minimal medium

Plate **C**: L agar

Strain **1**: TG-1

Strains **2** and **3**: AB2848/pPM23

Strains **4** and **5**: AB2848

via *S.lividans* into *S.coelicolor*, and transformation of *S.coelicolor* with non-methylated plasmid DNA.

For the first strategy, pPM23 was isolated from the methylating *E.coli* strain TG-1, and transformed into protoplasts of *S.lividans* TK64. The transformation mixture was plated on R2 regeneration medium and incubated at 30°C, with transformants being selected by overlaying with viomycin after approximately 18 hrs growth.

Two transformants were isolated, were grown in YEME liquid medium supplemented with viomycin, and plasmid DNA was isolated by alkaline lysis (see **2.3.2.4**). This plasmid DNA was transformed into protoplasts of *S.coelicolor* G216, which were again plated on R2 and overlaid with viomycin. Three colonies were selected for further analysis (G216/pPM23-1, -2, -4).

For the second strategy, pPM23 isolated from *E.coli* TG-1 was transformed into MB5386, a *dam*$^-$*dcm*$^-$ *E.coli* strain. Protoplasts of *S.coelicolor* G216 were transformed with pPM23 isolated from MB5386, and transformants selected by overlaying with viomycin. Two colonies were isolated (G216/pPM23-X, -Y).

These five isolates were grown in YEME supplemented with viomycin, and plasmid DNA was isolated using the Kieser method of alkaline lysis. As plasmid isolation from *S.coelicolor* was not particularly successful, and as pPM23 is a bifunctional *E.coli-S.coelicolor* shuttle vector, plasmid DNA rescued from these G216 isolates was transformed into competent *E.coli* TG-1 cells, and transformants selected by growth on L agar supplemented with ampicillin. Transformants were obtained with DNA rescued from the strains G216/pPM23-2, -X, and -Y. Plasmid DNA was isolated from these *E.coli* transformants, and digested with *Bam*H I, which removes the complete P_{tac}-*aroD* fragment from the vector. The restriction map produced from the sole transformant from pPM23-2 was not consistent with it being the desired construct, and so pPM23-2 was not investigated further. However, plasmid DNA from all of the putative pPM23-X and pPM23-Y transformants which were tested produced the anticipated restriction map. Therefore, G216/pPM23-X and G216/pPM23-Y isolates both apparently contained the correct pPM23 construct.

6.2.3 Biochemical analysis of *E.coli* and *S.coelicolor* containing pPM23

The two G216 isolates containing pPM23 were grown in YEME liquid medium supplemented with viomycin, glycine and $MgCl_2$. Untransformed G216 and TG-1 were grown in YEME and 2xYT respectively. TG-1/PM23-Y was grown in 2xYT

supplemented with ampicillin. The TG-1 strains were grown in 100 ml of culture for approximately 7 hours; the G216 strains were grown in 200 ml of culture, and cells were harvested by centrifugation, the pellets resuspended in extraction buffer, and the cells lysed using the French press method. The lysates were centrifuged to remove cell debris, and the supernatant assayed for dehydroquinase activity. The dehydroquinase activities in the crude extracts were assayed on a spectrophotometer by measuring the rate of increase in OD at a wavelength of 234 nm when the substrate (DHQ) is added to a sample of enzyme contained in assay buffer (**2.5.2**). Typically, 0·05 M Tris-Cl (pH 8·0) was used as the assay buffer (this is an assay buffer for a type II dehydroquinase enzyme) [White *et al.*, 1990]. The activities were calculated as an average of at least three assay measurements using differing volumes of supernatant in the assay.

The protein concentration in the crude extract was estimated from a standard curve of BSA (bovine serum albumin) using the Bradford's protein assay, measured at OD_{595} (**2.5.3**). From these estimated protein concentrations, the specific activity of dehydroquinase in the crude extracts (measured in Units of enzyme per mg of protein) was calculated.

The untransformed G216 mycelia contain only the heat-stable type II dehydroquinase, whereas the pPM23-containing G216 isolates were expected to also contain the heat-labile type I dehydroquinase activity. All three *E.coli* cultures were expected to contain only the type I dehydroquinase activity, but it was anticipated that the dehydroquinase would be overexpressed in the cells containing pPM23.

Type I dehydroquinase enzymes are known to be thermolabile, and so crude extracts were heated to 70°C for 15 mins to denature the type I dehydroquinase, followed by further assays of their dehydroquinase enzyme activity. Alternatively, the assays were performed in a potassium phosphate (pH 7·0) (the assay buffer for a type I dehydroquinase enzyme), in the presence or absence of 0·1 M KCl, which has previously been demonstrated to inhibit the type I dehydroquinase enzyme in *E.coli* [Chaudhuri *et al.*, 1986], but not the type II enzyme in *S.coelicolor* [White *et al.*, 1990].

The TG-1/pPM23-Y isolate showed an approximately 2968-fold overexpression of the type I dehydroquinase relative to the activity in the untransformed strain (**Table 6.1**). However, when the crude extract of TG-1/pPM23-Y was heated to 70°C, its dehydroquinase activity was reduced 1385-fold. When TG-1/pPM23-Y was assayed for activity in 0·1 M potassium phosphate buffer containing 0·1 M KCl, no activity was detected. These results confirm that the type I dehydroquinase enzyme, when

Strain	Assay conditions	Activity (Units/ml)	Protein concentration (mg/ml)	Specific Activity (Units/mg)	Standard Deviation (σ_{n-1})
TG-1	0·05 M Tris-Cl (pH 8)	1·3	9·0	0·14	0·11
TG-1/ pPM23-Y	0·05 M Tris-Cl (pH 8)	4155·0	10·0	415·5	N/A
	0·05 M Tris-Cl Heat: 70°C, 15 mins	3·0	10·0	0·30	0·005
	0·1 M pot. phos. (pH 7) + 0·1 M KCl	0·35	10·0	0·04	0·07
G216	0·05 M Tris-Cl (pH 8)	4·0	3·5	1·14	0·40
G216/ pPM23-X	0·05 M Tris-Cl (pH 8)	120·0	2·7	44·4	6·80
G216/ pPM23-Y	0·05 M Tris-Cl (pH 8)	8·0	1·5	5·33	0·45

Table 6.1 Activity of dehydroquinase in *E.coli* and *S.coelicolor* strains

N.B. pot. phos. denotes potassium phosphate

extracted from *E.coli*, is thermolabile and its activity is inhibited by KCl.

The G216 strains were grown until they formed a fairly dense culture but before they produced excessive amounts of pigment (which is indicative of them having undergone differentiation from primary to secondary metabolism). G216/pPM23-X cells showed an approximately 39-fold increase in activity, and G216/pPM23-Y an approximately 4·7-fold increase relative to the activity in crude extracts from the untransformed G216 strain (**Table 6.1**). These results suggested that the type I dehydroquinase enzyme from *E.coli* is functional in *S.coelicolor*, and is expressed in this host from the Pt_{ac} promoter.

Approximately 50 μg of protein from the crude extracts from these *S.coelicolor* strains, along with those from TG-1 and uninduced and IPTG-induced TG-1/pPM23-Y were run on a PAGE gel, using a 5% stacking gel and 12% running gel. As can be seen from **Figure 6.3**, the two TG-1 transformants overexpressed a single band of apparent molecular weight of approximately 25 kDa, which was not evident in the

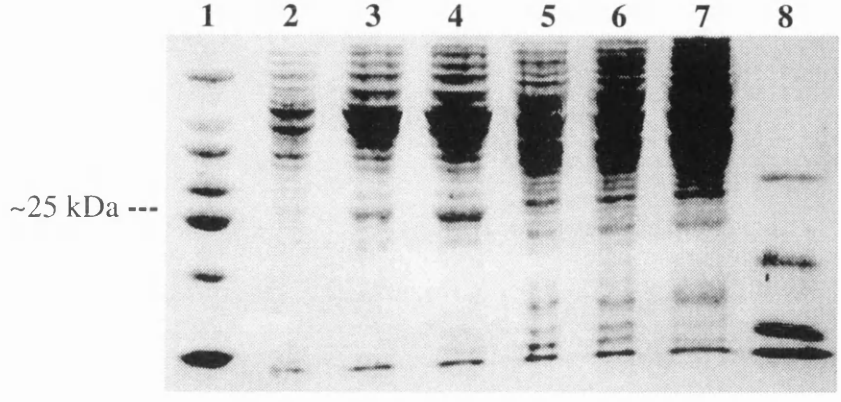


Figure 6.3 Crude extracts from *E.coli* and *S.coelicolor* with and without plasmid pPM23

Lane **1**: VII-L protein markers (14.2 / 20.1 / 24 / 29 / 36 / 45 / 66 kDa)
Lane **2**: Approx. 50 µg total protein from TG-1
Lane **3**: Approx. 50 µg total protein from TG-1/pPM23-Y (non-induced)
Lane **4**: Approx. 50 µg total protein from TG-1/pPM23-Y (induced)
Lane **5**: Approx. 50 µg total protein from G216
Lane **6**: Approx. 50 µg total protein from G216/pPM23-X
Lane **7**: Approx. 50 µg total protein from G216/pPM23-Y
Lane **8**: Low molecular weight protein markers

crude extract from the untransformed TG-1 strain. However, this protein was not visible in any of the crude extracts from the G216 strains.

When the assays were repeated using *S.coelicolor* cells grown to stationary phase (and producing considerable amounts of pigment), G216/pPM23-X exhibited the same level of activity as untransformed G216, and G216/pPM23-Y showed only an approximately 1·3-fold increase in activity (**Table 6.2**). Therefore, it appeared that the activity of dehydroquinase in the overexpressing strains was reduced by the cells having undergone differentiation.

Strain	Assay conditions	Activity (Units/ml)	Protein concentration (mg/ml)	Specific Activity (Units/mg)	Standard Deviation (σ_{n-1})
G216	0·05 M Tris-Cl (pH 8)	2·3	5·2	0·44	0·49
G216/ pPM23-X	0·05 M Tris-Cl (pH 8)	1·6	3·7	0·43	0·31
G216/ pPM23-Y	0·05 M Tris-Cl (pH 8)	1·5	2·6	0·58	0·15

Table 6.2 Dehydroquinase activity in *S.coelicolor* cultures having undergone differentiation

The three G216 strains were again grown in the appropriate liquid medium, but cells were harvested when they had only grown to a relatively low density. Their dehydroquinase activities were measured in 0·05 M Tris-Cl buffer, with G216/pPM23-X cells showing an approximate 3·5-fold reduction in activity, and G216/pPM23-Y cells exhibiting approximately the same activity as untransformed G216 cells (**Table 6.3**).

The three G216 strains were once again grown in the appropriate liquid medium, and cells were harvested at an OD_{600} of between 0·35 and 0·45. The G216 cells (OD_{600} of 0·35) had not excreted any pigment into the medium; the G216/pPM23-X cells (OD_{600} of 0·4) had secreted very little, if any, pigment; the G216/pPM23-Y cells (OD_{600} of 0·45) appeared to have just begun to excrete pigment. The dehydroquinase activities were assayed in 0·05 M Tris-Cl, with the G216/pPM23-Y cells exhibiting approximately a 5-fold increase in activity relative to the untransformed G216 and the G216/pPM23-X cells(**Table 6.4**).

Strain	Assay conditions	Activity (Units/ml)	Protein concentration (mg/ml)	Specific Activity (Units/mg)	Standard Deviation (σ_n)
G216	0·05 M Tris-Cl (pH 8)	0·16	0·75	0·21	0·03
G216/ pPM23-X	0·05 M Tris-Cl (pH 8)	0·31	5·0	0·06	0·06
G216/ pPM23-Y	0·05 M Tris-Cl (pH 8)	0·94	4·4	0·21	0·05

Table 6.3 Dehydroquinase activity in growing cultures of *S.coelicolor*

Strain	Assay conditions	Activity (Units/ml)	Protein concentration (mg/ml)	Specific Activity (Units/mg)	Standard Deviation (σ_n)
G216	0·05 M Tris-Cl (pH 8)	1·3	6·75	0·19	0·20
G216/ pPM23-X	0·05 M Tris-Cl (pH 8)	1·5	8·75	0·17	0·12
G216/ pPM23-Y	0·05 M Tris-Cl (pH 8)	8·0	8·5	0·94	0·24

Incubated at 70°C for 15 mins:

Strain	Assay conditions	Activity (Units/ml)	Protein concentration (mg/ml)	Specific Activity (Units/mg)	Standard Deviation (σ_n)
G216	0·05 M Tris-Cl (pH 8)	1·1	6·75	0·16	0·07
G216/ pPM23-X	0·05 M Tris-Cl (pH 8)	1·5	8·75	0·17	0·23
G216/ pPM23-Y	0·05 M Tris-Cl (pH 8)	1·4	8·5	0·16	0·11

Table 6.4 Dehydroquinase activity in cultures of *S.coelicolor* grown to an OD_{600} of 0·35 to 0·45

The crude extracts from the three strains were then heated to 70°C for 15 mins, and re-assayed (**Table 6.4**). Neither of the extracts from the G216 or G216/pPM23-X cultures showed any significant decrease in activity, but that from the G216/pPM23-Y culture showed a decrease in activity of approximately 80%, falling to a level similar to

that of the other two strains. This result suggests that G216/pPM23-Y was overexpressing the type I dehydroquinase from *E.coli*, but that the G216/pPM23-X strain was not, and contained only the native type II dehydroquinase activity.

The crude extracts were assayed for type I dehydroquinase activity in 0·1 M potassium phosphate buffer [Chaudhuri *et al.*, 1986] containing KCl (**Table 6.5**). It was anticipated that all three extracts would exhibit the same level of activity in the presence of 0·1 M KCl, as KCl inhibits the activity of the type I dehydroquinase enzyme. Although both of the strains not expressing the type I dehydroquinase under these conditions showed the same activity, the overexpressing strain had an approximately 17·7-fold higher level than these two strains. This appears to suggest that KCl was not exerting its usual inhibitory effects on type I dehydroquinase when this form of the enzyme was expressed in *S.coelicolor*. To quantify the inhibitory effect of KCl on the type I dehydroquinase in *S.coelicolor*, the activity in the crude extract from G216/pPM23-Y was assayed in 0.1 M potassium phosphate containing increasing concentrations of KCl. The dehydroquinase activity appears to be inversely proportional to the concentration of KCl in the assay buffer (**Figure 6.4**).

Strain	Assay conditions	Activity (Units/ml)	Protein concentration (mg/ml)	Specific Activity (Units/mg)	Standard Deviation (σ_{n-1})
G216	0·1 M pot. phos. (pH 7) + 0·1 M KCl	0·29	6·75	0·043	0·12
G216 / pPM23-X	0·1 M pot. phos. (pH 7) + 0·1 M KCl	0·38	8·75	0·043	0·10
G216 / pPM23-Y	0·1 M pot. phos. (pH 7) + 0·1 M KCl	6·5	8·5	0·76	0·10
	0·1 M pot. phos. (pH 7) + 0·2 M KCl	6·0	8·5	0·71	0·03
	0·1 M pot. phos. (pH 7) + 0·25 M KCl	5·4	8·5	0·64	0·27
	0·1 M pot. phos. (pH 7) + 0·35 M KCl	4·8	8·5	0·56	0·06
	0·1 M pot. phos. (pH 7) + 0·45 M KCl	4·2	8·5	0·49	0·15

Table 6.5 Dehydroquinase activity in crude extracts of *S.coelicolor* in the presence of KCl

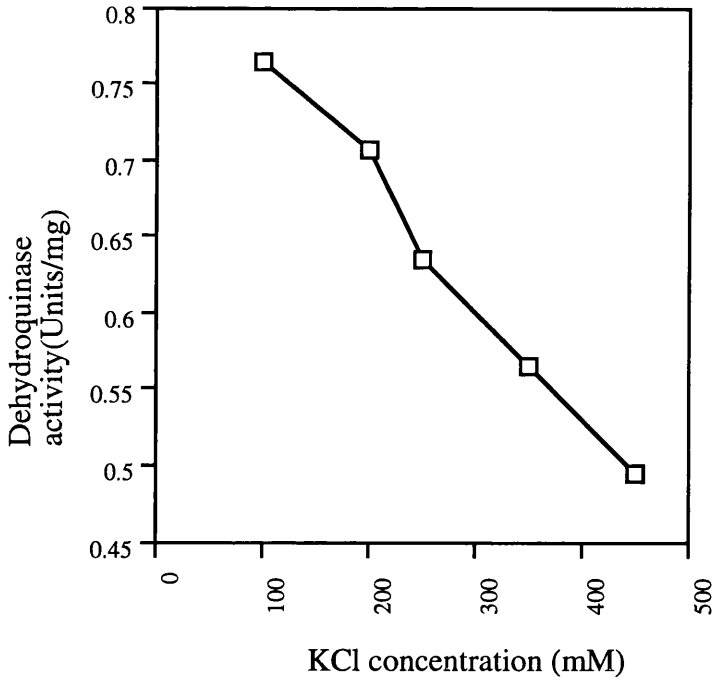


Figure 6.4 Inhibition of dehydroquinase activity in G216/pPM23-Y by KCl

The three crude extracts were frozen at -20°C for storage, but were thawed for assaying in the 0·1 M potassium phosphate (pH 7·0) buffer with and without KCl. They were also assayed in 0·05 M Tris-Cl (pH 8·0) buffer to allow comparison with the activities which had been determined previously. The crude extracts from both G216 and G216/pPM23-Y showed a significant decrease in activity in 0·05 M Tris-Cl relative to their activities before freezing (approximately 21% and 16% respectively) (**Table 6.6** relative to **Table 6.4**); the crude extract from G216/pPM23-X showed no significant decrease in activity (approximately 6%).

The crude extracts from all three strains showed no significant decrease in activity in the 0·1 M potassium phosphate (pH 7·0) buffer as they did before freezing (**Table 6.6** relative to **Table 6.5**).

The crude extracts from G216 and G216/PM23-X both showed slight increases in activity when assayed in 0·1 M potassium phosphate buffer containing 0·1 M KCl when compared to their activities in the absence of KCl, further evidence that neither culture is overexpressing the type I dehydroquinase. Although the overexpressing strain (G216/pPM23-Y) showed a decrease of approximately 17% due to the addition of 0·1 M KCl to the assay buffer, it still showed a 16-fold increase relative to that of

Strain	Assay conditions	Activity (Units/ml)	Protein concentration (mg/ml)	Specific Activity (Units/mg)	Standard Deviation (σ_n)
G216	0·05 M Tris-Cl (pH 8)	1·0	6·75	0·15	0·23
	0·1 M pot. phos. (pH 7)	0·24	6·75	0·04	0·04
	0·1 M pot. phos. (pH 7) + 0·1 M KCl	0·28	6·75	0·04	0·03
G216 / pPM23-X	0·05 M Tris-Cl (pH 8)	1·4	8·75	0·16	0·21
	0·1 M pot. phos. (pH 7)	0·17	8·75	0·02	0·06
	0·1 M pot. phos. (pH 7) + 0·1 M KCl	0·24	8·75	0·03	0·04
G216 / pPM23-Y	0·05 M Tris-Cl (pH 8)	6·7	8·5	0·79	0·24
	0·1 M pot. phos. (pH 7)	6·5	8·5	0·76	0·08
	0·1 M pot. phos. (pH 7) + 0·1 M KCl	5·4	8·5	0·64	0·16

Table 6.6 Effect of 0·1 M KCl on dehydroquinase activities in *S.coelicolor*

G216 under the same conditions (**Table 6.6**).

These results suggest that KCl does not have an inhibitory effect on the activity of the type II dehydroquinase in *S.coelicolor*; it does have an inhibitory effect on the activity of the type I dehydroquinase enzyme. However, although 0·1 M KCl effects 100% inhibition in *E.coli*, it only exerted approximately 16% inhibition in extracts of *S.coelicolor*.

6.2.4 Analysis of the shikimate intermediates in crude extracts and supernatants

It had been shown previously that the concentrations of six of the seven shikimate pathway intermediates (all except for DAHP) can be measured by high-performance liquid chromatography on an Aminex HPX-87H column, by following the absorbency

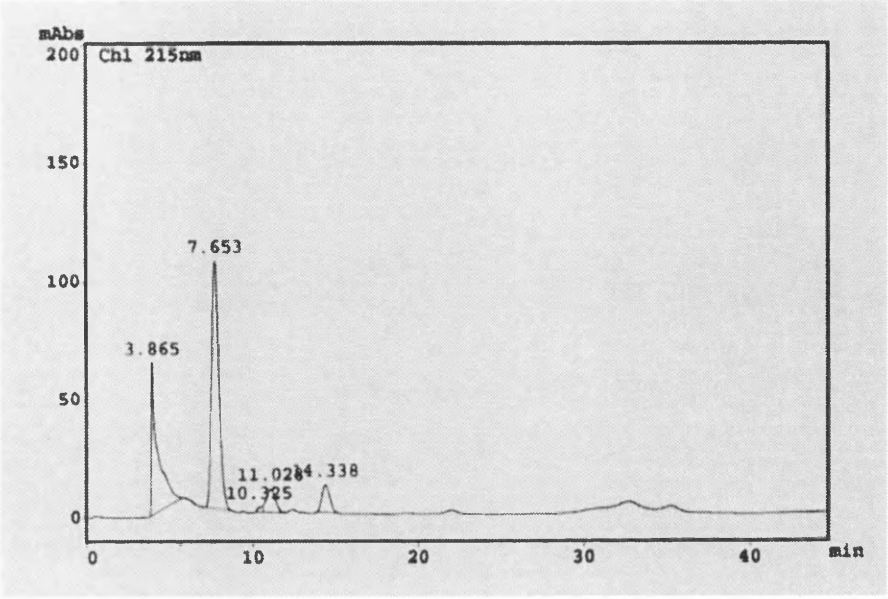
of the flow-through on a spectrophotometer at a wavelength of 215 nm [Mousdale & Coggins, 1985]. Therefore, approximately 10 μg of protein from the crude extract of the three cultures whose dehydroquinase activities were reported in **Table 6.5** and **6.6** were run down this column in 5 mM H_2SO_4. 100 pmoles of dehydroquinate and shikimate were also run separately on the column to correlate the retention times, which had been reported as 5·6 and 8·3 mins respectively, and were measured at 5·6 and 7·8 mins respectively in this experiment.

The only obvious difference between the three chromatograms was that the crude extract from G216 contained excessive amounts of shikimate (approx. 110 mAbs units relative to 15 mAbs units in the extract from G216/pPM23-X and 6 mAbs units in that from G216/pPM23-Y) (**Figure 6.5**).

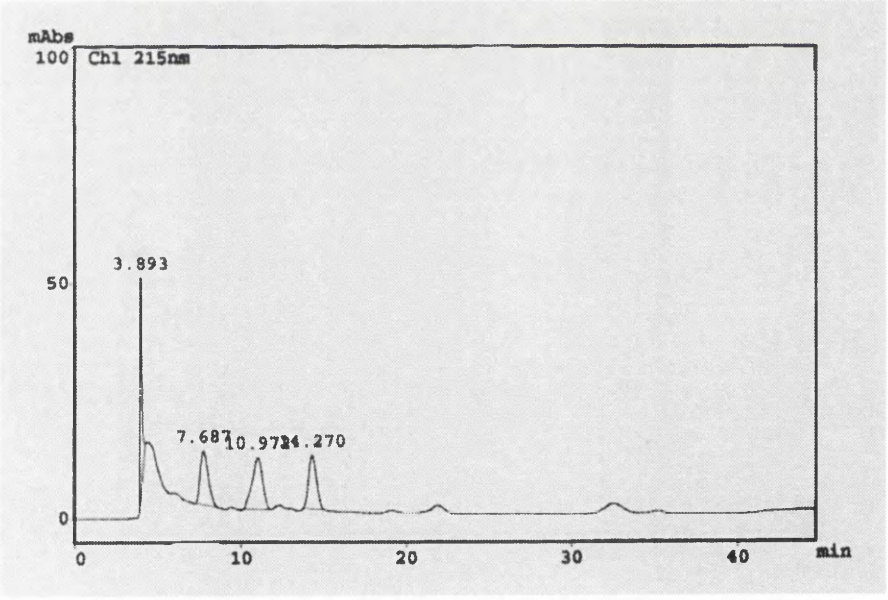
When the cultures were pelleted prior to French Pressing and assaying of their activity, they were diluted 1:1 with water to reduce the viscosity of the culture, and allow easier sedimentation of the cells. The supernatants of culture broth were frozen at -20°C for storage. 20 μl of the supernatant from each of the three cultures was also run down the Aminex column in 5 mM H_2SO_4 to discover whether any of the shikimate pathway intermediates were being excreted in large volumes from the cells.

Again, little discernible difference was observed between the chromatograms, except that the peak corresponding to shikimate 3-phosphate/EPSP from the supernatant from exponentially-growing G216 cells was approximately twice as high as that from the other supernatants (**Figure 6.6**).

G216:



G216/pPM23-X:



G216/pPM23-Y:

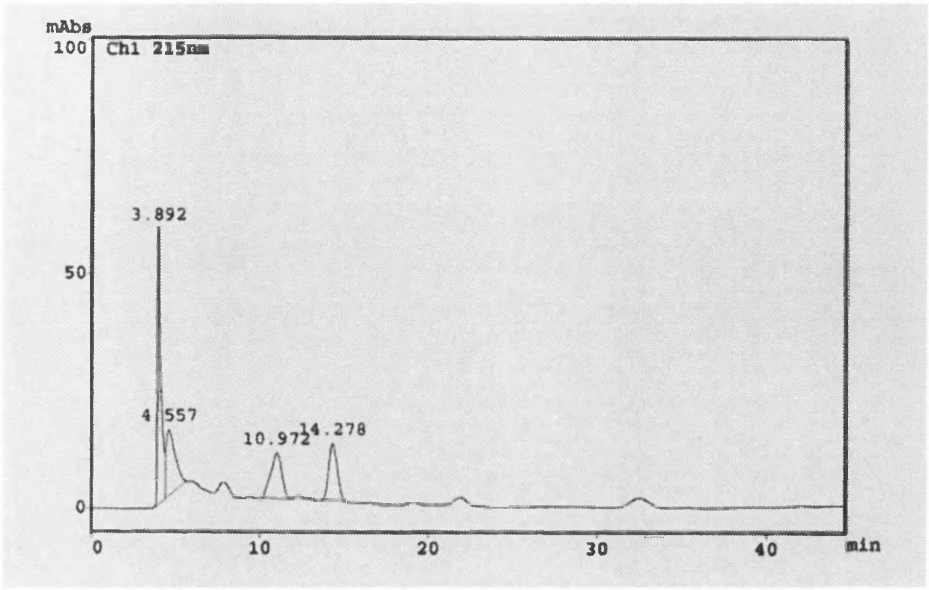
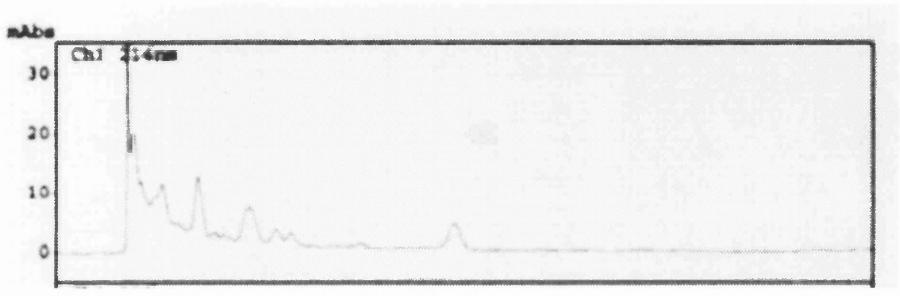


Figure 6.5 Chromatograms of crude extracts run on an Aminex amino exchange HPLC column in 5 mM H_2SO_4

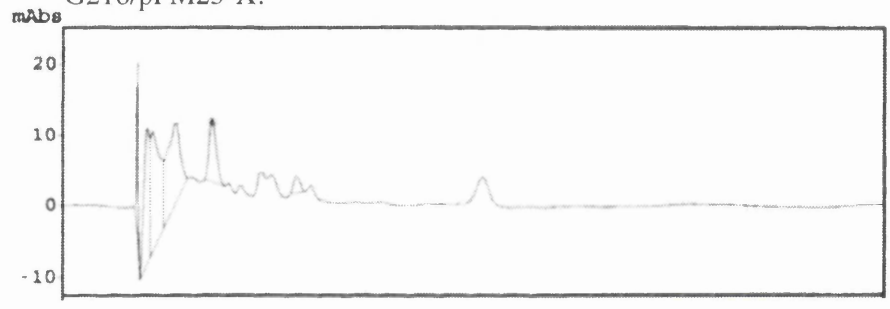
	S3P	EPSP	DHQ	DHS	Shikimate	Chorismate
reported retention times	4·0	4·0	5·6	8·3	11·0	27·6
measured retention times	N/D	N/D	5·6	7·8	N/D	N/D

Table 6.7 Retention times (in minutes) of shikimate pathway intermediates and product on an Aminex anion exchange HPLC column eluted with 5 mM H_2SO_4

G216:



G216/pPM23-X:



G216/pPM23-Y:

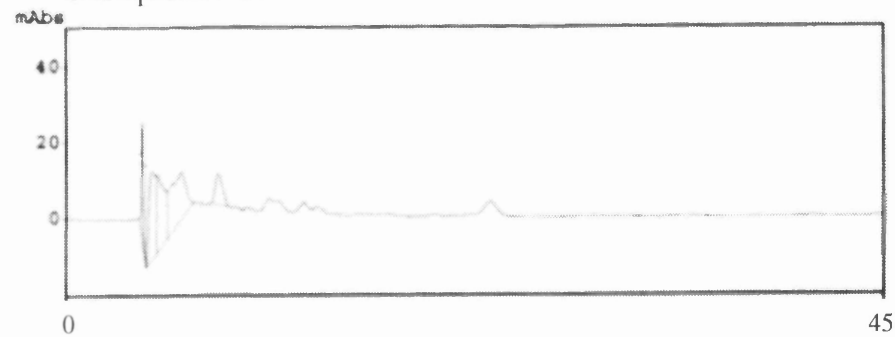


Figure 6.6 Chromatograms of supernatants run on an Aminex anion exchange HPLC column in 5 mM H_2SO_4

6.3 Discussion

The type I dehydroquinase from *E.coli* has been expressed in *S.coelicolor* G216 from the hybrid P_{tac} promoter, and forms an active protein. It has been overexpressed by virtue of the gene encoding this protein, *aroD*, being carried on the multicopy bifunctional vector pKL69. Despite the apparently low level of transcription from the Pt_{ac} promoter in *Streptomyces*, cells expressing this type I dehydroquinase enzyme have been measured as exhibiting an almost 40-fold increase in the levels of dehydroquinase relative to the untransformed parental strain.

As anticipated, the type I dehydroquinase was thermolabile in *S.coelicolor*, but it appeared to be inhibited to a much lower level by the presence of Cl^- anions in the assay buffer compared to the activity of the same enzyme in a crude extract from *E.coli* cells. It is feasible that the crude extract from *S.coelicolor* already contained significant levels of Cl^- anions, levels almost approaching saturation for this enzyme, causing a limited inhibition of the activity by the addition of KCl to the assay buffer. However, this appears unlikely, since this would imply that very large amounts of the type I dehydroquinase were present in these overexpressing cells, and no protein of the correct size was visible when crude extracts from such a strain were run on a SDS-PAGE gel. Also, 100 mM KCl appeared to completely inhibit the activity of the type I dehydroquinase in a crude extract from *E.coli* containing the overexpressing plasmid, suggesting that this level of KCl should also be sufficient to completely inhibit the type I enzyme in the crude extract from *S.coelicolor*.

Alternatively, it is possible that there is some other molecule sequestering Cl^- anions in the crude extract of *S.coelicolor* which is able to titrate out these anions. This would mean that the concentration of Cl^- anions available to inhibit the type I dehydroquinase would be significantly lower than that anticipated from the concentration of KCl in the assay buffer. However, since there is incomplete inhibition at a KCl concentration of up to 500 mM, this possibility appears to be unlikely. It is also possible that the type I dehydroquinase from *E.coli*, when expressed in *S.coelicolor*, is in some way protected from being bound by the Cl^- anions.

Although the results described in this chapter are only preliminary, they appear to imply that the type I dehydroquinase is only expressed at significant levels around the transition from exponential to stationary phase, which roughly coincides with the transition from primary to secondary metabolism (biochemical differentiation). It is feasible that, during the approach to differentiation, the primary metabolic pathways are upregulated in order to provide elevated levels of their products and intermediates,

which may then be used as substrates for the secondary metabolic pathways. The factors controlling such an upregulation event may induce the expression of the type I dehydroquinase.

Temporal regulation of the type I dehydroquinase in *S.coelicolor* would suggest either a transcriptional or translational control, or the inactivation of an inhibitor. The *Ptac* promoter apparently falls into the class of promoters which are capable of initiating transcription in both *E.coli* and *Streptomyces* [Horinouchi & Beppu, 1985]. Therefore, it is feasible that transcription from the P_{tac} promoter is regulated in a similar manner to a naturally-occurring *Streptomyces* promoter which it resembles, and which may be switched on around the time of the differentiation process. To test this hypothesis, it would be interesting to express the type I dehydroquinase from *Streptomyces* promoters which are known to initiate transcription at different stages of the life-cycle.

Translational regulation of the type I dehydroquinase in *S.coelicolor*, or the existence of a regulatable inhibitory substance, might imply that *S.coelicolor* contains a type I dehydroquinase enzyme, or that its ancestors contained such an enzyme, and that, while *S.coelicolor* may have lost this enzyme, it may have retained the controls for its expression. It is possible that *S.coelicolor* does contain a type I dehydroquinase, but that its expression is temporally regulated in the manner observed for the type I dehydroquinase from *E.coli* when expressed in *S.coelicolor*, and is expressed only transiently, and so has not yet been identified.

A more likely method for regulation of the activity of the type I dehydroquinase when expressed in *S.coelicolor* may be effected by the product of the *bldA* gene, which encodes the $tRNA_{UUA}$ which carries a leucine amino acid for insertion into the growing polypeptide chain during translation. This $tRNA_{UUA}$ is apparently extremely rare in *Streptomyces*, due to the temporal regulation of the expression from the *bldA* gene [Leskiw *et al.*, 1991]. UUA (leucine) codons occur only extremely rarely in *Streptomyces* genes, and are generally found only in genes concerned with antibiotic production or differentiation, i.e. genes which are expressed late in the growth cycle.

The mRNA produced by transcription of the *aroD* gene contains a UUA codon (TTA in the DNA sequence [Chaudhuri *et al.*, 1991]) at nucleotides 736-738. Therefore, it is possible that the type I dehydroquinase is only expressed at the transition from exponential to stationary phase because the *bldA* product is only available in the cell at a significant concentration at this point in the cell cycle.

Although G216/pPM23-Y is overexpressing the type I dehydroquinase, producing an approximately 5-fold increase in dehydroquinase activity relative to the untransformed G216 strain, there is no apparent build-up of any of the shikimate pathway intermediates in the crude extract of the overexpressing strain, and no excessive excretion of any of these intermediates into the growth medium. However, it is possible that an enzyme which functions earlier in the pathway is limiting the rate of flux through the pathway, preventing an increase in activity of dehydroquinase from having any effect on the pathway. The limiting enzyme may be DAHP synthase (which is believed in certain systems to be involved, through feedback inhibition, in regulation of the pathway) or dehydroquinate synthase. It is even possible that the availability of the precursors of DAHP (E4P and PEP) may be limiting the rate of flux through the shikimate pathway so that, under the growth conditions used, the whole of the shikimate pathway is not functioning at its maximum potential.

Another alternative is that increasing the dehydroquinase activity does cause an increase in the flux through the shikimate pathway, but that the product of the pathway, chorismate, is immediately converted to another metabolite as part of a different metabolic pathway, and is therefore not built up in, or excreted from, the overexpressing cells.

CHAPTER 7

Discussion and Future Work

7.1 Introduction

The overall aims of the experiments described in this thesis were to allow analysis of the rate of carbon flux through the primary metabolic shikimate pathway in *S.coelicolor*; to attempt to overexpress the dehydroquinase enzymes from both *E.coli* and *S.coelicolor* in *S.coelicolor* to ascertain whether any resultant increase in dehydroquinase activity produces an increase in the rate of flux through the pathway; and to clone the *aroA* gene, encoding EPSP synthase enzyme, to allow gene disruption and the subsequent analysis of flux through the majority of the pathway.

7.2 Attempted disruption of *aroQ*

To allow analysis of the flux through the first two steps of the pathway, it was intended to knock out the indigenous dehydroquinase enzyme by disruption of the *aroQ* gene of *S.coelicolor* via insertion mutagenesis. Initially, the *ermE* gene from *S.erythraea*, which encodes a gene providing resistance to erythromycin, was subcloned into the middle of the *aroQ* gene carried on a suicide plasmid. However, when protoplasts of *S.coelicolor* were transformed with this suicide plasmid, resultant transformants which showed the required antibiotic resistance profile (erythromycin resistant, thiostrepton sensitive) were found not to contain the *ermE* gene, and must have undergone spontaneous mutation to resistance.

The *ermE* gene was subsequently replaced with the *grmA* gene, which induces resistance to gentamicin. Again, transformants showing the anticipated resistance profile were found, by probing of genomic DNA, not to contain the *grmA* gene. Therefore, in order to disrupt the *aroQ* gene in *S.coelicolor*, another more reliable resistance marker must be found.

Disruption via integration of an internal fragment of the required gene with a single recombination event has been shown to be successful in *S.coelicolor* [Buttner & Lewis, 1992] using ∅C31 as the vector. This bacteriophage, or other unstable vectors, may be more useful delivery systems for promoting gene disruption than suicide *E.coli* vectors due to their relative longevity in the cell. However, disruption using gene replacement from a suicide vector has also shown to be successful in *S.coelicolor* [Khosla *et al.*, 1992].

Since three separate research groups have unsuccessfully attempted to disrupt the *aroD* gene in *E.coli* (which encodes the type I dehydroquinase enzyme) despite the

existence of *aroD* point mutations which cause auxotrophy for the three aromatic amino acids, it is possible that dehydroquinase is an essential enzyme, and therefore that null mutations in the genes encoding dehydroquinase are lethal. The *aroD* mutations which have arisen may be leaky, causing a low level of flux through the shikimate pathway to occur. This low level may be sufficient for the synthesis of another product (or by-product) of the shikimate pathway which is essential, and which has not been provided in the growth media used during the attempts at *aroD* and *aroQ* disruption. However, these point mutants do not allow sufficient flux through the shikimate pathway for the synthesis of the necessary quantities of the aromatic amino acids. It is anticipated that null mutations in the genes encoding dehydroquinase would prevent any flux through the shikimate pathway.

In order to test this theory, a copy of the *aroQ* gene from *S.coelicolor* could be subcloned onto a bacteriophage which subsequently infects a cell of *S.coelicolor* and integrates into the chromosome. Gene disruption could then be performed. Cells with a disruption in either the original chromosomal copy of *aroQ*, or the copy carried on the bacteriophage, could then be cured of the bacteriophage. If the only cells which survive are those with a functional *aroQ* gene on the chromosome, this would suggest that the *aroQ* gene product is essential. This method may also be performed on cells carrying an *aroQ* gene on a stably-replicating plasmid, and selection for the plasmid relaxed after gene disruption to allow loss of the plasmid.

Disruption of *aroQ* could also be attempted in the *S.coelicolor* strain expressing the type I dehydroquinase from *E.coli* (G216/pPM23). If the *aroD* gene is able to complement for a null mutation in *aroQ*, then the type I dehydroquinase would be capable of maintaining flux through the shikimate pathway, preventing disruption of the *aroQ* gene from being a lethal event.

Alternatively, disruption of the *aroD* gene may be performed in *E.coli* cells expressing the type II dehydroquinase from *S.coelicolor*. The *aroQ* gene from *S.coelicolor* has already been shown to complement an *aroD* point mutation in *E.coli* [G. Young, personal communication]. Therefore, if a functional dehydroquinase protein is essential to the cell, this would be supplied by the plasmid-borne *aroQ* gene, allowing the indigenous type I dehydroquinase to be knocked out.

7.3 Overexpression of the type II dehydroquinase in *S.lividans*

Overexpression of the type II dehydroquinase enzyme from *S.coelicolor* by

transformation of *S.lividans* with pPM27 has produced a 7-fold increase in activity relative to the untransformed *S.lividans* strain TK54. The plasmid construction was performed in *S.lividans* as this strain may be transformed with DNA isolated from *E.coli*. Apparently, only a single protein, the type II dehydroquinase, is overexpressed from the 3 kb insert contained in this plasmid. Not only does this indicate that the *aroQ* gene is transcribed from a promoter contained on this fragment, but also suggests that it is not transcriptionally coupled with any other gene.

In order to attempt to upregulate the conversion of dehydroquinate to dehydroshikimate in *S.coelicolor*, this overexpressing plasmid, pPM27, must be isolated from TK54/pPM27 cells, and transformed into protoplasts of *S.coelicolor*. Such overexpression may produce an increase in metabolic flux through the shikimate pathway. Isolation of pPM27 from the *S.lividans* strain TK54/pPM27 was attempted during this research, but was unsuccessful.

7.4 Attempted cloning of *aroA*

Oligonucleotide primers were designed against consensus regions of amino acid sequences of EPSP synthase enzymes from plants, Gram-negative bacteria and lower eukaryotes; and from Gram-positive bacteria and a cyanobacterium. These oligonucleotides were used in an attempt to clone a fragment of the *aroA* gene by PCR using genomic *Streptomyces* DNA as template, and for probing of genomic digests and lambda libraries to clone the *aroA* gene. However, none of these experiments was successful.

Ideally, oligonucleotides would have been designed against the DNA sequence deduced from the N-terminal (and possibly internal) amino acid sequence of the EPSP synthase from *S.coelicolor*. However, this protein has not been purified. Future work, therefore, would be to purify the protein and clone the gene encoding it via the above route. Alternatively, the *aroA* gene from *M.tuberculosis*, the closest relative to *S.coelicolor* from which the *aroA* gene has been cloned, could be used as a probe to clone the *aroA* gene of *S.coelicolor*. This method was considered, but the *aroA* gene from *M.tuberculosis* was not available. It would be possible to clone a region of this gene from genomic DNA isolated from *M.tuberculosis* by PCR using primers designed against the published DNA sequence [Garbe *et al.*, 1990], and to subsequently use this fragment for probing of genomic DNA or the cosmid library from *S.coelicolor*.

A further method for the cloning of the *aroA* gene would be to create a library of genomic DNA from *S.coelicolor*, and to attempt complementation of an *aroA* mutant of *E.coli*. This approach has been successful for the cloning of the *argCJB* cluster involved in arginine biosynthesis in *S.coelicolor* [Hindle *et al.*, 1994]. However, *Streptomyces* genes are usually only poorly expressed, or even not expressed at all, from their own promoters in *E.coli*. Indeed, the *aroQ* gene from *S.coelicolor* is not expressed from its own promoter in *E.coli* [P. White, Personal Communication]. When the *aroQ* gene was placed downstream of the T7 promoter, this construct did not complement an *aroD* mutant and it was only when the translation start site was engineered to have optimal spacing with regard to the *E.coli* ribosome binding site that complementation of the *E.coli aroD* mutant was achieved [G. Young, Personal Communication]. Also, attempts to clone genes involved in the shikimate pathway from *S.rimosus* by complementation of *E.coli* aromatic mutants were unsuccessful [Stuart, 1990]. Therefore, it is anticipated that this experiment would have a low probability of success, and was not undertaken.

7.5 Expression of a type I dehydroquinase in *S.coelicolor*

The type I dehydroquinase from *E.coli* was overexpressed from the hybrid P_{tac} promoter in *S.coelicolor* in an attempt to upregulate the conversion of dehydroquinate to dehydroshikimate, and possibly to increase the rate of flux through the shikimate pathway. Such expression produced an almost 40-fold increase in the dehydroquinase activity relative to that in the crude extract from untransformed parental cells. The type I dehydroquinase was thermolabile as anticipated, but was considerably more resistant to inhibition by Cl^- anions than expected.

The type I dehydroquinase appears to be expressed only around the point of transition from exponential to stationary phase, which coincides approximately with the onset of secondary metabolism (the biochemical differentiation process). Although this transient expression may a promoter-specific phenomenon, it is more likely that, as the *aroD* gene from *E.coli* contains a TTA codon at nucleotides 736-738, it is due to translational control by the product of the *bldA* gene in *S.coelicolor*, the *bldA* gene encoding the leucine-carrying $tRNA_{UUA}$, and which is apparently only expressed at significant levels towards the end of the growth cycle. To investigate this theory, it would be possible to mutate this TTA codon to a TTG codon, which also encodes a leucine, and so would function as a silent mutation. If expression of this type I dehydroquinase is indeed regulated by the product of the *bldA* gene, then it is anticipated that the mutant type I dehydroquinase would be expressed constitutively.

This would increase the possibility of the expression of this type I dehydroquinase enzyme having any effect on the flux through the shikimate pathway in *S.coelicolor*.

BIBLIOGRAPHY

Adamidis, T., Riggle, P. & Champness, W. (1990). Mutations in a new *Streptomyces coelicolor* locus which globally block antibiotic biosynthesis but not sporulation. J. Bact. **172**: 2962-2969.

Ahmed, S.I. & Giles, N.H. (1969). Organization of enzymes in the common aromatic synthetic pathway: evidence for aggregation in fungi. J. Bact. **99**: 231-237.

Alm, R.A., Dalrymple, B.P. & Mattick, J.S. (1994). Sequencing and expression of the *aroA* gene from *Dichelobacter nodosus*. Unpublished.

Alves, T.P.T.S. (1989). Development of plasmid vectors for regulated expression of cloned genes in *Streptomyces lividans* and *Escherichia coli*. Ph.D. Thesis. University of Glasgow.

Amrhein, N., Johänning, D., Schab, J. & Schulz, A. (1983). Biochemical basis for glyphosate-tolerance in a bacterium and a plant tissue culture. FEBS Lett. **157**: 191-196.

Anderson, K.S. & Sikorski, J.A. (1988). Isolation and structural elucidation of the tetrahedral intermediate in the EPSP synthase enzymatic pathway. J. Am. Chem. Soc. **110**: 6577-6579.

Anderson, K.S., Sikorski, J.A. & Johnson, K.A. (1988). A tetrahedral intermediate in the EPSP synthase reaction observed by rapid quench kinetics. Biochemistry **27**: 7395-7406.

Anton, D.L., Hedstrom, L., Fish, S.M. & Abeles, R.H. (1983). Mechanism of enolpyruvyl shikimate-3-phosphate synthase exchange of phosphoenolpyruvate with solvent protons. Biochemistry **22**: 5903-5908.

Anton, I.A. & Coggins J.R. (1984). Subcloning of the *Escherichia coli* shikimate dehydrogenase gene (*aroE*) from the transducing bacteriophage λspc1. Biochem. Soc. Trans. **12**: 275-276.

Anton, I.A. & Coggins, J.R. (1988). Sequencing and expression of the *Escherichia coli aroE* gene encoding shikimate dehydrogenase. Biochem. J. **249**: 319-326.

Anton, I.A., Duncan, K. & Coggins, J.R. (1987). A eukaryotic repressor protein, the *qa-1s* gene product of *Neurospora crassa*, is homologous to part of the Arom multifunctional enzyme. J. Mol. Biol. **197**: 367-371.

Anzai, H., Kumada, Y., Hara, O., Murakami, T., Itoh, R., Takano, E., Imai, S., Satoh, A. & Nagaoka, K. (1988). Replacement of *Streptomyces hygroscopicus* DNA segments with *in vitro* altered DNA sequences. J. Antibiotics **41**: 226-233.

Banerji, S., Wakefield, A.E., Allen, A.G., Maskell, D.J., Peters, S.E. & Hopkin, J.M. (1993). The cloning and characterization of the *arom* gene of *Pneumocystis carinii*. J. Gen. Micro. **139**: 2901-2914.

Bentley, R. (1990). The shikimate pathway - a metabolic tree with many branches. Crit. Rev. Biochem. & Mol. Biol. **25**: 307-384.

Beri, R.K., Whittington, H., Roberts, C.F. & Hawkins, A.R. (1987). Isolation and characterization of the positively acting regulatory gene *qutA* from *Aspergillus nidulans*. Nucleic Acids Res. **15**: 7991-8001.

Beri, R.K., Grant, S., Roberts, C.F., Smith, M. & Hawkins, A.R. (1990). Selective overexpression of the *qutE* gene encoding catabolic 3-dehydroquinase in multicopy transformants of *Aspergillus nidulans*. Biochem. J. **265**: 337-342.

Berlyn, M.B. & Giles, N.H. (1969). Organization of enzymes in the polyaromatic synthetic pathway: separability in bacteria. J. Bact. **99**: 222-230.

Berlyn, M.B. & Giles, N.H. (1973). Organisation of interrelated aromatic metabolic pathway enzymes in *Acinetobacter calcoaceticus*. J. Gen. Micro. **74**: 337-341.

Berlyn, M.B., Ahmed, S.I. & Giles, N.H. (1970). Organization of polyaromatic biosynthetic enzymes in a variety of photosynthetic organisms. J. Bact. **104**: 768-774.

Berson, A.E., Hudson, D.V. & Waleh, N.S. (1991). Cloning of a sequence of *Aquaspirillum magnetotacticum* that complements the *aroD* gene of *Escherichia coli*. Mol. Micro. **5**: 2261-2264.

Bibb, M.J., Findlay, P.R. & Johnson, M.W. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene **30**: 157-166.

Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**: 1513-1523.

Bode, R., Melo, C. & Birnbaum, D. (1984). Mode of action of glyphosate in *Candida maltosa*. Arch. Microbiol. **140**: 83-85.

Boocock, M.R. & Coggins, J.R. (1983). Kinetics of 5-enolpyruvylshikimate-3-phosphate synthase inhibition by glyphosate. FEBS Lett. **154**: 127-133.

Bradford, M.M., (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principal of protein dye binding. Annals Biochem. **72**: 248-254.

Brown, K.L., Wood, S. & Buttner, M.J. (1992). Isolation and characterization of the major vegetative RNA-polymerase of *Streptomyces coelicolor* A3(2) - renaturation of a single subunit using GroEL. Mol. Micro. **6**: 1133-1139.

Brown, K.D. & Somerville, R.L. (1971). Repression of aromatic amino acid biosynthesis in *Escherichia coli* K-12. J. Bact. **108**: 386-399.

Bugg T.D.H., Alefounder, P.R. & Abell, C. (1991). An amino acid sequence motif observed amongst enzymes of the shikimate pathway. Biochem. J. **276**: 841-845.

Buttner, M.J. & Lewis, C.G. (1992). Construction and characterization of *Streptomyces coelicolor* A3(2) mutants that are multiply deficient in the nonessential *hrd*-encoded RNA polymerase sigma factors. J. Bact. **174**: 5165-5167.

Buttner, M.J., Fearnley, I.M. & Bibb, M.J. (1987). The agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2): nucleotide sequence and transcriptional analysis. Mol. Gen. Genet. **209**: 101-109.

Buttner, M.J., Smith, A.M. & Bibb, M.J. (1988). At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). Cell **52**: 599-607.

Buttner, M.J., Chater, K.F. & Bibb, M.J. (1990). Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). J. Bact. **172**: 3367-3378.

Calcutt, M.J. & Schmidt, F.J. (1992). Conserved gene arrangement in the origin region of the *Streptomyces coelicolor* chromosome. J. Bact. **174**: 3220-3226.

Case, M.E. & Giles, N.H. (1971). Partial enzyme aggregates formed by pleiotropic mutants in the *arom* gene cluster of *Neurospora crassa*. Proc. Nat. Acad. Sci. **68**: 58-62.

Casjens, S. & Huang, W.M. (1993). Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. Mol. Micro. **8**: 967-980.

Catcheside, D.E.A., Storer, P.J. & Klein, B. (1985). Cloning of the *aro* cluster gene of *Neurospora crassa* and its expression in *Escherichia coli*. Mol. Gen. Genet. **199**: 446-451.

Charles, I.G., Keyte, J.W. & Brammar, W.J. (1985). Nucleotide sequence encoding the biosynthetic dehydroquinase function of the penta-functional *arom* locus of *Aspergillus nidulans*. Nucl. Acids Res. **13**: 8119-8128.

Charles, I.G., Keyte, J.W., Brammar, W.J., Smith, M. & Hawkins, A.R. (1986). The isolation and nucleotide sequence of the complex *arom* locus of *Aspergillus nidulans*. Nucl. Acids Res. **14**: 2201-2213.

Charles, I.G., Lamb, H.K., Pickard, D., Dougan, G. & Hawkins, A.R. (1990). Isolation, characterization and nucleotide sequences of the *aroC* genes encoding chorismate synthase from *Salmonella typhi* and *Escherichia coli*. J. Gen. Micro. **136**: 353-358.

Chater, K.F. (1989). Multilevel regulation of *Streptomyces* differentiation. TIG **5**:372-377.

Chater, K.F., Bruton, C.J., Plaskitt, K.A., Buttner, M.J., Méndez, C. & Helmann, J.D. (1989). The developmental fate of *Streptomyces coelicolor* hyphae depends upon a gene product homologous with the motility σ factor of *Bacillus subtilis*. Cell **59**: 133-143.

Chatfield, S., Dougan, G. & Charles, I. (1990). Complete nucleotide sequence of the *aroA* gene from *Salmonella typhi* encoding 5-enolpyruvylshikimate 3-phosphate synthase. Nucl. Acids Res. **18**: 6133.

Chaudhuri, S., Lambert, J.M., McColl, L.A. & Coggins, J.R. (1986). Purification and characterization of 3-dehydroquinase from *Escherichia coli*. Biochem. J. **239**: 699-704.

Chaudhuri, S., Duncan, K., Graham, L.D. & Coggins, J.R. (1991). Identification of the active site lysine residues of two biosynthetic 3-dehydroquinases. Biochem. J. **275**: 1-6.

Comai, L., Facciotti, D., Hiatt, W.R., Thompson, G., Rose, R.E. & Stalker, D.M. (1985). Expression in plants of a mutant *aroA* gene from *Salmonella typhimurium* confers tolerance to glyphosate. Nature **317**: 741-744.

Da Silva, A.J.F., Whittington, H., Clements, J., Roberts, C. & Hawkins, A.R. (1986). Sequence analysis and transformation by the catabolic 3-dehydroquinase (*qutE*) gene from *Aspergillus nidulans*. Biochem. J. **240**: 481-488.

Dalla Chiesa, M., Mayes, S.R., Maskell, D.J., Nixon, P.J. & Barber, J. (1994). An *aroA* homologue from *Synechocystis* sp. PCC6803. Gene **144**: 145-146.

Davies, J., Von Ahsen, U., Wank, H. & Schroeder, R. (1992). Evolution of secondary metabolite production: potential roles for antibiotics as prebiotic effectors of catalytic RNA reactions. In "Secondary metabolites: their evolution and function." Chadwick, D.J. & Whelan, J. (eds.). CIBA Foundation Symposium **171**: pp 24-44.

De Boer, H.A., Comstock, L.J. & Vasser, M. (1983). The *tac* promoter - a functional hybrid derived from the *trp* and *lac* promoters. Proc. Nat. Acad. Sci. **80**: 21-25.

De Feyter, R.C. & Pittard, J. (1986). Genetic and molecular analysis of *aroL*, the gene for shikimate kinase II in *Escherichia coli* K-12. J. Bact. **165**: 226-232.

De Feyter, R.C., Davidson, B.E. & Pittard, J. (1986). Nucleotide sequence of the transcription unit containing the *aroL* and *aroM* genes from *Escherichia coli* K-12. J. Bact. **165**: 233-239.

Deka, R.K., Kleanthous, C. & Coggins, J.R. (1992). Identification of the essential

histidine residue at the active site of *Escherichia coli* dehydroquinase. J. Biol. Chem. **267**: 22237-22242.

Deka, R.K., Anton, I.A., Dunbar, B. & Coggins, J.R. (1994). The characterisation of the shikimate pathway enzyme dehydroquinase from *Pisum sativum*. FEBS Lett. **349**: 397-402.

Demain, A.L. (1983). New applications of microbial products. Science **219**: 709-714.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. & Mattick, J.S. (1991). "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucl. Acids Res. **19**: 4008.

Duncan, K. & Coggins, J.R. (1984). Subcloning of the *Escherichia coli* genes *aroA* (5-enolpyruvylshikimate 3-phosphate synthase) and *aroB* (dehydroquinate synthase). Biochem. Soc. Trans. **12**: 274-275.

Duncan, K. & Coggins, J.R. (1986). The *serC-aroA* operon of *Escherichia coli*: a mixed function operon encoding enzymes from two different amino acid biosynthetic pathways. Biochem. J. **234**: 49-57.

Duncan, K., Lewondon, A. & Coggins, J.R. (1984a). The purification of 5-enolpyruvylshikimate 3-phosphate synthase from an overproducing strain of *Escherichia coli*. FEBS Lett. **165**: 121-127.

Duncan, K., Lewondon, A. & Coggins, J.R. (1984b). The complete amino acid sequence of *Escherichia coli* 5-enolpyruvylshikimate 3-phosphate synthase. FEBS Lett. **170**: 59-63.

Duncan, K., Chaudhuri, S., Campbell, M.S. & Coggins, J.R. (1986). The overexpression and complete amino acid sequence of *Escherichia coli* 3-dehydroquinase. Biochem. J. **238**: 475-483.

Duncan, K., Edwards, R.M. & Coggins, J.R. (1987). The pentafunctional Arom enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. Biochem. J. **246**: 375-386.

Errington, J. (1993). *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. Micro. Rev. **57**: 1-33.

Euverink, G.J.W., Hessels, G.I., Vrijbloed, J.W., Coggins, J.R. & Dijkhuizen, L. (1992). Purification and characterization of a dual function 3-dehydroquinate dehydratase from *Amycolatopsis methanolica*. J. Gen. Micro. **138**: 2449-2457.

Fernández-Moreno, M.A., Martín-Triana, A.J., Martínez, E., Niemi, J., Kieser, H.M., Hopwood, D.A. & Malpartida, F. (1992). *abaA*, a new pleiotropic regulatory locus for antibiotic production in *Streptomyces coelicolor*. J. Bact **174**: 2958-2967.

Gaertner, F.H. & Cole, K.W. (1977). A cluster-gene: evidence for one gene, one polypeptide, five enzymes. Biochem. Biophys. Res. Comm. **75**: 259-264.

Ganson, R.J. & Jensen, R.A. (1988). The essential role of cobalt in the inhibition of the cytosolic isozyme of DAHP synthase from *Nicotiana silvestris* by glyphosate. Arch. Biochem. Biophys. **260**: 85-93.

Garbe, T., Jones, C., Charles, I., Dougan, G. & Yound, D. (1990). Cloning and characterization of the *aroA* gene from *Mycobacterium tuberculosis*. J. Bact. **172**: 6774-6782.

Garbe, T., Servos, S., Hawkins, A., Dimitriadis, G., Young, D., Dougan, G. & Charles, I. (1991). The *Mycobacterium tuberculosis* shikimate pathway genes: evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases. Mol. Gen. Genet. **228**: 385-392.

Gasser, C.S. & Klee, H.J. (1990). A *Brassica napus* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase. Nucl. Acids Res. **18**: 2821.

Gasser, C.S., Winter, J.A., Hironaka, C.M. & Shah, D.M. (1988). Structure, expression and evolution of the 5-enolpyruvylshikimate-3-phosphate synthase genes of petunia and tomato. J. Biol. Chem. **263**: 4280-4287.

Geever, R.F., Huiet, L., Baum, J.A., Tyler, B.M., Patel, V.B., Rutledge, B.J., Case, M.E. & Giles, N.H. (1989). DNA sequence, organization and regulation of the *qa* gene cluster of *Neurospora crassa*. J. Mol. Biol. **207**: 15-34.

Gibson, T.J. (1984). Studies on the Epstein-Barr virus genome. PhD thesis, Cambridge, England.

Giles, N.H., Case, M.E., Partridge, C.W.H. & Ahmed, S.I. (1967). A gene cluster in

Neurospora crassa coding for an aggregate of five aromatic synthetic enzymes. Proc. Natl. Acad. Sci. **58**: 1453-1460.

Giles, N.H., Partridge, C.W.H., Ahmed, S.I. & Case, M.E. (1967). The appearance of two dehydroquinases in *Neurospora crassa,* one constitutive and one inducible. Proc. Natl. Acad. Sci. **58**: 1930-1937.

Glazer, N.A. & Nikaido, H. (1995). Microbial Biotechnology: Fundamentals of Applied Microbiology. W. H. Freeman & Co. pp 66.

Gormley, E.P. & Davies, J. (1991). Transfer of plasmid RSF1010 by conjugation from *Escherichia coli* to *Streptomyces lividans* and *Mycobacterium smegmatis*. J. Bact. **173**: 6705-6708.

Griffin, H.G. & Griffin, A.M. (1991). Cloning and DNA sequence analysis of the *serC-aroA* operon from *Salmonella gallinarum*; evolutionary relationships between the prokaryotic and eukaryotic *aroA*-encoded enzymes. J. Gen. Micro. **137**: 113-121.

Gygax, D., Christ, M., Ghisalba, O. & Nüesch, J. (1982). Regulation of DAHP synthase in *Nocardia mediterranei*. FEMS Mic. Lett. **15**: 169-173.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**: 557-580.

Harris, J., Kleanthous, C., Coggins, J.R., Hawkins, A.R. & Abell, C. (1993). Different mechanistic and stereochemical courses for the reactions catalysed by type I and type II dehydroquinases. J. Chem. Soc. Chem. Commun. **2**: 1080-1081.

Hasan, N. & Nester, E.W. (1978). Dehydroquinate synthase in *Bacillus subtilis*: an enzyme associated with chorismate synthase and flavin reductase. J. Biol. Chem. **253**: 4999-5004.

Hautala, J.A., Jacobson, J.W., Case, M.E. & Giles, N.H. (1975). Purification and characterization of catabolic dehydroquinase, an enzyme in the inducible quinic acid catabolic pathway of *Neurospora crassa*. J. Biol. Chem. **250**: 6008-6014.

Hawkins, A.R. (1987). The complex *arom* locus of *Aspergillus nidulans*. Curr. Genet. **11**: 491-498.

Hawkins, A.R. & Smith, M. (1991). Domain structure and interaction with the pentafunctional Arom polypeptide. Eur. J. Biochem. **196**: 717-724.

Hawkins, A.R., Giles, N.H. & Kinghorn, J.R. (1982). Genetical and biochemical aspects of quinate breakdown in the filamentous fungus *Aspergillus nidulans*. Biochem. Genet. **20**: 271-286.

Hawkins, A.R., Da Silva, A.J.F. & Roberts, C.F. (1984). Evidence for two control genes regulating expression of the quinic acid utilization (*qut*) gene cluster in *Aspergillus nidulans*. J. Gen. Micro. **130**: 567-574.

Hawkins, A.R., Da Silva, A.J.F. & Roberts, C.F. (1985). Cloning and characterization of the three enzyme structural genes *qutB*, *qutC*, and *qutE* from the quinic acid utilization gene cluster in *Aspergillus nidulans*. Curr. Genet. **9**: 305-311.

Hawkins, A.R., Lamb, H.K., Smith, M., Keyte, J.W. & Roberts, C.F. (1988). Molecular organisation of the quinic acid utilization (*qut*) gene cluster in *Aspergillus nidulans*. Mol. Gen. Genet. **214**: 224-231.

Hawkins, A.R., Lamb, H.K. & Roberts, C.F. (1992). Structure of the *Aspergillus nidulans qut* repressor-encoding gene: implications for the regulation of transcription initiation. Gene **110**: 109-114.

Hawkins, A.R., Lamb, H.K., Moore, J.D., Charles, I.G. & Roberts, C.F. (1993a). The pre-chorismate (shikimate) and quinate pathways in filamentous fungi: theoretical and practical aspects. J. Gen. Micro. **139**: 2891-2899.

Hawkins, A.R., Lamb, H.K., Moore, J.D. & Roberts, C.F. (1993b). Genesis of eukaryotic transcriptional activator and repressor proteins by splitting a multidomain anabolic enzyme. Gene **136**: 49-54.

Hawkins, A.R., Moore, J.D. & Adeokun, A.M. (1993c). Characterization of the 3-dehydroquinase domain of the pentafunctional Arom protein, and the quinate dehydrogenase from *Aspergillus nidulans*, and the overproduction of the type II 3-dehydroquinase from *Neurospora crassa*. Biochem. J. **296**: 451-457.

Heatwole, V.M. & Somerville, R.L. (1992). Synergism between the Trp repressor and Tyr repressor in repression of the *aroL* promoter of *Escherichia coli* K-12. J. Bact. **174**: 331-335.

Henner, D.J., Band, L., Flaggs, G. & Chen, E. (1986). The organization and nucleotide sequence of the *Bacillus subtilis hisH*, *tyrA* and *aroE* genes. Gene **49**: 147-152.

Hillemann, D., Pühler, A. & Wohlleben, W. (1991). Gene disruption and gene replacement in *Streptomyces* via single stranded DNA transformation of integration vectors. Nucl. Acids Res. **19**: 727-731.

Hindle, Z., Callis, R., Dowden, S., Rudd, B.A.M. & Baumberg, S. (1994). Cloning and expression in *Escherichia coli* of a *Streptomyces coelicolor* A3(2) *argCJB* gene cluster. Microbiology **140**: 311-320.

Hood, D.W., Heidstra, R., Swoboda, U.K. & Hodgson, D.A. (1992). Molecular genetic analysis of proline and tryptophan biosynthesis in *Streptomyces coelicolor* A3(2): interaction between primary and secondary metabolism - a review. Gene **115**: 5-12.

Hoiseth, S.K. & Stocker, B.A.D. (1981). Aromatic dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. Nature **291**: 238-239.

Hoiseth, S.K. & Stocker, B.A.D. (1985). Genes *aroA* and *serC* of *Salmonella typhimurium* constitute an operon. J. Bact. **163**: 355-361.

Homchampa, P., Strugnell, R.A. & Adler, B. (1992). Molecular analysis of the *aroA* gene of *Pasteurella multocida* and vaccine potential of a constructed *aroA* mutant. Mol. Micro. **6**: 3585-3593.

Hong, S-K., Kito, M., Beppu, T. & Horinouchi, S. (1991). Phosphorylation of the AfsR product, a global regulatory protein for secondary-metabolite formation in *Streptomyces coelicolor* A3(2). J. Bact. **173**: 2311-2318.

Hopwood, D.A. & Kieser, T. (1990). The *Streptomyces* chromosome. In "The Bacterial Chromosome." Drlica, K. & Riley, M. (eds.). Am. Soc. Micro., Washington DC. pp 147-162.

Hopwood, D.A. & Wright, H.M. (1983). CDA is a new chromosomally-determined antibiotic from *Streptomyces coelicolor* A3(2). J. Gen. Micro. **129**: 3575-3579.

Hopwood, D.A., Harold, R.J., Vivian, A. & Ferguson, H.M. (1969). A new kind of

fertility variant in *Streptomyces coelicolor*. Genetics **62**: 461-467.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. & Schrempf, H. (1985). Genetic manipulation of *Streptomyces*: a laboratory manual. The John Innes Foundation, Norwich, UK.

Horinouchi, S. & Beppu, T. (1985). Construction and application of a promoter-probe plasmid that allows chromogenic identification in *Streptomyces lividans*. J. Bact. **162**: 406-412.

Horinouchi, S. & Beppu, T. (1990). Autoregulatory factors of secondary metabolism and morphogenesis in actinomycetes. Biotechnology **10**: 191-204.

Horinouchi, S., Kito, M., Nishiyama, M., Furuya, K., Hong, S-K., Miyake, K. & Beppu, T. (1990). Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). Gene **95**: 490-56.

Huang, L., Montoya, A.L. & Nester, E.W. (1975). Purification and characterization of shikimate kinase enzyme activity in *Bacillus subtilis*. J. Biol. Chem. **250**: 7675-7681.

Hunter, I.S. (1985). Gene cloning in *Streptomyces*. In DNA Cloning Volume II: a practical approach. Glover, D.M. (ed.) IRL Press. pp 19-44.

Hunter, I.S. & Baumberg, S. (1989). Molecular genetics of antibiotic formation. In Microbial products: new approaches. Baumberg, S., Hunter, I.S. & Rhodes, M. (eds.) Society for General Microbiology Symposium **44**. Cambridge University Press. pp 122-162.

Huynh, Q.K. (1987). Reaction of 5-enol-pyruvoyl shikimate-3-phosphate synthase with diethyl pyrocarbonate: evidence for an essential histidine residue. Arch. Biochem. & Biophys. **258**: 233-239.

Huynh, Q.K. (1988). Evidence for a reactive γ-carboxyl group (Glu-418) at the herbicide glyphosate biding site of 5-enolpyruvylshikimate-3-phosphate synthase from *Escherichia coli*. J. Biol. Chem. **263**: 11631-11635.

Huynh, Q.K., Kishore, G.M. & Bild, G.S. (1988a). 5-enolpyruvyl shikimate 3-phosphate synthase from *Escherichia coli*: identification of Lys-22 as a potential active site residue. J. Biol. Chem. **263**: 735-739.

Huynh, Q.K., Bauer, S.C., Bild, G.S., Kishore, G.M. & Borgmeyer, G.R. (1988b). Site-directed mutagenesis of *Petunia hybrida* 5-enolpyruvylshikimate-3-phosphate synthase: Lys-23 is essential for substrate binding. J. Biol. Chem. **263**: 11636-11639.

Ingham, C.I., Hunter, I.S. & Smith, M.C.M. (1995). Rho-independent terminators without 3' poly-U tails from the early region of actinophage ØC-31. Nucl. Acids Res. **23**: 370-376.

Ishizuka, H., Horinouchi, S., Kieser, H.M., Hopwood, D.A. & Beppu, T. (1992). A putative two-component regulatory system involved in secondary metabolism in *Streptomyces* spp. J. Bact **174**: 7585-7594.

Jaurin, B. & Cohen, S.N. (1984). *Streptomyces lividans* RNA polymerase recognises and uses *Escherichia coli* transcriptional signals. Gene **28**: 83-91.

Kacser, H. & Porteous, J.W. (1987). Control of metabolism: what do we have to measure? TIBS **12**: 5-15.

Kelemen, G.H., Cundliffe, E. & Financsek, I. (1991). Cloning and characterization of gentamicin-resistance genes from *Micromonospora purpurea* and *rosea*. Gene **98**: 53-60.

Kendall, K. & Cullum, J. (1984). Cloning and expression of an extracellular-agarase gene from *Streptomyces coelicolor* A3(2) in *Streptomyces lividans* 66. Gene **29**: 315-321.

Khosla, C., Ebert-Khosla, S. & Hopwood, D.A. (1992). Targeted gene replacements in a *Streptomyces* polyketide synthase gene cluster: role for the acyl carrier protein. Mol. Micro. **6**: 3237-3249.

Kieser, H.M., Kieser, T. & Hopwood, D.A. (1992). A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. J. Bact. **174**: 5496-5507.

Kinghorn, J.R. & Hawkins, A.R. (1982). Cloning and expression in *Escherichia coli* K-12 of the biosynthetic dehydroquinase function of the *arom* cluster gene from the eukaryote, *Aspergillus nidulans*. Mol. Gen. Genet. **186**: 145-152.

Kinghorn, J.R., Schweizer, M., Giles, N.H. & Kushner, S.R. (1981). The cloning and analysis of the *aroD* gene of *Escherichia coli* K-12. Gene **14**: 73-80.

Kleanthous, C. & Coggins, J.R. (1990b). Reversible alkylation of an active site methionine residue in dehydroquinase. J. Biol. Chem. **265**: 10935-10939.

Kleanthous, C., Campbell, D.G. & Coggins, J.R. (1990a). Active site labelling of the shikimate pathway enzyme, dehydroquinase. J. Biol. Chem. **265**: 10929-10934.

Kleanthous, C., Deka, R., Davis, K., Kelly, S.M., Cooper, A., Harding, S.E., Price, N.C., Hawkins, A.R. & Coggins, J.R. (1992). A comparison of the enzymological and biophysical properties of two distinct classes of dehydroquinase enzymes. Biochem. J. **282**: 687-695.

Klee, H.J., Muskopf, Y.M. & Gasser, C. (1987). Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. Mol. Gen. Genet. **210**: 437-442.

Kormanec, J. & Farkasovsky, M. (1993). Differential expression of principal sigma factor homologues of *Streptomyces aureofaciens* correlates with the developmental stage. Nucl. Acids Res. **21**: 3647-3652.

Kormanec, J., Farkasovsky, M. & Potúcková, L. (1992). Four genes in *Streptomyces aureofaciens* containing a domain characteristic of principal sigma factors. Gene **122**: 63-70.

Koshiba, T. (1978). Purification of two forms of the associated 3-dehydroquinate hydro-lase and shikimate:$NADP^+$ oxidoreductase in *Phaseolus mungo* seedlings. Biochim Biophys. Acta **522**: 10-18.

Kovalic, D., Kwak, J.H. & Weisblum, B. (1991). General method for direct cloning of DNA fragments generated by the polymerase chain reaction. Nucl. Acids Res. **19**: 4560.

Laemmli U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**: 680-685.

Lalonde, G., O'Hanley, P.D., Stocker, B.A.D. & Denich, K.T. (1994). Characterization of a 3-dehydroquinase gene from *Actinobacillus pleuropneumoniae* with homology to the eukaryotic genes *qa-2* and *qutE*. Mol. Micro. **11**: 273-280.

Lamb, H.K., Hawkins, A.R., Smith, M., Harvey, I.J., Brown, J., Turner, G. &

Roberts, C.F. (1990). Spatial and biological characterisation of the complete quinic acid utilisation gene cluster in *Aspergillus nidulans*. Mol. Gen. Genet. **223**: 17-23.

Lamb, H.K., Bagshaw, C.R. & Hawkins, A.R. (1991). In vivo overproduction of the pentafunctional arom polypeptide in *Aspergillus nidulans* affects metabolic flux through in the quinate pathway. Mol. Gen. Genet. **227**: 187-196.

Lamb, H.K., Roberts, C.F. & Hawkins, A.R. (1992a). A second gene (*qutH*) within the *Aspergillus nidulans* quinic acid utilisation gene cluster encodes a protein with a putative zinc cluster motif. Gene **112**: 219-224.

Lamb, H.K., Van Den Hombergh, J.P.T.W., Newton, G.H., Moore, J.D., Roberts, C.F. & Hawkins A.R. (1992b). Differential flux through the quinate and shikimate pathways: implications for the channelling hypothesis. Biochem. J. **284**: 181-187.

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*. Mol. Cell. Biol. **3**: 1609-1614.

Leblond, P., Redenbach, M. & Cullum, J. (1993). Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). J. Bact. **175**: 3422-3429.

Leskiw, B.K., Bibb, M.J. & Chater, K.F. (1991). The use of a rare codon specifically during development? Mol. Micro. **5**: 2861-2867.

Lewendon, A. & Coggins, J.R. (1983). Purification of 5-enolpyruvylshikimate 3-phosphate synthase from *Escherichia coli*. Biochem. J. **213**: 187-191.

Lim, C-J., Hwang, W., Park, E-H. & Fuchs, J.A. (1994). Cyclic AMP-dependent expression of the *Escherichia coli serC-aroA* operon. Biochim. Biophys. Acta **1218**: 250-253.

Limaoro, D., Avitabile, S., Capellano, M., Puglia, A.M. & Bruni, C.B. (1990). Cloning and characterisation of the histidine biosynthetic gene cluster of *Streptomyces coelicolor* A3(2). Gene **90**: 31-41.

Lin, Y-S., Kieser, H.M., Hopwood, D.A. & Chen, C.W. (1993). The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol. Micro. **10**: 923-933.

Løbner-Olesen, A. & Marinus, M.G. (1992). Identification of the gene (*aroK*) encoding shikimic acid kinase I of *Escherichia coli*. J. Bact. **174**: 525-529.

Lowe, D.A. & Westlake, D.W.S. (1971). Regulation of chloramphenicol synthesis in *Streptomyces* sp. 3022a DAHP synthase. Can. J. Biochem. **49**: 448-455.

MacNeill, D.J. (1988). Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. J. Bact. **170**: 5607-5612.

Mandelstam, J., McQuillen, K. & Dawes, I. (1982). Biochemistry of bacterial growth pp161-162.

Maskell, D. (1993). Cloning and sequencing of the *Haemophilus influenzae aroA* gene. Gene **129**: 155-156.

Maskell, D.J., Morrisey, P. & Dougan, G. (1988). Cloning and nucleotide sequence of the *aroA* gene of *Bordetella pertussis*. J. Bact. **170**: 2467-2471.

Mayes, S.R., Dalla Chiesa, M., Zhang, Z. & Barber, J. (1993). The genes *aroA* and *trnQ* are located upstream of *psbO* in the chromosome of *Synechocystis* 6803. FEBS Lett. **325**: 255-261.

Mazodier, P., Petter, R. & Thompson, C. (1989). Intergenic conjugation between *Escherichia coli* and *Streptomyces* species. J. Bact. **171**: 3583-3585.

Mendez C. & Chater, K.F. (1987). Cloning of *whiG*, a gene critical for sporulation of *Streptomyces coelicolor* A3(2). J. Bact. **169**: 5715-5720.

Millar, G. & Coggins, J.R. (1986). The complete amino acid sequence of 3-dehydroquinate synthase from *Escherichia coli* K-12. FEBS Lett. **200**: 11-17.

Mizosawa, S., Nichimora, S. & Seda, F. (1986). Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucl. Acids Res. **14**: 1319-1324.

Moore, J.D. & Hawkins, A.R. (1993). Overproduction of, and interaction within, bifunctional domains from the amino- and carboxy-termini of the pentafunctional Arom protein of *Aspergillus nidulans*. Mol. Gen. Genet. **240**: 92-102.

Moore, J.D., Lamb, H.K., Garbe, T., Servos, S., Dougan, G., Charles, I.G. & Hawkins, A.R. (1992). Inducible overproduction of the *Aspergillus nidulans* pentafunctional Arom protein and the type-I and -II 3-dehydroquinases from *Salmonella typhi* and *Mycobacterium tuberculosis*. Biochem. J. **287**: 173-181.

Moran, C.P., Lang, N., Lehrice, S.F.C., Lec, G., Stephens, M., Sonenshein, A.L., Pero, J. & Losick, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186**: 339-346.

Mousdale, D.M. & Coggins, J.R. (1984). Purification and properties of EPSP synthase from seedlings of *Phaseolus mungus*. Planta **160**: 78-83.

Mousdale, D.M., Campbell, M.S. & Coggins, J.R. (1987). Purification and characterization of bifunctional dehydroquinase-shikimate:NADP oxidoreductase from pea seedlings. Phytochemistry **26**: 2665-2670.

Muday, G.K., Johnson, D.I., Somerville, R.L. & Herrmann, K.M. (1991). The tyrosine repressor negatively regulates *aroH* expression in *Escherichia coli*. J. Bact. **173**: 3930-3932.

Muth, G., Nußbaumer, B., Wohlleben, W. & Pühler, A. (1989). A vector system with temperature-sensitive replication for gene disruption and mutational cloning in streptomycetes. Mol. Gen. Genet. **219**: 341-348.

Nakanishi, N. & Yamamoto, M. (1984). Analysis of the structure and transcription of the *aro3* cluster gene in *Saccharomyces pombe*. Mol. Gen. Genet. **195**: 164-169.

Niederberger, P., Prasad, R., Miozzari, G. & Kacser, H. (1992). A strategy for increasing an *in vivo* flux by genetic manipulations: the tryptophan system of yeast. Biochem. J. **287**: 473-479.

O'Connell, C., Pattee, P.A. & Foster, T.J. (1993). Sequence and mapping of the *aroA* gene of *Staphylococcus aureus* 8325-4. J. Gen. Micro. **139**: 1449-1460.

O'Gaora, P., Maskell, D., Coleman, D., Cafferkey, M. & Dougan, G. (1989). Cloning and characterisation of the *serC* and *aroA* genes of *Yersinia enterocolitica*, and construction of an *aroA* mutant. Gene **84**: 23-30.

Ogino, T., Garner, C., Markley, J.L. & Herrmann, K.M. (1982). Biosynthesis of

aromatic compounds: ^{13}C NMR spectroscopy of whole *Escherichia coli* cells. Proc. Natl. Acad. Sci. **79**: 5828-5832.

Padgette, S.R., Huynh, Q.K., Aykent, S., Sammons, R.D., Sikorski, J.A. & Kishore, G. (1988a). Identification of the reactive cysteines of *Escherichia coli* 5-enolpyruvylshikimate-3-phosphate synthase and their nonessentiality for enzyme catalysis. J. Biol. Chem. **263**: 1798-1802.

Padgette, S.R., Smith, C.E., Huynh, Q.K. & Kishore, G.M. (1988b). Arginine chemical modification of *Petunia hybrida* 5-enol-pyruvylshikimate-3-phosphate synthase. Arch. Biochem. Biophys. **266**: 254-262.

Padgette, S.R., Biest Re, D., Gasser, C.S., Eichholtz, D.A., Frazier, R.B., Hironaka, C.M., Levine, E.B., Shah, D.M., Fraley, R.T. & Kishore, G.M. (1991). Site-directed mutagenesis of a conserved region of the 5-enolpyruvylshikimate-3-phosphate synthase active site. J. Biol. Chem. **266**: 22364-22369.

Palmer, B.R. & Marinus, M.G. (1994). The *dam* and *dcm* strains of *Escherichia coli*-a review. Gene **143**: 1-12.

Patel, V.B. & Giles, N.H. (1979). Purification of the Arom multienzyme complex from *Euglena gracilis*. Biochim. Biophys. Acta **567**: 24-34.

Pigac, J. & Schrempf, H. (1995). A simple and rapid method of transformation of *S.rimosus* R6 and other streptomycetes by electroporation. Appl. & Env. Micro. **61**: 352-356.

Polley, L.D. (1978). Purification and characterization of 3-dehydroquinate hydrolase and shikimate oxidoreductase: evidence for a bifunctional enzyme. Biochim. Biophys. Acta **526**: 259-266.

Powell, H.A., Kerby, N.W., Rowell, P., Mousdale, D.M. & Coggins, J.R. (1992). Purification and properties of a glyphosate-tolerant 5-enolpyruvylshikimate 3-phosphate synthase from the cyanobacterium *Anabaena variabilis*. Planta **188**: 484-490.

Promega Corporation, USA. (1991). Promega Protocols and Applications Guide, 2nd edition.

Rubens, C.E., McNeill, W.F. & Farrar Jr., W.E. (1979). Transposable plasmid DNA

sequence in *Pseudomonas aeruginosa* which mediates resistance to gentamicin and four other antimicrobial agents. J. Bact. **139**: 877-882.

Rubin, J.L., Gaines, C.G. & Jensen, R.A. (1982). Enzymological action for herbicidal action of glyphosate. Plant Physiol. **70**: 833-839.

Sambrook, J., Maniatis, T. & Fritsch, I. (1989). Molecular cloning: a laboratory manual. Vol **1-3**.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain termination inhibitors. Proc. Nat. Acad. Sci. **74**: 5463-5467.

Schmidt, J., Bubunenko, M. & Subramanian, A.R. (1993). A novel operon organization involving the genes for chorismate synthase (aromatic biosynthesis pathway) and ribosomal GTPase center proteins (L11, L1, L10, L12:*rplKAJL*) in Cyanobacterium *Synechocystis* PCC 6803. J. Biol. Chem. **268**: 27447-27457.

Schulz, A., Krüper, A. & Amrhein, N. (1985). Differential sensitivity of bacterial 5-enolpyruvylshikimate-3-phosphate synthases to the herbicide glyphosate. FEMS Micro. Lett. **28**: 297-301.

Servín-González, L., Jensen, M.R., White, J. & Bibb, M. (1994). Transcriptional regulation of the four promoters of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). Microbiology **140**: 2555-2565.

Servos, S., Chatfield, S., Hone, D., Levine, M., Dimitriadis, G., Pickard, D., Dougan G., Fairweather, N. & Charles, I. (1991). Molecular cloning and characterization of the *aroD* gene encoding 3-dehydroquinase from *Salmonella typhi*. J. Gen. Micro. **137**: 147-152.

Shneier, A., Kleanthous, C., Deka, R., Coggins, J.R. & Abell, C. (1991). Observation of an imine intermediate on dehydroquinase by electrospray mass spectrometry. J. Am. Chem. Soc. **113**: 9416-9418.

Shneier, A., Harris, J., Kleanthous, C., Coggins, J.R., Hawkins, A.R. & Abell, C. (1993). Evidence for opposite stereochemical courses for the reactions catalysed by type I and type II dehydroquinases. Bioorganic & Med. Chem. Lett. **3**: 1399-1402.

Skeggs, P.A., Holmes, D.J. & Cundliffe E. (1987). Cloning of the aminoglycoside-

resistance determinants from *Streptomyces tenebrarius* and comparison with related genes from other actinomycetes. J. Gen. Micro. **133**: 1915-1923

Smith, D.D.S. & Coggins, J.R. (1983). Isolation of a bifunctional domain from the pentafunctional Arom enzyme complex of *Neurospora crassa*. Biochem. J. **213**: 405-415.

Smith, C.P. (1992). Personal Communication to the European Union Laboratories Without Walls Project on *Streptomyces*.

Sost, D. & Amrhein, N. (1990). Substitution of Gly-96 to Ala in the 5-enolpyruvylshikimate 3-phosphate synthase of *Klebsiella pneumoniae* results in a greatly reduced affinity for the herbicide glyphosate. Arch. Biochem. Biophys. **282**: 433-436.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**: 503-517.

Stalker, D.M., Hiatt, W.R. & Comai, L. (1985). A single amino acid substitution in the enzyme 5-enolpyruvylshikimate-3-phosphate synthase confers resistance to the herbicide glyphosate. J. Biol. Chem. **260**: 4724-4728.

Stallings, W.C., Abdel-Meguid, S.S., Lim, L.W., Shieh, H.-S., Dayringer, H.E., Leimgruber, N.K., Stegeman, R.A., Anderson, K.S., Sikorski, J.S., Padgette, S.R. & Kishore, G.M. (1991). Structure and topological symmetry of the glyphosate target 5-enol-pyruvylshikimate-3-phosphate synthase: a distinctive protein fold. Proc. Natl. Acad. Sci. **88**: 5046-5050.

Steinrücken, H.C. & Amrhein, N. (1980). The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate-3-phosphate synthase. Biochem. Biophys. Res. Comm. **94**: 1207-1212.

Steinrücken, H.C. & Amrhein, N. (1984). 5-enolpyruvylshikimate-3-phosphate synthase of *Klebsiella pneumoniae*: 2. Inhibition by glyphosate [N-(phosphonomethyl)glycine]. Eur. J. Biochem. **143**: 351-357.

Stragier, P., Parsot, C. & Bouvier, J. (1985). Two functional domains conserved in major and alternate bacterial sigma factors. FEBS **187**: 11-15.

Strauss, A. (1979). The genetic fine structure of the complex fine locus *aro3* involved

in early aromatic amino acid biosynthesis in *Saccharomyces pombe*. Mol. Gen. Genet. **172**: 233-241.

Stuart, F. (1990). An investigation of aromatic amino acid biosynthesis in *Streptomyces rimosus*. PhD Thesis. University of Glasgow.

Stuart, F. & Hunter, I.S. (1993). Purification and characterisation of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Streptomyces rimosus*. Biochim. et Biophys. Acta **1161**: 209-215.

Takahashi, H., Tanaka, K. & Shiina, T. (1988). Genetic constituent of *rpoD* gene homologues in *Streptomyces* strains. In Biology of Actinomycetes '88. Okami, Y., Beppu, T. & Ogawara, H. (eds.) Japan Scientific Societies Press, Tokyo. pp 58-63.

Tanaka, K., Shiina, T. & Takahashi, H. (1988). Multiple principal sigma factor homologues in eubacteria: identification of the "*rpoD* box". Science **242**: 1040-1042.

Tanaka, K., Shiina, T. & Takahashi, H. (1991). Nucleotide sequence of genes *hrdA*, *hrdC*, and *hrdD* from *Streptomyces coelicolor* A3(2) having similarity to *rpoD* genes. Mol. Gen. Genet. **229**: 334-340.

Tatum, F.M., Briggs, R.E. & Halling, S.M. (1994). Molecular gene cloning and sequencing and construction of an *aroA* mutant of *Pasteurella haemolytica* serotype A1. Appl. Env. Micro. **60**: 2011-2016.

Taylor, R.D. (1992). Purification and characterisation of the isocitrate dehydrogenase from *Streptomyces coelicolor* and cloning of its gene. PhD thesis. University of Glasgow.

Thompson, C.J., Ward, J.M. & Hopwood, D.A. (1980). DNA cloning in *Streptomyces*: resistance genes from antibiotic-producing species. Nature **286**: 525-527.

Thompson, C.J., Kieser, T., Ward, J.M. & Hopwood, D.A. (1982a). Physical analysis of antibiotic-resistance genes from *Streptomyces* and their use in vector construction. Gene **20**: 51-62.

Thompson, C.J., Ward, J.M. & Hopwood, D.A. (1982b). Cloning of antibiotic resistance and nutritional genes in *Streptomyces*. J. Bact. **151**: 668-677.

Thompson, C.J., Skinner, R.H., Thompson, J., Ward, J.M., Hopwood, D.A. & Cundliffe, E. (1982c). Biochemical characterization of resistance determinants cloned from antibiotic-producing streptomycetes. J. Bact. **151**: 678-685.

Thompson, C.J., Skeggs, P.A. & Cundliffe, E. (1985). Methylation of 16S ribosomal RNA and resistance to the aminoglycoside antibiotics gentamicin and kanamycin determined by DNA from the gentamicin producer *Micromonospora purpurea*. Mol. Gen. Genet. **201**: 168-173.

Vallee, B.L., Coleman, J.E. & Auld, D.S. (1991). Zinc fingers, zinc clusters, and zinc twists in DNA-binding domains. Proc. Nat. Acad. Sci. **88**: 999-1003

Van Den Hombergh, J.P.T.W., Moore, J.D., Charles, I.G. & Hawkins, A.R. (1992). Overproduction in *Escherichia coli* of the dehydroquinate synthase domain of the *Aspergillus nidulans* pentafunctional Arom protein. Biochem. J. **284**: 861-867.

Van Wezel, G.P., Buttner, M.J., Vijgenboom, E., Bosch, L., Hopwood, D.A. & Kieser, H.M. (1995). Mapping of genes involved in macromolecular synthesis on the chromosome of *Streptomyces coelicolor* A3(2). J. Bact. **177**: 473-476.

Vara, J., Lewandowska-Skarbek, M., Wang, Y-G., Donadio, S. & Hutchinson, R.C. (1989). Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). J. Bact. **171**: 5872-5881.

Vaughan, L.M., Smith, P.R. & Foster, T.J. (1993). An aromatic-dependent mutant of the fish pathogen *Aeromonas salmonicida* is attenuated in fish and is effective as a live vaccine against the Salmonid disease Furunculosis. Infect. Immun. **61**: 2172-2181.

Vögtli, M. & Cohen, S.N. (1992). The chromosomal integration site for the *Streptomyces* plasmid SLP1 is a functional $tRNA^{Tyr}$ gene essential for cell viability. Mol. Micro. **6**: 3041-3050.

Walker, G.E., Dunbar, B., Hunter, I.S., Nimmo, H.G. & Coggins, J.R. (1995). Evidence for a novel type of microbial 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus*, and *Neurospora crassa*. Microbiology : in press.

Walker, M.C., Ream, J.E., Sammons, R.D., Logusch, E.W., O'Leary, M.H., Somerville, R.L. & Sikorski, J.A. (1991). Structural requirements for PEP binding to EPSP synthase. Bioorganic & Med. Chem. Lett. **1**: 683-688.

Wang, Y.X., Jones, J.D., Weller, S.C. & Goldsbrough, P.B. (1991). Expression and stability of amplified genes encoding 5-enolpyruvylshikimate-3-phosphate synthase in glyphosate-tolerant tobacco cells. Plant Mol. Biol. **17**: 1127-1138.

Ward, J.M., Janssen, G.R., Kieser, T., Bibb, M.J. & Buttner, M.J. (1986). Construction and characterisation of a series of multicopy promoter-probe vectors for *Streptomyces* using the aminoglycoside phosphotransferase gene from Tn*5* as an indicator. Mol. Gen. Genet. **203**: 468-478.

Westpheling, J. & Brawner, M. (1989). Two transcribing activities are involved in expression of the *Streptomyces* galactose operon. J. Bact. **171**: 1355-1361.

Westpheling, J., Ranes, M. & Losick, R. (1985). RNA polymerase heterogeneity in *Streptomyces coelicolor*. Nature **313**: 22-27.

White, P.J., Millar, G. & Coggins, J.R. (1988). The overexpression, purification and complete amino acid sequence of chorismate synthase from *Escherichia coli* K-12 and its comparison with the enzyme from *Neurospora crassa*. Biochem. J. **251**: 313-322.

White, P.J., Young, J., Hunter, I.S., Nimmo, H.G. & Coggins, J.R. (1990). The purification and characterization of 3-dehydroquinase from *Streptomyces coelicolor*. Biochem. J. **265**: 735-738.

Whittington, H., Grant, S., Roberts, C.F., Lamb, H. & Hawkins, A.R. (1987). Identification and isolation of a putative permease gene in the quinic acid utilization (*qut*) gene cluster of *Aspergillus nidulans*. Curr. Genet. **12**: 135-139.

Wright, F. & Bibb, M.J. (1992). Codon usage in the G+C-rich *Streptomyces* genome. Gene **113**: 55-65.

Woese, C.R. (1987). Bacterial evolution. Micro. Rev. **51**: 221-271.

Zakrzewska-Czerwinska, J. & Schrempf, H. (1992). Characterization of an

autonomously replicating region from the *Streptomyces lividans* chromosome. J. Bact. **174**: 2688-2693.