Studies on the Gene Cluster for Oxytetracycline Biosynthesis from *Streptomyces rimosus*

A thesis submitted for the degree of
Doctor of Philosophy at the University of Glasgow

by

Sheila R. Garven

Institute of Biomedical and Life Sciences
Robertson Laboratory of Biotechnology
University of Glasgow
54 Dumbarton Road
Glasgow, G11 6NU

January 1995

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

Sheila R. Garven,
January, 1995.

Dedicated to Mom and Dad.

CONTENTS

Abbreviations

ap	ampicillin
APS	ammonium persulphate
ATP	Adenosine 5'-triphosphate
bp	base pairs
Ci	curie
CIAP	calf intestinal alkaline phospate
D	daltons
DEPC	diethyl pyrocarbonate
DMSO	dimethylsulphoxide
DMF	demethylformamide
DTT	dithiothreitol
dNTP	deoxynucleotide 5'-triphosphate
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytosine 5'-triphosphate
dGTP	deoxyguanosine 5'-triphosphate
dTTP	deoxythymidine 5'-triphosphate
DNA	deoxyribonucleic acid
ssDNA	single stranded deoxyribonucleic acid
dH_20	distilled water
E	core RNA polymerase
EDTA	ethylene diaminetetra-acetic acid (disodium salt)
EtBr	ethidium bromide
EtOH	ethanol
g	centrifugal force equivalent to gravitional acceleration
gmr	gentamicin
IPTG	isopropylthio-β-D-galactoside
hrs	hours
kb	kilo bases
knt	kilonucleotides
l	litres
M	molar

mins	minutes
Mr	molecular weight
OD	optical density
ORF	open reading frame
OTC	oxytetracycline
^{0}C	degrees celsius
nt	nucleotide
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pi	orthophosphate
PIPES	piperazine-NN'-bis-2-ethane sulphonic acid
PPCR	potential protein coding region
RBS	ribosome binding site
RNA	ribonucleic acid
mRNA	messenger ribonucleic acid
rRNA	ribosomal ribonucleic acid
tRNA	transfer ribonucleic acid
rpm	revolutions per minute
RT	room temperature
σ	sigma
S.	*Streptomyces*
SDS	sodium dodecylsulphate
ssi	second strand initiation
TEMED	NNN-N'-tetramethylethylenediamine
TES	N- tris (hydroxymethyl) methy-2-aminoethanesulphonic acid
transf.μg	transformants per microgram
TRIS	tris (hydroxymethyl) methylamine
TSB	tryptone soya broth
tsr	thiostrepton
U	units
UV	ultraviolet
V	volts
v/v	volume by volume
w/v	weight by volume
W	watts
Xgal	5-bromo-4-chloro-3-indoyl-β-galactoside
YEME	yeast extract-malt extract

Acknowledgements

Would like to thank Iain Hunter for his guidance over the last two years.

A special thank you to Helena for her friendship and taking the time to struggle through the manuscript.

Marion was the one you could always count on to kept the lab running smoothly, thank you, and to the Prep. room ladies, for the expedient return of autoclaved tips and solutions, they were always so cheerful!

I would like to thank all my relatives in the UK, some of which I'm sure I still haven't met, for their support, blankets, tables and the list goes on. It was wonderful to get to know my "Scottish cousins" after all these years, and those bi-weekly treks up to Gran's. I will miss the stories she would tell over afternoon tea. A special thank you also goes to Walter for generously sharing "the flat with the e x p a n d a b l e walls".

To all the new friends I've made, especially Sheila and Angus and Colin, for introducing me to a touch of Scotland!

Lastly, and most importantly, I would like to thank my family for all their encouragement and love. The weekly phone calls were a godsend without which I would not have kept in touch with reality. The monthly letters from Jennifer and Lauren were a welcome sight and yes Jennifer I am coming home soon!

Summary

Prior to the work described in this thesis, a great deal of the oxytetracycline cluster (*otc*) had been sequenced and the putative functions of deduced gene products assigned, based on similarity to known gene products in the databases.

The 3.0 kb *Kpn*I$_{23}$-*Mlu*I$_{26b}$fragment located between the *otc*D and *otc*X loci was sequenced. Three open reading frames with the same direction of transcription were found, *otcD*-ORF3, -ORF4 and -ORF5. *OtcD*-ORF3 has good similarity to oxidoreductases reported previously in other polyketide gene clusters. The biosynthesis of OTC requires three oxidoreductase steps, one at position C9 which removes the keto function, one at position C6 in the conversion of 4-keto-ATC to ATC and another at the last step [OTC dehydrogenase]. A ketoreductase has been reported previously within the *otcY* locus that shows stronger similarity to *act*III of *S. coelicolor*. *OtcD*-ORF3 has been tentatively assigned as catalysing the last step of the pathway. *OtcD*-ORF4 shows good end-to-end similarity to a coding region of unknown function within the gene cluster for daunorubicin biosynthesis in *S. griseus*. *OtcD*-ORF5 is clearly defined as a potential protein coding region that shows good similarity to a small region of an open reading frame within the *B. subtilis* genome which again has unknown function. The *B. subtilis* coding region is very hydrophobic and carries a series of repeat sequences, the significance of which is not yet understood.

Preliminary transcriptional analysis of this 3.0 kb *Kpn*I$_{23}$-*Mlu*I$_{26b}$ fragment using low-resolution S1 mapping has revealed a single transcript which encompasses *OtcD*-ORF3 and *OtcD*-ORF4. A stable stem-loop immediately after the translational stop codon of *OtcD*-ORF4 was located with high-resolution mapping. A potential promoter sequence was located immediately preceding *OtcD*-ORF5, however no transcript was detected with low-resolution mapping at the time point when the RNA was isolated.

Previous attempts to disrupt gene products in the *otc* cluster using suicide-type vectors were unsuccessful possibly due to poor recombination efficiencies. Various aspects of the biology the *S. rimosus* bacterium were investigated. DNA isolated from a *dcm*$^-$ *E. coli* strain transformed at a much higher rate than DNA

isolated from *E. coli* strains which were capable of full methylation or *dam*$^{-}$. The technique to transform protoplasts was improved by electrotransformation. 'Electrotransformants' regenerate much faster and at a higher frequency than those prepared by more conventional methods. Another method of DNA transfer, conjugation between *E. coli* and some *Streptomyces* strains has been very successful. However, attempts at conjugal transfer between *E. coli* and *S. rimosus* were unsuccessful.

A strategy to further analyse the *OtcD*-ORF3 gene product was undertaken by attempting gene disruption and replacement. The gentamicin-resistance gene (*gmr*) was initially inserted into the *Kpn*I$_{26}$ site of the reductase gene which was carried on a plasmid. When the construct was introduced into *S. rimosus,* a homologous recombination event between the *otc* DNA carried on the plamid and homologous regions on the chromosome allowed integration of this DNA and produced gentamicin-resistant colonies. The delivery vector has an unstable *Streptomyces* replicon and was lost through a non-selective round of growth. Hybridisation analysis confirmed the integration event but a high degree of amplification was also observed. Further study showed the disrupted genome also contained deletions of the *otc* cluster including the *otrB* gene. A second integration attempt resulted in complete loss of the *otc* cluster. The *otc* cluster is situated towards one end of the linear chromosome of *S. rimosus* between large direct repeats. Amplifications and sequential chromosomal deletions have been reported in other *Streptomyces* strains. Typically, the nonessential genes are lost in favour of the primary metabolite genes. Due to the location of the cluster within the chromosome the integration strategy developed was not successful. However relocating the cluster could make this strategy feasible.

Introduction

1.1 General Introduction

Streptomyces are producers of more than 60% of the known antibiotics. The commerical importance of this has lead to intense investigations into how these antibiotics are produced. The work described in this thesis investigates various aspects of the oxytetracycline gene cluster from *Streptomyces rimosus*. A general introduction of the Gram-positive genus *Streptomyces*, it's complex life cycle and it's regulation and differentiation, will be discussed. Particular emphasis will be given to the co-ordination of production of antibiotics within this life cycle.

1.2 Medical and Commerical Interests in *Streptomyces*

Secondary metabolites, produced primarily by streptomycetes and fungi, are not essential for the growth of the organism but may give a competitive advantage, although in some cases it is not clearly evident. Secondary metabolites show great diversity in chemical structures and biological activities. The majority of secondary metabolites are biologically-active and based on their function, have been categorised into two major groups; the first group is seen to serve a purpose within the producing species and the second, to act on targets external to the producer. Broadly speaking, the intrinsic functions of the first group involve mediation of growth, reproduction or differentiation. Numerous examples fit this pattern and are reviewed in Vanek *et al.*, (1981). The secondary metabolites which act on targets external to the producing species include antibiotics, mycotoxins, insecticides and herbicides, all of which have found uses in human society.

Streptomyces have produced a great deal of industrial and medical interest because they produce a multitude of chemically-distinct secondary metabolites such as aminoglycosides, macrolides, tetracyclines, polyethers, ansamycins, β-lactams and oligopeptides. The major role of antibiotics, in both human and animal health, is against bacterial infection. The mechanisms of action of the secreted antibiotics on the infective bacteria include; inhibition of protein (e.g. tetracyclines and macrolides), and interference of RNA (e.g. rifamycins) and DNA (e.g. anthracylines) synthesis (Lewin; 1994). Anthracylines, (such as tetracenomycin and daunorubicin) have also been used as cytostatic agents in the treatment of certain kinds of tumours (Ye *et al.*, 1993). As well as being anti-infective agents, antibiotics have also been used as food preservatives to maintain

freshness in fish, meat and poultry by incorporation into ice during shipping (Crueger and Crueger; 1989).

Antibiotics have also been used in animal husbandry, both as growth promoters and in veterinary medicine. Animal feed is processed more efficiently by the animal's digestive system if the antibiotic additive is used at subtherapeutic concentrations. Weight gain may also be accelerated (Crueger and Crueger; 1989). The cause of this growth promotion may be traced to changes in the microflora of the gastrointestinal tract. Originally, large quantities of therapeutically-useful antibiotics such as penicillins, tetracyclines and erythromycins were added to the feed but such extensive use of common antibiotics enhanced the risk of rapid development of antibiotic resistance in consumers of animal products. Government regulations were established therefore aimed at eliminating parallel use of antibiotics in human medicine and in animal feed. Specific antibiotics were subsequently assigned for use solely in veterinary work. Bacterial resistance to a wide range of antibiotics has heightened the need for either the emergence of previously undiscovered natural antibiotics or the production of synthetic molecules, which are non-toxic to the consumer (reviewed in Crueger and Crueger; 1989).

A number of non-antibiotic but bioactive metabolites which have emerged from *Streptomyces* in modern screening procedures may be of clinical importance. An example is triacsin, an inhibitor of acetyl-CoA synthetase, isolated from *Streptomyces* sp. SK-1894, capable of potentiating platelet-activating factor production of A23187-treated rat polymorphonuclear leukocytes (Omura; 1992). These non-antibiotic metabolites could prove to be more useful than antibiotics in the future as their spectrum of uses grow. At present there are more new discoveries of these new kinds of metabolites than of antibiotics.

Besides the above mentioned major research areas, the last few years has seen the emergence of an interest in *Streptomyces* as a possible host organism for the production and secretion of recombinant proteins. The free-living saprophytic *Streptomyces* species are well-known producers of a great variety of extracellular enzymes, including amlyases, cellulases, xylanases, phosphatases, lipases, proteinases and nucleases, as well as numerous secreted proteinaceous enzyme inhibitors (Molnár; 1994). Their simple, Gram-positive type, cell wall structure allows the exported proteins to be localised directly into the culture medium instead of being trapped in the periplasmic space, in an insoluble form, as in *E. coli* (Malek *et al.*, 1990). *Streptomyces* cultures remain metabolically-active

in the stationary phase and continue to synthesise and secrete secondary metabolites in large amounts. The natural ability of *Streptomyces* strains to secrete extracellular enzymes derived from their heterotrophic existence is, however, likely to limit the commerical exploitation of heterologous protein secretion by these microbes. *Streptomyces lividans* is the most commonly used host, due primarily to the relatively low amount of extracellular and intracellular proteases (Aretz *et al.*, 1989). Limited success has already been achieved in the production of large recombinant proteins and work is actively underway to produce a *Streptomyces* host which is free of the major protease products observed in the culture media (Malek; pers. comm.). Classified as a 'safe' microrganism, the handling of recombinant streptomycete systems makes it an attractive host organism well worth further investigation.

1.3 The *Streptomyces* Life Cycle

In contrast to the extensively-studied bacterium *Escherichia coli*, which has a virtually synonymous cell cycle and life cycle, certain bacteria have a life cycle which goes beyond the cell cycle. This is illustrated with *Streptomyces* species. *Streptomyces* undergo a complex lifestyle, the initial phase of which is growth as substrate mycelium. Growth occurs mainly by cell-wall extensions at the hyphal tips. The hyphal compartments (especially the tip cells) contain many copies of the genome with very few sites of cell division. This produces a coherent multicellular mat of branching hyphae on a solid medium. As the colonies grow, the parts farthest from the advancing edge may accumulate various kinds of storage materials such as glycogen, lipids, and polyphosphate (Brana *et al.*, 1986). These parts of the colony probably grow more slowly, either because of nutrient limitation or because conditions have become inhibitory to growth. It is at this point that the colony reaches its first stationary phase.

The second stage of growth begins with the formation first of an aerial mycelium growing (both physically and nutritionally) on the substrate mycelium. Aerial growth is accompanied by physiological changes in the colony, typified by the production of a range of secondary metabolites that include a multitude of antibiotics, extracellular enzymes (described in the previous section) and various polymeric storage compounds such as glycogen. Aerial growth follows a short period of reduced macromolecular synthesis (Granozzi *et al.*, 1990) and seem to involve the reuse of material first assimilated into the substrate mycelium either as macromolecules such as DNA and proteins or as storage compounds (Brana *et al.*, 1986). Thus, as aerial hyphae grow, many

cells in the substrate mycelium die (Wildermuth; 1970). The growth of aerial hyphae is thought to be driven by the osmotic potential generated through the solubilisation of macromolecules such as glycogen (Chater; 1993).

The extension aerial hyphae growth stops eventually as the cell enters the second stationary phase. Regularly-spaced and simultaneously-formed crosswalls subdivide the orderly fragmentation of individual aerial hyphae into coiled chains of hydrophobic spores. The sporulation septa typically consist of two membrane layers separated by a double layer of cell-wall material, an arrangement which eventually permits the separation of adjacent spores. During the rounding up of spores, the spore wall thickens, and in most species, a pigment is deposited in the wall. Spore chains may consist of many tens of spores, each spore compartment containing one nucleoid (Figure 1.1). In adverse conditions, the spores will mature with increasing grey pigment and be released into the air in order to colonise more favourable habitats and begin a new micro cycle (reviewed in Chater; 1993).

1.4 Regulation of Differentiation in *Streptomyces*

The complex regulatory mechanisms governing the temporally- and spatially-arranged processes of morphogenesis and biochemical differentiation in *Streptomyces,* and the closely-linked switch from primary to secondary metabolism have received a great deal of interest. Many scientists have focussed on specific aspects of the wide range of secondary metabolic processes. However, to understand fully the switch from primary to secondary metabolism, the regulation of primary metabolism also needed to be investigated. The most intensively studied strain, *S. coelicolor* A3(2) provides examples of global regulation. Two series of mutants have been isolated, bald (*bld*) mutants, which do not form aerial mycelium and do not produce, with the exception of *bldC*, any antibiotics (Piret and Chater; 1985), and white (*whi*) mutants, which form aerial mycelium and produce antibiotics but do not form pigmented, mature spores (Chater; 1989b). Such mutant phenotypes suggest that antibiotic production and spore formation may be regulated by common molecular mechanisms.

At least seven different classes of bald mutants fail to produce aerial mycelium. The rarity of *S. coelicolor* mutants defective only in aerial mycelium development (only *bldC)*, suggests that either very few biochemical features are peculiar to, and necessary for, aerial hyphae to grow, or there is genetic

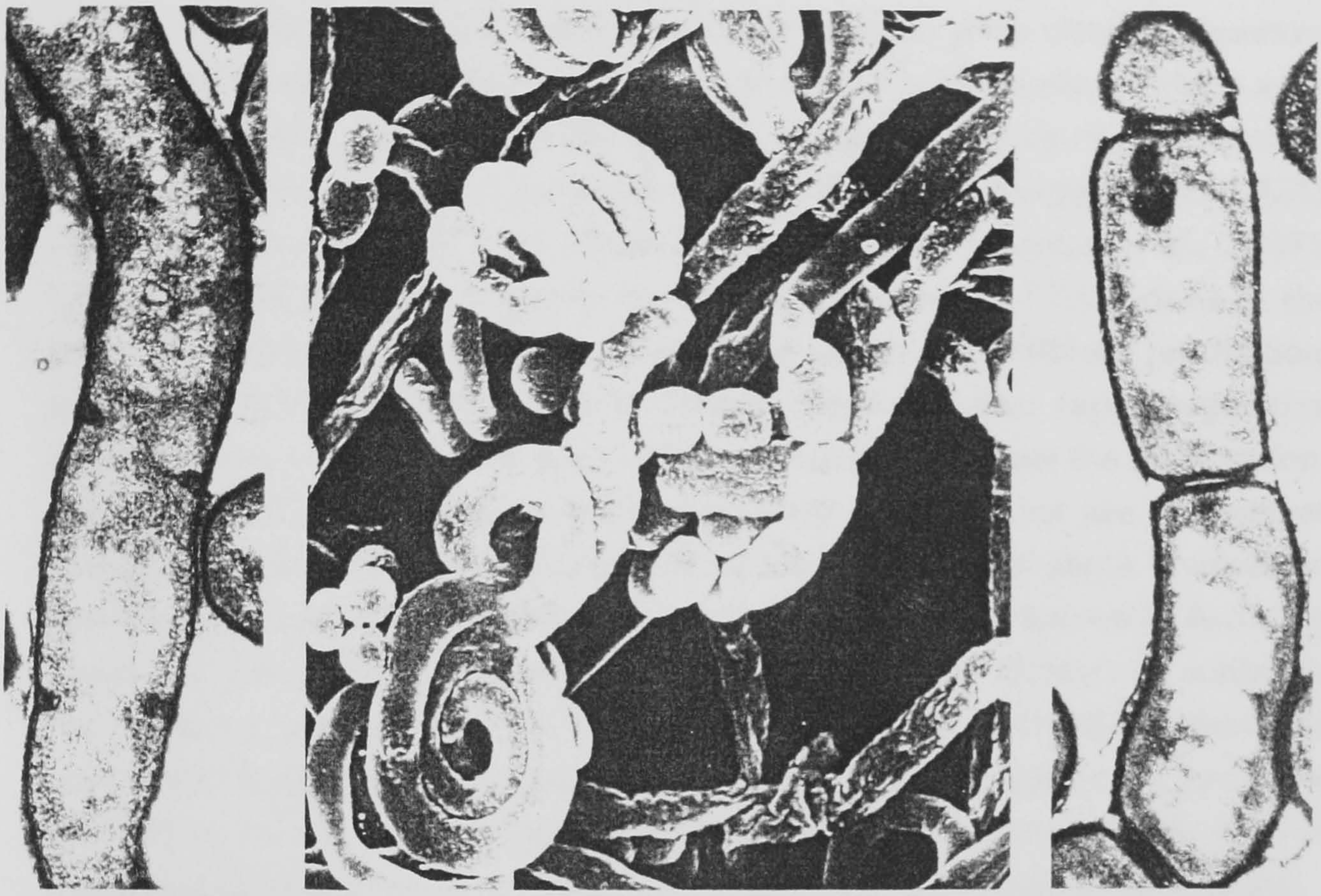


Figure 1.1
Spore chains developing on aerial hyphae. The central panel is a scanning electron micrograph of aerial hyphae *S. lividans* before, during and after the process of sporulation septation. The outer panels are thin sections of *S. coelicolor* at later stages. Photographs were taken from Chater (1989a).

redundancy for such features so that mutants are not isolated readily by normal genetic procedures. Mutants *bldA,D,G* and*H* have been termed 'conditionally bald'. They are capable of producing normal sporulating aerial mycelium when grown on mannitol (instead of the normal glucose source). However, only *bldH* mutants recover antibiotic production in these conditions (reviewed in Chater; 1989a). The *bldA* loci of five independent *bldA* mutants were cloned, sequenced and found to carry mutations within 16 bp of each other, none of which were within a potential protein coding region. Computer analysis revealed that all five encode a putative tRNA needed for the translation of a leucine codon (UUA), rarely seen in the G+C rich *Streptomyces* genome (Lawlor *et al.,* 1987). Circumstantial evidence suggests that, since there are no TTA codons in the known genes involved in normal growth, sporulation and antibiotic production, the *bldA*-specified tRNA, which is mainly produced after rapid vegetative growth is completed (Lawlor *et al.,* 1987), is needed to translate the UUA codon-containing mRNA for one or more regulatory proteins that are themselves directly or indirectly involved in activating the expression of genes involved in sporulaton and antibiotic production (Chater; 1990). The few known TTA-coding genes are nearly all resistance or regulatory genes from clusters of antibiotic biosynthetic genes (Leskiw *et al.,* 1991), including *act*II-ORF4 from the actinorhodin cluster (Fernández-Moreno *et al.,* 1991). This theory has lost some credibility in recent years however when Bibb (1990) showed that *bldA* is expressed during growth. It's utility of the tRNA may be subject to maturation.

Growth next to a wild-type strain can cause a number of the *bld* mutants to produce aerial mycelium, providing evidence for a diffusible extracellular factor. A possible candidate may be the SapB protein. Spore associated proteins (Saps) are associated with the surface of aerial hyphae and spores, serving a variety of functions including providing directionality in hyphal growth, preventing the desiccation of aerial hyphae and providing a hydophobic surface to the spores (Guijarro *et al.,* 1988). The *sapB* gene has proved difficult to clone. However, antibodies were made and two mutants were found that did not produce SapB on rich medium. These mutants also did not produce aerial mycelium. Subsequent investigation with the previously-identified *bld A,B, C, D, G, H,* and *I* mutants revealed that they failed to produce SapB. A colony overproducing SapB grown in close proximity to these *bld* mutants resulted in the production of aerial mycelium (Willey *et al.,* 1991). Further complementation studies using pairwise combinations of different *bld* mutants grown near to each other on rich medium showed extracellular complementation, such that one of the pairs produced both aerial mycelium and SapB (Willey *et al.,* 1991). Four

complementation groups have been found and are interpreted in terms of a cascade of extracellular signals reminiscent of those exchanged during fruiting body development in *Myxococcus xanthus* (Kaiser *et al.,* 1989). Following this hypothesis, the *bld* mutants might conceivably be defective primarily in elements of this cascade (Wiley *et al.,* 1991).

Formation of aerial hyphae in *S. coelicolor* is accompanied by a change in the surface of the colony, which adopts a hairy white appearance. When sporulation takes place, this aerial mycelium becomes grey, because of a spore-associated pigment. Thus mutants (*whi*) blocked in sporulation retain the immature white colony colour even after prolonged incubation (Hopwood *et al.,* 1970). The process of sporulation of aerial hyphae involves at least eight sporulation-specific genetic loci (the *whi* genes). The switch from continued extension of aerial hyphae to their development into chains of spores is controlled by a σ factor, σ^{whiG}. Sequence analysis confirmed the general organisation typical of sigma factors (Chater *et al.,* 1989b). Inactivation of *whiG* leads to complete failure of the sporulation process. Conversely, a high-copy number of *whiG* leads to abundant sporulation, and causes the spores to be produced even in the substrate hyphae growing into the agar and in cultures growing in liquid medium (Chater *et al.,* 1989c). Throughout its length σ^{whiG} closely resembles another factor, σ^{D}, of *B. subtilis,* especially in the so-called 2.4 and 4.2 regions (Chater *et al.,* 1990), which in other sigma factors appear to make base-specific contacts with cognate promoters (Helmann *et al.,* 1988). This suggests that the σ^{whiG} and σ^{D} RNA polymerase holoenzymes might recognise similar promoters. Mutants of *B. subtilis* deficient in σ^{D} remain filamentous and non-motile, very similar phenotypically to the *whiG* mutants, which are unable to develop beyond the aerial hyphal stage. This family of σ factors may play a role in the dispersal of bacteria during nutrient limitation.

The *whiE* locus consists of a gene cluster of seven open reading frames, most encoding recognisable homologues of proteins involved in the biosynthesis of polyketide antibiotics such as actinorhodin, oxytetracycline, tetracenomycin and granaticin (Hopwood and Sherman; 1990). Disruption of the wild-type *whiE* locus with cloned DNA generated from the mutant phenotype resulted in non-pigmented spores suggesting that the gene product of the*whiE* locus is responsible for spore pigment synthesis. The remarkable number of conserved genes with the biosynthesis of polyketide antibiotics and spore pigments suggests that an ancient evolutionary split occurred between the two classes of

genes before the antibiotic clusters diverged from each other (Blanco *et al.*, 1993). The regulatory mechanisms probably also diverged at about the same time.

The remainder of this chapter will focus on the large group of secondary metabolites, the antibiotics. Particular emphasis will be placed on the polyketide class of antibiotics, culminating in a discussion of the specific gene functions of oxytetracycline cluster produced by *S. rimosus*.

1.5 The Architecture of Antibiotic Biosynthetic Gene Clusters

It is a common feature of antibiotic biosynthetic and antibiotic resistance genes that they are situated adjacent to each other as a cluster on the chromosome, implying that they have evolved as a unit (Maplestone *et al.*, 1992). An exception to this rule is the genes encoding methylenomycin production. Although the production genes are clustered, they are very unusual in being plasmid-located; in two different hosts they are found on different extrachromosomal elements, one of which, SCP1, is a 350 kb linear plasmid in *S. coelicolor* (Kirby and Hopwood; 1977). The pathways for the biosynthesis of antibiotics are complex and consist of around 10-30 steps, requiring a corresponding number of gene products. Therefore, the clustering of structural genes increases the probabilities of them being passed on as a single unit to subsequent generations or to other species (Stone and Williams, 1992). Evidence so far suggests that gene clusters for antibiotic biosynthesis will rarely be found to represent single transcription units (Chater; 1993). Instead, the gene clusters that have been studied are organised as groups of operons, some monocistronic and others polycistronic, and some transcribed from DNA in each direction. Examples are seen for the biosynthesis of methylenomycin in *S. coelicolor* A3(2) (Chater *et al.*, 1989b), streptomycin in *S. griseus* (Piepersberg *et al.*, 1988) and tylosin in *S. fradiae* (Hershberger *et al.*, 1989). The gene clusters for actinorhodin and methylenomycin in *S. coelicolor* A3(2) are shown in figure 1.2 and discussed in detail below.

For the actinorhodin cluster, six possible transcription units were identified by *in vivo* disruption of the actinorhodin cluster and S1 nuclease mapping (Fernández-Moreno *et al.*, 1991). Mutational cloning, using *att*-deleted ϕC31 derivative phage, to integrate into the chromosome *via* a homologous DNA fragment cloned in the vector, revealed the approximate location of transcripts (Fernández-Moreno *et al.*, 1991). S1 mapping was then used to define the direction and length of transcripts. A complete description of all mutant classes

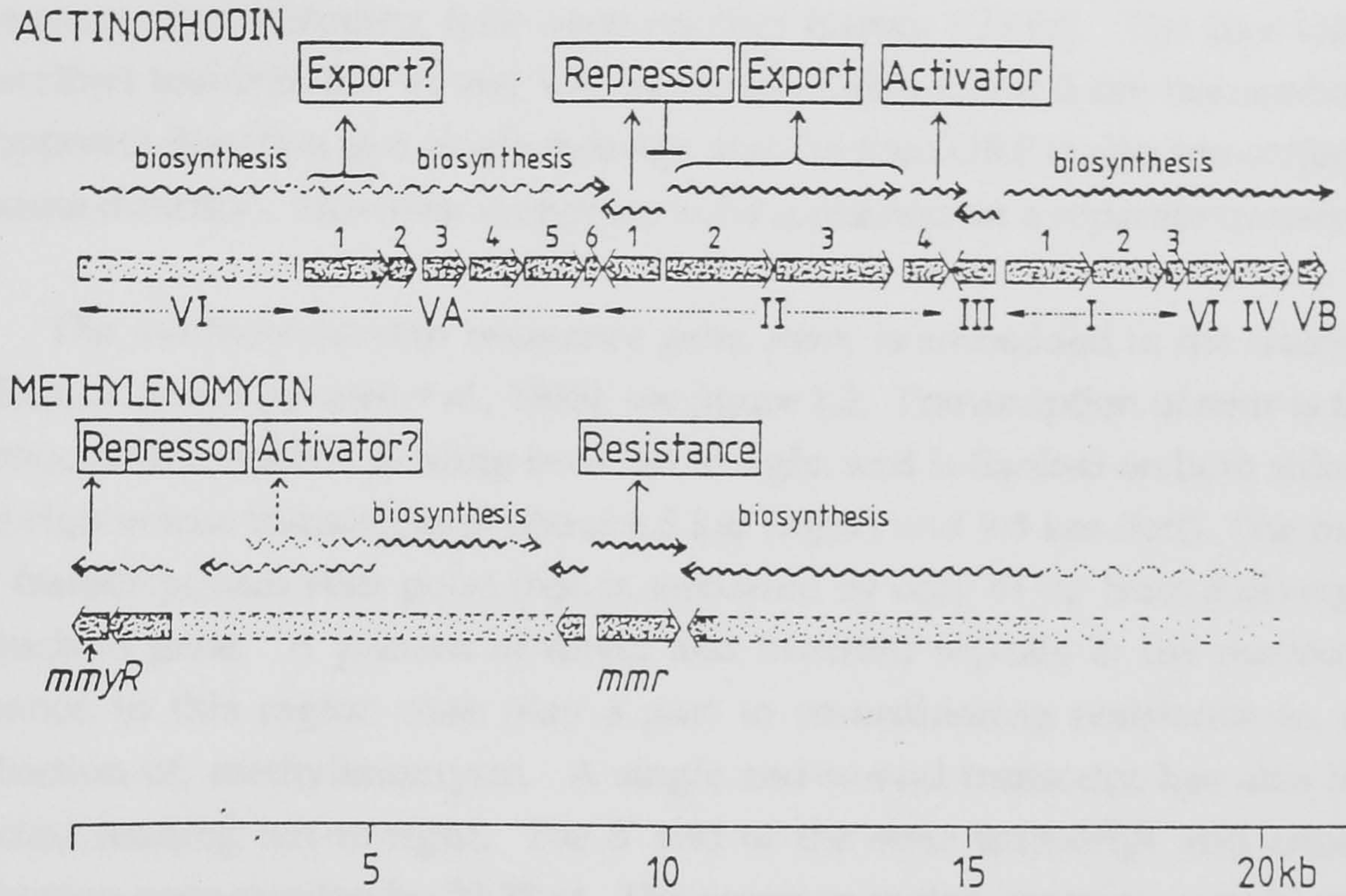


Figure 1.2
The organisation of two gene clusters for antibiotic production in *S. coelicolor* A3(2). Open reading frames are represented by stippled arrows and the position and extent of transcription is represented by wavy lines. Mutant classes in the actinorhodin cluster are represented by roman numerals. Data taken from Chater, 1992.

mentioned here can be found in section 1.7.2. On the right side, reading from left to right, is a large (6.5 knt) polycistronic transcript corresponding to genes of mutant classes I, VII and IV (as well as V_B). A small transcript (0.8 knt) corresponding to the gene of class III was found to read in the opposite direction. A long transcript (~9 knt) covers the two genes (VI and VA) on the left side, reading towards the centre of the cluster. The class II genes appear to contain three transcripts encoding four open reading frames (ORFs). The first ORF is transcribed towards the VI and VA transcript, ORFs 2 and 3 are transcribed in the opposite direction as a single message and the final ORF is also transcribed in the same direction. However it appears to be contained on a separate transcript.

The methylenomycin resistance gene, *mmr,* is embedded in the cluster of production genes (Chater *et al.,* 1985), see figure 1.2. Transcription of *mmr* is from a monocistronic mRNA reading from left to right, and is flanked on both sides by large rightwards transcripts of about 6.5 knt (right) and 9.5 knt (left). The major *mmr* transcriptional start point (*tsp*) is separated by only 81 bp from a divergent production gene. A pattern of direct and inverted repeats in the nucleotide sequence in this region may play a part in co-ordinating resistance to, and production of, methylenomycin. A single end-to-end transcript has also been detected reading left-to-right. The 3' end of the *mmr* transcript and another production gene overlap by 20-30 nt. The common region contains a stable stem-loop structure and has shown *in vitro* termination activity in at least one direction (Neal and Chater; 1991).

While a gene for actinorhodin resistance is presumed to be present in the *act* cluster, it has not been localised. On the other hand, it is clear that resistance to methylenomycin is encoded by a gene placed approximately in the centre of the cluster and is transcribed divergently from a presumed biosynthetic transcript. This feature is seen in several other resistance genes; *ermE* gene of the eythromycin producer *Saccharopolyspora erythraea* (Bibb *et al.,* 1986); the *aph* gene of the neomycin producer, *S. fradiae* (Janssen *et al.,* 1989); and the *sph* gene of the hydroxystreptomycin producer, *S. glaucescens* (Vögtli *et al.,* 1987). The possible significance of such an arrangement may be to ensure that resistance is always operational when the organism is exposed to its own product, that is, when synthesis of the biosynthetic gene products is occurring, and may even play an intimate part in the complex regulation of antibiotic production (Chater *et. al* 1989a).

1.6 Pathway-Specific Regulation

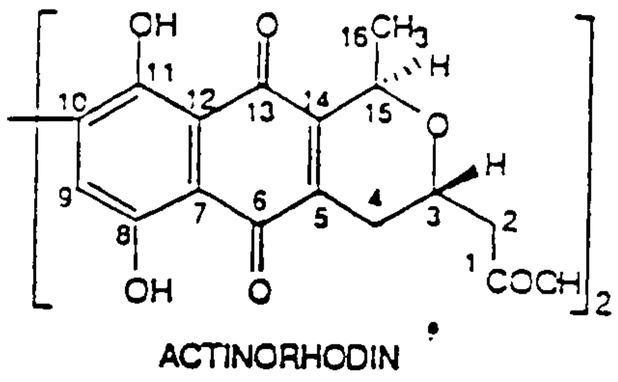
Insight into the regulation of trancription of closely-linked genes has been obtained from studies of mutations within clusters of production genes that interfere with the functions of most or all of the other genes in the region. For instance, in the *act* cluster, *actII* seems to encode a positive regulator, since point mutations in *actII* appear to abolish (or at least reduce significantly) transcription of the other *act* genes, while extra copies of the *actII* region enhance production of actinorhodin (Malpartida *et al.*, 1986). In contrast, studies of pleiotropic mutations have provided information about the regulation by unlinked genes, since these mutations affect antibiotic production as well as other characteristics that are known to be developmentally-regulated, for example the formation of aerial mycelia and spores. As described above (Section 1.5), *BldA* mutants of *S. coelicolor* strains exhibit defects in both the formation of aerial mycelia and the production of antibiotics. Lawlor *et al.*, (1987) suggested that these properties may be mediated by the *bldA* product in the wild-type strain *via* a novel type of translational control. The *S. coelicolor afsR* gene seems to bridge the actions of *actII* and *bldA*, since *afsR* can modulate the function of two antibiotic clusters in *S. coelicolor,* actinorhodin (*act*) and undecylprodigiosin (*red*) (and possibly A-factor production [Horinouchi *et al.*, 1984]), but has no proven role in development (Horinouchi *et al.*, 1990).

The *afsR* gene of *S. coelicolor,* like *act*II-ORF4, can stimulate the transcription of the *act* genes. The *afsR* gene, in turn, appears to be induced by the kinase activity of *afsK*, located immediately downstream of *afsR* (Horinouchi and Beppu; 1992). This was demonstrated when purified AfsR became radiolabelled when incubated with [γ-^{32}P] ATP and cell-free extracts of *S. coelicolor* or *S. lividans* (Hong *et al.*, 1991). This suggests that an uncharacterised extracellular signal is perceived by AfsK, which is stimulated to phosphorylate, and thereby activate, AfsR. Similar regulatory systems have been detected in *S. griseus* (*strR*) (Piepersberg *et al.*, 1988); *S. hygroscopicus* (*brpA*) (Anzai *et al.*, 1987); and *S. peucetius* (*dnrR*) (Stutzman-Engwall *et al.*, 1992). These systems exhibit the characteristic features of two-component regulatory systems. Such systems, made up of genes for sensor (protein kinase) and response regulator (transcriptional activator) proteins, are common in bacteria and reflect the ability of an organism to orchestrate multifaceted responses to specific environmental changes (Stock *et al.*, 1989).

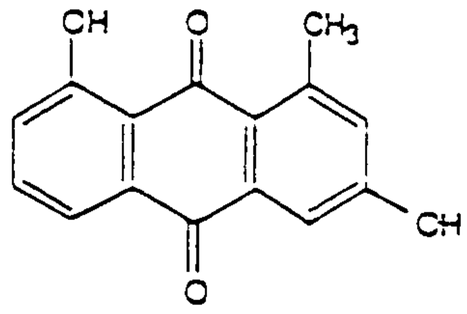
1.7 Polyketide Biosynthesis

Polyketides represent a large class of secondary metabolites. This very diverse group include flavonoids from plants, aflatoxins from fungi and hundreds of compounds of different structures that exhibit antibacterial, antifungal and antihelminthic properties (figure 1.3). The formation of polyketides was first elucidated by Birch (1967). The process is catalysed by polyketide synthases (PKS) in a manner that is conceptually similar to the biosynthesis of long-chain fatty acids catalysed by the fatty acid synthases (FAS), found in all organisms. Comprehensive reviews can be found in Hopwood and Sherman (1990) and Katz and Donadio (1993). The known mechanism of FAS will be described briefly to provide a basis for the discussion of polyketide biosynthesis.

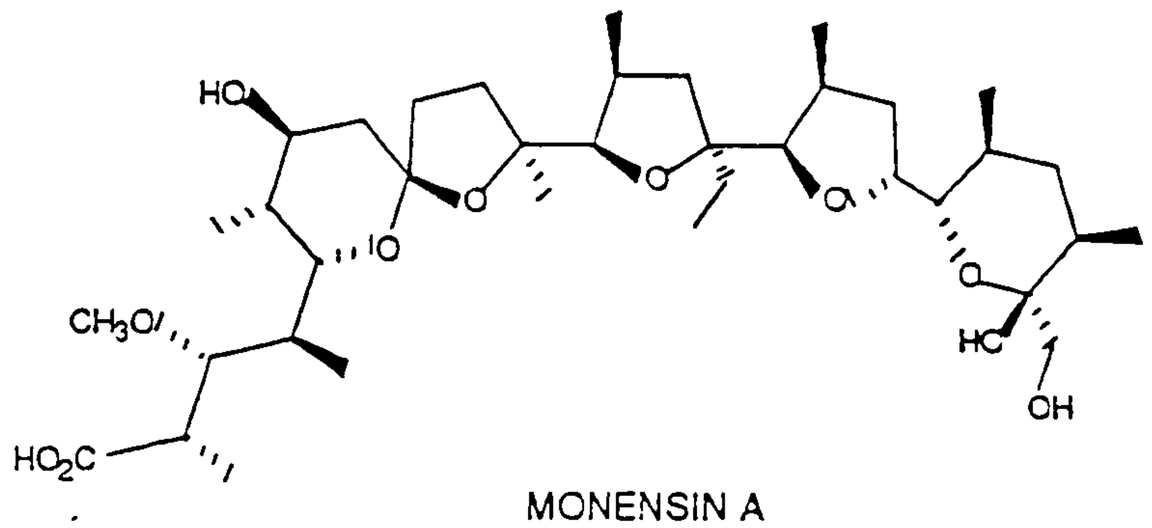
The essence of fatty acid biosynthesis (and polyketide biosynthesis) is a repeated decarboxylative condensation in which simple carboxylic acids are joined in a head to tail fashion to produce a carbon chain of 6-50 atoms in length, with each building unit contributing 2 carbon atoms to the growing chain. The general reactions carried out by the FAS are outlined in figure 1.4. The starter unit, acetyl-CoA, is condensed with a malonyl-extender unit, to yield a 4 carbon intermediate and CO_2. For condensation to occur, an acyltransferase moves the starter acetyl unit from the 4'-phosphopantetheine thiol of CoA to that of the acyl carrier protein (ACP), also part of the FAS. The acetyl-ACP is then transferred to the active site cysteine of the β-ketoacyl ACP synthase (KS), the condensing enzyme, leaving the thiol of the ACP free to receive a malonyl extender unit, transferred from malonyl-CoA by malonyltransferase (MT). The resulting ACP-bound β-ketoacyl chain undergoes three successive processing steps; a β-ketoreduction, a dehydration and an enoylreduction catalysed by β-ketoreductase (KR), dehydratase (DH), and enoylreductase (ER), respectively. The growing chain returns to the thiol of the KS and a new cycle begins. After the correct number of cycles, the completed carbon chain leaves the FAS, either by transfer back to CoA or by hydrolysis to the free acid catalysed by a thioesterase (TE). Therefore, the FAS complex, along with the 4'-phosphopantetheine containing ACP, is made up of eight enzymatic processes, the condensing enzyme β-ketoacyl ACP synthase (KS), malonyltransferase (MT), β-ketoreductase (KR), dehydratase (DH), enoylreductase (ER), thioesterase (TE) and two acyltransferases.



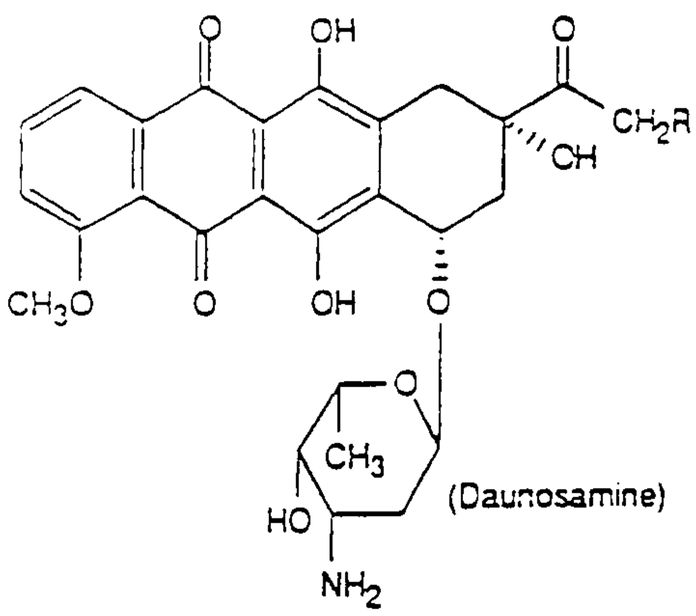
ACTINORHODIN



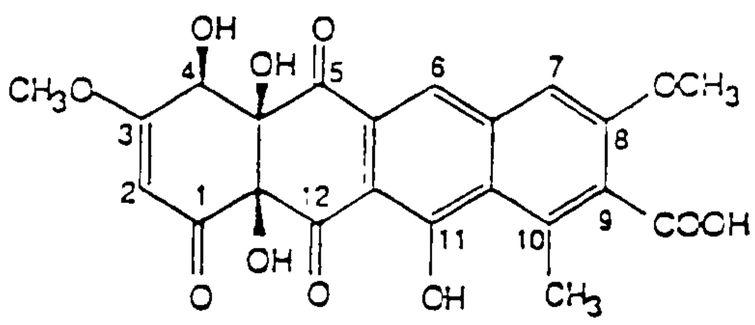
ALOESAPONARIN II



MONENSIN A



	R
DAUNORUBICIN	H
DOXORUBICIN	OH



TETRACENOMYCIN C

Figure 1.3
The diverse structures of selected polyketides.

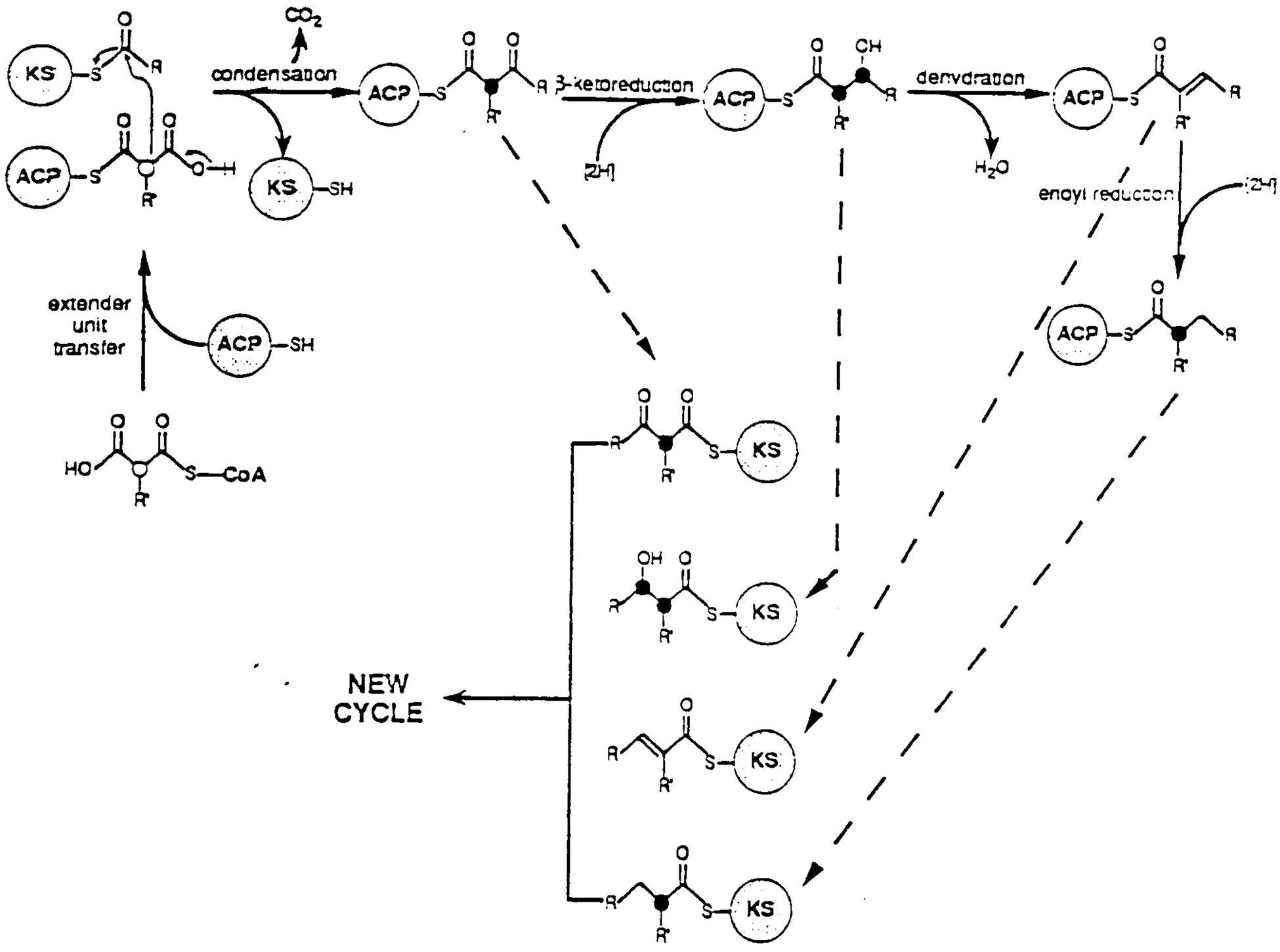


Figure 1.4
Schematic representation of the reactions in fatty acid and polyketide formation. The acyl carrier protein (ACP) and β-ketoacyl synthase (KS) both carry thiol groups. Chiral centres are denoted by a dot. From Katz and Donadio (1993).

Polyketides are formed by enzymatic reactions similar to those responsible for the formation of long-chain fatty acids. However, there are a number differences which set them apart. Firstly, in the simpliest example, long-chain fatty acids are generally derived by the sequential addition of acetate units from an acetyl unit. In reality, alternative starter units such as isopropyl or isobutyryl, are employed in some of the branched fatty acids (reviewed in Hopwood and Sherman; 1990). In many of the polyketides the starter unit can also be quite complex with linear or branched carboxylic acids, aromatic and aliphatic rings. Examples include a propionyl-CoA starter for daunorubicin (Kitamura *et al.*, 1981) and malonamyl-CoA in oxytetracycline (Gatenbeck, 1961; Thomas and Williams, 1983). Secondly, in the synthesis of saturated fatty acids, all of the keto groups brought to the growing chain after condensation (except that of the terminal carboxyl) are removed during carbon chain assembly by the reduction-dehydration-reduction cycle. In contrast, the synthesis of polyketides requires a highly-programmed PKS which may or may not 'choose' to fully complete the cycle following condensation. The most cited example of the complexity of polyketide biosnthesis is that of the PKS responsible for the formation of the *Streptomyces* ionophore, monensin A (Hopwood and Sherman; 1990). This enzyme uses an acetate starter unit, but the extender units vary and are used in a defined sequence. They include not only acetate residues but propionate from methylmalonyl-CoA (thus introducing a methyl side chain) and butyrate from ethylmalonyl-CoA, leaving a ethyl side chain. The PKS for monensin A also handles the post-condensation 'programming' of 12 keto groups; in this case two remain intact implying that the entire reduction-dehydration-reduction cycle has been aborted; three are reduced to hydroxyls, by curtailment of the cycle before the dehydration step; three go on to form double bonds, indicating ketoreduction and dehydration but no enoyl reduction and four keto groups are modified to an alkyl function by completion of the entire cycle. Finally, termination of synthesis is usually accompanied by further modifications of functional groups. These can include folding and cyclisation, lactonisation, or the formation of an amide bond with an amino acid, all of which are believed to occur while the acyl chain is still bound to the PKS. It is these features of the PKSs that account for the huge structural diversity found in polyketide metabolites. Figure 1.5 shows the varied architecture of the genes which encode both the fatty acid and polyketide synthases.

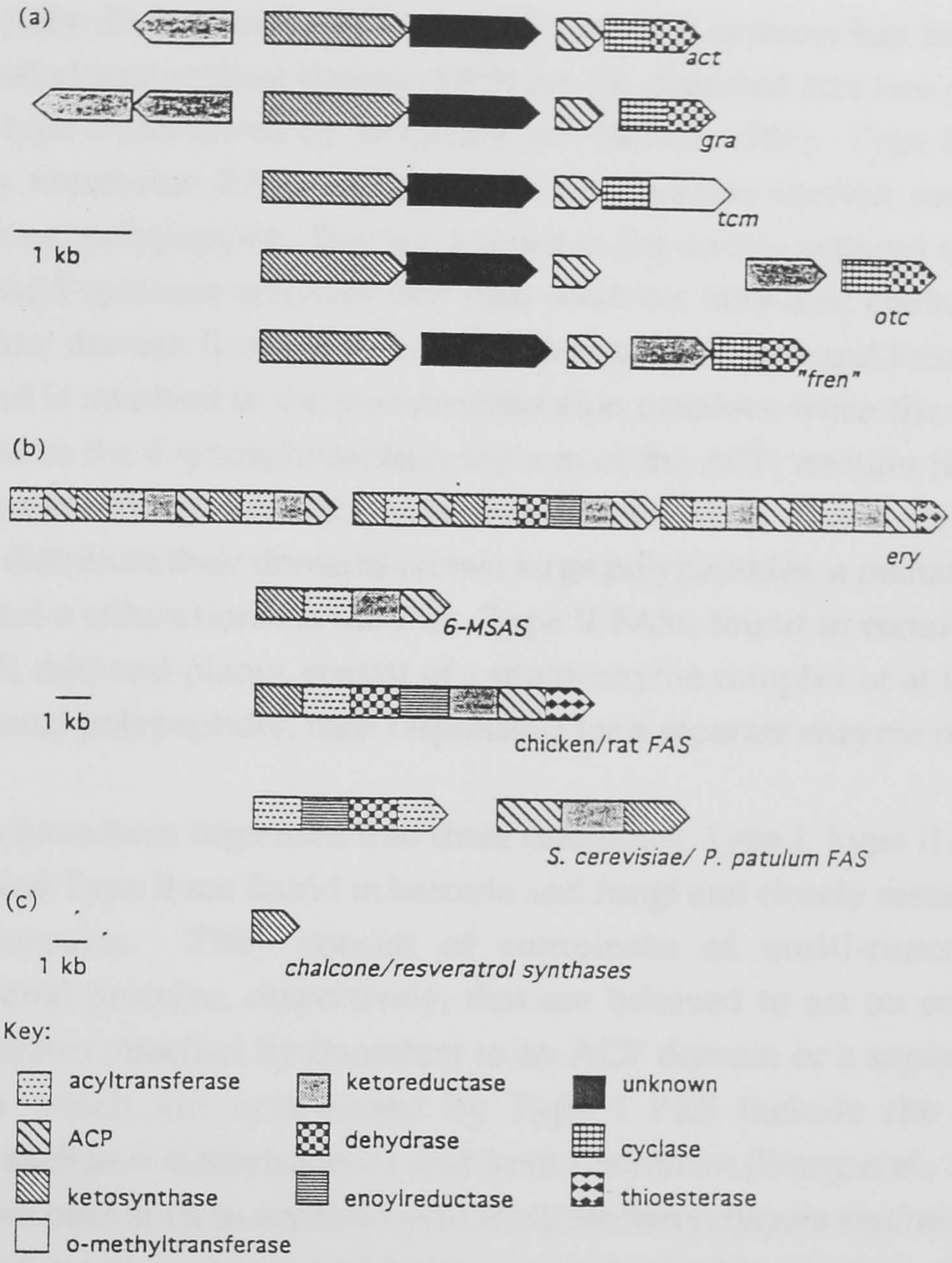


Figure 1.5
The architecture of genes encoding fatty acid and polyketide synthases. (a) Type II PKSs. The clusters are drawn approximately to scale except for the gap between the ACP and ketoreductase of the *otc* cluster, which is longer than shown. (b) Type I FASs and PKSs, arranged to show the conservation of the linear order of active sites between vertebrate Type I FASs and the PKSs. (c) Chalcone/resveratrol synthases, drawn to the same scale as the Type I FASs in (b). They each may represent a single condensing enzyme that may be evolutionarily unrelated to other ketosynthases. (a) Type II PKSs (*act*, actinorhodin; *gra*, granaticin; *tcm*, tetracenomycin; *otc*, oxytetracycline and *fren*, frenolicin); (b) Type I PKSs (*ery*, erythromycin and *6-MSAS*, 6-methylsalicylic acid). Taken from Hopwood and Khosla (1992), and references within.

1.8 Classification of FAS and PKS

The very diverse range of both FAS and PKS systems has led them to being separated into various classes. FASs can be classified into two categories, type I and type II (reviewed by McCarthy and Hardie; 1984). Type I enzymes, typified by vertebrate FASs, consist of three domains carried on a single, multifunctional polypeptide. Domain I contains the acetyl/malonyl transferase and β–ketoacyl synthase activities and thus catalyses fatty acid chain initiation and extension; domain II contains the ACP, dehydrase, enoyl and ketoreductase activities and is involved in the post-condensation reactions when the acyl chain is still bound to the 4'-phosphopantetheine arm of the ACP; domain III contains the thioesterase activity that terminates chain construction. The Type I FASs of *S. cerevisiae* distribute their domains on two large polypeptides, a pentafunctional β-subunit and a trifunctional α-subunit. Type II FASs, found in certain bacteria (including *E. coli*) and plants, consist of a multienzyme complex of at least eight monofunctional polypeptides, each responsible for a separate enzyme reaction.

PKSs have been organised into three categories, Type I, Type II and Type III. Type I and Type II are found in bacteria and fungi and closely resemble their FAS counterparts. They consist of complexes of multi-functional or monofunctional proteins, respectively, that are believed to act on covalently-bound substrates attached by thioesters to an ACP domain or a separate ACP. Polyketides which are synthesised by Type I PKS include the simplest polyketides such as 6-methylsalicylic acid from *Penicillium* (Wang *et al.*, 1990) and more complex ones such as erythromycin from *Saccharopolyspora erythraea* (Cortes *et al.*, 1990). Several aromatic antibiotics are synthesised by the type II enzyme system, for example, actinorhodin from *S. coelicolor*, tetracenomycin from *S. glaucescens* and oxytetracycline from *S. rimosus* (Hopwood and Sherman, 1990). Type III PKSs, distributed in plants, do not appear to require an ACP. The successive condensation of malonyl-CoA onto the starter unit is catalysed by a single polypeptide very similar to the condensing enzyme alone in FAS systems (Hahlbrock; 1981). The two sections that follow will show the organisation and "programming" of two distinct types of enzyme systems, the Type I PKS of the erythromycin producing *Saccharopolyspora erythraea* and the Type II PKS of the actinorhodin producing *S. coelicolor*. (For a comprehensive review, see Katz and Donadio; 1993).

1.8.1 Erythromycin, A Type I PKS

Erythromycin is composed of the 14-membered polyketide-derived macrolactone ring, 6-deoxyerythronolide B (6dEB), to which are attached two deoxysugars. The biosynthesis of erythromycin occurs by condensation of a starter unit of propionyl-CoA with six methylmalonyl-CoA extender units to produce the earliest 14-membered macrolactone ring intermediate, (6dEB), by a Type I PKS multienzyme complex. Following C-6 hydroxylation of 6dEB, addition of glucose-derived deoxysugars, mycarose and desosamine takes place at the C-3 and C-5 hydroxyls, respectively; then *O*-methylation of the mycarosyl moiety and C-12 hydroxylation of the ring produces erythromycin A.

The Type I PKS enzyme required for the synthesis of the macrolide antibiotic erythromycin is very complex. The genes thus far characterised consist of three repeat units, eryA1, eryAII, and eryAIII. These repeat units have been designated "modules" and the corresponding protein segments, synthase units (SU). Each repeat unit carries two "modules" of information (for a total of six), each carrying sites for the ACP, ketosynthase and acyltransferase activity. Each of the six modules, also carries a unique combination of post-condensation enzymes. Modules 1, 2, 5 and 6 encode only a KR domain, module 3 lacks a functional KR although much of the sequence is present, and module 4 carries the full complement of KR, DH and ER domains. In addition, extra AT and ACP domains are found at the N-terminus of module 1 and a TE domain is located at the C-terminus of module 6 (Donadio *et al.,* 1991) [Figure 1.6]. Bearing in mind that the assembly of erythromycin requires 6 successive condensation reactions and that there are 6 modules, the modular hypothesis suggests the genetic order of the modules corresponds to the sequence in which the SU's are employed in the synthesis of complex polyketides such as macrolides. The modular hypothesis, has recently lost credibility as illustrated with the avermectin PKS, which consists of 12 SU's (and the FK506 enzyme, which carries an additional ten), whose genetic order reads from the left for 6 SU's and then jumps right to the end and reads the remaining six from the right (Katz and Donodio; 1993). After each chain extension and any programmed post-condensation reactions specific for that particular module, the acyl substrate would be transferred to the active site thiol of the β-ketoacyl synthase of the next module, where condensation would occur with the specific chain elongation unit required for that particular step in the pathway (reviewed in Katz and Donadio; 1993).

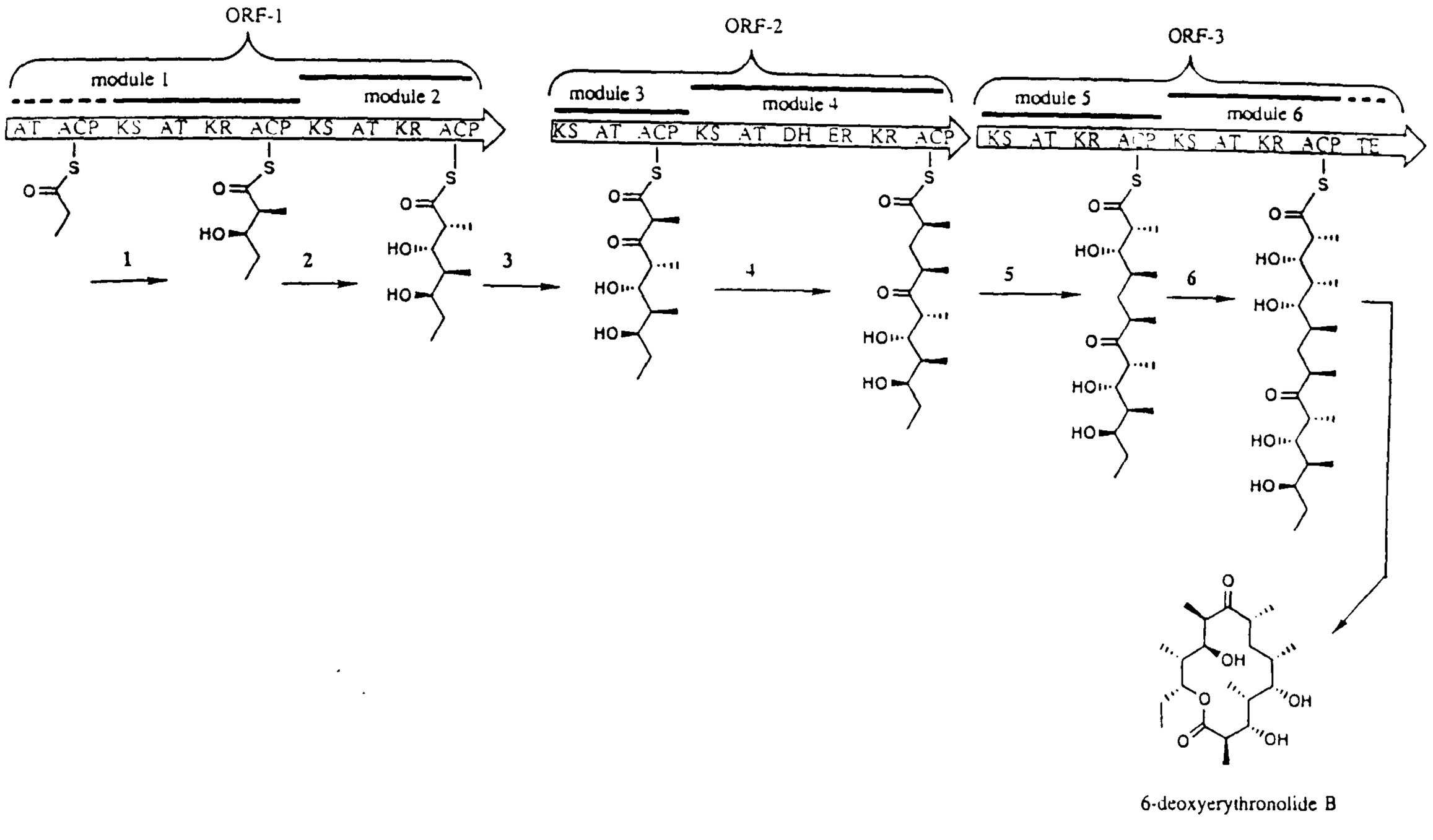


Figure 1.6
Organisation of Type I systems. The 'programmed' assembly of 6-deoxyerythronolide B (the precursor of erythromycin) on the three multifunctional proteins, encoded by open reading frames 1, 2 and 3 of the *eryA* locus of *Saccharopolyspora erythraea*. AT, acyltransferase; ACP, acyl carrier protein; KS, β-ketoacyl-ACP synthase; KR β-ketoacyl-ACP reductase; DH, dehydrase; ER, enoyl reductase; and TE, thioesterase. (The figure is based on data of Cortes *et al.*, 1990 and Donodio *et al.*, 1991, taken from Hopwood and Khosla, 1992).

The pathway of erythromycin biosynthesis has been the subject of extensive biochemical and genetic studies (Donadio *et al.*, 1992). The biosynthetic genes for erythromycin(*ery*) are clustered in a ca. 45 kb region of the *Sac. erythraea* chromosome that also contains *ermE*, the erythromycin resistance gene (Donadio *et al.*, 1991). The PKS genes responsible for the formation of 6dEB make up more than 70% of this cluster and through hybridisations with specific probes, the components of each module were identified. If the final molecular structure is programmed by the number and composition of the different SU's then it follows that alterations in a single SU would affect only that portion of the PKS template. This was confirmed when Donadio *et al.* (1991) introduced a mutation in the KR of the SU5 and the unaltered SU6 acted upon the "unnatural" SU5 derivative. If any SU can be altered without affecting the synthetic makeup of the other SUs, the potential then exists for creating a whole new series of hybrid antibiotics with yet uncharted antibacterial capabilities.

1.8.2 Actinorhodin, A Type II PKS

The most-studied aromatic polyketide is actinorhodin. This is due primarily to its characteristic properties as a red-blue acid-base indicator which allows for easy visual inspection of mutants and clones altered in its production. Actinorhodin is one of at least four antibiotics produced by *S. coelicolor* A3(2), the others being undecylprodigiosin (red) methylenomycin (mmy) and calcium dependent antibiotic (cda) (Hopwood *et al.*, 1986). The biochemical and genetic pathways of the biosynthesis of actinorhodin have both been elucidated (figure 1.7).

The genetic map of the *act* cluster was determined by a series of seven blocked mutants. Pairwise cosynthesis studies in which precursors secreted by later-blocked mutants are converted by earlier-blocked mutants into actinorhodin, permitted the mutants to be placed in sequential order; I, III, VII, IV, VI, and V (Rudd and Hopwood; 1979), class V was later subdivided into VA and VB (Cole *et al.*, 1987). The seventh class, II, has a phenotype consistent with a mutation in a positively-acting regulatory gene for the biosynthesis of actinorhodin, as described earlier. The *act*I and *act*III mutants did not secrete any biosynthetic intermediate which could be converted to actinorhodin by any of the other mutant classes, but were able to convert intermediates secreted by all the other mutant classes to actinorhodin. This suggested that both *act*I and *act*III mutants were defective in the earliest enzymatic steps of the pathway. A red

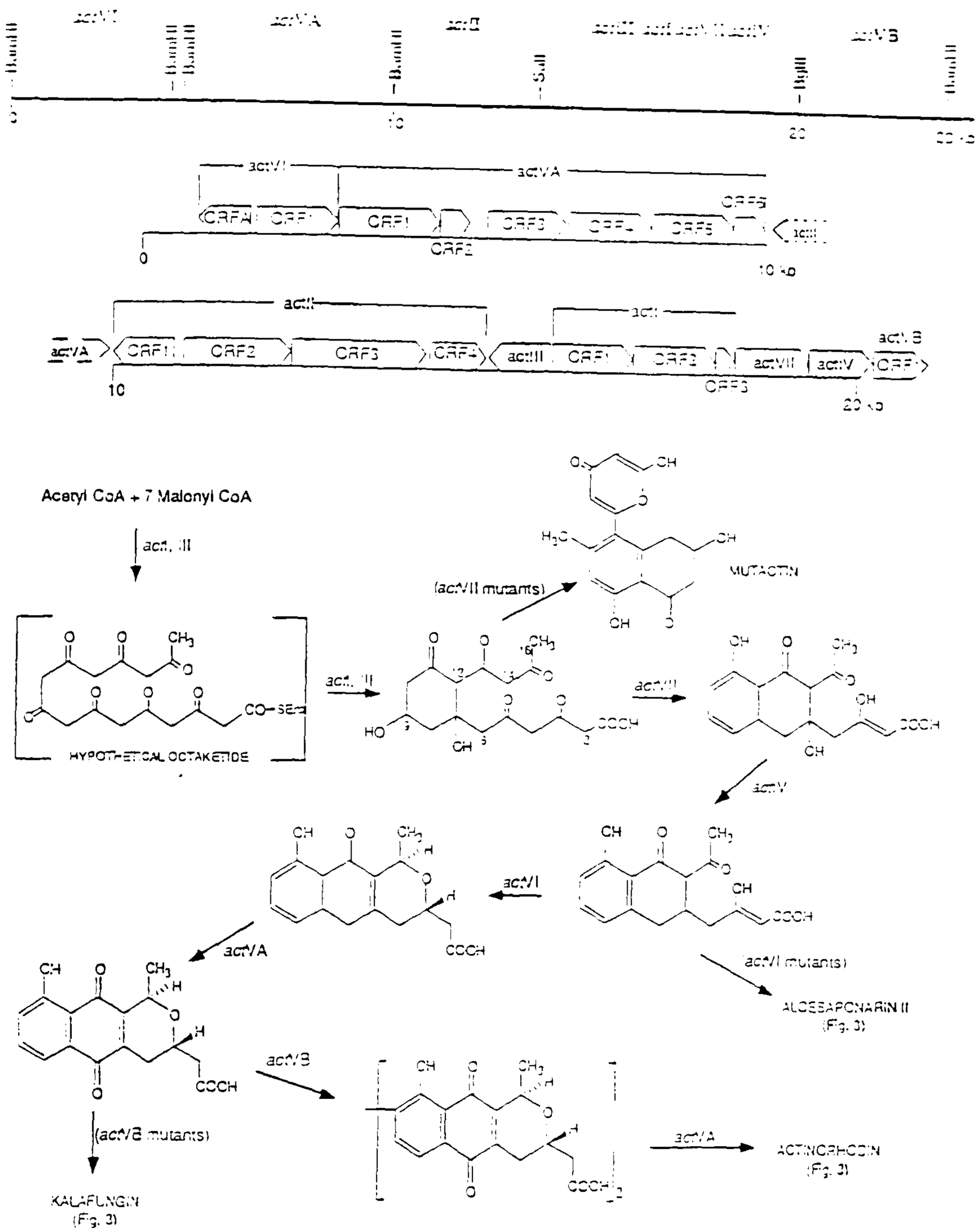


Figure 1.7
Organisation of the *act* cluster and pathway for actinorhodin formation. ORFs are indicated by empty arrows in the two enlargements. From Hopwood *et al.*, 1989.

pigment was accumulated by the *act*III mutants whereas no pigment at all was seen with the *act*I mutants. From this information, it was concluded that *act*I mutants were blocked at an earlier step than the *actIII* mutants of the actinorhodin pathway (Rudd and Hopwood; 1979). The genes for the entire synthesis of actinorhodin, including the *act*II gene, were mapped and shown to be clustered in a region of the chromosome of approximately 26 kb (Malpartida and Hopwood, 1984; 1986).

Shunt metabolites produced by the blocked mutants have been used to the elucidate the biochemical pathway and potential gene functions of each class of mutant. The starter unit, acetyl-CoA, requires the addition of seven acetate units to produce the first isolatable intermediate, a hypothetical octaketide with a reduced keto group at C-9, that has undergone a spontaneous C-7:C-12 ring closure through an aldol condensation. The *act*III mutant has been assigned the KR function. *Act*I mutants were thought to occur even earlier in the pathway and were therefore assigned with the KS function. A number of post-PKS functions are also required and include; a dehydration at C-8:C-9 and C-5:C-14 ring closure with the concomitant aromatisation of the first ring through loss of water at C-7:C-12 by a putative cyclase/dehydratase (*act* VII); a C-5:C-14 dehydration (*act*IV); followed by reduction at C-3 to form the pyran ring (*act*Va). Prior to dimerization, the molecule is hydroxylated at C-6 and a final hydroxylation at C-8 yields actinorhodin (*act*VA). *Act*VA is composed of 6 ORFs, ORF1 appears to encode a transmembrane protein responsible for export the antbiotic, ORFs 2, 4, 5, and 6 are likely involved in the two hydroxylation steps. The dimerization function has been assigned to the *act*VB class (reviewed in Katz and Donadio, 1993).

The Type II PKS required for the synthesis of actinorhodin is composed of three polypeptides, each reponsible for a discrete step, ORF1, 2 and 3. The predicted protein sequences of both ORF1 and 2 show convincing similarity to the fatty acid β-ketoacyl synthase of *E. coli*, encoded by *fasB* (reviewed by Katz and Donadio; 1993). Although ORF1 and ORF2 are very similar, ORF1 carries an active-site serine, characteristic of β-ketoacyl:ACP synthase, not found in ORF2 and thus was assigned the role of catalysing the condensation reactions between the growing acyl-CoA chain and the malonyl-CoA extenders. ORF1 and 2 are translationally coupled and the role of ORF2 is thought to be in dictating the chain length. This was demonstated elegantly by McDaniel *et al.*, (1993), from hybrid studies of *act* and *fren* PKSs. Various combinations of *act*- and *fren*-ORF1, 2, and 3 were cloned into an ACT⁻ mutant of *S. coelicolor*. When the *act*-ORF2

was used an actinorhodin-like product was detected. However, when *fren*-ORF2 was employed with *fren*-ORF1 and *act*-ORF3 frenolicin-like products were detected. The predicted protein sequence of ORF3 is less than 100 amino acids in length and contains the highly conserved DLxGyDS motif characteristic of the 4'-phospopantotheine attachment site of the ACP. Two post-condensation enzymes are the KR and bifunctional cyclase/dehydrogenase which flank the ORF1, 2, and 3 cluster. The similarity to FAS type I is striking. However, no separate AT or TE ORFs were located in the actinorhodin PKS. Fernández-Moreno *et al.*, (1992) have located a GHSxG motif, typical of AT functions, in the C-terminal half of ORF1, but it has not been established whether an AT or TE (which also carries this motif) is associated with ORF1.

Notwithstanding the sequence conservation across the PKSs and FASs, a fundamental difference has emerged in the mechanisms by which the enzymes determine the structure for fatty acids and simple aromatic polyketides on one hand and those for complex polyketides, such as macrolides, on the other. In the macrolide case, each round of carbon chain assembly and modification is catalysed by a distinct active site (Donadio *et al.*, 1991). Thus, the sequence of biosynthetic reactions is overtly programmed into the DNA that encodes the PKS. In contrast, for the aromatic polyketides, a single set of reiteratively used active sites of the PKS must determine the number of extender units to add and the extent of post-condensation modifications. The programming of this class still remains obscure. In an attempt to elucidate the programming mechanism of aromatic polyketides, a number of groups have constructed a series of hybrid polyketide synthases (Khosla *et al.*, 1993; Sherman *et al.*, 1992). The conservation of the PKS genes within the Type II aromatic antibiotics is striking (figure 1.8) with good end to end similarity. This does not extend to the Type I PKSs however. A DNA probe containing *act*I has been used for identification and cloning of PKS genes for 13 different actinomycetes including granaticin (Sherman *et al.*, 1989), and frenolicin (Malpartida *et al.*, 1984). The replacement of the ACP homologues for the *act* ACP in *S. coelicolor* has led to the expression of functional synthases, and the recombinants synthesized derivatives with actinorhodin-like properties, albeit, with varying levels of different metabolites. These studies are only in their preliminary stages and with further work by Shen and Hutchinson (1993), using cell free systems to demonstrate the potential to reconstitute a functional Type II PKS *in vitro* from its constitutent proteins, the biochemical properties of the PKSs may soon be elucidated.

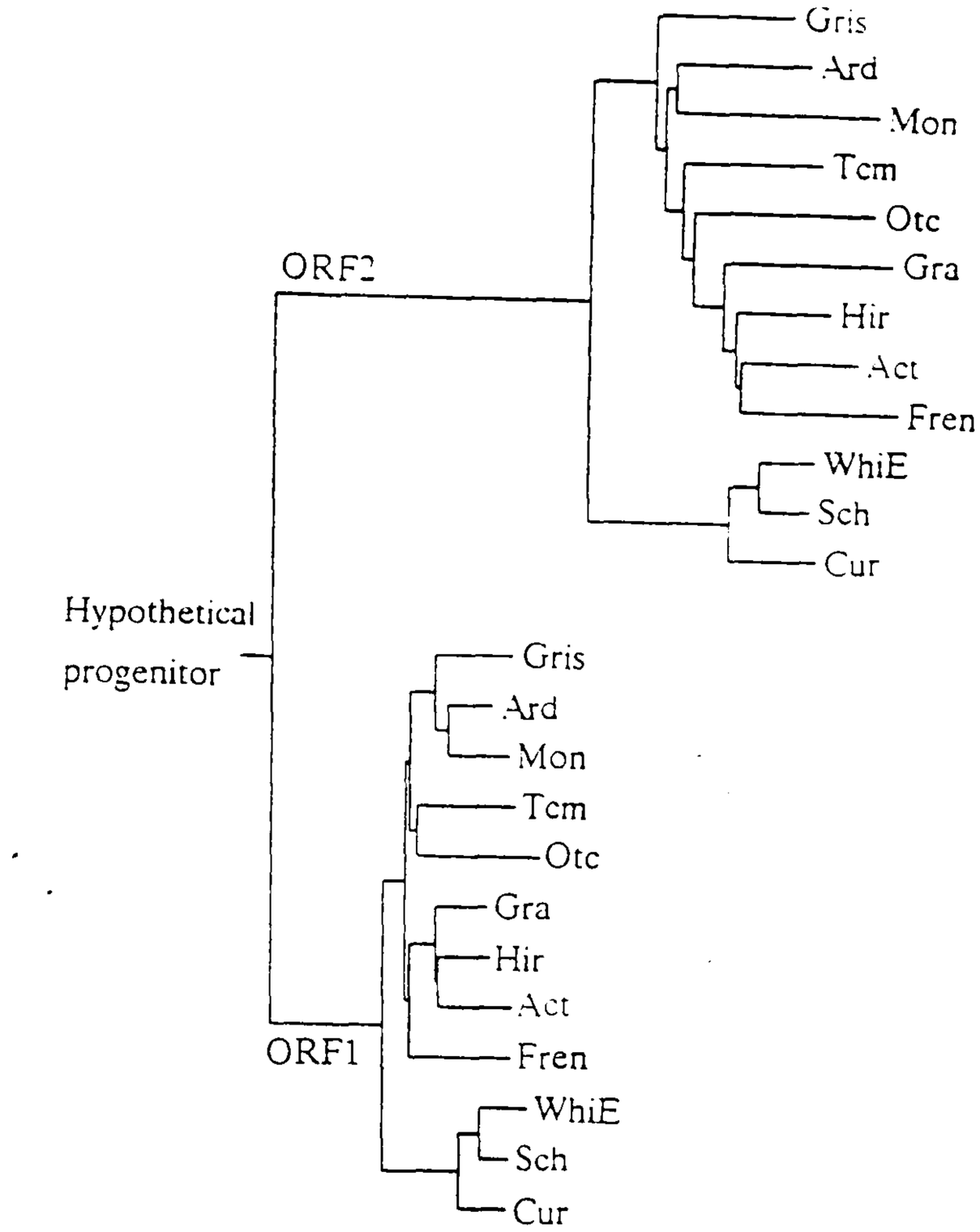


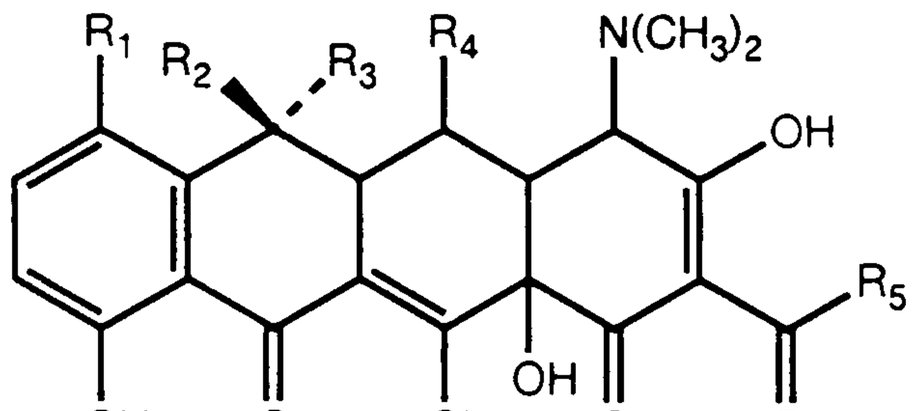
Figure 1.8
Dendrogram, using the PILEUP program of GCG (Devereux *et al.*, 1984), showing the relationships of Type II PKS components ORF1 and ORF2. Data compiled from Yu *et al.*, 1994.

1.9 The tetracycline family

The tetracyclines were the first major group of antimicrobial agents for which the term 'broad-spectrum' was used, i.e. they exhibit activity against both Gram-positive and Gram-negative bacteria. The broad spectrum activity and relative lack of toxicity of tetracyclines has led to their wide-spread use in fighting infections in man, animals and even certain plants. All members of the tetracycline family are synthesized by actinomycetes. The main producers are *S. aureofaciens* and *S. rimosus*. *S. aureofaciens* produces a mixture of chlortetracycline (CTC) and tetracycline (TC) whereas *S. rimosus* produces oxytetracycline (OTC) and a small amount of tetracycline (see recent review from Behal and Hunter, 1994).

Structurally, tetracyclines are composed of a 4-ring carbocyclic structure which differ from each other in the functional groups present at carbons C5, C6, and/or C7 (Figure 1.9), whereby each possesses a different pharmacological property (Levy; 1984). Slight structural variations make these chemicals unique pharmacologically, but all possess the same basic mode of action. The tetracyclines inhibit bacterial growth primarily by inhibiting protein synthesis at the level of the ribosome (Ohnuki *et al.*, 1985). Inhibition of protein synthesis results from the disruption of the codon-anticodon interactions between tRNA and mRNA so that binding of aminoacyl-tRNA to the ribosomal acceptor site is prevented. The precise mechanism by which they prevent binding is not fully understood. However, it is likely to result from interactions with the 30S ribosomal subunit because the majority of the tetracyclines are known to bind strongly to a single site on the 30S subunit (Chopra; 1985).

Bacterial resistance to tetracyclines, which is often transposon- or plasmid-borne, can be acquired by a population as a result of the selective pressure placed on it by the administration of tetracyclines for chemotherapy in humans and animals. Tetracycline resistance is the most frequently encountered in nature. In Gram-negative bacteria, five genetically different classes of tetracycline resistance determinants have been defined. In Gram-positive streptococci, three distinct tetracycline resistance genes have been identified (Chopra *et al.*, 1992). This may ultimately limit the usefulness of the family. Therefore a complete biochemical and genetic understanding of the organism and the antibiotic that it produces, may eventually led to the construction of hybrid molecules. The biochemical and genetic knowledge of the tetracycline family of antibiotics accumulated to date is



	R1	R2	R3	R4	R5
Tetracycline	H	OH	CH_3	H	NH_2
6-Demethyltetracycline	H	OH	H	H	NH_2
2-Acetyl-2-decarboxamide-tetracycline	H	OH	CH_3	H	CH_3
Chlortetracycline	Cl	OH	CH_3	H	NH_2
6-Demethylchlortetracycline	Cl	OH	H	H	NH_2
Oxytetracycline	H	OH	CH_3	OH	NH_2
6-deoxytetracycline	H	H	CH_3	H	NH_2
2-Acetyl-2-decarboxamide-oxytetracycline	H	OH	CH_3	OH	CH_3
Minocycline	$N(CH_3)2$	H	H	H	NH_2

Figure 1.9
Chemical structures of tetracycline family.

outlined below. Particular emphasis will be placed on the oxytetracycline (OTC) pathway in later sections.

A pausible biosynthetic pathway for the tetracylines was first proposed in the 1960's by McCormick and Jensen at the Lederle Laboratories (Figure 1.10). By using blocked mutants, intermediates in the pathway were obtained and purified. The earliest intermediate that could be detected was 6-methylpretetramid (6-MPT). Prior to this stage, the PKS catalyses the condensation of the starter unit, malonamyl-CoA, and eight actetate units (which are derived from malonyl-CoA) to form the primary carbon skeleton of OTC, (Gatenbeck; 1961). Post-condensation reactions include a reduction and a dehydration at C-8 to form the 7-8 double bond and subsequent cyclisation to form a naphthacene nucleus. This was thought to take place in a stepwise fashion as more than one shunt product was detected and cosynthesis activity of each these mutants converted tetracycline biosynthetic intermediates effectively. A third post-condensation reaction, which introduces a methyl group at the C-6 position, was confirmed to occur before the first isolatable intermediate since it was observed that both wild-type and non-methylating mutants were able to convert the 6-methylated precursors into 6-methyl tetracyclines (McCormick *et al.*, 1963).

The biochemical analysis of a further six mutant classes allowed elucidation of the remainder of the oxytetracycline biosynthetic pathway; (i) 4-hydroxy-6MPT, formed from the hydroxylation at C-4 of 6-MPT; (ii) 4-ketodedimethyl anhydrotetracycline (4-keto-ATC), formed from a C-12 hydroxylation of 4-keto-ATC; (iii) 4-amino-ATC, produced by C-4 transamination of 4-keto-ATC; (iv) ATC, produced stepwise by methylation of the amino group, methylation of the methylamino group and reduction of a double bond at C-4-4a; (v) 5a,11a-dehydrotetracycline (DHTC), produced by hydroxylation at C-6 (vi) 5-hydroxy-5a,11a-dehydrotetracycline (DHOTC), formed from a 5-hydroxylation of DHTC; and finally, 5-hydroxytetracycline (oxytetracycline), by reduction at the 5a,11a of DHOTC (McCormick *et al.*, 1962; Miller *et al.*, 1965). At least 13 stages are required in the formation of OTC. An in-depth discussion of each individual step can be found in Thamchaipenet (1994).

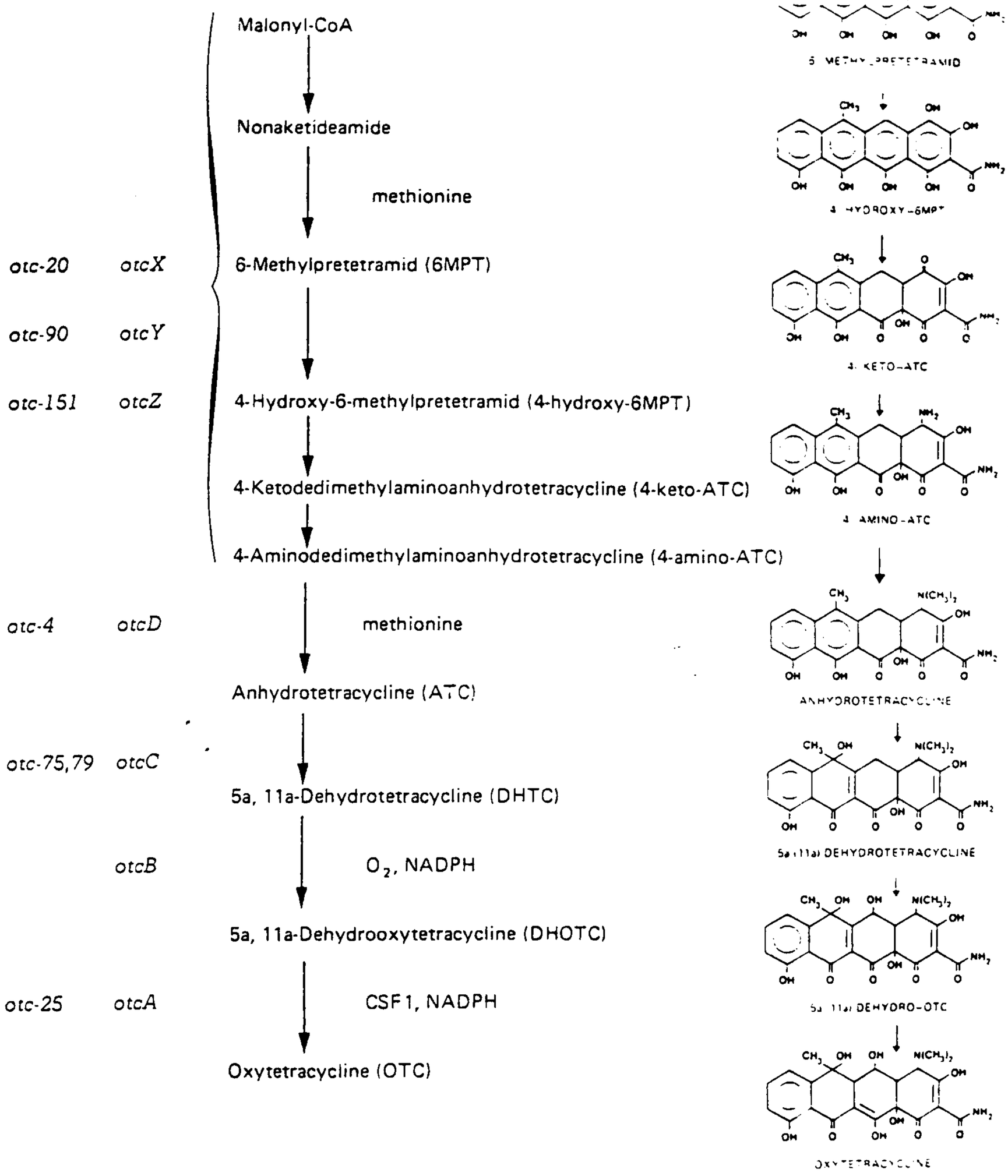


Figure 1.10
Oxytetracycline pathway and assignment of blocked mutants (data of Rhodes *et al.*, 1981).

1.10 Biosynthesis of OTC

A number of independent groups have reported mutants impaired in OTC production (Delic *et al.,* 1969; Alikanian *et al.,* 1961) and a preliminary genetic map was produced (Boronin and Mindlin; 1971, Alacevic, 1976). However, it was not until 1981 that Rhodes *et al.,* carried out a combined biochemical and genetic study. This group from Pfizer laboratories produced a series mutants, unable to produce OTC, by exposure to chemicals or UV. Classification of 57 mutants was done employing the principles of cosynthesis, in much the same way as Rudd and Hopwood (1979) had employed in the elucidation of the biosynthetic pathway for actinorhodin. Pairwise combination of the mutants grown both on the surface of agar plugs and together in liquid culture, resulted in nine different types of mutations. The mutations were all thought to be located in the structural genes encoding biosynthetic enzymes, since all the mutants could cosynthesise OTC with other mutants. Representatives of the nine classes (*otc-4, -19, -20, -25, -56, -75, -90, -151* and *-155*) were chosen to try to determine the sequence of the biosynthetic blocks. The mutants could be divided into two classes based on their cosynthetic patterns. Class I mutants (*otc-19, -56* and *-155*) were shown to produce OTC when fed culture filtrate of the Coenzyme F_{420}-producing strain *S. aureofaciens* ATCC 13190. This suggests that these *otc* mutants were blocked in the synthesis of this cosynthetic factor (CSF1), which is essential for the final step on OTC biosynthesis. The class II mutants (*otc-4, -20, -25, -75, -90,* and *-151)* did not produce OTC when CSF1 was added and were, therefore, presumed to be blocked at six different steps in the main OTC pathway (Rhodes *et al.,* 1981).

The sequential order of class II mutants was elucidated by supplying OTC precursors to liquid cultures of all class II mutants. If the supplied compound was converted to OTC, the mutant was assumed to be blocked at a step in the OTC biosynthetic pathway prior to that intermediate. The nature of the impaired step was defined in three mutants in this way. Mutant *otc-25* was the only mutant which could not convert a derivative of the penultimate intermediate, 7-chloro-DHTC, into 7-chloro-OTC. DHTC is an unstable intermediate and *S. rimosus* had previously shown the ability to convert 7-chloro-DHTC, into 7-chloro-OTC (Martin *et al.,* 1966). Although there were two stages in which the mutant could be blocked, it was argued that if the defect occurred between DHTC and DHOTC, a shunt antibiotic, tetracycline, would have been detected, which was not the case. Mutant *otc-25* was presumed to be blocked in the reduction step which converts DHOTC to OTC. Mutant *otc-25* and *-75* were the

only two mutants which were unable to convert ATC to OTC. Mutant *otc-75* was however able to convert 7-chloro-DHTC to 7-chloro-OTC and therefore must be blocked in the C-6 hydroxylation that forms DHTC. Through similar methods of elimination, mutant *otc-4* was found to be blocked in the N-methylation of 4-amino-ATC because it was unable to convert 4-amino-ATC to OTC but was able to convert ATC to OTC. The other three mutant classes were all unable to convert the earlier precursors, 4-hydroxy-6MPT or 6-MPT, to OTC and it was concluded that they were blocked before the formation of 4-amino-ATC. Previous cosynthetic tests in liquid culture showed that they were blocked at different stages, the pale non-pigmented mutant class, *otc-90* , was assigned the earliest blockage. The mutant classes represented by *otc-4, -20, -25, -75, -90* and *-151* were assigned to putative loci *otcD, otcX, otcA, otcC, otcY* and *otcZ*, respectively (Figure 1.10) [Rhodes *et al.*, 1981].

Genetic mapping experiments, employing representatives of each class, suggested that the production genes for OTC were clustered on the chromosome. At the time, the interpretation of the data implied there were two clusters of genes (Rhodes *et al.*, 1981). One cluster, located at 4 o'clock on the circular genetic map, was responsible for the 'early' genes in the pathway, *otc*D, *otc*X, *otc*Y, *otc*Z and the OTC-resistance determinants (see below). The other cluster, including *otc*A, *otc*B, *otc*C and CSF1 biosynthetic genes, the 'late' genes, was located at 10 o'clock on the map.

Early experiments to clone the OTC resistance gene were undertaken in an OTC-sensitive mutant of *S. rimosus* by simply constructing a gene bank in a high-copy number vector and selecting for recombinants that were now resistant to the drug. Surprisingly two resistance determinants were cloned and shown, independently, to encode resistance to OTC (Rhodes *et al.*, 1984). Two resistance genes were also cloned from another strain of *S. rimosus* (Ohnuki *et al.*, 1985) by screening a genomic library in *S. griseus*, a host which was sensitive to OTC. Advances in recombinant DNA techniques for use in the genetic manipulation of *Streptomyces* (Hopwood *et al.*, 1985), made it feasible for the Pfizer group to clone the *otc* genes as a prelude to overexpressing them by first locating the resistance genes and then "chromosome walking" to find the linked biosynthetic genes (Butler *et al.*, 1990). A producer of an antibiotic must be resistant to the antibiotic it makes; otherwise it will be suicidal. Butler *et al.*, (1989) went about cloning the OTC production genes on the premise that production genes and resistance genes are frequently linked. Southern blotting of a cosmid library revealed that the two resistance genes were located around 30 kb apart on the genome of *S.*

rimosus. The structural genes for the bioysnthesis were confirmed to be located between them when mutants blocked in the 'early' part of the biochemical pathway (Rhodes *et al.*, 1981) were capable of producing OTC when complemented with different segments of DNA from within the region between the two resistance markers (Butler *et al.*, 1989).

The Pfizer group then turned their attention to those functions involved in the conversion of ATC to OTC. The product of the *otcC* gene is ATC oxygenase whose activity can be assayed easily. A 'reverse genetics' approach was taken to clone this gene, thought to be located at the 10 o'clock position of the genomic map of *S. rimosus*. The enzyme was purified and the N-terminus of the protein sequenced. Oligonucleotides were then designed from this sequence and used as probes to clone the corresponding *otcC* genomic DNA. Hybidisation results contradicted earlier genetic mapping data of Rhodes *et al.*, (1981), when it was discovered that *otcC* mapped within the 'early' cluster for the OTC biosynthesis (Binnie *et al.*, 1989). The formal possibility that all the structural genes for the production of OTC were located within a single cluster was then addresssed by Binnie *et al.*, (1989). They cloned a large segment of DNA, including both the resistance genes, into a low-copy number vector and transformed it into both *S. lividans* and *S. albus*. The resulting transformants produced OTC, confirming all of the structural genes necessary to specify OTC biosynthesis reside in a single cluster (Figure 1.11).

The 'late' cluster is now assumed to encode the genes for the synthesis of the 5-deazaflvin cofactors necessary for the last step in the biosynthesis (see earlier section). The ability of *S. lividans* and *S. albus* to produce OTC without these genes being specifically introduced on the plasmid implies that they already carries CSF genes (Binnie *et al.*, 1989). It has already been observed that this unusual class of cofactor has a widesread distribution with the *Actinomyces* (Daniels *et al.*, 1985).

1.10.1 Gene Functions

Recombinational analysis by Rhodes *et al.*, (1981) concluded that the genes for the biosynthesis of OTC were clustered. In fact, the entire cluster has been located on a 34 kb *Eco*RI fragment (Binnie *et al.*, 1989). The nucleotide sequence has been obtained for much of the cluster. Two resistance genes flank the cluster, *otrA* and *otrB*. The resistance gene product of *otrA* protects the ribosomes from

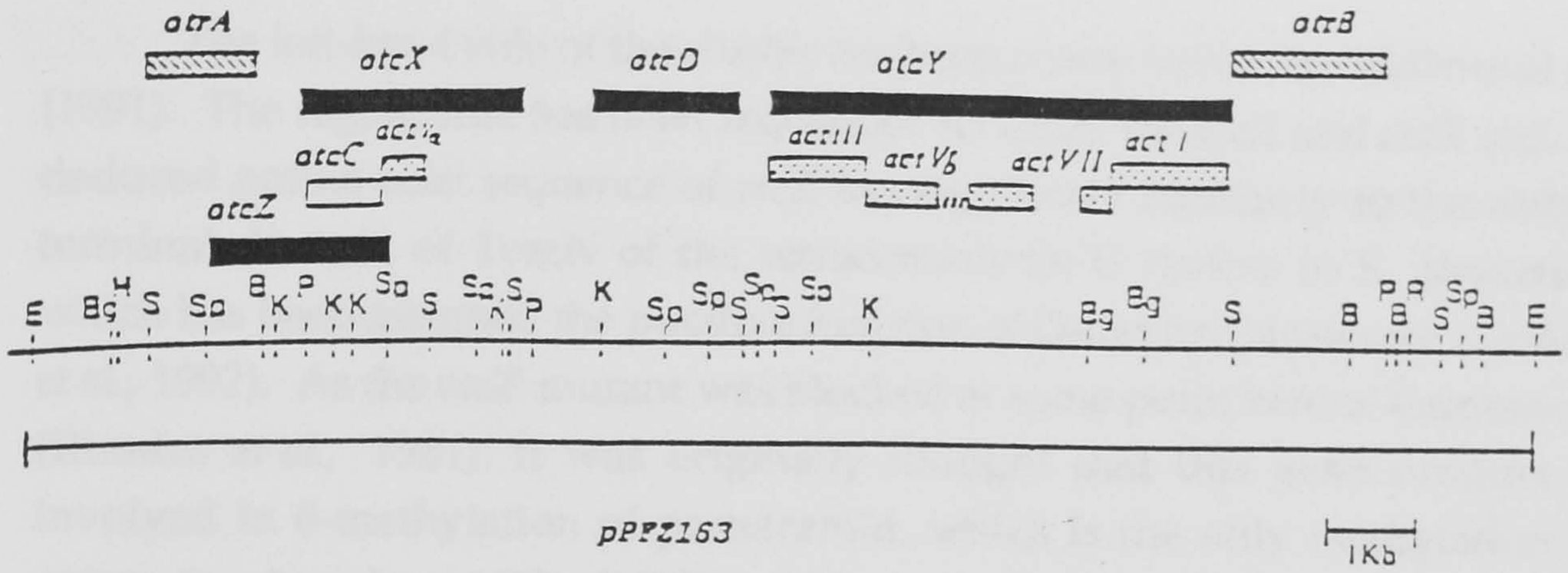


Figure 1.11
Restriction map of the oxytetracycline cluster of *S. rimosus*. The closed boxes show the minimum DNA segments that can complement the OTC-negative classes of mutants (Butler *et al.* 1989). The line delineates the region of DNA which confers the ability to produce OTC when introduced in *S. lividans* and *S. albus* (Binnie *et al.* 1989). Abbreviations: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; R, *Eco*RV; S, *Sst*I; Sp, *Sph*I. Modified from Butler*et al.* 1990.

translational arrest by OTC, while the product of *otrB* resistance gene results in decreased accumulation (increased efflux) of the antibiotic from the mycelium (Ohnuki *et al.*, 1985). The nucleotide sequence of *otrA* showed a deduced amino acid sequence very similar to that of the elongation factor (EF-G) from *E. coli* (Zengel *et al.*, 1984). Doyle *et al.*, (1991) proposed that the *otrA* gene may function as an alternative EF-G, which substitutes for the usual EF-G that is presumed to be inhibited by tetracycline. The nucleotide sequence for *otrB* has also been determined and has strong similarity to other exporter genes (McGregor; pers. comm.).

The left-hand side of the cluster has been characterised by McDowall *et al.*, (1991). The region that has been sequenced includes the *otcZ* and *otcX* loci. The deduced amino acid sequence of *otcZ* has significant similarity to the carboxy-terminal domain of TcmN of the tetracenomycin C cluster in *S. glaucescens*,, which has been assigned the putative function of *O*-methyltransferase (Summers *et al.*, 1992). As the *otcZ* mutant was blocked at some point before 4-amino-ATC (Rhodes *et al.*, 1981), it was originally thought that this gene product was involved in 6-methylation of pretetramid, which is the only methylation step prior to 4-amino-ATC (McDowall *et al.*, 1991). However, the recent characterisation of the genes in chlorotetracycline biosynthetic cluster (Ryan *et al.*, 1994) suggest that *otcZ* may be involved in the amino-dimethylation in the conversion of 4-amino-ATC to ATC (see figure 1.11). Ryan *et al.*, (1994) were able to isolate separate mutants deficient in both the 6-methylation of the pretetramid and the amino-dimethylation. The *otcZ* mutant was unable to convert either 6MPT or 4-hydroxy-6MPT to OTC (Rhodes *et al.*, 1981) which implied that it may contain mutations at more than one locus. Separate mutants of both the 6-methylation of the pretetramid and the amino-dimethylation were not isolated in the *otc* cluster. Recent sequence analysis within the *otcY* locus, see below, has revealed a second methylase gene (Thamchaipenet; pers. comm.). It is proposed that this may be the gene involved in the 6-methylation of pretetramid, its physical location among the early biosynthetic gene locus gives it credibility. The *otcX* locus encodes two gene products. One, *OtcX*-ORF1 is similar to the gene product *ActVA*-ORF2, which is involved in ring-hydroxylation in the biosynthesis of actinorhodin (McDowall *et al.*, 1991; Caballero *et al.*, 1991). The second, *OtcX*-ORF2 is significantly similar to an ORF (of unknown function) immediately downstream of the ferredoxin gene of *Sac. erythraea* (Donadio *et al.*, 1991).

Moving right to the *OtcD* loci, previous sequencing (Thamchaipenet; 1994) has revealed two ORFs. Both *OtcD*-ORF1 and ORF2 are transcribed from the right toward the *otcD* locus. The *OtcD*-ORF1 shows good similarity to other cyclase/dehydrases and may catalyse at least one of the subsequent cyclisation mechanisms to yield an aromatic ring product. The gene product of *OtcD*-ORF2 appears to be involved in steps after pretetramid formation and was similar in deduced amino acid sequence to several bacterial flavin-dependent hydroxylases.

The *OtcY* loci can be divided into two regions, *otcY1* and *otcY2*. The left-hand region, *otcY2*, is composed of *OtcY2*-ORF1, which reads right to left along with the two *otc*D gene products, and *OtcY2*-ORF2, ORF3 and ORF4, which are transcribed in that order in the opposite direction. Through predicted amino acid sequence determination and similarities to genes of known function, the following functions have been assigned. The *OtcY2*-ORF1 gene product is thought to encode the β-ketoacyl reductase, responsible for a ketoreduction at the C-9 position on the OTC polyketide chain. The *OtcY2*-ORF2, a putative dioxygenase, was similar to the β-subunit of toluene and benzene dioxygenases from *Pseudomonas* and the *actVI*-ORFA and *fren*-ORFX gene products from the actinorhodin and frenolicin gene clusters. *OtcY2*-ORF3, a putative combination malonamate:CoA ligase enzyme. The *OtcY2*-ORF4 gene product was significantly similar to one of the subunits of tetracycline dehydrogenase which is required in the late stages of tetracycline biosynthesis in *S. aureofaciens* (Thamchaipenet; 1994).

At the right end of the *otcY* locus is a coding region with strong similarity to *actI* genes of the actinorhodin cluster. This region has been sequenced (Kim *et al.,* 1994) and the deduced amino acid sequences correspond to the presumed heterodimeric ß-ketoacyl synthase and the acyl carrier protein found in other type II (multi-component) PKS systems that specify construction of acetate-derived polyketide antibiotics including tetracenomycin (Bibb *et al.,* 1989) and granaticin gene clusters (Sherman *et al.,* 1989).

Situated between the *OtcY* loci and the *OtrB* loci is a putative repressor (McGregor; per. comm). It surprisingly does not look similar to other tetracyline-type repressors but is more similar to the *E. coli* multi-drug resistant repressors. One or two gaps were left to be sequenced in the *otc* cluster, most notably between the *OtcX* and *OtcD* loci, before the entire pathway was to be eludicated.

Definitive gene functions had also to be determined, this could be accomplished genetically by gene disruption experiments.

1.11 Scope of Present Work

The work presented in this thesis is mainly concerned with the development of a 'workable' integration strategy to help deduce putative gene functions of the oxytetracycline pathway of *S. rimosus*. Specific aims were to sequence the region between the *OtcX* and *OtcD* loci and determine the transcriptional pattern of this region. One of the genes in this region was then the subject of gene disruption and replacement experiments, by integration in the *S. rimosus* chromosome.

During the course of this work various methods to improve the level of foreign DNA accepted by the *S. rimosus* host were also investigated.

Chapter 2

Materials and Methods

INTRODUCTION

Experiments which form the basis of this thesis were carried out by the following general procedures. The chapter is laid out in three main sections. The first two sections divide the manipulations related to *Escherichia coli* and *Streptomyces*. The third section contains bacterial strains, plasmids and bacteriophages.

2.1 STANDARD MICROBIOLOGICAL TECHNIQUES FOR *E.coli*

All *E.coli* procedures can be found in "Molecular Cloning, a Laboratory Manual" (Sambrook *et al.,* 1989) unless otherwise stated.

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, Poole, Dorset; Difco Laboratories, Detriot, Michigan, USA and Sigma Chemical Co. Ltd, Poole, Dorset.

2.1.1 STANDARD MEDIA FOR THE PROPAGATION OF *E. coli*

L-Broth: 10g tryptone, 5g yeast extract, 5g NaCl and 1g glucose made up to 1 litre in distilled water and adjusted to pH 7.0 with NaOH.

L-Agar: As L-Broth without glucose and the addition of 15g.L^{-1} agar.

2YT-Broth: 10g bacto-tryptone, 10g yeast extract, 5g NaCl made up to 1 litre in distilled water.

Minimal Agar: 7g K_2HPO_4, 2g KH_2PO_4, 4g $(NH_4)_2SO_4$, 0.25M trisodium citrate, 0.1g $MgSO_4.7H_2O$, 17.5g agar made up to 1 litre in distilled water.

Minimal Medium Supplements: glucose and thiamine were added to give concentrations of 2mg.mL^{-1} and 20µg.mL^{-1}, respectively.

Davis and Mingioli (D&M) Salts (X4): 28g K_2HPO_4, 8g KH_2PO_4, 16g $(NH_4)_2SO_4$, 1g trisodium citrate, 0.4g $MgSO_4.7H_2O$, made up to 1 litre with distilled water.

2.1.2 STERILISATION

All growth media were sterilised by heating to 120^{0}C for 15 mins in an autoclave. Supplements and buffer solutions were heated to 108^{0}C and $CaCl_2$

to 114^{0}C for 10 mins. Heat-labile solutions, such as amino acids, were sterilized by filtration through Nalgene 0.22 μm pore membranes.

2.1.3 GROWTH CONDITIONS FOR *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 (usually ampicillin at 100 μg.mL^{-1}). The volume of broth inoculated depended on the quantity of plasmid required. Routinely, 2mL and 100mL cultures were used for small and large scale plasmid preparations, respectively (see section 2.1.9). For the preparation of competent cells, liquid cultures of *E. coli* TG1 were grown in 2YT. To maximise aeration of the culture, the volume of the Ehrlenmeyer flask was at least five times that of the broth. All cultures were incubated at 37^{0}C in an orbital shaker at ca. 250 rpm. *E. coli* strains were also propagated on L-agar or Minimal Medium agar plates (containing supplements) with the appropriate selection. Plates contained ca. 25ml of agar and were incubated overnight at 37^{0}C.

2.1.4 ANTIBIOTICS AND INDICATORS

The antibiotic concentrations used throughout for both liquid and plate selection were as follows:

DRUG	FINAL CONCENTRATION	STOCK SOLUTION	STORAGE TEMP.
Ampicillin	100 μg.mL^{-1}	20 mg.mL^{-1} (water)	-20^{0}C
Kanamycin	50 μg.mL^{-1}	20 mg.mL^{-1} (water)	-20^{0}C
Tetracycline	50 μg.mL^{-1}	4 mg.mL^{-1} (ethanol)	-20^{0}C
Neomycin	10 μg.mL^{-1}	10mg.mL^{-1} (water)	-20^{0}C
Nalidixic Acid	200 μg.mL^{-1}	20mg.mL^{-1} (water)	-20^{0}C
Chloramphenicol	25 μg.mL^{-1}	34mg.mL^{-1} (ethanol)	-20^{0}C
Streptomycin	50 μg.mL^{-1}	20mg.mL^{-1} (water)	-20^{0}C

Stock solutions were added to molten agar, which had been cooled to 55^{0}C.

5-bromo-4-chloro-3-indolyl-ß-galactosidase (X-gal) was used in conjunction with IPTG to identify *E. coli* strain TG1 containing pIBI or M13mp vectors with inserts in their multiple cloning sites. The insert containing recombinants are generally white while those lacking inserts are blue. X-gal was stored at a concentration of 20mg.mL^{-1} in dimethylformamide (DMF) at -20^{0}C while IPTG was stored at a concentration

of 20mg.mL^{-1} in d$H_2$0 at -20^{0}C. X-gal and IPTG were added to L-agar plates to a final concentration of 20μg.mL^{-1} and 50 μg.mL^{-1} respectively.

2.1.5 PRESERVATION OF *E. coli* STRAINS

E. coli strains were stored in glycerol. An 800μl aliquot of an overnight culture was mixed with an equal volume of 40% (v/v) glycerol, 2% (w/v) peptone and frozen at -70^{O}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.

2.1.6 INTRODUCTION OF PLASMID DNA INTO *E. coli*

Two methods of preparing competent cells were routinely carried out; the procedure described by Hanahan (1983) gave higher efficiencies but could not be stored. Cells made competent by $CaCl_2$ were stored at -70^{0}C routinely. An efficiency of 10^6-10^7 ampicillin-resistant transformants.μg^{-1} plasmid was typically achieved using either method.

Preparation of competent cells by $CaCl_2$ treatment: An overnight culture of the recipient strain was diluted 1 in 100 into 50mL 2YT broth and incubated for 120-150 min to a density of approximately 10^8.mL^{-1} cells (OD_{600} 0.45-0.55). The cells were harvested by a centrifugation (12000g, 5 min, 4^{0}C) and resuspended in 10mL of ice-cold 100mM $MgCl_2$. The cells were kept on ice for 5 min and pelleted again, resuspended in 10mL of ice-cold 100mM $CaCl_2$ and kept on ice for at least 15 min before being pelleted a third time. The cells were resuspended in 2 mLs of ice-cold $CaCl_2$. At this stage glycerol could be added to the cells to 20% (v/v) and 200μl aliquots frozen at -70^{0}C. The cells retained their competence over several months without any significant decrease in transformation efficiency.

Preparation of competent cells by Hanahan (1983) -a modified version:

Reagents: TFB; 10mM 2[N-morpholino]ethanesulphonic acid (MES) buffer (pH 6.3), 10mM KCl, 45mM $MnCl_2.4H_2O$, 10mM $CaCl_2.2H_2O$, **DTT/KAc;** 2.25mM dithiothreitol, 40 mM potassium acetate (pH 6.0).

An overnight culture of the recipient strain was diluted 1 in 100 into 50mL 2YT broth and incubated for 2.5 hrs to a density of ca. 10^{8}·.mL^{-1} cells (OD_{600}0.45-0.55). The cells were cooled on ice for 60 min and then harvested by centrifugation (7000g, 10 min, 4^{O}C). The pellets were resuspended in 10mL of TFB and chilled for a further 30 min. The cells were pelleted again, resuspended in 2mL of TFB and 70 μl of fresh DMSO. The cells were left on

ice for 5 mins before 155µl of DTT/KOAc was added. After a 10 min incubation on ice, 70 µl of fresh DMSO was added and the incubation continued for a further 5 minutes.

Transformation with plasmid DNA: Plasmid DNA (1-100ng) in a volume less than 10µl, was added to 100µl aliquots of the competent cells, mixed gently by inverting the microfuge tube and incubated on ice for up to 1 hour. The cells were heated to 42^0C for 1 min and returned to ice for a further minute. A 300µl aliquot of 2YT broth was added and the cell suspension incubated at 37^0C for 30 min to allow expression of the plasmid resistance genes. For selection with ampicillin, an expression stage was not required. The cells were spread on L-agar plates containing the appropriate antibiotic.

2.1.7 CONJUGATION OF *E. coli*

The donor culture was grown to mid log phase and the recipient to a late log or stationary phase. The two cultures were mixed at a 1:1 ratio and plated on a nutrient agar plate (stiff and dry). The plate was incubated at 37^0C for 3 hours. The nutrient agar plate was then flooded with 1mL of 2YT media and the culture removed. A series of dilutions were made and plated onto selective media. The resulting exconjugants were counted after overnight incubation at 37^0C.

2.1.8 TRANSFECTION OF *E. coli* TG1 WITH BACTERIOPHAGE M13

The modified version of the protocol developed by Hanahan (1983) was used to prepare competent cells of *E. coli* TG1 for transfection with the replicative form of M13.

Transfection of *E. coli* TG1 with bacteriophage M13mp: The procedure followed the plasmid transformation protocol described above, up to and including the heat-shock stage. After this step, 200µl of a fresh exponential TG1 culture was added to the transfected cells, followed by 4µl of IPTG ($200mg.mL^{-1}$) and 40µl X-gal ($20mg.mL^{-1}$). The cells were then mixed and added to 2.5mL of molten water-agar (0.7% w/v, pre-cooled to 45^0C), which were poured onto thoroughly-dried L-agar plates.

2.1.9 ISOLATION OF PLASMID DNA FROM *E. coli*

Protocols based on the alkaline lysis method (Birnboim and Doly, 1979) were used for the isolation of plasmid from both small (5 mL) and large (50-200 mL) culture volumes of *E. coli*.

Reagents: Birnboim Doly I (BDI); 50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA. Lysozyme was added immediately before use to a final concentration of 1-4mg.mL^{-1}. **Birnboim Doly II (BDII);** 0.2M NaOH, 1% (w/v) SDS which was stored in a plastic container. **Birnboim Doly III (BDIII);** 5M KOAc (pH 4.8); prepared by mixing equal volumes of 3M CH_3COOK and 2M CH_3COOH. **Phenol;** The phenol was redistilled, buffered against 0.5M Tris.HCl (pH 8.0) and contained 0.1% (w/v) 8-hydroxyquinoline. **Phenol.Chloroform;** 50 parts phenol, 49 parts chloroform, 1 part isoamyl alcohol. **DNAase-free RNAase;** Pancreatic RNAase (RNAase A) was dissolved at a concentration of 10mg.mL^{-1} in dH_20, heated to 100^{o}C for 15min and allowed to cool slowly to room temperature. The RNAase was then aliquoted and stored at -20^{o}C.

Large-scale plasmid preparations: 200mL cultures of stationary phase cells were harvested using a centrifuge (12000g, 5min at 4^{0}C). The pellet was resuspended in 4mL of Birnboim-Doly I solution. Cultures were incubated on ice for 5min. Then 8mL of Birnboim-Doly II solution was added and the solution was left on ice for 5-10 min before 6mL of cold Birnboim-Doly III solution was added. The suspension was mixed gently and left on ice for 15-30 min. The cell debris and most of the chromosomal DNA was removed by centrifugation (32000g, 5 min at 4^{0}C). The remaining nucleic acid was precipitated by the addition of an equal volume of isopropanol and then harvested by centrifugation (39200g, 15 min). 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 $dH_20$and 4.5g of CsCl dissolved in 3.5mL of dH_20. The DNA and CsCl solutions were combined with 250µl of EtBr (10mg.ml^{-1}), creating a solution with a density of 1.58g.mL^{-1}. The nucleic acid-CsCl solution was spun in a Beckman Ti70 angled rotor at 289000g for 16 hours at 20^{0}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 1mL syringe and the EtBr removed by repeated extractions with water-saturated butanol. After dilution with 3 volumes of dH_20, 9 volumes of absolute ethanol were added. The precipitate was recovered by centrifugation (27000g, 4^{0}C for 30min). The resulting plasmid pellet was washed twice with 70% (v/v) ethanol and dried *in vacuo* before being redissolved in 1mL dH_20. This procedure yielded very large amounts of pure plasmid DNA (up to 1mg), suitable for all *in vitro* manipulations.

Small-scale plasmid preparations: Routinely, plasmids were isolated from 1.5mL of *E. coli* cultures. The cells were harvested by centrifugation in a 1.5mL microfuge tube (12000g for 30s) and resuspended in 100µl of BDI, containing lysozyme at a concentration of $1mg.mL^{-1}$, using a vortex mixer. This was followed by the addition of 200µl of BDII and repeated inversion of the microfuge tube to mix the suspension thoroughly. Immediately afterwards, 150µl of prechilled BDIII was added to the viscous bacterial lysate, mixed gently on the vortex mixer and placed on ice for 5-10 min. The cell debris and most of the chromosomal material was removed by centrifugation (12000g, 4^0C for 10min). 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 incubated at room temperature for 3-5 min. The precipitate was harvested by centrifugation in a microfuge (12000g, 4^0C for at least 15min). 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 50µl dH_20 containing DNAase-free RNAase ($20\mu g.mL^{-1}$). The typical yield of high copy number plasmids such as pUC from *E. coli* was 2-5µg.

2.1.10 QUANTIFICATION OF NUCLEIC ACID

The concentration and purity of the nucleic acid was determined by spectrophotometry; an A_{260} of 1 is equivalent to $40\mu g.mL^{-1}$ RNA and $50\mu g.mL^{-1}$ double stranded DNA. Pure preparations of RNA and DNA have an A_{260}/A_{280} of 1.8 and 2.0, respectively. Contaminating protein or phenol lowers these values significantly.

2.1.11 DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASES

Digestion of DNA was performed in a total volume of, typically, 20µl containing 0.25-1.0µg of DNA and 1-10 units of restriction endonuclease per µg DNA in the appropriate buffer (provided by BRL the main enzyme supplier). For the digestion of larger amounts of DNA, the volumes were scaled up accordingly. The reactions were allowed to proceed for 1 to 2 hours at the temperature recommended by the supplier. Reactions were stopped either by the addition of gel loading buffer or by heating to 70^0C for 5 min, followed by cooling on ice.

2.1.12 PARTIAL DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASE USING ETHIDIUM BROMIDE

A series of tubes were set up all containing the same amount of DNA and reaction buffer. Varying amounts of EtBr (100ng - 500ng) in a constant volume of dH_2O were added to the DNA. Finally, the appropriate restriction enzyme was added last. An uncut and completely-cut sample (without EtBr) were always run as controls. The reactions were digested for 1 hr at the optimal digestion temperature for the restriction enzyme.

2.1.13 FILLING IN OF 5' OVERHANGING BASES RESULTING FROM RESTRICTION DIGESTIONS

After digestion, 1µl of a solution containing all 4 dNTP's at a 2mM concentration and 1-2U of T_4 DNA polymerase was added to the 20µl reaction and incubated at 12^0C for 15 minutes. The dNTP's were be removed before ligations by gel electrophoresis.

2.1.14 LIGATION OF DNA FRAGMENTS

The ligation of DNA fragments was carried out at a DNA concentration of 6mg.mL^{-1}. The molar ratio of insert fragment to vector was 2:1, when the vector could not ligate to itself, but only with the insert fragment. A molar ratio of 10:1 was used when the ends of the vector could ligate to each other. Ligations were performed in 20µl of 1X BRL ligation buffer, containing 1U of T_4 DNA ligase.µg^{-1} of DNA. The reactions were incubated for 4 hours at room temperature or overnight at 16^0C.

2.1.15 ADDITION OF LINKERS TO BLUNT-ENDED FRAGMENTS

The molar ratio of linkers to blunt-end vector fragment must be 200:1, starting with 2-5pmoles of vector. The ligations were set up in 20µl of 1X BRL ligation buffer containing 1mM ATP, 50µg/mL BSA and 2U of T_4DNA ligase µg^{-1} of DNA. The reaction was incubated overnight at room temperature. 80µl of H_20 was added before the reaction was heated at 70^0C for 15min to inactivate the enzyme. A routine restriction digestion was then carried out to remove any concatemerized linkers.

2.1.16 REMOVAL OF THE 5' PHOSPHATE FROM LINEARISED DNA

10X CIP Buffer: 200mM Tris.HCl (pH 8.0), 10mM $MgCl_2$, 10mM $ZnCl_2$ and 0.5mg.mL^{-1} Bovine Serum Albumin.

Procedure: Calf Intestinal Alkaline phosphatase (CIP) was used to remove the 5' phosphate from DNA. Five pmoles of 5'-terminal

phosphorylated DNA with 5' protruding ends (approximately 7μg of a 5kb molecule) were incubated in 1X CIP buffer, containing 0.1U of CIP at 37^0C for 30min. The reaction was terminated by heating to 65^0C in 1X gel loading buffer for 10min. The dephosphorylated DNA was purified from an agarose gel after electrophoresis.

2.1.17 REMOVAL OF PROTEIN FROM NUCLEIC ACID SOLUTIONS USING ORGANIC SOLVENTS

Reagents: Phenol; All phenol used in the purification of DNA or RNA was redistilled, buffered against 500mM Tris.HCl (pH 8.0) and contained 0.1% (w/v) 8-hydroxyquinoline.

Procedure: Proteins were removed from nucleic acid solutions by extracting first with phenol.chloroform and then chloroform. Half volumes of the solvents were added to the nucleic acid solution, mixed using a vortex mixer and the phases separated by centrifugation (12,000g for 3min). Proteins were retained in the organic phase. Subsequent precipitation of the nucleic acid in the aqueous phase using isopropanol or ethanol removed any residual solvent.

2.1.18 AGAROSE GEL ELECTROPHORESIS

Both DNA and RNA were visualized 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 <1.5kb. Gels were routinely prepared and run in TAE buffer.

10X TAE Buffer pH 8.2; 48.4g Tris, 16.4g Na acetate, 3.6g $Na_2EDTA.2H_20$, made up to 1 litre in distilled water, pH adjusted to 8.2 with glacial acetic acid.
5X Agarose gel loading (AGL) buffer (pH 7.4); 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol, 25% (w/v) ficoll, 0.5% (w/v) SDS, 50mM EDTA.

2.1.18.1 MINI GELS

BRL model H6 gel kits were used for the rapid analysis of DNA after digestion with restriction enzymes or precipitation steps. 0.16g agarose was added to 20mL of 1X TAE , boiled then cooled to 60^0C. EtBr was added to 200ng.mL^{-1} and the molten agarose poured into a 7.6cm X 5.1cm gel caster with an 8 well slot former (4.1 X 0.8mm wells). After the gel had set, the slot former was removed and the gel placed in the tank with 500mL of 1X TAE. Depending on the time available and the level of resolution required the

DNA was separated by electrophoresis for 30-45min with an applied voltage of 2-10V.cm^{-1}. The separated DNA molecules were visualised on a 302nm UV transilluminator.

2.1.18.2 LARGE GELS

100mL gels were also used to ensure good separation of DNA fragments for accurate sizing and/or Southern analysis. They were made by pouring 100ml of molten agar containing 200µg EtBr, into a 16.5 X 23cm gel former with a 14 space slot former. The gels were run overnight at 20V in 1X TAE buffer in gel tanks with a capacity of 1 litre. DNA samples were mixed with 1/5 volume of 5X AGL buffer before loading onto the gel.

2.1.19 PHOTOGRAPHY OF RESOLVED NUCLEIC ACIDS

Gels stained with ethidium bromide were viewed on a 302nm UV transilluminator and photographed using Polaroid type 67 land film or using a Pentax 35mm SLR loaded with Ilford HP5 film. Both cameras were fitted with Kodak Wratten filters (No. 23A).

2.1.20 RECOVERY OF DNA FROM AGAROSE GELS

The DNA band were exised from the agarose gel and loaded into the top chamber of a Spin-X™ column (0.45 µm cellulose acetate, Costar). The tube was incubated at -20°C for 15 min. After 15 min the tube was spun at 10,000 rpm for 5 min. The fragment DNA was now in the bottom chamber and the top chamber could be discarded. Recovery levels ranged between 57 and 69%.

2.1.21 NON-RADIO-LABELLED PROBES AND DETECTION

Analysis of DNA on agarose gels by Southern blotting employed the digoxygenin (DIG)-DNA labelling and luminescent detection kit from Boehringer Mannheim Biochemica. In principle, DIG is bound *via* a spacer arm to uridine-nucleotides (DIG-dUTP) and incorporated enzymatically into DNA probes by random-primed DNA labelling (see below). After blocking and hybridisation, DIG-labelled probes are detected by a high affinity anti-DIG -antibody Fab-fragments conjugated to alkaline phosphatase. Alkaline phosphatase dephosphorylates the chemiluminescent substrate AMPPD which upon decomposition *via* an unstable intermediate emits light which can be detected on X-ray film.

2.1.21a RANDOM PRIMED DNA LABELLING METHOD

Labelling of DNA fragments with digoxigenin 11-dUTP was set up following the procedure of Boehringer Mannheim:

1 μg DNA of denatured DNA fragment in a volume less than 10μl (denatured by heating for 10min at 95^0C with subsequent cooling on ice).
2μl of hexanucleotide mix (10X concentrated).
2μl of dNTP labeling mix (10X concentrated)
1μl (1U) of Klenow enzyme, labelling grade

Made up to 20μl total volume with dH_2O.

The reaction mixture was incubated at 37^oC for at least 60 minutes (overnight was preferred to increase the amount of labelled DNA), and stopped by addition of 2 μl of 0.2M EDTA (pH 8.0). The labelled DNA fragments were precipitated with 2.5 μl of 4M lithium chloride and 75 μl of prechilled (-20^oC) ethanol and harvested by centrifugation in a microfuge (12,000g 4^oC for 10min). The DNA pellet was washed twice with 70% (v/v) ethanol, dried in an open microfuge tube at room temperature and dissolved in 50 μl of TE buffer (ca. 250 ng of labelled DNA was expected). The labelled-DNA could be kept at -20^oC for a very long period.

2.1.22 TRANSFER OF DNA FROM *E. coli* COLONIES TO AMERSHAM HYBOND-N™ MEMBRANES

Reagents: 20X SSC; 3M NaCl, 0.3M Tri-sodium citrate. **Denaturing solution;** 1.5M NaCl, 0.5M NaOH. **Neutralising solution;** 1.5M NaCl, 0.5M Tris.HCl (pH 7.2), 1mM EDTA.

Procedure: A Hybond-N™ membrane was cut to the correct size and placed onto the surface of a L-agar plate containing the appropriate antibiotic. The *E. coli* clones to be screened were streaked out on top of the membrane and the plate incubated at 37^0C overnight. The membrane was then removed and placed, colony side up, on a pad of absorbent filter paper soaked in denaturing solution and left for 7min. Next the membrane was transferred to a pad of filter paper soaked in neutralising solution and left for 3min. This step was then repeated, with a fresh pad soaked in neutralising solution. The membrane was washed in 2X SSC, transferred to dry filter paper and air dried,

colony side up. Finally, the samples were fixed to the membrane by baking in an oven at 80^0C for 2 hours.

2.1.23 CAPILLARY TRANSFER OF DNA TO HYBOND™

After electrophoresis and photography, the resolved DNA fragments were transferred under alkali conditions to Hybond-N™ (adapted from Southern; 1975).

Reagents: **Alkali transfer buffer;** 0.25M NaOH, 1.5M NaCl. **Denaturing solution and 20X SSC;** as described in Section 2.1.22.

Procedure: The gel was rinsed in dH_20, placed in enough denaturing solution to immerse it completely and left for 30min. The gel was removed, excess liquid removed by blotting and equilibriated for 10-15min in alkaline transfer buffer. The membrane, cut to the appropriate size, was placed on top of the gel on a raised platform with wicks passing from underneath the gel to a reservoir of transfer buffer. The filter was then overlaid with absorbant pads and weighted down. After blotting for a least 4 hours the membrane was washed briefly in 2X SSC to remove any adhering agarose. The membrane was then air dried for at least 30 minutes.

2.1.24 CONDITIONS FOR AQUEOUS PREHYBRIDIZATION AND HYBRIDIZATION

The procedures which follow were suitable for DNA immobilised on filters from colony lifts and Southern transfer and follow protocols outlined in "The DIG System User's Guide for Filter Hybridization" produced by Boehringer Mannheim.

Prehybridization and hybridization conditions: The hybridization temperature is dependent on the application. However all hybridizations in this work were done at 42^0C. The appropriate volume of DIG Easy HyB™ (approx. 20mL/100cm^2) was prewarmed to 42^0C and then incubated with the blot for 30 minutes. The prehybridization solution was poured off and the hybridization solution added. 5-25ng/mL of DIG-labeled probe was denatured at 95^0C for 5 minutes and then added to prewarmed DIG Easy HyB™. The volume of the hybridization solution depended on the size of the filters; 2.5ml was added for every 100cm^2 of membrane. Ensuring that all air bubbles were removed, the blot was incubated overnight with gentle agitation.

Washing of membranes after hybridization: The nylon membrane was washed twice in 200mL of 2X SSC, 0.1% (w/v) SDS at room temperature for

5min each. It was then washed in 200mL of 0.1X SSC, 0.1% (w/v) SDS at 68^{0}C for 15 min each under constant agitation.
The probe could be stripped from the filter (provided the filter was kept moist after hybridization) by washing in 0.4M NaOH at 45^{0}C for 30min. The filter was then washed in 0.1X SSC, 0.2M Tris.HCl (pH 7.4) and 0.1% (w/v) SDS for 15min and stored in a sealed plastic bag until required.

2.1.25 DETECTION OF DIG-LABELLED NUCLEIC ACIDS

Chemiluminescent detection was carried out with Lumigen™ PPD. **Reagents: Buffer1** 100mM maleic acid, 150 mM NaCl; pH 7.5 **Buffer 2**: 1% Blocking Reagent dissolved in buffer 1. **Buffer 3**: 100mM Tris-HCl, pH9.5, 100mM NaCl, 50mM $MgCl_2$. **Washing Buffer:** Buffer 1, 0.3% Tween® 20. **Blocking Reagent:** A 10% stock solution is made up in Buffer 1. The following solutions are supplied with the Detection kit: Blocking Reagent, Anti-digoxigenin-AP Fab fragments and Lumigen™ PPD.
Procedure: After the post-hybridization washes the membrane was equilibrated in Buffer 1 for 1 minute. The membrane was then transferred to a clean dish and incubated in Buffer 2 for 30 min. The Anti-digoxigenin-AP Fab fragments were diluted 1:10,000 in Buffer 2. Buffer 2 was poured off and the membrane was incubated in the antibody solution for 30 min. The antibody solution was then discarded and the membrane washed twice for 15 min each in washing buffer. The membrane was equilibrated for 2 minutes in Buffer 3. Lumigen™ PPD was diluted 1:100 in buffer 3 and 500µl was added directly to the filter and incubated for 5 min. The excess liquid was poured off and the filter sealed in a plastic bag. It was incubated for 15 minutes at 37^{0}C to initiate the reaction catalysed by alkaline phosphatase. The membrane was then exposed to standard X-ray film.

2.1.26 PREPARATION OF SINGLE-STRANDED M13 DNA

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

Minipreparations: A single M13 plaque was used to infect 5ml of 2X YT broth containing 200µl of an overnight culture of *E. coli* TG1. This was grown at 37^{0}C for 4.5 hours with vigorous shaking, then transferred to a microfuge tube and harvested by centrifugation at room temperature for 5min. The supernatant, containing the phage particles, was recovered and respun. The supernatant (1mL) that remained was mixed with 300µl of a solution of 20%

(w/v) PEG (6000)/2.5M NaCl and left to stand at room temperature for 15min to precipitate the phage particles. These were recovered by centrifugation at room temperature in a microfuge for 15min. The supernatant was discarded, the pellet respun and all traces of supernatant removed. The pellet was then resuspended in 100µl dH_2O and extracted twice with phenol.chloroform and twice with chloroform. The DNA was then ethanol precipitated with ethanol and redissolved in 30 µl dH_20.

2.1.27 DNA SEQUENCING

Dideoxy sequencing (Sanger *et al.*, 1975) was carried out on single-stranded M13 templates using a Sequenase™ kit supplied by United States Biochemical Corporation (USB). T7 DNA polymerase was supplied by Pharmacia Ltd.

Annealing template and primer: In a 1.5mL microfuge tube, 1µg of single-stranded template, 1µl of sequencing primer (0.5pmoles) and 2µl of reaction buffer were combined and the volume made up to 10µl with dH_20. After denaturing at 85^0C for 2min, the primer was annealed to the template by incubating at 45^0C for 10min. After annealing was complete, the tube was placed on ice and used within ca. 2 hours.

Labelling reaction: The T7 DNA polymerase enzyme was diluted 1:6 with reaction buffer containing 0.5µl of pyrophosphatase.
To the annealed template-primer, the following were added on ice:

	Volume (µl)
7-deaza-dGTP labelling mix	2
[^{35}S]dATP (1000-1500Ci.mmol^{-1})	0.5
diluted T7 DNA polymerase (2 units)	2
DTT (0.1M)	1

The above were mixed and incubated at room temperature for 5min. To read sequences close to the primer annealing site, Manganese Buffer (supplied with the kit), was used.

Termination reactions: Immediately after the labelling reaction was completed, 3.5µl aliquots were dispensed into four microfuge tubes containing one of the four 7-deaza-dGTP termination mixtures; ddATP, ddCTP, ddGTP and ddTTP, respectively. The contents were mixed as quickly as possible and incubated at 37^0C for 5min. After the reactions were complete, 4µl of stop loading buffer was added and then they were stored on ice until

the sequencing gel was ready to load. The samples were heated to 70^0C for 2-5 min, cooled rapidly on ice and loaded immediately onto the gel (see below). Approximately 2-3µl of the sequencing products were loaded per track.

2.1.28 DENATURING POLYACRYLAMIDE GEL ELECTROPHORESIS

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

Reagents: 5X Sequencing gel loading buffer (pH 8.2); 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol, 10mM Na_2EDTA, 95% (v/v) formamide (de-ionized with a mixed-bed resin). **40% Acrylamide stock solution:** purchased from Severn Biotech. Ltd containing a 19:1 acrylamide : bis ratio. **10X TBE Buffer:** 108gm Tris Base, 55 gm Boric acid, 40ml 0.5M EDTA pH 8.0 and dH_2O to 1 litre.

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
	ca. 60 mL

The urea was dissolved by heating the mix to 37^0C and then cooling to room temperature. The sequencing gel solution could be stored at 4^0C for several weeks before use. Before pouring the gel, 300µl of freshly prepared 10% (w/v) ammonium persulphate and 30µl of TEMED were added to 60mL of the stock solution.

Preparation of glass plates and pouring the gel: The plates (40cm X 33cm) were cleaned thoroughly with water and alcohol. One plate was treated with a commercial non-stick baking grease and then assembled using three spacers (0.4mm thick) along the vertical sides and the bottom of the gel. The entire assembly was held in place by clamps. The gel solution was poured using a 60mL syringe down one edge of the plates while tilting the plates at an angle of about 30^0. The plates were then laid at an angle of 5^0 and the sharks tooth combs inserted upside down to give a flat top. The gel polymerized usually within 30 min at room temperature.

Electrophoresis of sequencing gels: The gel was pre-electrophoresed for 30min at a constant power of 60W. Prior to loading, the samples prepared as described in Section 2.1.27 were heated to 95^0C for 5 min, placed on ice and loaded on to the gel. 6% (w/v) gels were run for 1.75-2hrs to read the first 100 nucleotides and for 4.5-5hrs to read up to 400 nucleotides. At the completion of the runs, the plates were disassembled and the gel soaked in 12% methanol, 10% acetic acid for 30 min. The gels were then dried under vacuum for 30 min before being exposed to autoradiographic film.

2.1.29 OLIGONUCLEOTIDE DEPROTECTION

Oligonucloides were synthesized on a Applied Biosystems International (ABI) 391 DNA synthesizer PCR-Mate™. The columns and chemicals were purchased from Cruachem Ltd.

Procedure: The columns were carefully broken open with the use of a spatula and the powdery glass bead support was tapped into a 2.0mL screw-capped tube. One mL of 30% aqueous ammonia was added and the tube left at room temperature for 1-2 hours. The ammonia solution must be fresh, kept tightly sealed and stored at -20^0C. After 2 hours the contents of the tube were mixed and then centrifuged in a microfuge for a few seconds. The supernatant was then transfered to a fresh screw-capped tube. A further 1mL of 30% ammonia was added and the tube sealed firmly with parafilm™. The tube was incubated overnight at 50^0C and allowed to cool before opening. The DNA was then precipitated with 2 volumes of ethanol after making the solution to 0.5M ammonium acetate. Precipitation could be completed at -70^0C in 1 hour. The pellet was washed after spinning for 15 min at 4^0C. The dried pellet was resuspended in 100µl of sterile dH_2O. The appropriate dilutions were made after quantification.

2.1.30 AUTORADIOGRAPHY

Autoradiography was performed in metal cassettes (medical chest X-ray type) using Kodak X-OMATS film. Enhancement in the intensity of the autoradiographic images was obtained when the film was exposed at low temperature (-70^0C) in close contact with a du Pont Cronex Lightening Plus intensifying screen. The X-ray films were developed using a Kodak X-OMAT automatic processor, Model ME-I.

2.1.31 POLYMERASE CHAIN REACTION

The polymerase chain reactions (PCR) were carried out in a Perkin -Elmer Cetus DNA Thermal Cycler. A typical reaction requires:

DNA template	1µl
10X Taq reaction buffer	5µl
20% Glyercol	25µl
2X primers at 100pmol each	2µl
dNTP's (@1.25mM)	8µl
dH_2O	9µl
	ca. 50µl

One drop of oil was added to the tube. The thermal cycler was programmed in a series of files. The first file was set to give one cycle at 94^0C for 4 mins to denature the template. The second file was set to denature at 94^0C for 30 sec., anneal at Tm-5^0C for 30 sec., and extend for 2mins at 72^0C for a total of 30 cycles. The third file was one cycle of denaturing at 94^0C for 30 sec., annealling at Tm-5^0C for 1 min. and extending at 72^0C for 5 mins. The final file cooled the machine to 4^0C for 15 mins.

The Tm was calculated using the following equation (Sambrook; 1989):

$$Tm = 81.5\text{-}16.6(\log_{10}[Na^+]) + [0.41(G+C)]\ \% - \frac{600}{L} = \ ^0C$$

The PCR products were analyzed by running 5µl on an agarose gel.

2.2 STANDARD MICROBIOLOGICAL TECHNIQUES FOR *Streptomyces*

All procedures can be found in "Genetic Manipulation of *Streptomyces*: A laboratory manual" (Hopwood *et al.*, 1985), unless otherwise stated.

2.2.1 STANDARD MEDIA FOR THE PROPAGATION OF *Streptomyces*

Emersons agar: purchased as a powder and prepared as directed by the manufacturer (Difco).

Hopwood's Minimal Medium (HMM): 0.5g L-asparagine, 0.2g K_2HPO_4, 0.2g $MgSO_4.7H_20$, 0.01g $FeSO_4.7H_20$, 10g glucose, 10g agar, made up to 1 litre with distilled water.

Liquid Complete Medium (LCM): 10g glucose, 5g yeast extract, 15g lactalbumin hydrolysates, 2.8g sucrose, 1g calcium chloride, made up to 1 litre with distilled water.

R2 Medium: R2A- 44g agar, 0.5g K_2SO_4, 20.2g $MgCl_2.6H_20$, 5.9g $CaCl_2.2H_2O$, 20g glucose, 6g proline, 0.2g casamino acids, 4ml trace elements solution (Hopwood et al., 1985a), made up to 1 litre in distilled water. R2B- 11.5g MOPS, 10g yeast extract, 203g sucrose, adjusted to pH 7.4 with NaOH, made up to 1 litre in distilled water. Equal volumes of R2A (melted and cooled to 55^{o}C) and R2B plus 1mL of 1% (w/v) KH_2PO_4 per 200mL of agar, were combined prior to use.

R9 Medium: R9A- 44g agar, 0.5g K_2SO_4, 8.2g $MgSO_4.6H_20$, 4.7g $CaCl_2.H_20$, 4g $NaNO_3$, 1g KCl, 0.4g $MgSO_4$, 20g glucose, 0.8g casamino acids, 4ml trace elements solution (Hopwood et al., 1985a), 2ml $FeSO_4$ (1% [w/v] solution), made up to 1 litre in distilled water. R9B- 11.5g MOPS, 10g yeast extract, 410g sucrose, adjusted to pH7.4 with NaOH, made up 1 litre with distilled water. Equal volumes of R9A (melted and cooled to 55^{0}C) and R9B were mixed with 1mL of 1% (w/v) KH_2PO_4 per 200mL of agar, prior to use.

Soya Mannitol Agar (SM): 20g mannitol, 20g soya bean flour, 16g agar, made up to 1 litre using tap water.

Tryptone Soya Broth (TSB): 30g of Oxoid tryptone soya broth powder made up to l litre in distilled water.

Yeast extract-Malt extract (YEME): 3g Difco yeast extract, 5g Difco bacto peptone, 3g Oxoid malt extract, 10g glucose, 340g sucrose, made up to 1 litre in distilled water.

Inoculum Medium for OTC Production: The exact chemical composition of this medium cannot be revealed as it is used commercially by Pfizer for strain improvement.

Tetracycline Soluble Medium (TSM6): glucose, MOPS, $Ca(NO_3)_2.4H_20$, $NaNO_3$, NH_4NO_3, KCl, $MgSO_4.7H_20$, $ZnSO_4.7H_20$, $NH_4Fe(SO_4)_2.12H_20$, $MnSO_4.4H_20$, $CuSO_4.5H_20$, $Na_2MoO_4.2H_20$, and $CaCl_2.6H_20$, (adjusted to pH 6.9 with NaOH) with addition of K_2HPO_4 and KH_2PO_4 prior to use. The exact composition of this medium cannot be revealed as it is used commercially by Pfizer for strain improvement.

2.2.2 PREPARATION OF *Streptomyces* SPORE SUSPENSIONS

Concentrated spore suspensions were required for inoculating liquid cultures of *S. lividans* and *S. rimosus*. The protocol described by Hopwood *et al.*, (1985a) was followed with minor modifications.

Procedure: *S. lividans* and *S. rimosus* strains used during this work sporulated exceptionally well on Soya Mannitol (SM) agar. A boiling tube containing a slant of SM agar (produced by pouring ca. 15mL of molten agar into the tube and allowing it to solidify with the tube held at 5^0 from the horizontal) was inoculated with 150µl of a suspension of spores or mycelial fragments and incubated at 30^0C. After 4-7 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 -20^0C. Spores could then be harvested immediately or stored at -20^0C indefinitely. The spores were harvested by adding 5mL of 20% (v/v) glycerol to the frozen slant and rubbing the surface of the slant with a 10mL glass pipette. When the surface had been scraped clean of spores, the spore suspension was decanted into a sterile universal. The slant was rinsed with an additional 5mL of 20% (v/v) glycerol to remove any spores remaining from the first wash. The spore suspension (ca. 10mL) thus produced was dark grey in colour. Contaminating agar or mycelial fragments were removed by a single passage through a cotton wool filter, as described in Hopwood *et al.*, (1985a). The filtered spore suspension was then divided into aliquots and frozen at -20^0C. Aliquots were thawed and thoroughly mixed using a vortex, prior to use. The titre of the spore suspension was determined after storage at -20^0C overnight by spreading out serial dilutions onto plates containing Emersons agar. Titres remained stable over several months. Solid growth media other than SM agar were conducive to sporulation. R2 agar was used sometimes for *S. lividans* strains and R9 agar was used occasionally for *S. rimosus* strains.

2.2.3 PREGERMINATION OF *Streptomyces* SPORES

Pregermination media (double strength); 1% Difco yeast extract; 1% Difco Casamino acids; autoclave and then add sterile 0.01M $CaCl_2$. 0.05M TES buffer; pH 8.0.

Procedure; A dense spore suspension was prepared and the spores pelleted in a bench top centrifuge. Spores were recovered in 5mL of TES buffer and heated to 50^0C for 10 mins. The tube was then cooled quickly under a cold tap. An equal volume of double strength pregermination media was added and the mixture incubated in a 37^0C shaker for 2-3 hours. The pregerminated

spores were recovered in a bench top centrifuge and resuspended in water or TES buffer. The suspension was agitated on a vortex mixer to disperse the clumps and was then inoculated in growth media to give an OD_{450} of 0.03-0.05.

2.2.4 GROWTH OF *Streptomyces* MYCELIUM

Ehrlenmeyer flasks were used for growing liquid cultures of *Streptomyces*. YEME broth was used for growing *Streptomyces lividans* strains and TSB broth used for growing *Streptomyces rimosus* strains. Cultures required for formation of *S. lividans* or *S. rimosus* protoplasts also contained 0.5% (w/v) glycine and 5mM $MgCl_2$. The volume of the culture depended on the final use of the mycelium. The formation of protoplasts required 30 to 50mL of broth, while 200mL of broth or multiples thereof were used for plasmid isolation. Routinely, 100µl of a dense spore suspension (10^5-10^6 spores) were used to inoculate 50mL of broth. The flask volume was at least five times the volume of the broth to facilitate good aeration when incubated in an orbital shaker (30^0C at ca. 250 rpm).

2.2.5 ANTIBIOTICS AND INDICATORS

The antibiotic concentrations used throughout for both liquid and plate selection were as follows:

DRUG	FINAL CONCENTRATION	STOCK SOLUTION	STORAGE TEMP.
Thiostrepton	50 µg.mL	10mg.mL^{-1} (DMSO)	4^0C
Gentamicin	200 µg.mL	200mg.mL^{-1} (water)	-20^0C
Erythromycin	200 µg.mL	10% solution (EtOH)	4^0C

Stock solutions were added to molten agar, which had been cooled to 55^0C.

2.2.6 PRESERVATION OF *Streptomyces* STRAINS

Streptomyces strains were preserved by freezing agar slants covered in spores at -20^0C. The *Streptomyces* strains used in this work remained viable indefinitely under these conditions. Small slants made in 5mL bijous were used for the long-term storage of strains. However, few spores could be harvested from these slants and it was therefore necessary to inoculate large slants (ca. 15mL) in order to generate sufficient spores for most purposes.

2.2.7 INTRODUCTION OF PLASMID DNA INTO *Streptomyces*

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

Reagents: Trace element solution; Per litre, 40mg $ZnCl_2$, 200mg $FeCl_2.6H_20$, 10mg $CuCl_2.2H_20$, 10mg $MnCl_2.4H_20$, 10mg $Na_2B_4O_7.10H_20$ and 10mg $(NH_4)_6Mo_7O_{24}.4H_20$. **Medium P**; 5.73g N-tris[Hydroxymethyl]methyl-2-aminoethanesulfonic acid (TES), 103g sucrose, 2.93g $MgCl_2.7H_20$, 0.5g K_2SO_4, 3.68g $CaCl_2.2H_20$, 2mL trace element solution. Adjust to pH 7.4 with NaOH and made up to 1 litre in distilled water. **Lysozyme solution**; 10% (w/v) sucrose, 25mM TES buffer (pH 7.2), 2.5mM K_2SO_4, 2mL trace elements (Hopwood et al., 1985a), 2.5mM $MgCl_2$, 2.5mM $CaCl_2$, K_2HPO_4 (0.005% [w/v]) and lysozyme (0.3mg.mL^{-1}) were added immediately prior to use. **PEG solution;** 1g of polyethylene glycol 1540 (supplied by BDH) was melted in a 50^{0}C waterbath and then mixed with 3mL of medium P.

Preparation of protoplasts: 30mL cultures were grown in the appropriate medium at 30^{0}C. The optimal time to harvest the mycelium in order to recover the most competent protoplasts was species-dependent i.e. *S. lividans* mycelium was harvested after 65hr and *S. rimosus* after 40hr. The mycelium was pelleted at 3000g for 15 minutes and washed twice in 10.3% (w/v) sucrose. The pellet was then resuspended in 4mL of lysozyme solution and incubated at 37^{0}C for 15-30 min. The formation of protoplasts was monitored using a microscope, the reaction terminated by adding 5mL of P medium and the protoplasts triturated twice. The protoplasts were then filtered through cotton wool (Hopwood *et al.*,, 1985a), pelleted using a centrifuge (3000g for 15 minutes) and washed twice in P medium. Finally, they were resuspended in 4mL of medium P, dispensed into 200µl aliquots and frozen at -70^{0}C.

Transformation of protoplasts: The protoplasts were thawed at 37^{0}C and placed on ice. DNA was added in a volume of less than 10µl and the mixture incubated on ice for 30s. PEG solution (400µl) was added, the solution incubated for a further 1min on ice and finally, medium P (800µl) was added. Dilutions of the transformation mix were then made in medium P.

Regeneration of transformed protoplasts: The method used to prepare the regeneration medium for the protoplasts was standardised. The medium was stored in two parts, RA and RB ; the former solid and the latter liquid. Both parts were placed in a steam bath until the RA portion melted. The two components were allowed to cool to 50^0C before they were combined and 1mL of 1%(w/v)K_2HPO_4 (pH 7.0) added. The complete regeneration medium was mixed by swirling and then poured into petri dishes (diameter 9cm). On average 8 plates were obtained from 200mL of medium. The plates were dried in a laminar flow hood to minimise air borne contamination by leaving the lids half open for 90min. The plates were covered and left overnight at 30^0C. The next day, any plates which showed any signs of contaminating growth were discarded and the rest used for the regeneration of protoplasts.

The only drug resistance used for plasmid selection after transformation into *Streptomyces* was thiostrepton (obtained from E.R. Squibb, New Jersey, USA). It was dissolved in DMSO to make a 1% (w/v) stock solution. Transformed protoplasts were selected after 16-22 hrs of non-selective growth at 30^0C by overlaying the regeneration plates with 1mL of a 220μg.mL^{-1} thiostrepton solution in 10.3% (w/v) sucrose solution.

S. lividans protoplasts were regenerated on R2 agar plates and R9 plates were used for the regeneration of *S. rimosus*.

2.2.8 ELECTROTRANSFORMATION OF *Streptomyces* STRAINS

2.2.8.1 PREPARATION OF MYCELIUM

Electrotransformation of *Streptomyces* following the procedure outlined by Pigac and Schrempf (1995).

Reagents: CRM; 10 g Glucose, 15g Tryptic(ase) soya-broth, 103g Sucrose, 10.15g $MgCl_2.6H_2O$, 5g yeast extract. **ET-Buffer;** 30% PEG 1000(or 1500), 10% glycerol, 6.5% Sucrose.

Preparation of Mycelium: 100 mLs of CRM or YEME + Glycine was inoculated with 1/100 volume of a fresh 24hr culture. The mycelium was then allowed to grow for 48hrs at 30^0C with vigorous shaking. To harvest, the flask was chilled on ice for 15-30 mins, and centrifuged in a cold rotor (10,000 rpm) for 15 mins. The pellets were resuspended in a total of 100mL of a cold, low ionic strength wash medium such as 10% sucrose. The mycelium was again pelleted and washed in 50 mL of ice-cold 10% sucrose. The pellet was washed in 20mL of 15% ice-cold glycerol, and the resuspended in a final volume of 10mL of 15% glycerol. Lysozyme was added to give a final

concentration of 400μg/mL (for *S. rimosus*) and the tube was incubated at 37^{0}C for 30 mins. The mycelium was washed once again in 20mL of 15% glycerol and resuspended in a final volume of 1-5 mL of E-T buffer. 50μl aliquots could be stored by immediately freezing in a dry ice/MeOH bath before storing at -70^{0}C.

Electrotransformation: The cells were thawed at room temperature before placing them on ice. DNA (1-2μl in a low ionic strength buffer, such as TE) was added to the cell suspension and the mixture allowed to stand on ice for 1 minute. The Gene Pulser (Biorad) apparatus was set to 25μF and 2.0kV. The pulse controller was set at 400Ω. The mixture of cells and DNA was transfered to a cold, electroporation cuvette (0.2cm size). The cuvette was the placed in the chilled safety chamber slide and the slide pushed into the chamber until the cuvette was set between the contacts in the base of the chamber. One pulse was administered with a time constant of 6-9msec, with a field strength of 10kV/cm. The cuvette was removed from the chamber and 750μl of CRM was added to the cuvette. This cell suspension was then transferred to a bijou bottle and incubated at 30^{0}C with shaking for 4 hours. After incubation, 200μl of CRM was added to the bijou bottle and mixed. The suspension was then ready to plate on selective media, either TSB or SM, with 50μg/mL of thiostrepton.

2.2.9 CONJUGATION OF *Streptomyces*

2.2.9.1 Conjugation with Spores

Fresh *Streptomyces* spores were pregerminated (See Section 2.2.3), pelleted by centrifugation and resuspended in 2YT to a concentration of 10^{8}spores/mL. The *E. coli* donor was prepared by overnight incubation at 37^{0}C in 2YT supplemented with the appropriate antibiotic. The cells were washed twice in 2YT before resuspending in the original volume of 2YT. An equal volume of donor and recipient were mixed and plated on LB plates. The plates were incubated at 30^{0}C from 6 hrs to overnight (duration of incubation was strain-dependent). The plates were flooded with 1mL of LB broth and the *E. coli* was carefully scraped off the plate, ensuring that the streptomycetes remained attached to the agar. The plates were then overlayed with 100μg of neomycin and 200μg of nalidixic acid. Exconjugants were counted 6-10 days later.

2.2.9.2 Conjugation with Mycelium

The *Streptomyces* mycelial cultures were inoculated into TSB supplemented with glycine and $MgCl_2$, to weaken the cell walls, and grown at 30^0C for 72 hrs. Two mLs of culture was homogenised and transferred to 18 ml of fresh medium. This was grown at 30^0C for a further 16 hrs. The culture was again homogenised, 1 mL transferred to 9 ml of fresh medium and grown for a further 3 hrs. The mycelium was recovered by centrifugation and washed once in TSB medium and resuspended in a final volume of 2 mL. The *E. coli* donor was prepared as described above. Equal volumes of donor and recipient were mixed and plated onto TSB plates for *S. rimosus* and LB plates for *S. lividans*. The plates were incubated overnight at 30^0C and then covered with 3-4 mLs of soft nutrient agar containing 1.5 mg nalidixic acid and 1.5 mg of thiostrepton. Dilutions of the recipient and donor cells were plated as controls.

2.2.10 ISOLATION OF PLASMID DNA FROM *Streptomyces*

Plasmid isolation by Alkaline Lysis was outlined by Kieser (1984) without modification.

Reagents: **Lysozyme solution**; 2mg/ml lysozyme (Sigma) in 0.3M sucrose, 25mM Tris (pH 8.0), 25mM EDTA, 0.02% bromocresol green, 50µg/ml RNase (heat treated). **Alkaline SDS**; 0.3M NaOH, 2% SDS. **Acid Phenol/Chloroform**; 5g phenol, 5mL chloroform, 1mL H_2O, 5mg 8-hydroxyquinoline (AnalaR), **Neutral Phenol/Chloroform**; acid phenol/ chloroform equilibrated first with 0.5 vol 1M Tris pH 8.8 and then 1vol 0.1M Tris pH 8.0. **Unbuffered 3M Sodium Acetate, 100mM Spermine-HCl;** sterilized by filtration. **Extraction Buffer**; 0.3M sodium acetate, 10mM $MgCl_2$. **TE Buffer**; 10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0.

Procedure for Standard version: The standard version is based on the volume of bacteria spun down in a 1.5mL tube. The bacteria were resuspended in a total volume of 500µl of lysozyme solution. The tube was incubated at 37^0C for 30 min. After incubation, 250µl of the alkaline SDS solution was added. Complete and immediate mixing was essential. The tubes were then incubated at 70^0C for 15 min or 30min at 55^0C for plasmids larger than 20 kb. The tubes were allowed to cool before adding 80µl of acid phenol/chloroform and mixing thoroughly. A 2 min spin separated the phases. The upper aqueous phase was then removed to a fresh tube containing 70 µl of 3M unbuffered sodium acetate and 700µl of isopropanol. The tube was mixed by inversion and allowed to stand at room temperature

for 5 mins before being spun for 2 min. After centifugation, all of the liquid was removed and the pellet dissolved was in 50µl of TE. Unbuffered 3M sodium acetate (5µl) was added to the tube along with 25µl of neutral phenol/chloroform. The tube was vortexed to mix the phases and spun for 2 min. The upper phase was then transferred to a fresh tube containing 50µl of isopropanol. The DNA was again precipitated by spinning for 2 min. All of the liquid was removed and the pellet dissolved in 500µl of TE buffer and 25µl of 100mM spermine-HCl. The mixture was allowed to stand at room temperature to precipitate the DNA and centrifuged again for 2 min to remove all the liquid. The pellet was then dispersed in 300µl of extraction buffer, 700µl of ethanol was added, mixed and the tube left at room temperature for 60 min. A final centrifugation of 5 min pelleted the DNA. The pellet was resuspended in a final volume of 50µl of TE. This procedure was easily scaled up 10-fold.

The typical yield of high-copy-number plasmids such as pIJ486 from *Streptomyces* was 2-5µg of plasmid. Plasmid prepared in this way could be used for most *in vitro* manipulations.

Large-scale plasmid preparations: Reagents: as defined in Section 2.1.9.

Procedure; 200mL cultures of stationary phase cells were harvested using a centrifuge (12000g, 5min at 4^0C). The pellet was resuspended in 4mL of Birnboim-Doly I solution. At this stage, *Streptomyces* preparations were treated with lysozyme. The amount of lysozyme that resulted in the optimum recovery of plasmid was determined for each strain. *S. lividans* and *S. rimosus* strains were incubated at 37^0C for 20min with 4mg.mL^{-1} and 1mg.mL^{-1} lysozyme solutions, respectively. Birnboim-Doly II solution was added (8mL) and the solution left on ice for 5-10min. Six mL of cold Birnboim-Doly III solution was then added. The suspension was mixed gently and left on ice for 15-30 min. The cell debris and most of the chromosomal DNA was removed by centrifugation (32000g, 5 min at 4^0C). The remainder of the procedure is identical to the isolation of *E. coli* DNA, and can be found in Section 2.1.9. Large amounts of pure plasmid DNA (up to 100µg from *Streptomyces* cultures) can be obtained, suitable for all *in vitro* manipulations.

2.2.11 ISOLATION OF CHROMOSOMAL DNA FROM *S. rimosus*

The small scale total DNA preparation developed by S. Fisher (1984) starts with 50 mg of wet mycelia in a 1.5ml microfuge tube. The mycelia was

resuspended in 500µl of a lysozyme solution containing 0.3M sucrose, 25mM Tris (pH 8.0), 25mM EDTA (pH 8.0), 50µg/mL RNase and 4 mg/mL lysozyme. It was incubated at 37^{0}C until the cells cleared, approximately 1 hour. After 1hr, 250µl of 2% SDS was added and the mixture vortexed for 1 min. 250µl of a neutral 1:1 phenol/chloroform mixture was added to the tube and again vortexed for 30 seconds. The phases were separated by spinning at 13,000 rpm for 4min in a microcentrifuge. The phenol/chloroform extracts were repeated until no interface was seen. The DNA was precipitated by adding 0.1 volume of 3M NaOAc pH 4.8 and 1 volume of isopropanol. The solution was mixed thoroughly, left at room temperature for 5 min and then spun 10 min at 13,000 rpm. The DNA was resuspended in TE (10:1).

2.2.12 ISOLATION OF TOTAL RNA FROM *S. rimosus*

This procedure is based on that of Kirby *et al.,* (1976) with extensive modifications by Covey and Smith (see Hopwood *et al.,* , 1985a). RNAase is a very persistent enzyme and precautions were taken against contamination of equipment and solutions. Prior to use, all glassware was incubated overnight at 200^{0}C. All solutions were prepared from DEPC-treated water and chemicals reserved for RNA work.

Reagents: Phenol; as described previously. **Phenol.Chloroform;** as described previously. **Kirby mixture;** 1g tri-isopropylnapthalene sulphonate, 6g 4-amino salycilate (Na salt), 50mM Tris-HCl (pH 8.3), 6mL phenol, made up to 100mL in distilled water.

Procedure: RNA was always isolated from 50mL liquid cultures of the appropriate organism, grown in 250mL Ehrlenmeyer flasks. When the culture reached the appropriate stage of growth, the contents were immediately decanted into a 50mL flask containing 20mL of dH_20, which had been frozen at -20^{0}C , rapidly lowering the temperature to ca. 0^{0}C. As much of the chilled culture as possible, including any residual ice, was then transferred to an unused polypropylene centrifuge tube. The mycelium were then harvested using a centrifuge in a prechilled rotor. The pellet was resuspended in 5mL Kirby mix (prechilled on ice) containing 12g of 0.45mm diameter glass beads and agitated thoroughly on a vortex mixer for 1-2 min. Five mL of phenol.chloroform (stored at room temperature) was then added and the mixture agitated for a further 1-2 min. The homogenate was then

spun in the centifuge (9000g, 4^0C for 10 min) to separate the phases. The aqueous phase was removed, using a baked Pasteur pipette, to a fresh centrifuge tube containing 5mL of phenol.chloroform and mixed for 1 min on the vortex mixer. The phases were then separated as before. At this stage, the nucleic acid solution could be stored at -70^0C without any degradation being detectable. To remove DNA from the preparation, the RNA was pelleted by centrifugation in caesium chloride .

Pelleting the RNA through a dense caesium chloride solution: Homogenate (ca. 3.5mL) from the Kirby lysis was layered carefully on top of a 5.7M CsCl-0.1M EDTA cushion (1.5mL, density 1.707g.mL^{-1}) in a ultracentrifuge tube (5mL, Beckman, Polyallomer™). The RNA was then pelleted by centrifugation (35,000rpm, 20^0C for 16 hours) using a Beckman SW50.1 rotor. After centrifugation, the supernatant was removed using a baked pasteur pipette. The tube was then inverted to drain away the remaining fluid. Using a fresh scalpel blade to cut the centrifuge tube, the bottom was isolated in the form of a small cup. The pellet was dissolved in 400µl of dH_20 and extracted with half volumes of phenol.chloroform and chloroform. The RNA was precipitated by the addition of 1/50 volume of 5M NaCl and two volumes of absolute ethanol, and harvested by centrifugation in a microfuge (12,000g, 4^0C for 10 min). The RNA pellet was washed twice with 70^0C (v/v) ethanol, dried in the open microfuge tube at room temperature and dissolved in 100 to 200µl of dH_20. The RNA concentration was estimated spectrophotometrically, reprecipitated and stored at -70^0C.

2.2.13 TRANSCRIPT MAPPING USING COMPLEMENTARY SINGLE-STRANDED DNA PROBES

Mapping of RNA transcripts was undertaken using single-stranded probes derived from M13 (Calzone *et al.*, , 1987), as described by Sambrook *et al.*, (1989).

Reagents: 1X Hybridization buffer: 40mM PIPES adjusted to pH 7.4 with NaOH, 400mM NaCl, 1mM EDTA, 80% (v/v) formamide (de-ionized with mixed-bed resin). Stored at -70^0C. **10X S1 nuclease digestion buffer:** 2.8M NaCl, 500mM $NaCH_3COO$ (pH 4.5), 45mM Tris.HCl (pH 7.8), 500mM KCl and 100mM EDTA (pH 8.0). **Denatured Salmon sperm DNA:** prepared by dissolving in dH_20, extracting any protein with phenol.chloroform, shearing by passage through a narrow gauge hypodermic needle and precipitating the DNA with ethanol. The fragmented DNA was then redissolved in dH_20 to a

concentration of 10mg.mL^{-1}, boiled for 10 min and stored at -20^{0}C in small aliquots. Prior to use, the DNA was denatured partially by heating to 100^{0}C for 5 min and chilling rapidly on ice. **Carrier tRNA:** prepared by dissolving commerically-available yeast tRNA in dH_20, extracting any protein with phenol.chloroform, precipitating and redissolving in dH_20 to a final concentration of 10mg.mL^{-1}. **S1 nuclease stop buffer:** 500mM Tris.HCl (pH 8.0), 125mM EDTA.

Synthesis of Probe: 1.0 pmole of template ssDNA was mixed with 6 pmoles of an oligonucleotide primer in a volume of 16µl, containing 2µl of 10X BRL REact 2 buffer, and then annealed by heating to 85^{0}C for 5 min and cooled slowly to 37^{0}C. The extension reaction was then carried out at 37^{0}C for 30 min by adding:

	Volume (µl)
100mM dithiothreitol	2
50µCi of [α-^{32}P] dATP (sp. act. 3000 Ci.mmole^{-1})	5
40µM solution of dATP	1
20µM solution of dCTP, dGTP and dTTP	1
Klenow (1 U.µl^{-1})	5

The above reaction was then "chased", by adding 1µl of a 20mM dATP solution and incubating for a further 20 min. After synthesis, the reaction was heated to 68^{0}C for 10 min to inactivate the Klenow enzyme. The probe was then digested at a restriction site, within or beyond the insert, by the addition of 20µl of restriction digest solution containing the appropriate restriction endonuclease (20U) and buffer. The reaction was incubated for 2 hours at the optimal temperature for the enzyme. The nucleic acid was then precipitated with ethanol and recovered by centrifugation in a microfuge.

Hybidization: The single-stranded probe (ca. 0.1-0.2 pmol) and the *Streptomyces* RNA (5-20µg) were co-precipitated, washed with 70% (v/v) ethanol and dried by leaving the tube open on the bench, until the last traces of ethanol had evaporated. The probe/RNA pellet was resuspended in 20µl of 1X hybidization buffer. The solution was pipetted up and down many (20-30) times and heated to 60^{0}C for 30 min. to ensure that the pellet was dissolved completely. The hybridization solution was then heated to 85^{0}C for

10 min, to denature the nucleic acids and incubated at 55-65^{0}C overnight. After overnight incubation, the hybridization mixture was treated with S1 nuclease.

S1 Nuclease digestion: 280µl of ice-cold 1X S1 digestion buffer, containing 10µg of partially-denatured Salmon sperm DNA, was added to the hybridization reaction and mixed by inverting the microfuge tube several times. As quickly as possible, 250 units of S1 nuclease (1-2 µl) were added, the reaction mixture pooled by centrifugation and incubated at 37^{0}C for 30 min. The reaction was terminated by the addition of 75µl stop buffer and 10µg carrier tRNA. The reaction mixture was then extracted with phenol.chloroform and chloroform, ethanol precipitated, washed with 70% (v/v) ethanol, dried *in vacuo* and resuspended in 7µl 1X sequencing loading buffer. The sizes of products were analysed by electrophoresis in 1% (w/v) agarose gels.

COMPUTER-ASSISTED SEQUENCE ANALYSIS

The DNA sequences in this thesis were analysed using the "Sequence Analysis Software Package, Version 7.3" from the Genetics Computer Group at the University of Wisconsin (UWGCG) (Devereux *et. al*, 1984). See "Program Manual for GCG" for references.

CODON PREFERENCE: is a frame-specific gene finder that tries to recognize protein coding sequence by virtue of the similarity of their codon usage to a codon frequency table or by the bias of their composition (usually GC) in the third position of each codon.

FASTA: searches for similarity between a query sequence and any group of sequences using a Pearman and Lipman algorithm (1985).

FOLD: finds an optimum secondary structure for RNA molecules using the method of Zucker (1981).

FRAME analysis: uses the algorithm of Bibb *et. al.* (1984) to find potential protein coding regions for nucleotide sequences of *Streptomyces*.

GAP: uses the algorithm of Needleman and Wunsch to find the alignment of two complete sequences that maximises the number of matches and minimises the number of gaps.

PILEUP: creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment.

STEMLOOP: finds inverted repeats within a sequence which could form secondary structure.

TFASTA: translates nucleotide sequences held in a database in all six frames, before searching for any sequence similarity to a query peptide sequence using a Pearman and Lipman algorithm (1985).

2.3 BACTERIAL STRAINS AND VECTORS

The bacterial stains, plasmids and bacteriophages used in this study are listed in Tables 2.1, 2.2 and 2.3 respectively.

TABLE 2.1 Bacterial strains

STRAIN	GENOTYPE	REFERENCE/SOURCE
***Escherichia coli* strains**		
CB51	*dam*-3, *ara*-14, *thi*-1, Δ(*lac-pro*)	C. Boyd, University of Edinburgh
DS649	R751 in C600	D. Sherratt, University of Glasgow
DS801	*thr*-1,*leu*6, *thi* 1, *lac*YI, *gal*K2, *ara*14, *xyl*1-5, *mtl*1, *pro*A2, *his*4, *arg*E2, str31, tsx33 Lambda$^-$Lambda^S	D. Sherratt, University of Glasgow
DS941	*recF*143, *proA*7, *str*31, *thr*1, *leu*6, *tsx*33, *mt12*, *his*4, *argE*3, *lacY*$^+$, *lacZ*ΔM15, *lacI*^{q}, *galK*2, *ara*14, *supE*44, *xyl*5.	D. Sherratt, University of Glasgow
GM31	*thr* -1, *ara*-14,*leuB*6, *tonA*31, *lacY*1, *tsx*-78, *glnV*44, *galT*22, *thi*-1, *rpsL*136, *hisG*4, *mtl*-1, *xyl*-5,*dcm*-6 (F$^-$)	Marinus *et. al.*. (1994)
ET12567	F$^-$ *dam*-13::Tn 9, *dcm*-6, *hsdM*, *hsdR*, *recF*143, Tn 10, *galk*2, *galT*22, *ara*-14, *lacY*1, xyl-5, *leuB*6, *thi*-1, *tonA*31, *rpsL*136 *hisG*4, *tsx*-78, *mtl*-1, *gln*44	MacNeil (1992)
TG1	*supE*, *hsd*Δ5, *thi*, Δ(*lac-proAB*), F$'$ [*traD*36, *proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15].	Gibson (1984)
***Streptomyces lividans* strains**		
TK24	*str*6.	Hopwood *et al.*, (1985a)
TK64	*pro*2, *str*6.	Hopwood *et al.*, (1985a)
***Streptomyces rimosus* strains**		
M4018	OTC^R	Rhodes *et al.* (1981)
MV9	OTC^S	J. Pigac (Pliva Inc.)

TABLE 2.2 Plasmids

PLASMID	DESCRIPTION*	REFERENCE/SOURCE
pGEM7		Promega
pGM160		Muth *et al.*, (1989)
pIBI24/25	--	Dente *et al,* (1983)
pIJ486	--	Ward *et al.*, (1986)
pIJ680	--	Hopwood *et. al.* (1985)
pIJ699	--	Kieser and Melton (1988)
pIJ702	--	Hopwood *et. al.* (1985)
pLST14	--	Skeggs *et. al.* (1987)
pLUC801		C. Chen (pers. comm.)
pPM803		Mazodier *et. al.* (1989)
pPZ103		Pfizer UK Ltd.
pPZ105		Pfizer UK Ltd.
pT7BLUE(R)		Novagen
pUC18/19	--	Yanisch-Perron *et al.*, (1985)
pXED3-13		Ingram *et. al.* (1989)
pSRG1	5.4kb *Sst*I$_{24\text{-}29}$ fragment from pPZ103 in pIBI24	Chapter 3
pSRG2	1.7kb *ermE* fragment in the *Kpn*I site of pSRG1	
pSRG3	5.9kb *Pst*I$_{17\text{-}25}$ fragment from pPZ103 in pIBI24	Chapter 3
pSRG4	2.2kb *Kpn*I$_{19\text{-}23}$ fragment from pPZ105	Chapter 3
pSRG5	2.1kb *Sst*I fragment from pSRG3	Chapter 3
pSRG6	3.6kb *Sst*I- *Sph*I fragment from pSRG1	Chapter 3
pSRG6X	*Kpn*I site converted to *Xba*I site in pSRG6	Chapter 6
pSRG7	1.5kb *Kpn*I- *Sph*I fragment from pSRG1	Chapter 3
pSRG9	1.8kb *Sal*I- *Sph*I fragment from pSRG1	Chapter 3
pSRG10	0.7kb *Pvu*II- *Sst*I fragment from pSRG3	Chapter 3
p501		Chapter 6
p502		Chapter 6
p503		Chapter 6
p504		Chapter 6

p505	Chapter 6
p506	Chapter 6
p507	Chapter 6
pmrA	Chapter 6
pmrB	Chapter 6
pCMR	Chapter 6
p509	Chapter 6
p510	Chapter 6
p511	Chapter 6
p512	Chapter 6

*the numbering system used to designate restriction sites in this thesis is not that of Butler *et al.*, (1989)

TABLE 2.3 Bacteriophages

BACTERIOPHAGE	DESCRIPTION	REFERENCE/SOURCE
M13mp18/19	--	Yanisch-Perron *et al.*, (1985)
18-SRG1	1.5kb *Pstt*I-*Kpn*I fragment in M13mp18	Chapter 3
19-SRG1	1.5kb *Pstt*I-*Kpn*I fragment in M13mp19	Chapter 3
18-SRG1.1	0.6kb *Sst*I- *Pst*I fragment in M13mp18	Chapter 3
19-SRG1.1	0.6kb *Sst*I- *Pst*I fragment in M13mp19	Chapter 3
18-SRG5	0.22kb *Sst*I- *Kpn*I fragment in M13mp18	Chapter 3
19-SRG5	0.22kb *Sst*I- *Kpn*I fragment in M13mp19	Chapter 3
18-SRG5.1	0.9kb *Sph*I- *Sst*I fragment in M13mp18	Chapter 3
19-SRG5.1	0.9kb *Sph*I- *Sst*I fragment in M13mp19	Chapter 3
18-SRG6	0.9kb *Pvu*II- *Sal*I fragment in M13mp18	Chapter 3
19-SRG6	0.9kb *Pvu*II- *Sal*I fragment in M13mp19	Chapter 3
18-SRG7	0.35kb *Kpn*I-*Pvu*II fragment in M13mp18	Chapter 3
19-SRG7	0.35kb *Kpn*I-*Pvu*II fragment in M13mp19	Chapter 3
19-SRG8	1.8kb*Sal*I- *Sph*I fragment in M13mp19	Chapter 3
18-SRG9	1.3kb *Sph*I- *Pst*I fragment in M13mp18	Chapter 4
19-SRG9	1.3kb *Sph*I- *Pst*I fragment in M13mp1	Chapter 4
18-PCR1	0.28kb fragment P1-P9 inclusive, in M13mp18	Chapter 4
19-PCR1	0.28kb fragment P1-P9 inclusive, in M13mp18	Chapter 4

Chapter 3

Deduction of Functions of Genes Surrounding the *otcD* locus

3.1 Introduction

Rhodes *et al.,* (1981) isolated a series of OTC non-producing mutants. The basis of classification with respect to the biochemical pathway is shown in figure 1.10. The pathway clearly shows more steps are involved in the synthesis of OTC, than the number of mutant classes recovered. Some mutants would not be selected in the screening process used by Rhodes *et al.,* (1981) because an antibiotic-negative phenotype was sought. This was demonstrated when no mutants which were unable to hydroxylate DHTC in the 5-position were found. A blockage at this step would have produced the shunt metabolite, tetracycline. An intensified investigation into the cloning and physical characterisation of the *otc* cluster by Butler *et al.,* (1989) resulted in the cloning of two resistance determinants *otrA* and *otrB*, and led subsequently to the isolation of the biosynthetic genes and the formation of a genetic map (figure 1.11). The precise location and function of each locus is described elsewhere (see Chapter 1.10.1). The physical characterisation of the cluster, along with hybridisation studies which compared the ACT and OTC clusters was then sufficiently advanced to characterise the genes further by DNA sequencing.

A concerted effort was then undertaked in Iain Hunter's laboratory (Glasgow) to sequence the OTC cluster. Chapter 1 gives a detailed description of all the genes uncovered to date. A 3.0 kb region between the *otcX* and *otcD* loci is the subject of the current chapter. No OTC non-producing mutant mapped to this region and, consequently, no preliminary clues into possible functions could be gained. Hence the genes were to be sequenced. Nucleotide sequence and deduced amino acid sequences would be compared with other known sequences and possible gene functions assigned. It was hoped that knowledge of gene functions would lead to a better understanding of the genes in the OTC biosynthetic pathway of *S. rimosus.*

3.2 Results and Discussion

3.2.1 Nucleotide Sequencing of the region: *Kpn*I$_{23}$ to right of *Kpn*I$_{26}$

A summary of the strategy for sequencing of the 3.0 kb DNA fragment extending from *Kpn*I$_{23}$ to right of *Kpn*I$_{26}$ in the OTC cluster is shown in figure 3.1

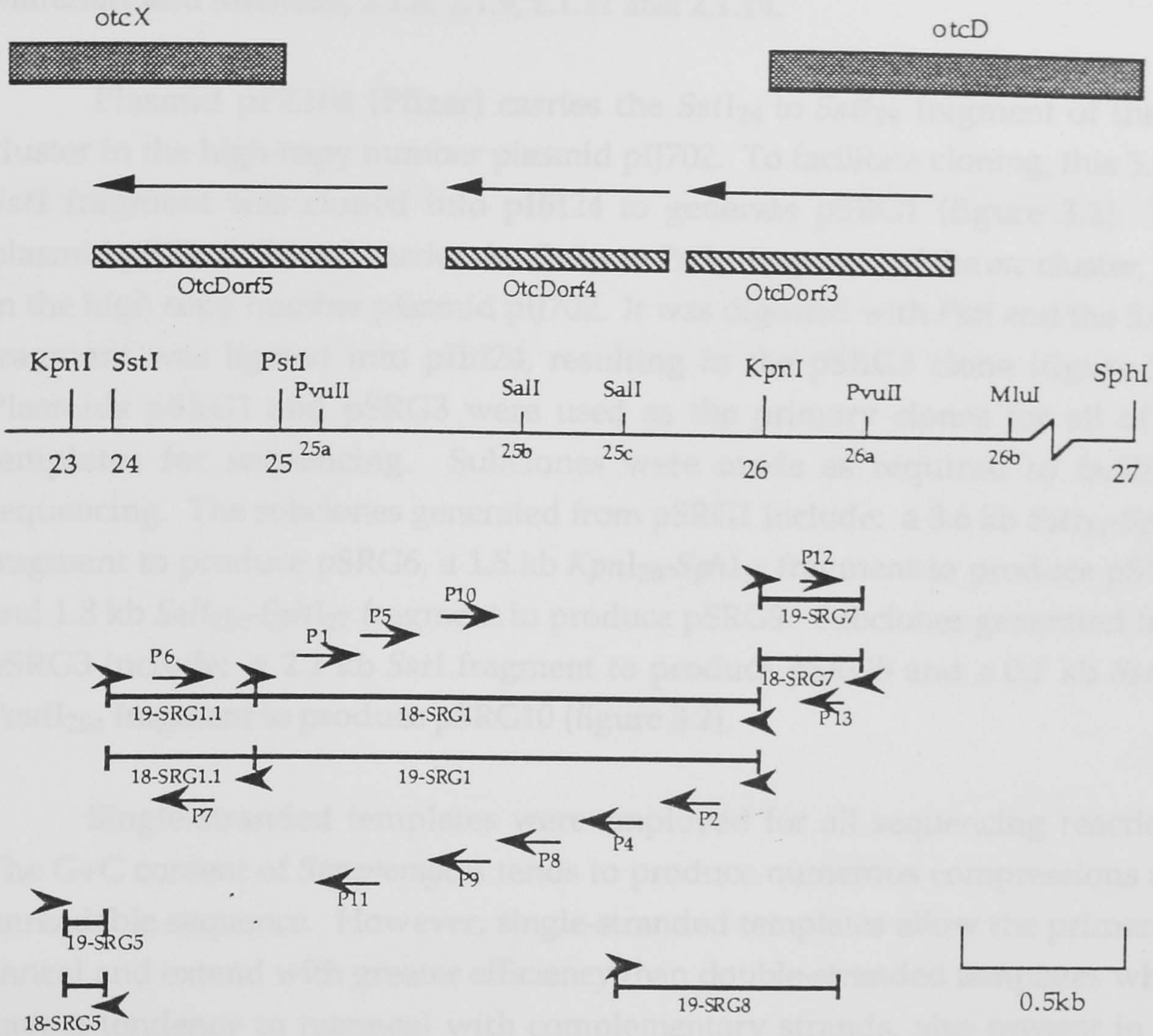


Figure 3.1
M13 templates used to sequence the *Kpn*I$_{23}$ to *Mlu*I$_{26b}$ region of the *otc* cluster. Above the restriction map are the (▭) regions of the *otc* cluster which complement the *otc* mutants (Butler *et al.*, 1989; Rhodes *et al.*, 1981), and the deduced open reading frames (▨) resulting from analysis of sequence. The long arrows above the deduced ORFs show the direction of translation. The template designations are given below the lines and the short arrows indicate the direction and length of primed synthesis during the nucleotide sequence analysis. Arrowheads alone represent the universal primer of the sequence.

The subcloning of DNA fragments was carried out routinely, as described in Materials and Methods, 2.1.6, 2.1.9, 2.1.11 and 2.1.14.

Plasmid pPZ104 (Pfizer) carries the *Sst*I$_{24}$ to *Sst*I$_{29}$ fragment of the *otc* cluster in the high-copy number plasmid pIJ702. To facilitate cloning, this 5.4 kb *Sst*I fragment was cloned into pIBI24 to generate pSRG1 (figure 3.2). The plasmid pPZ105 (Pfizer) carries the *Pst*I$_{17}$ to *Pst*I$_{25}$ fragment of the *otc* cluster, also in the high copy number plasmid pIJ702. It was digested with *Pst*I and the 5.0 kb fragment was ligated into pIBI24, resulting in the pSRG3 clone (figure 3.2). Plasmids pSRG1 and pSRG3 were used as the primary clones for all of the templates for sequencing. Subclones were made as required to facilitate sequencing. The subclones generated from pSRG1 include: a 3.6 kb *Sst*I$_{24}$-*Sph*I$_{27}$ fragment to produce pSRG6, a 1.5 kb *Kpn*I$_{26}$-*Sph*I$_{27}$ fragment to produce pSRG7 and 1.8 kb *Sal*I$_{25c}$-*Sph*I$_{27}$ fragment to produce pSRG9. Subclones generated from pSRG3 include: a 2.1 kb *Sst*I fragment to produce pSRG5 and a 0.7 kb *Sst*I$_{24}$-*Pvu*II$_{25a}$ fragment to produce pSRG10 (figure 3.2).

Single-stranded templates were employed for all sequencing reactions. The G+C content of *Streptomyces* tends to produce numerous compressions and unreadable sequence. However, single-stranded templates allow the primers to anneal and extend with greater efficiency than double-stranded templates which have a tendency to reanneal with complementary strands, also present in the reaction, rather than the primer. The largest contiguous template generated was a 1.5 kb *Pst*I$_{25}$-*Kpn*I$_{26}$ fragment cloned into both M13mp18 and M13mp19 (18-SRG1 and 19-SRG1 respectively). Primers P1, P5, P10 and the universal (figure 3.1) were used to obtain sequence from 18-SRG1 and primers P2, P4, P8, P9 and the universal (figure 3.1) were used to obtain sequence from 19-SRG1. A further series of templates was created to confirm sequence within this region. These include, 18-SRG6 and 19-SRG6 (0.9 kb *Pvu*II$_{25a}$-*Sal*I$_{25b}$ fragment) and 19-SRG8 (1.8 kb *Sal*I$_{25c}$-*Sph*I$_{27}$ fragment). To the left of this region, the templates 18-SRG1.1 and 19-SRG1.1 were generated which contain a 0.55 kb *Sst*I$_{24}$-*Pst*I$_{25}$ fragment. Primary sequence with the universal primer was later confirmed using P6 oligo on 18-SRG1.1 and P7 oligo on 19-SRG1.1. A 0.22 kb *Kpn*I$_{23}$-*Sst*I$_{24}$ fragment (18-SRG5 and 19-SRG5), was sequenced through completely using the universal primer on both strands. Templates 18-SRG5.1 and 19-SRG5.1 (containing a 0.9 kb *Sph*I$_{22}$-*Sst*I$_{24}$ fragment) confirmed this sequence. The right hand region was sequenced off a series of templates which carry the 0.35 kb *Kpn*I$_{26}$-*Pvu*II$_{26a}$ fragment. The resulting templates were 18-SRG7 and 19-SRG7. The universal

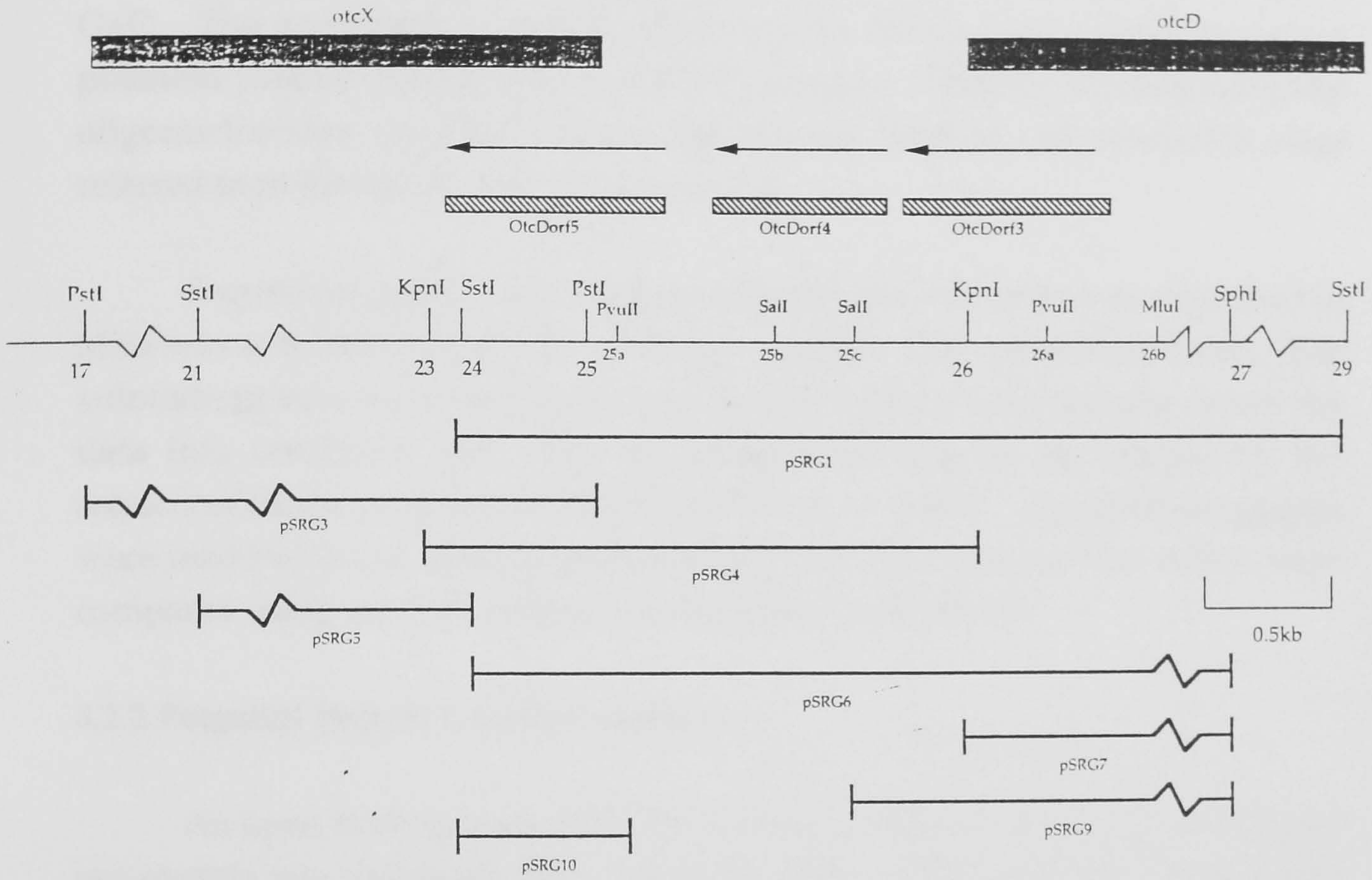


Figure 3.2
Subclones generated from the *Kpn*I$_{23}$ to *Mlu*I$_{26b}$ region of the *otc* cluster. Above the restriction map are the ( ) regions of the *otc* cluster which complement the *otc* mutants (Butler *et al.*, 1989; Rhodes *et al.*, 1981), and the deduced open reading frames ( ) resulting from analysis of sequence. The long arrows above the deduced ORFs show the direction of translation. The subclone designations are given below the lines.

primer was used initially and the sequence was confirmed with P12 and P13 primers. The region to the right of the *Pvu*II$_{26a}$ site was initially sequenced by K. Linton and confirmed with a long run using P12 primer and 19-SRG7.

The sequence consists of 3012 bp with a base composition of 72.8 mol% G+C. The annotated sequence, including the amino acid sequences of the potential protein coding regions (PPCRs), putative ribosome binding sites, the oligonucleotides used for sequencing and the locations of restriction sites referred to in the text are shown in figure 3.3.

Sequencing reactions and polyacrylamide gels were set up as described in Materials and Methods (2.1.27). All the templates were sequenced twice. The autoradiographs were read using a gel digitiser which automatically enters the data into computer files. The sequences were aligned and edited by the sequence/editor program SEQED (Devereux *et. al.,* 1984). The autoradiographs were read twice and entered independently into two separate files which were compared using the GAP program to eliminate "input" errors.

3.2.2 Potential Protein Coding Sequences

An open reading frame (ORF) is defined as a string of triplets, which does not contain any translation stop codons (i.e. UAA, UGA or UAG). Only ORF's which include an 'in-frame' translation start codon at the 5' end can be considered as potential protein coding regions (PPCRs). In *Streptomyces*, as in *E. coli,* the most common initiation codon is AUG (82%). Some genes in *Streptomyces* initiate with the codon GUG. With the high G+C content, it is understandably more frequent, being present in approximately 18% of genes compared to the 3% in *E. coli.* Translational initiation at other codons has not been identified in *Streptomyces* genes. However, in *E. coli,* a very small number of proteins are initiated by the UUG codon and the AUU codon (Gold *et al.,* 1981; Stromo; 1986). Of the *Streptomyces* genes which have been sequenced completely, 67% are terminated by UGA, 29% by UAG and 4% by UAA (Bibb; pers. comm.).

Potential protein coding regions (PPCRs) within the 3.0 kb region were identified using two independent programs. The first program, CODONPREFERENCE (Devereux *et al.,* 1984), exploits the non-random distribution of bases within codons to identify PPCRs. The nature of this distribution depends on the base composition of the coding sequence. The high

```
    CCCGCCCTCGCCGACCCCGACGACGCTCTGCGGCAGACCCTCGGCGTCCCTCCCGGCGGC
  1 ---------+---------+---------+---------+---------+---------+ 60
    GGGCGGGAGCGGCTGGGGCTGCTGCGAGACGCCGTCTGGGAGCCGCAGGGAGGGCCGCCG

    TGGGCGCTGATCAGGCCGGACGGCTATCTGGCCGCGAAGGGGCAGCGGTCCGGCACCACC
 .61 ---------+---------+---------+---------+---------+---------+ 120
    ACCCGCGACTAGTCCGGCCTGCCGATAGACCGGCGCTTCCCCGTCGCCAGGCCGTGGTGG

    ACCCTGACCGCCCGGCTCCAGGCACTGCACCTGCTCCCGGAGGACACCGCGCCCGGCGCC
121 ---------+---------+---------+---------+---------+---------+ 180
    TGGGACTGGCGGGCCGAGGTCCGTGACGTGGACGAGGGCCTCCTGTGGCGCGGGCCGCGG

    GGTGACTCTGCCGGCCGGCCCGCGCCGGACGGCACACGCCGGGGCGTGACGACCGAATGA
181 ---------+---------+---------+---------+---------+---------+ 240
    CCACTGAGACGGCCGGCCGGGCGCGGCCTGCCGTGTGCGGCCCCGCACTGCTGGCTTACT

          rbs    MluI otcDorf3
    CCAAGGAGGACACGCGTATGACAGCGATCAACGGCGCACCGGCGGCAGGCGGTTCGGTGG
241 ---------+---------+---------+---------+---------+---------+ 300
    GGTTCCTCCTGTGCGCATACTGTCGCTAGTTGCCGCGTGGCCGCCGTCCGCCAAGCCACC

                       M  T  A  I  N  G  A  P  A  A  G  G  S  V  A

    CACTGGTCATCGGCGGCACCCGCGGCATCGGGCTGGCCGCCGCCCGCAAACTGAGCGCGG
301 ---------+---------+---------+---------+---------+---------+ 360
    GTGACCAGTAGCCGCCGTGGGCGCCGTAGCCCGACCGGCGGCGGGCGTTTGACTCGCGCC

      L  V  I  G  G  T  R  G  I  G  L  A  A  A  R  K  L  S  A  A

    CGGGCAGCGAGGTGCTGCTCAACTACGCGCACGACGAGGACGGCGCGCTGGCGGCCGAAC
361 ---------+---------+---------+---------+---------+---------+ 420
    GCCCGTCGCTCCACGACGAGTTGATGCGCGTGCTGCTCCTGCCGCGCGACCGCCGGCTTG

      G  S  E  V  L  L  N  Y  A  H  D  E  D  G  A  L  A  A  E  R

       PvuII
    GGCAGCTGTCCGAGGAGGGCGGCAAGGTACGGCTGATGCGCGCGGACATCGGCCGCCCCG
421 ---------+---------+---------+---------+---------+---------+ 480
    CCGTCGACAGGCTCCTCCCGCCGTTCCATGCCGACTACGCGCGCCTGTAGCCGGCGGGGC

      Q  L  S  E  E  G  G  K  V  R  L  M  R  A  D  I  G  R  P  A

    CCGGTGTGGTGCGGCTGCTGGAGGACATCCGGCGCACCCACGGCCGGCTGGACGTCCTGG
481 ---------+---------+---------+---------+---------+---------+ 540
    GGCCACACCACGCCGACGACCTCCTGTAGGCCGCGTGGGTGCCGGCCGACCTGCAGGACC

      G  V  V  R  L  L  E  D  I  R  R  T  H  G  R  L  D  V  L  V

                     P13
    TGCACGCAGCCGGTTCCTTCCATCCGGCGCCGACCGCGTCCCCGCACATCGGCAAGTACC
541 ---------+---------+---------+---------+---------+---------+ 600
    ACGTGCGTCGGCCAAGGAAGGTAGGCCGCGGCTGGCGCAGGGGCGTGTAGCCGTTCATGG

      H  A  A  G  S  F  H  P  A  P  T  A  S  P  H  I  G  K  Y  L

    TGGGCGACGGCGCCGTGGCGGTCGGACCGCTGCTGTACGGGGCGGCGCGCCTGGGCACCC
601 ---------+---------+---------+---------+---------+---------+ 660
    ACCCGCTGCCGCGGCACCGCCAGCCTGGCGACGACATGCCCCGCCGCGCGGACCCGTGGG
                          P12

      G  D  G  A  V  A  V  G  P  L  L  Y  G  A  A  R  L  G  T  L
```

```
    TGATGACCCCGGACACGGGCCGTATCGTCGCCGTCTCCAGCATCGGCGCCCGGACGGTCG
661 ---------+---------+---------+---------+---------+---------+---------+ 720
    ACTACTGGGGCCTGTGCCCGGCATAGCAGCGGCAGAGGTCGTAGCCGCGGGCCTGCCAGC

      M   T   P   D   T   G   R   I   V   A   V   S   S   I   G   A   R   T   V   V

                                                                      KpnI
    TCCCCGGCTACGCGGGTCTGGGCATGGCCAAGGCGGCCTTGGAGACGCTGGTGCGGTACC
721 ---------+---------+---------+---------+---------+---------+---------+ 780
    AGGGGCCGATGCGCCCAGACCCGTACCGGTTCCGCCGGAACCTCTGCGACCACGCCATGG

      P   G   Y   A   G   L   G   M   A   K   A   A   L   E   T   L   V   R   Y   L

    TGGCGGTGGAACTCGCCGGGAAGGGCGTCGCGGTGAACGCGGTGGCCGCGGGCAAGATCG
781 ---------+---------+---------+---------+---------+---------+---------+ 840
    ACCGCCACCTTGAGCGGCCCTTCCCGCAGCGCCACTTGCGCCACCGGCGCCCGTTCTAGC

      A   V   E   L   A   G   K   G   V   A   V   N   A   V   A   A   G   K   I   A

    CCGACGGCGGTCCCGTACCGGCCCAGGTACTGGAGGGCCTGCTGCGGCGGACCCCGACGG
841 ---------+---------+---------+---------+---------+---------+---------+ 900
    GGCTGCCGCCAGGGCATGGCCGGGTCCATGACCTCCCGGACGACGCCGCCTGGGGCTGCC

      D   G   G   P   V   P   A   Q   V   L   E   G   L   L   R   R   T   P   T   G

    GGCGGCTGGCGACTGCCGAAGAGGTGGCCGACGTGGTGGCGCTGCTGTGCCGCCCCGAGG
901 ---------+---------+---------+---------+---------+---------+---------+ 960
    CCGCCGACCGCTGACGGCTTCTCCACCGGCTGCACCACCGCGACGACACGGCGGGGCTCC

      R   L   A   T   A   E   E   V   A   D   V   V   A   L   L   C   R   P   E   A

    CCGGCGGCCTGCACGGACAGGTGCTCACCGTGGACGGCGGGGCCTGCCTGCGGTAGGCGC
961 ---------+---------+---------+---------+---------+---------+---------+ 1020
    GGCCGCCGGACGTGCCTGTCCACGAGTGGCACCTGCCGCCCCGGACGGACGCCATCCGCG

      G   G   L   H   G   Q   V   L   T   V   D   G   G   A   C   L   R   *

        rbs          otcDorf4      P2    SalI
    CCGGAAGGACCGACATGCGCATCATCGATCTGTCGACAACCGTGGACGCCGGCCGGTGGG
1021---------+---------+---------+---------+---------+---------+---------+ 1080
    GGCCTTCCTGGCTGTACGCGTAGTAGCTAGACAGCTGTTGGCACCTGCGGCCGGCCACCC

                  M   R   I   I   D   L   S   T   T   V   D   A   G   R   W   E

    AGGTCGATCCGGTCGAACACGAAATCCTCACCCCGGCCGAGGGCGGCCGGCACATGGCCG
1081---------+---------+---------+---------+---------+---------+---------+ 1140
    TCCAGCTAGGCCAGCTTGTGCTTTAGGAGTGGGGCCGGCTCCCGCCGGCCGTGTACCGGC

      V   D   P   V   E   H   E   I   L   T   P   A   E   G   G   R   H   M   A   E

    AGGGGATGCGGCGCCACCACGGCATCGACTTCGACCCCGCCGACCTCCCGGACGGCGAAC
1141---------+---------+---------+---------+---------+---------+---------+ 1200
    TCCCCTACGCCGCGGTGGTGCCGTAGCTGAAGCTGGGGCGGCTGGAGGGCCTGCCGCTTG

      G   M   R   R   H   H   G   I   D   F   D   P   A   D   L   P   D   G   E   L

                                                          SalI
    TGCTGTCCCTCGACACCCTGCGGCTGACCACCCACACCGGCACCCACGTCGACGCCCCGT
1201---------+---------+---------+---------+---------+---------+---------+ 1260
    ACGACAGGGAGCTGTGGGACGCCGACTGGTGGGTGTGGCCGTGGGTGCAGCTGCGGGGCA

      L   S   L   D   T   L   R   L   T   T   H   T   G   T   H   V   D   A   P   S
```

```
                                                                  P4
    CGCACTACGGGTCCCGCGCGCGCTACGGCACGCCGCGCCACATCGACCAGATGCCACTGG
1261----------+---------+---------+---------+---------+---------+---------+ 1320
    GCGTGATGCCCAGGGCGCGCGCGATGCCGTGCGGCGCGGTGTAGCTGGTCTACGGTGACC

      H   Y   G   S   R   A   R   Y   G   T   P   R   H   I   D   Q   M   P   L   D

    ACTGGTTTCTGCGGCCGGGCCTCAAGCTCGACCTTACCGACGAGCCGGTGGGCGCCATCG
1321----------+---------+---------+---------+---------+---------+---------+ 1380
    TGACCAAAGACGCCGGCCCGGAGTTCGAGCTGGAATGGCTGCTCGGCCACCCGCGGTAGC

      W   F   L   R   P   G   L   K   L   D   L   T   D   E   P   V   G   A   I   G

    GAGCCGACCGGATACGGCGGGCCCTCGACGAAGCGGGCCGCGACCCGCGTCCGTACGACA
1381----------+---------+---------+---------+---------+---------+---------+ 1440
    CTCGGCTGGCCTATGCCGCCCGGGAGCTGCTTCGCCCGGCGCTGGGCGCAGGCATGCTGT

      A   D   R   I   R   R   A   L   D   E   A   G   R   D   P   R   P   Y   D   I

    TCGTGCTGCTGCACACCGGCGCGGACCGGCGGGCCGGAAAGCCGGAGTACTTCACCGAGT
1441----------+---------+---------+---------+---------+---------+---------+ 1500
    AGCACGACGACGTGTGGCCGCGCCTGGCCGCCCGGCCTTTCGGCCTCATGAAGTGGCTCA

      V   L   L   H   T   G   A   D   R   R   A   G   K   P   E   Y   F   T   E   F

    TCGCCGGCCTGGACGCCGCGGCGACCCACCTGCTGCTCGACTTCGGCGTCCGGGTCATCG
1501----------+---------+---------+---------+---------+---------+---------+ 1560
    AGCGGCCGGACCTGCGGCGCCGCTGGGTGGACGACGAGCTGAAGCCGCAGGCCCAGTAGC
                                                          P10

      A   G   L   D   A   A   A   T   H   L   L   L   D   F   G   V   R   V   I   G

              P8
    GCACCGACGCCTTCAGCCTCGACGCCCCGTTCGGGCACATGATCAAGGAGTACCAGGAGC
1561----------+---------+---------+---------+---------+---------+---------+ 1620
    CGTGGCTGCGGAAGTCGGAGCTGCGGGGCAAGCCCGTGTACTAGTTCCTCATGGTCCTCG

      T   D   A   F   S   L   D   A   P   F   G   H   M   I   K   E   Y   Q   E   R

    GCACCGGCGACCGCGGTGTGCTGTGGCCCGCCCACTTCGTGGGACGGGAGCGCGAGTACT
1621----------+---------+---------+---------+---------+---------+---------+ 1680
    CGTGGCCGCTGGCGCCACACGACACCGGGCGGGTGAAGCACCCTGCCCTCGCGCTCATGA

      T   G   D   R   G   V   L   W   P   A   H   F   V   G   R   E   R   E   Y   C

    GCCAGATCGAGCGGCTGGACAACCTCGGCGCGCTGCCCGGACCGGACGGCTTCTTGGTCT
1681----------+---------+---------+---------+---------+---------+---------+ 1740
    CGGTCTAGCTCGCCGACCTGTTGGAGCCGCGCGACGGGCCTGGCCTGCCGAAGAACCAGA

      Q   I   E   R   L   D   N   L   G   A   L   P   G   P   D   G   F   L   V   S

          P9
    CCTGCTTTCCGTTCAAGATCGCGGGAGCGGGGGCGGGGTGGACGCGGGCGGTGGCGGTGG
1741----------+---------+---------+---------+---------+---------+---------+ 1800
    GGACGAAAGGCAAGTTCTAGCGCCCTCGCCCCCGCCCCACCTGCGCCCGCCACCGCCACC

      C   F   P   F   K   I   A   G   A   G   A   G   W   T   R   A   V   A   V   V
```

```
    TGGAGGAGTAGGGACGGACGGGTGAACGGGCGCCCCCGGAGCATCCTGGGGCGCCCGTCC
1801---------+---------+---------+---------+---------+---------+ 1860
    ACCTCCTCATCCCTGCCTGCCCACTTGCCCGCGGGGGCCTCGTAGGACCCCGCGGGCAGG

      E   E   *

                                                    rbs      otcD
    GCACGGCTTCCACGGAATCCGCCACGGCCCACCTGCGATGGCGGTCGAATGGCAGTATGC
1861---------+---------+---------+---------+---------+---------+ 1920
    CGTGCCGAAGGTGCCTTAGGCGGTGCCGGGTGGACGCTACCGCCAGCTTACCGTCATACG

                                                        M   R

    orf5
    GAAACCGGCGGAGTCAGCAGTTGCACCGCGTGGTCCGGCGCGCTTCCCGGGCCGGGAACG
1921---------+---------+---------+---------+---------+---------+ 1980
    CTTTGGCCGCCTCAGTCGTCAACGTGGCGCACCAGGCCGCGCGAAGGGCCCGGCCCTTGC

      N   R   R   S   Q   Q   L   H   R   V   V   R   R   A   S   R   A   G   N   G

                                P11
    GCCCGCCACCGGCCGGCAGCGTCCCTCCGCGTCCGGAGCTGGGCTGGGCCCGTCTGATCG
1981---------+---------+---------+---------+---------+---------+ 2040
    CGGGCGGTGGCCGGCCGTCGCAGGGAGGCGCAGGCCTCGACCCGACCCGGGCAGACTAGC
                                P1

      P   P   P   A   G   S   V   P   P   R   P   E   L   G   W   A   R   L   I   E

    AGCTGACTTCGGTACTGCTCGCCTCGGTCGTCGCCGTGCTGAGCCTCTGGTTCTCCAACC
2041---------+---------+---------+---------+---------+---------+ 2100
    TCGACTGAAGCCATGACGAGCGGAGCCAGCAGCGGCACGACTCGGAGACCAAGAGGTTGG

      L   T   S   V   L   L   A   S   V   V   A   V   L   S   L   W   F   S   N   R

    GCCAGGTGGGCAACCAGTTGCGGATCGCCCAGGGCGAGCTGAGCACCACCAGGGAAGGTC
2101---------+---------+---------+---------+---------+---------+ 2160
    CGGTCCACCCGTTGGTCAACGCCTAGCGGGTCCCGCTCGACTCGTGGTGGTCCCTTCCAG

      Q   V   G   N   Q   L   R   I   A   Q   G   E   L   S   T   T   R   E   G   Q

                                          PvuII
    AGATCACCGAGCGGTACACCGCGGCCGTGGGCCAGCTGGGCGAGGATTCGGTGGACGTGC
2161---------+---------+---------+---------+---------+---------+ 2220
    TCTAGTGGCTCGCCATGTGGCGCCGGCACCCGGTCGACCCGCTCCTAAGCCACCTGCACG

      I   T   E   R   Y   T   A   A   V   G   Q   L   G   E   D   S   V   D   V   R

                                  PstI
    GCCTGGGCGGCATCTACGCCCTGCAGGGCAACATGGAGGACCTCCCCCGCGACCACCCCA
2221---------+---------+---------+---------+---------+---------+ 2280
    CGGACCCGCCGTAGATGCGGGACGTCCCGTTGTACCTCCTGGAGGGGGCGCTGGTGGGGT

      L   G   G   I   Y   A   L   Q   G   N   M   E   D   L   P   R   D   H   P   T

    CCATCGCCGACGTCCTGGCCGCCTATATCCGCACCCACGCCGGCACCCCGCCGAAAGGCG
2281---------+---------+---------+---------+---------+---------+ 2340
    GGTAGCGGCTGCAGGACCGGCGGATATAGGCGTGGGTGCGGCCGTGGGGCGGCTTTCCGC

      I   A   D   V   L   A   A   Y   I   R   T   H   A   G   T   P   P   K   G   G
```

```
    GCAAACGGGTGCCACCCGATGTCCTGGCCGCGCTCGACGTGATCAGCGGCCGGAACTCCG
2341--------+---------+---------+---------+---------+---------+---------+ 2400
    CGTTTGCCCACGGTGGGCTACAGGACCGGCGCGAGCTGCACTAGTCGCCGGCCTTGAGGC

      K   R   V   P   P   D   V   L   A   A   L   D   V   I   S   G   R   N   S   A

                                                      P7
    CCCATGACGGCTCCTTCGTTCCCGACCTCCGCTCCACCCACCTGCCCGGCGTCGAACTGG
2401--------+---------+---------+---------+---------+---------+---------+ 2460
    GGGTACTGCCGAGGAAGCAAGGGCTGGAGGCGAGGTGGGTGGACGGGCCGCAGCTTGACC

      H   D   G   S   F   V   P   D   L   R   S   T   H   L   P   G   V   E   L   G

    GGTGGGACCGTTCGTCTCCCGTCGCCGCCGACCACAAGCGGGCCCAGTTGCGCGGAGCGA
2461--------+---------+---------+---------+---------+---------+---------+ 2520
    CCACCCTGGCAAGCAGAGGGCAGCGGCGGCTGGTGTTCGCCCGGGTCAACGCGCCTCGCT

      W   D   R   S   S   P   V   A   A   D   H   K   R   A   Q   L   R   G   A   I

    TCCTCCGTGACTGCGACCTGAGGGACGCGCGTCTGAGCGGCGCGGACCTGCGAGGCGCGA
2521--------+---------+---------+---------+---------+---------+---------+ 2580
    AGGAGGCACTGACGCTGGACTCCCTGCGCGCAGACTCGCCGCGCCTGGACGCTCCGCGCT
          P6

      L   R   D   C   D   L   R   D   A   R   L   S   G   A   D   L   R   G   A   R

    GGCTCGCACGCGCCGATCTGGCCGACGCGGATCTGCGCGAGGCGGACCTCCGGGAAGCCA
2581--------+---------+---------+---------+---------+---------+---------+ 2640
    CCGAGCGTGCGCGGCTAGACCGGCTGCGCCTAGACGCGCTCCGCCTGGAGGCCCTTCGGT

      L   A   R   A   D   L   A   D   A   D   L   R   E   A   D   L   R   E   A   T

    CTCCGGCGCGGGCGAACCTGCGCGACGCGGACTTGTCCGACGCCAACGTGCGCAAGGCGG
2641--------+---------+---------+---------+---------+---------+---------+ 2700
    GAGGCCGCGCCCGCTTGGACGCGCTGCGCCTGAACAGGCTGCGGTTGCACGCGTTCCGCC

      P   A   R   A   N   L   R   D   A   D   L   S   D   A   N   V   R   K   A   D

    ACCTGAGGTTCGCGGATCTTCGCGGCGTGGACCTGTGGGGCACCGATCTGCGCGGCGCGG
2701--------+---------+---------+---------+---------+---------+---------+ 2760
    TGGACTCCAAGCGCCTAGAAGCGCCGCACCTGGACACCCCGTGGCTAGACGCGCCGCGCC

      L   R   F   A   D   L   R   G   V   D   L   W   G   T   D   L   R   G   A   V

                                                  SstI
    TCCTGTACCGGGCGAAACTGGCCGGCCTGGAGCTCAGCGAGGCGCACCTGGACGGAGCCG
2761--------+---------+---------+---------+---------+---------+---------+ 2820
    AGGACATGGCCCGCTTTGACCGGCCGGACCTCGAGTCGCTCCGCGTGGACCTGCCTCGGC

      L   Y   R   A   K   L   A   G   L   E   L   S   E   A   H   L   D   G   A   D

    ACCTCCGCGGTGCTGATCTCACGGACGCGGCGGTGGCACGGAGCCAGGTCCTCGCGAGCC
2821--------+---------+---------+---------+---------+---------+---------+ 2880
    TGGAGGCGCCACGACTAGAGTGCCTGCGCCGCCACCGTGCCTCGGTCCAGGAGCGCTCGG

      L   R   G   A   D   L   T   D   A   A   V   A   R   S   Q   V   L   A   S   R
```

```
    GCATCGACGCGGAGACCAGGCTCCCACCGGACATCGCCCGCGATCCGGAGGTGCGGGCCC
2881-----------+---------+---------+---------+---------+---------+ 2940
    CGTAGCTGCGCCTCTGGTCCGAGGGTGGCCTGTAGCGGGCGCTAGGCCTCCACGCCCGGG

      I   D   A   E   T   R   L   P   P   D   I   A   R   D   P   E   V   R   A   R

    GGATCGCTCAGGCCGAGGCGGAGGACGATGGTTTCTGATCGCCGCCTCACCCTCCTCAGG
2941-----------+---------+---------+---------+---------+---------+ 3000
    CCTAGCGAGTCCGGCTCCGCCTCCTGCTACCAAAGACTAGCGGCGGAGTGGGAGGAGTCC

      I   A   Q   A   E   A   E   D   D   G   F   *

              KpnI
    GAATCCGGTACC
3001-----------+-- 3012
    CTTAGGCCATGG
```

Figure 3.3

Annotated reverse complement sequence of the 3,012 bp *Kpn*I$_{23}$-*Mlu*I$_{26b}$ region of the oxytetracycline cluster. The predicted amino acid sequences of *OtcD*-ORF3, *OtcD*-ORF4 and *OtcD*-ORF5 are shown corresponding to the nucleotide sequences. The oligonucleotides designed as primers for sequencing are bold. The putative ribosome binding sites (rbs) for the predicted proteins are underlined. The restriction sites referred to in the text are underlined. Sequence 1-425 was originally Sequenced by K. Linton. It was confirmed ~~written down~~ by sequencing in the other direction.

G+C content of *Streptomyces* (typically 73% mol G+C) allows for particularly reliable predictions of the reading frame based on position-specific differences. It was proposed that a reading frame could be established if the there is a high mol% G+C content at the third codon position, a much lower one at the second position and an intermediate one at the first (Bibb *et al.*, 1984). The intervening regions between the coding sequences have a random distribution of bases within consecutive triplets. The CODONPREFERENCE program scans the DNA sequence and determines the mol% G+C composition at the first, second and third positions of a number of triplets from a "window", which usually contains 25 to 50 triplets. A "window" is a fixed length of sequence within which a statistical calculation is performed. The mol% G+C for each of the three base positions within the "window" are then plotted against a linear representation of the sequence, above the first nucleotide of the "window". The "window" is then moved along one triplet and the calculation is repeated. This procedure is continued until the entire sequence has been analysed. The resulting output presents all the three reading frames graphically and any potential ORF's are also displayed (Bibb *et al.*, 1984). The second program, FRAME, utilises much the same principles of base distribution within the codons. However, the results could only be displayed in tabular form with current computer software. The size of the "window" can also be adjusted to further investigate smaller regions of the sequence. This program was utilised primarily to determine the orientation of any PPCRs and the CODONPREFERENCE program to determine the reading frame.

The 3.0 kb *Kpn*I$_{23}$-*Pvu*II$_{26b}$ sequence was subjected independently to these two frame-analysis programs. The FRAME program determined that there were three PPCRs. Starting from the right-hand side of the sequence, the 0.63 kb *Pvu*II-*Sal*I fragment showed a clearly-defined PPCR, the orientation of which was determined by the mol% G+C in each position. The series of "windows" showed a low mol% G+C at position 1, a medium mol% G+C at position 2 and a high mol% G+C at position 3. This indicated that the PPCR read from right to left in the second frame. This was confirmed with a plot using the CODONPREFERENCE program shown in figure 3.4. For simplicity, the reverse complement is displayed in both figure 3.3 and 3.4 to parallel the direction of translation. The stop codon (UAG) was located at position 1014 (figure 3.3). The start of the PPCR was not within this sequence. A look 167 bp to the right of the *Pvu*II site (position 258) revealed in a AUG codon 'in frame' with the rest of the sequence. Also 'in frame' a further 21 bp upstream (position 237) was another

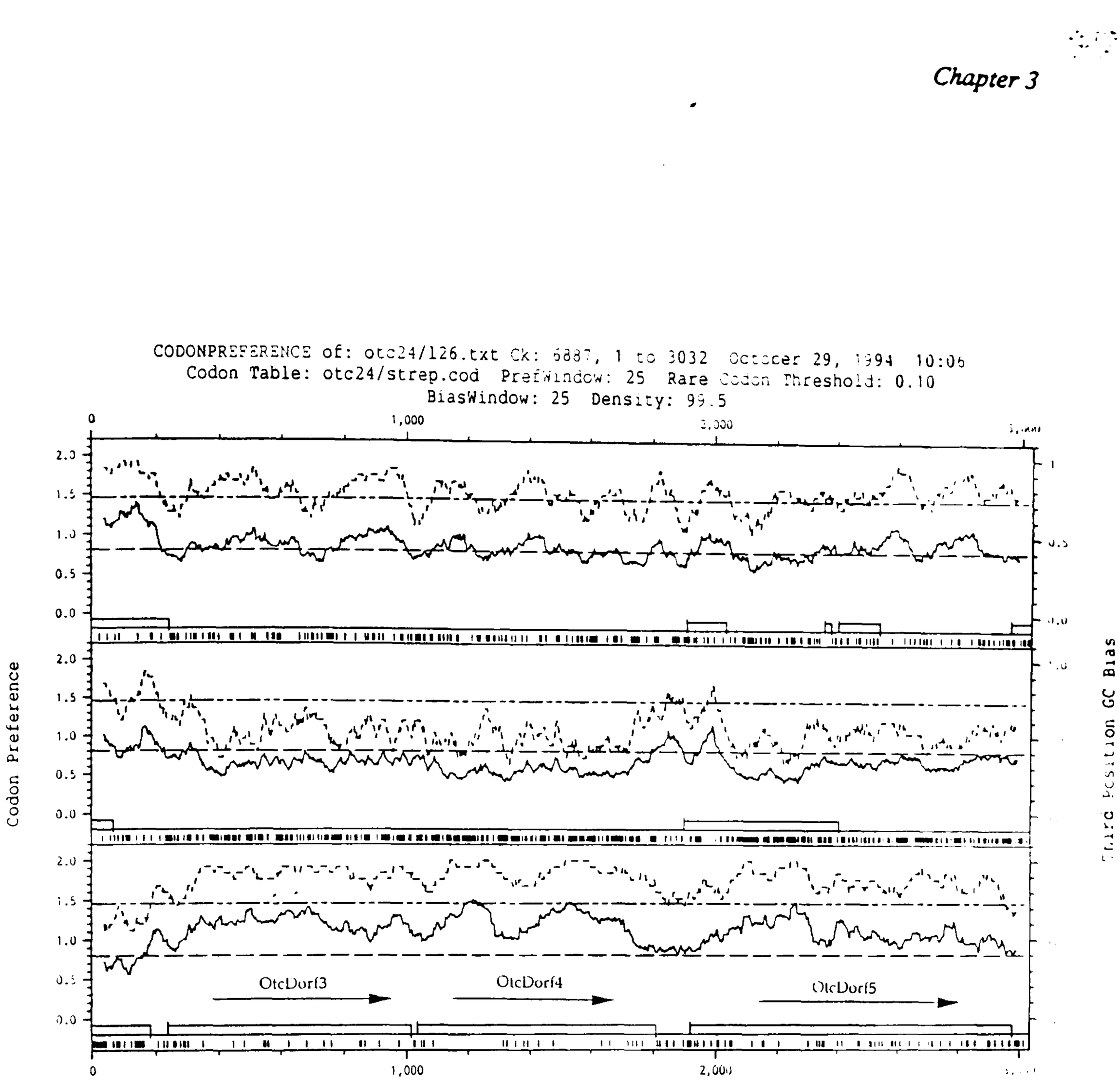


Figure 3.4
Open reading frames, *OtcD*-ORF3 and ORF-4 ORF-5 deduced by CODONPREFERENCE analysis from nucleotide sequence from Figure 3.3. The arrows indicate the probable direction of transcription.

potential start codon (AUG). The AUG at position 258 was preceded by a reasonable ribosome binding site (AGGAGG) with a high degree of complementarity to a region close to the 3' end of the 16S rRNA from *S. lividans* (sequenced by Bibb and Cohen; 1982). This putative ribosome binding site (rbs) is separated from the AUG triplet by 8 nt and is within the limits of 5-9 nt for Shine-Dalgarno interactions (Gold *et al.*, 1987). No significant complementarity to the 3' terminus of 16S rRNA of *S. lividans* was found within the sequence immediately upstream of the second AUG codon. Therefore, it was concluded that the AUG at position 258 nt was the probable translation initiation codon for *OtcD*-ORF3.

The FRAME analysis proceeding left from the stop codon of *OtcD*-ORF3 (position 1014) was initially diverse but quickly fell into a recognisable PPCR. The region between the two PPCRs was only 18 nt. The pattern produced from the FRAME analysis program was similar to the previous PPCR, a low mol% G+C at position 1, a medium mol% G+C at position 2 and a high mol% G+C at position 3. Again there were two possible initiation start sites, an AUG start at position 1035 and a GUG start at position 1062. The AUG codon was preceded by a reasonable rbs (GGAAGG) 6 nt upstream from the AUG start. The GUG codon, although 'in frame', was not preceded by a potential rbs. 774 nt downstream from the AUG start was an UAG stop codon, located at position 1809. Therefore it was concluded that *OtcD*-ORF4 reads from right to left from an AUG initiation start site (position 1035) to an UAG stop site (position 1809).

The FRAME analysis of the left most region of the 3.0 kb sequence showed great diversity immediately following the UAG of *OtcD*-ORF4. There were three potential AUG start sites, all in different reading frames. The FRAME program showed that the PPCR read in the same direction and reading frame as the previous two PPCRs. A possible rbs was located 4 nt upstream of the AUG and the PPCR extended to a UGA stop codon at position 2976. It was concluded that the third PPCR, *OtcD*-ORF5 was 1059 nt in length running from position 1917 to 2976. The region sequenced was completed at the *Kpn*I$_{23}$ site, 37 nt downstream of the *OtcD*-ORF5 translational stop codon..

3.2.3 Comparisons of Amino Acid Sequence

Analysis of the predicted amino acid sequences for all three PPCRs was done by comparison with sequences in the databases. Each PPCR will be

discussed separately. To ensure that the comparison is a valid one, the databases used should be updated frequently and for maximum sensitivity, the database searches should be performed using a range of scoring tables and other parameters. It was hoped that the result of database searches would bring an indication of a structural, functional or evolutionary relationship with some other sequence for which additional, direct experimental information was readily available. The two systems employed here for data searching were TFASTA (Devereux *et al.,* 1984) and BLAST (Altschul *et al.,* 1990). TFASTA is a family of programs based on the idea that strong similarities will be marked by high concentrations of short runs of exact matches. If there are gaps in an alignment, these matches will not lie on a single diagonal of the comparison surface but if the length of gap is small then they will lie on adjacent diagonals. TFASTA operates in several phases; a rapid initial scan identifies regions of the database in which there are diagonal bands of good match to the query sequence. A second scan applies the complete Needleman-Wunsch-Sellers (NWS) (Sellers; 1980) algorithm to this limited subset. The shortfall to this program is that the first scan is sensitive only to exact matches and for more distantly-related sequences, their similarity may be expressed significantly in conservative substitutions of amino acids. The second program, BLAST, attempts to take the conservative substitution into account. It also works by scanning rapidly for good matches in short, fixed-length 'words' but, in addition to looking for exact matches to the words in the query sequence, BLAST also looks at a larger set derived from them by generating all words that score above a threshold when accessed with a general similarity table. Local similarities discovered in this way are then added in the same registry until the aggregate score falls below the threshold level. Gaps are not allowed directly. However when several strong, ungapped alignments are found in one database sequence, the probability that this has occurred by chance is assessed in terms of a Poisson distribution. For a comprehensive discussion of high-performance searching of biosequence databases, see Coulson (1994).

3.2.3.1 *OtcD*-ORF3

The predicted amino acid sequence of *OtcD*-ORF3 is 252 amino acids in length with a predicted molecular weight of 25,274 D. The *OtcD*-ORF3 gene product has an average 21.8% identity and 43.6% similarity with FAS-type ketoreductases, such as, NodG (*nod*G) from *Rhizobium meliloti* (Fisher *et al.*, 1987) and *E. coli* FabG(β-ketoacyl acyl carrier protein reductase) (Cronon *et al.*, 1992). A much greater similarity to type II PKS ketoreductases is seen, with an amino

acid sequence identity of 31% and similarity of 51.0% for *OtcY*-ORF1 (Thamchaipenet; 1994), 32.2% identity, 55.6% similarity for DauB (Ye *et al.*, 1994), 30.8% identity, 56.2% similarity for ActIII (Hallam *et al.*, 1988), 30% identity, 52.4% similarity for Gra-ORF5 and 32% identity, 53.7% similarity for Gra-ORF6 (Sherman *et al.*, 1989). All are aromatic polyketide reductases, some with greater sequence identity to *OtcD*-ORF3 than others. The dendrogram shown in figure 3.6 shows *OtcY*-ORF1, also from the *otc* cluster, is grouped with those showing a lower sequence identity suggesting that there is more than one ketoreductase function in the biosynthesis of OTC. A comparison of the amino acid alignments of the deduced peptide sequence of *OtcD*-ORF3 with the other enzymes was displayed using the program PILEUP (figure 3.5). The degree of similarity and the conservation of several core sequences in a defined order shows that all of these proteins are homologous. A consensus NADP(H) binding motif, Gly-X-Gly-X-X-Ala (where X is any amino acid) (Scrutton *et al.*, 1990), is found in the gene product of *Otc*Y2-ORF1 but not *OtcD*-ORF3. A similar motif is also present at the same position in ActIII, DauB and Gra-ORF5 (Hopwood and Sherman; 1990)[underlined in figure 3.5]. It was proposed as a consensus NADP(H) binding motif of those putative ketoreductases (Hopwood and Sherman; 1990). NADPH may be a cofactor for *act-*, *gra-*, *dau-* and *otc*-ketoreductase systems. However, this motif is apparently not conserved at the same position in other reductases (figure 3.5). A second conserved motif, Gly-3(X)-Gly-X-Gly-3(X)-Ala-6(X)-Gly (where X is any amino acid) was found at residues 18-35 of the *OtcD*-ORF3 gene product (figure 3.5). This motif is also conserved in the sequences from the other seven proteins compared in figure 3.5. (marked by asterisks; figure 3.5). It situated close to the N-terminal region, is very glycine rich, and was previously proposed as the potential NADP(H)-binding fold in acetoacetyl-CoA reductase (*phb*B; Peoples and Sinskey; 1989) and 3-ketoacyl-ACP reductase (*fab*G; Rawlings and Cronan; 1992). Thus, this motif is also likely to be a binding site for NADPH. The GXGXXA motif that is conserved only in those ketoreductases underlined in figure 3.5, is associated with Type II PKS's.

Representative of this class of ketoreductases is the ActIII reductase from the actinorhodin cluster of *S. coelicolor*. Biochemical genetic analysis of the actinorhodin (*act*) biosynthetic pathway identified seven classes of mutants blocked at different stages which allowed an ordered biosynthetic sequence to be deduced: I, III, VII, IV, VI, VA and VB (Rudd *et al.*, 1979). (A full description of the pathway is given in Chapter 1.9.2). *ActI* and *III* mutants failed to secrete any

```
                                    *   * *    *        *
           1                                                   50
  act111   .......... .MATQDSEVA LVTGATSGIG LEIARR.LGK EGLRVFV.CA
    gra5   MTTATATATA TPGTAAKPVA LVTGATSGIG LAIARR.LAA LGARTFL.CA
    daub   .......... .MSDAADRVA LVTGGTSGIG LAVGRRKLAL DGTRVFL.CA
otcYorf1   .......... MTTSATPRTA LVTGGTSGIG AVV..KTLAA RGLRVFL.CA
    fabg   .......... ..MNFEGKIA LVTGASRGIG RAIA.ETLAA RGGKVIG.TA
    nodg   .......... .MFELTGRKA LVTGASGAIG GAIA.RVLHA QGAIVGL.HG
    gra6   .........M ATDAPEAPVA LVTGSSSGIG QTVAQR.LAA EGYRVVVNSA
otcDorf3   .....MTAIN GAPAAGGSVA LVIGGTRGIG LAAARK.LSA AGSEVLLNYA

           51                                                 100
  act111   RGEEGLRTTL KELREAGVEA DGRTCDVRSV PEIEALVAAV VERYGPVDVL
    gra5   RDEERLAQTV KELRGEGFDV DGTVCDVADP AQIRAYVAAA VQRYGTVDIL
    daub   RDESAVTGTV KELQASGLEV EGTSCDVRST DAVDRLVRTA RNPLRAIDIV
otcYorf1   RSRENVDAVV KELRDQGLQV AGCEGDVRSR ESVERVVRAA VDRFGPLSVL
    fabg   TSENGAQAIS DYL...GANG KGLMLNVTDP ASIESVLEKI RAEFGEVDIL
    nodg   TQIEKLETLA TEL...GDRV KLFPANLANR DEVKALGQRA EADLEGVDIL
    gra6   RSVEDGEKTA AALPDALY.. ..VRADVSEE ADARRLVDTA VEHYGRLDVL
otcDorf3   HDEDGALAAE RQLSEEGGKV RLMRADIGRP AGVVRLLEDI RRTHGRLDVL

           101                                                150
  act111   VNNAGRPGGG ATAELAD... ELWLDVVETN LTGVFRVTKQ VLKAGGMLER
    gra5   VNNAGRSGGG ATAEIAD... ELWLDVITTN LTSVFLMTKE VLNAGGMLAK
    daub   VNNAGRGGGG VTAQITD... DLWSDVVDTN LGGAFRVTRA VLTGGGMQEH
otcYorf1   VNNAGRSGGG VTARIPD... ELWYDVIDTN LNSAFLVTRE ALTTGGLERA
    fabg   VNNAGITRDN LLMRMKD... EEWNDIIETN LSSVFRLSKA VMRA...MMK
    nodg   VNNAGITKDG LFLHMAD... PDWDIVLEVN LTAMFRLTRE ITQQ...MIR
    gra6   VNNAGRTRAI PHADLAAATP EVWREILGLN VIGTWQTTVA AMPH...LAR
otcDorf3   VHAAGSFHPA PTA...SPHI GKYLGDGAVA VGPLLYGAAR L....GTLMT

           151                                                200
  act111   GT.GRIVNIA STGGKQGVVH AAPY.SASKH GVVGFTKALG LELARTGITV
    gra5   KR.GRIINIA STGGKQGVVH AVPY.SASKH GVVGLTKALG LELARTGITV
    daub   GW.GRIINIA STGGKQGVAL GAPY.SASKS GLIGFTKAVA LELARTGITV
otcYorf1   GVDGRIISIA STGGKQGVPL GAPY.SASKA GLIGFTKAXG QGLAPTGVTV
    fabg   KRHGRIITIG SVVGTMGNGG QANYAAAAKA GLIGFSKSLA REVASRGITV
    nodg   RRNGRIINVT SVAGAIGNPG QTNY.CASKA GMIGFSKSLA QEIATRNITV
    gra6   SGNGSVVNVS SIAGSRPAGS SIPY.AVSNG GHRAQTRLLA NTVGP.AVRV
otcDorf3   PDTGRIVAVS SIGARTVVPG YAGL.GMAKA ALETLVRYLA VELAGKGVAV
```

```
          201                                                        250
  act111  NAVCPGFVET PMAASVREHY SDIWEVSTEE AFDRITARVP IGRYVQPSEV
    gra5  NAVCPGFVET PMAERVREHY AGIWQVSEEE TFDRITNRVP LGRYVETREV
    daub  NAVCPGYVET PMAQGVRQRY AAFWGITEDD VLEKFRAKIP LGRYSTSDEV
otcYorf1  NAVCPGYVET PMAVRVRQAY AATWDTTEED VLSRFNDKIP LGRYSTPEEV
    fabg  NVVAPGFIET DMTRAL.... .......SDD QRAGILAQVP AGRLGGAQEI
    nodg  NCVAPGFIES AMT.DL.... .......NHK QKEKIMVAIP IHRMGTGTEV
    gra6  NAVAPGLIET P.....WTQN SDFFAPIAEH ....VRQTTP LRRTGRPEDV
otcDorf3  NAVAAGKIAD G......... ....GPVPAQ VLEGLLRRTP TGRLATAEEV

          251                              280
  act111  AEMVAYLIGP GAAAVTAQAL NVCGGLGNY.
    gra5  AAMVEYLVAD DAAAVTAQAL NVCGGLGNY*
    daub  AGMVHYLVSD SADSITAQAI NVCGGLGSY.
otcYorf1  AALVGHLAGD EAGSITAQAL NVCGGLGIY.
    fabg  ANAVAFLASD EAAYITGETL HVNGGMYMV.
    nodg  ASAVAYLASD HAAYVTGQTI HVNGGMAMI.
    gra6  AEAVLGLV.. RATYTTGQVL LVDGGAHLL.
otcDorf3  ADVVALLCRP EAGGLHGQVL TVDGGACLR*
```

Figure 3.5
Linear alignment of amino acid sequences of ActIII from *S. coelicolor*, GraORF5 and GraORF6 from *S. violaceoruber*, 3-oxoacyl-[acyl carrier protein] reductase (FabG) from *E. coli*, nodulation protein G (NodG) from *Rhizobium meliloti*, DauB from *Streptomyces sp.* C5 and *OtcY2*-ORF1 and *OtcD*-ORF3 from *S. rimosus* with amino acid numbering indicated. The motif Gly-3(X)-Gly-X-Gly-3(X)-Ala-6(X)-Gly is indicated by asterisks and the motif Gly-X-Gly-X-X-Ala is underlined (see text for detail).

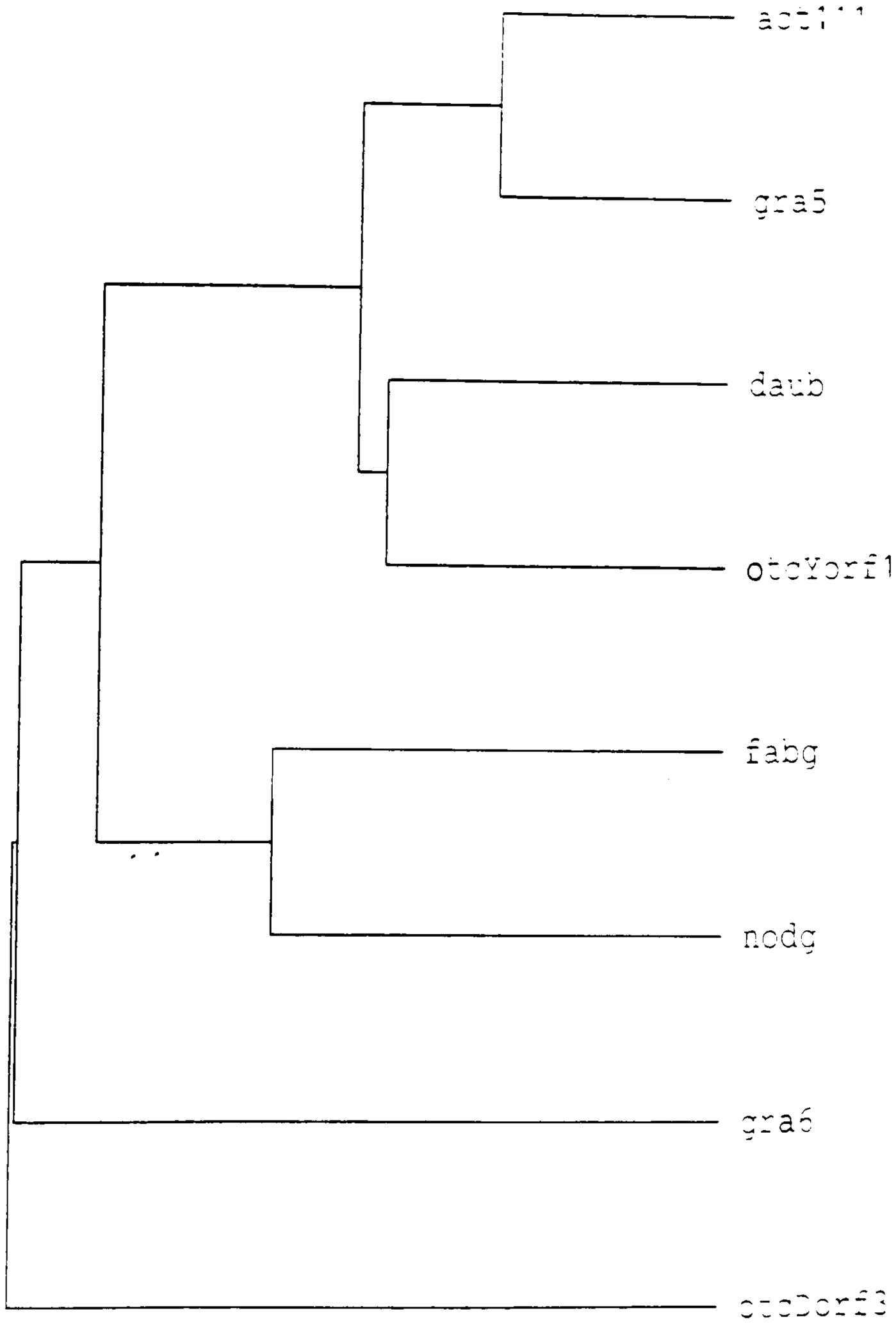


Figure 3.6
Dendrogram deduced from the alignment of amino acid sequences derived by the PILEUP analysis (Fig. 3.5) of ActIII, GraORF5 and GraORF6, NodG, FabG, DauB, *OtcY*-ORF1 and *OtcD*-ORF3 (see also legend of Figure 3.5 and see text for references).

biosynthetic intermediate active in co-synthesis with the other classes of mutants, while being able to convert actinorhodin intermediates secreted by the other mutant classes. This indicates that *actI* and *III* mutants are likely defective in the genes which encode enzymes catalysing the construction of the polyketide carbon skeleton of actinorhodin. The nucleotide sequence of *actIII* was determined and database searches showed similarity to the ribitol dehydrogenase gene from *Klebsiella aerogenes* and the alcohol dehydrogenase gene from *Drosophila melanogaster* (Hallam *et al.*, 1988). Malpartida *et al.*, (1987) had previously hypothesised that the *actIII* gene encoded a polyketide reductase which reduces a keto group of the polyketide intermediate in the biosynthesis of actinorhodin. After this reduction a hydroxyl group would be lost from this position during ring closure and aromatisation, and the resulting compound would lack an oxygen in that position (Chapter 1.9.2). This hypothesis was substantiated in work by Bartel *et al.*, (1990b), who, in an attempt to generate hybrid antibiotics by transforming discrete restriction fragments of the actinorhodin biosynthetic gene cluster into various *Streptomyces* strains, found that anthraquinone was produced by the recombinant *Streptomyces* that normally do not produce the compound. The function of the *actIII* gene product was shown therefore to be a polyketide reductase which reduces a keto group at position C-9 at the carboxy terminus of the assembled polyketide to the corresponding secondary alcohol. According to the theories of polyketide formation (See Chapter 1 for a full description), these compounds are formed by the successive additions of acetyl groups (from malonyl-CoA) to a starter unit (7 additional acetates for actinorhodin and 8 additional acetates for oxytetracycline). The fact that catalysis by the *actIII* gene product occurs at the ninth carbon from the end of the molecule, implies that this enzyme does not reduce the carbonyl group until the entire chain is assembled. Hypothetically then, the polyketide reductase might 'measure' the distance from the carboxyl end of the completely assembled chain and reduce the proper keto group. During *otc* biosynthesis, reduction occurs at the ninth carbon, counting from the carboxyl end of the mature polyketide chain. If no reduction were to take place at C-9 before ring cyclisation, an analogue of OTC would be predicted to contain a hydroxyl group in that position. Indeed, some of the antibiotics of the tetracycline family have an oxygen remaining at the C-9 position in a methoxy form, for example 2'-N-methyl-8-methoxychortetracycline produced by *Actinomadura brunnea* (Patel *et al.*, 1987a) and 4a-hydroxy-8-methoxychlortetracycline from *Dactylosporangium sp.* (Patel *et al.*, 1987b). In both cases mentioned here, the reduction occurs at a carbon which is part of the first

ring formed during the biosynthesis of these various polyketide compounds. Bartel *et al.,* (1990) also showed that anthracyclines are also reduced at the ninth carbon position by a enzyme showing strong similarity to the *actIII* gene, although this may just be coincidence. It was therefore, on the basis of the strong similarity to the *actIII* gene product, the two conserved motifs and the position of ketoreduction within the cluster that *OtcY2*-ORF1 was assigned as the polyketide reductase which reduces a keto group at the C9 position of the otc intermediate.

In the biosynthesis of OTC, three oxidoreductase steps are required. One, already mentioned above, at position C9, one at position C6 in the conversion of 4-keto-ATC to ATC and one at the last step of the pathway, OTC dehydrogenase. Mutants blocked in the final step of the pathway require the addition of CSF1. CSF1 (cosynthetic factor 1) a flavin-like cofactor, is required along with NADPH to reduce 5a, 11a-dehydrooxytetracycline to oxytetracycline. This was first demonstrated by Miller *et al.,* (1960), with certain mutants of *S. aureofaciens* which were unable to produce chlortetracycline because they lacked this cofactor and then subsequently with *S. rimosus* as the identical biochemial steps occur in both chlorotetracycline and oxytetracycline biosynthesis. By amino acid comparisons of the available polyketide reductase genes, there is strong similarity of *OtcD*-ORF3 to a group of reductases which carry a conserved motif at the N-terminus, see figure 3.5. More than one reductase has been found in a number of strains including granaticin (Sherman *et al.,* 1989) and daunorubicin (Strohl; unpublished). *OtcD*-ORF3 shows slightly greater similarity to Gra-6 (55.7 %) then Gra-5 (49.3 %). This leads to the conclusion that Gra-5 is involved in the PKS region of the granaticin biosynthetic pathway and that Gra-6 is a tailoring enzyme which carries out a role similar to *OtcD*-ORF3. Sequence similarity to two other reductases found in the daunorubicin producing strain C5 (Strohl; unpublished) would help to establish a second reductase sub-family.

3.2.3.2 *OtcD*-ORF4

The predicted amino acid sequence of *OtcD*-ORF4 consists of 258 amino acids with a deduced molecular weight of 28,571 D. When the deduced amino acid sequence was searched using the BLAST program, a strong similarity was found to ORF1 in the *lkmB* region of the daunorubicin biosynthetic gene cluster of *Streptomyces griseus* (Krügel *et al.,* 1993). Direct comparison with the amino acid sequences employing the BESTFIT program (Devereux *et al.,* 1984), revealed

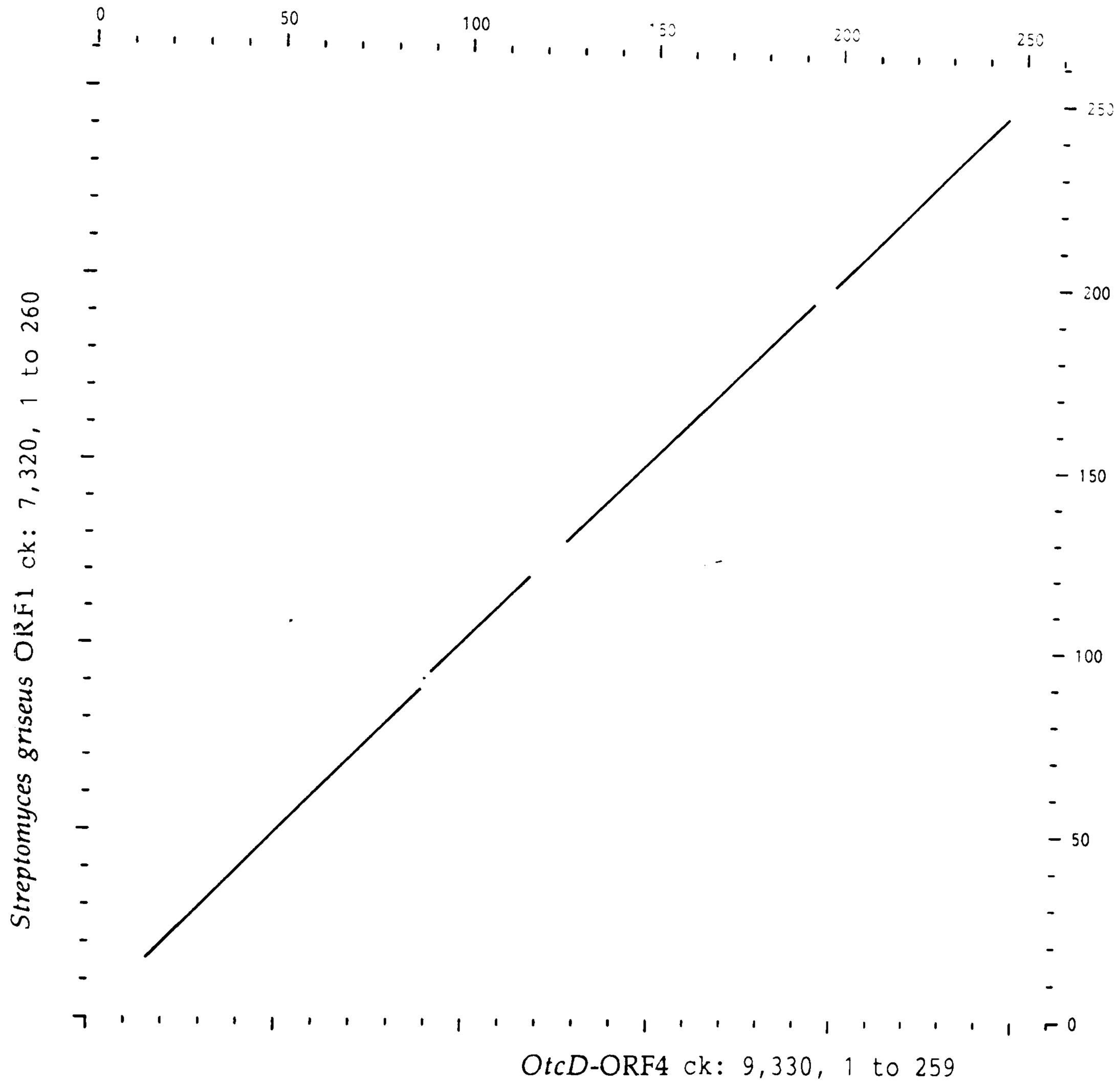


Figure 3.7
Dot-plot analysis of the PILEUP comparison (Devereux *et al.*, 1984) of the deduced amino acidsequence of *OtcD*-ORF4 with *Streptomyces griseus* ORF1 (stringency, 21; window, 30).

a 79.7% similarity and 66.9% identity with *lkmB* (referred later in this text as Dau-ORF1). The Dau-ORF1 consists of 259 amino acids with a molecular weight of 28.7 kD, similar to *OtcD*-ORF4. A dot-plot analysis of the PILEUP comparison (Devereux *et al.*, 1984) between *OtcD*-ORF4 and Dau-ORF1 is shown in figure 3.7 (stringency 21, window, 30). *S. griseus* strains IMET JA3933, JA5142 and JA5570 produce a complex of anthracycline antibiotics that has been named leukaemomycin (Krügel *et al.*, 1993). The main component of this complex is daunorubicin. Wagner *et al.*, (1984) showed that the first stable intermediate in the biosynthetic pathway, after cyclisation and aromatization, is aklanonic acid. Aklanonic acid is then converted to aklavinone and ε-rhodomycinone. After UV and nitrosoguanidine treatment 39 mutants blocked in different steps of the pathway were isolated. *LkmB* mutants fail to make any detectable intermediate of the daunorubicin pathway and show no typical red colour. They are able however, to convert aklanonic acid and the compounds secreted by other classes of *lkm* mutants into ε-rhodomycinone and daunorubicin. This suggests that the early steps of the pathway have been disrupted in all these mutants. Sequence analysis of ORF1 in the mutant shows a C to T transition at position 1172 causing an amino acid change from glycine to aspartic acid. No function has been assigned to this gene product. Dau-ORF1 was shown to complement *in trans* the *lkm* mutation which seems to affect an early step in daunorubicin biosynthesis. ORF1 of the *lkm* region also shows significant similarity to *dnr*Y in the daunorubicin pathway of *S. peucetius*. This gene also complements a mutation is *S. griseus* that blocks an early step in the biosynthetic pathway (CR. Hutchinson; pers. comm.).

3.2.3.3 *OtcD*-ORF5

The predicted amino acid sequence of *OtcD*-ORF5 is a longer peptide of 353 amino acids with a deduced molecular weight of 38,113 D. When the deduced amino acid sequence was searched using the BLAST program, similarity to an ORF from the *Bacillus subtilis* genome was obtained. It was sequenced by Ogasawara *et al.*, (1994) as part of the systematic sequencing project of the *B. subtilis* genome. It is located within a 180 kb region of the chromosome containing the replication origin. Due to the nature of this project, the elucidation of functions for all ORFs sequenced is not a priority if no homology is found immediately. Interestingly located within the same 180 kb region are a number of other tetracycline-like biosynthetic gene products, namely a tetracycline resistance gene and an acyltransferase gene. An acyltransferase gene product has recently been located immediately downstream of the *OtcD*-

```
 OtcD-ORF5                                                    SS PVAADHKRAQ
B.sub-ORF1   LTRLSKERLL TLQVDEHRAE VNEWLLKTSE LVRAQARHPK
```

```
 OtcD-ORF5   LRGAILRDCD LRDARLSGAD LRGARLARAD LADADLREAD
             |       |  |  | | | | |||| |  |  |  |||| ||
B.sub-ORF1   LPKKVSRGVS LIGAKLKGLD LRGANLRGAL LIAADLRNAD
```

```
 OtcD-ORF5   LREATPARAN LRDADLSDAN VRKADLRFAD LRGVDLWGTD
             ||      |  |||||| |
B.sub-ORF1   LRMTDFIGAD MRDADLSGAD LTG
```

```
OtcD-ORF5   LRGAVLYRAK LAGLELSEAH LDGADLRGAD LTDAAVARSQ
```

Figure 3.8
Alignment of *OtcD*-ORF5 with a *B. subtilis* ORF of unknown function. The sequence similarity is confined to a small region of repeat sequences with an overall consensus sequence **LRGAD**.

ORF5, with the opposite direction of translation (Hunter; pers. comm.). The sequence similarity was confined to a short stretch of the sequence located at position 2472 through to position 2927. A BESTFIT comparison of this region showed a 57% similarity and 46% identity (figure 3.8). Similarity is seen over 76 amino acids in a series of direct repeats of 5 amino acids with a consensus sequence of '**LRGAD**'. This region was run through both the PROSITE and MOTIFS (Devereux *et al.*, 1984) programs with no matches. This amino acid region is very hydrophobic and secondary structure analysis has revealed a helix-turn-helix. This preliminary information may suggest some sort of binding site in which the leucine residues may play an important part. It is apparent that this region is significantly unusual to warrant further investigation, and will be discussed further in Chapter 4.

A series of cosmid libraries of *S. peucetius* and *S. galilaeus* were generated at Glasgow University by Ranjan Prasad. Several *Streptomyces* species, including *S. peucetius* and *S. galilaeus* produce 7,8,9,10-tetrahydro-5,12-naphthacene quinone glycosides, commonly known as anthracyclines (Ye *et al.*, 1994). Three such anthracyclines are doxorubicin, daunorubicin, and aclacinomycin. The biosynthetic pathway of the anthracyclines has been determined with the use of blocked mutants (Eckardt and Wagner; 1988). Aklavinone is produced by the condensation of nine acetyl equivalents with a propionyl-CoA starter molecule. Aklavinone is glycosylated directly to produce aklavin and aclacinomycin A (Oki *et al.*, 1979; Strohl *et al.*, 1989). In the formation of daunorubicin, however, aklavinone is hydroxylated to form ε-rhodomycin, a red-pigmented compound (McGuire *et al.*, 1980), which is then converted to the final product via several reactions. Mutants which accumulate aklanonic acid can be found in both *S. peucetius* and *S. galilaeus*. It thus appears that aklanonic acid is a primary intermediate in the aklavinone pathway of several organisms (Eckardt and Wagner; 1988).

Hybridisation studies were carried out with these *Streptomyces* spp. to determine if similar gene products to *OtcD*-ORF5 could be found. Positive clones were obtained from Ranjan Prasad. The cosmid library was constructed from DNA of the *S. galilaeus*, which produces aclacinomycin A. An 800 bp *Pvu*II-*Sst*I fragment from pSRG4 (See figure 3.2) labelled with digoxigenin was used as a probe against 6 cosmid colony blots. Clone number 6 hybridised well against the probe. The colony was restreaked to obtain single colonies, DNA was purified from two colonies and digested with *Bam*HI. As controls, *S. peucetius*, *S. galilaeus*

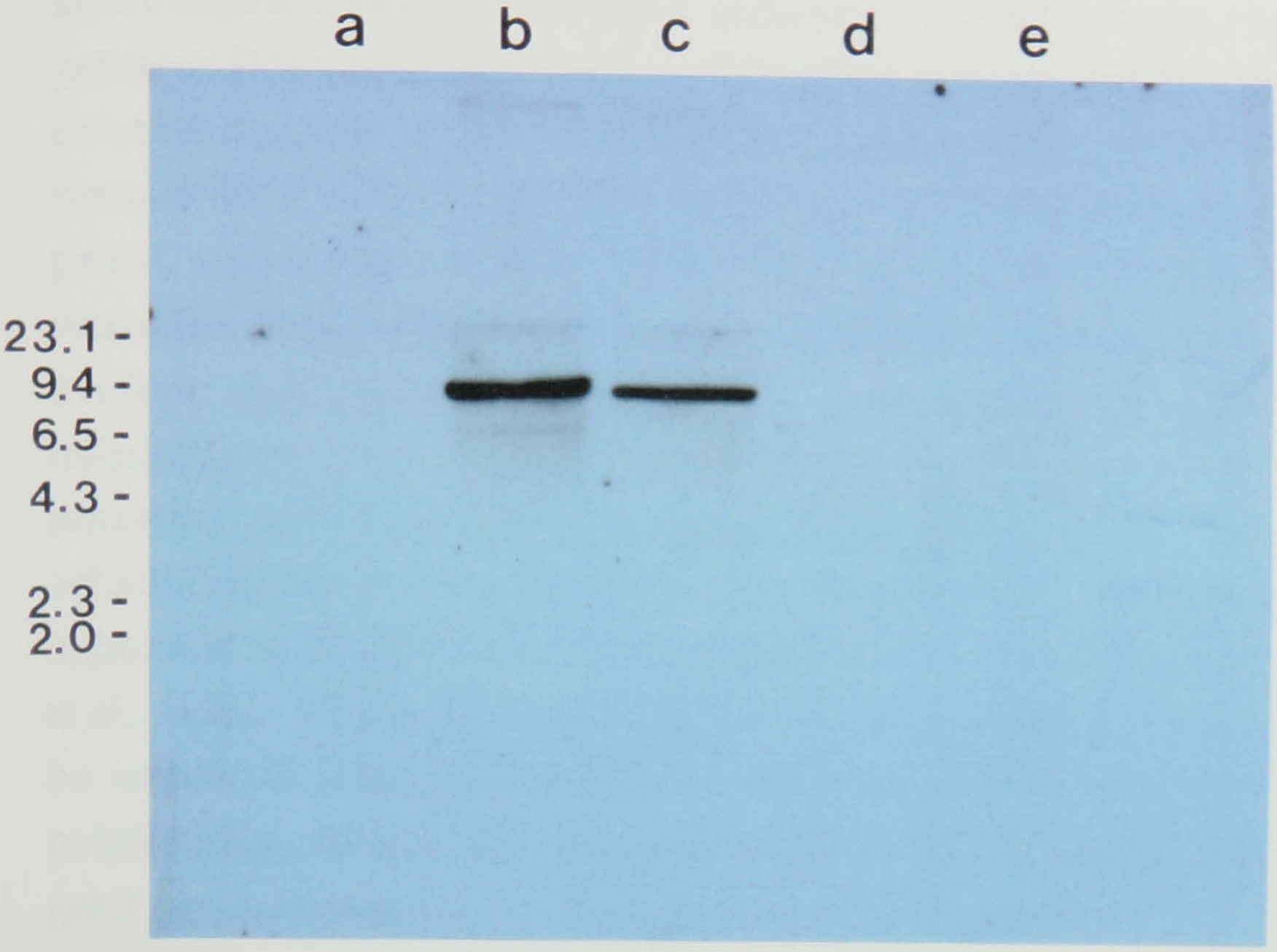


Figure 3.9
Hybridisation analysis of the a cosmid clone from a *S. galilaeus* cosmid library using the*Sst*I_{24}-*Pvu*II_{25a}DNA fragment from the *otc* cluster as a probe (lanes b & c). Genomic DNA's from *S. peucetius* (lane a) *S. galilaeus* (lane d), and *S. lividans* (lane e) were used as controls.

and *S. lividans* genomic DNA were also digested. The same probe was employed and a band of approximately 9 kb was detected in both cosmid clones (figure 3.9). Although not visible on this blot, bands were seen in both *S. peucetius* and *S. galilaeus* genomic lanes after a much longer exposure. However, no band was seen in the *S. lividans* genomic lane. The differences in intensities of the bands are explained by the fact that the cosmid clones produce multiple copies of the DNA segment whereas the DNA segment is present only in a single copy in the genomic DNA samples. Confirmation of the positive results was obtained with a dot-blot analysis of genomic DNA of *S. peucetius*, *S. galilaeus*, *S. lividans* and *S. rimosus* 4018 as well as pSRG1, which carries the homologous sequence of the probe, and pIBI24 vector sequence. Signals were seen in DNA from pSRG1, *S. peucetius*, *S. galilaeus,* and *S. rimosus* 4018 but not *S. lividans* or pIBI24. It is evident that this region of the OTC cluster contains DNA homologous in hybridisation with antibiotic producing strains *S. peucetius* and *S. galilaeus*. *S. peucetius* and *S. galilaeus* produce the anthracyclines, daunorubicin and aclacinomycin A, respectively. An anthraquinone intermediate, similar to aklanonic acid, also has been implicated in oxytetracycline biosynthesis (Thomas *et al.,* 1983). This suggests that aklanonic acid or analogous anthraquinones may be universal intermediates in the formation of various naphthacenequinone polyketides. At this point it is pure speculation to suggest that the *OtcD*-ORF5 gene product may be involved in production of this intermediate, as there are other gene products in common with the biosynthetic pathways. Once again, disruption of this gene product, or overexpression and purification of the protein, would give more credibility to any further speculations.

Chapter 4

Transcriptional Analysis of the *otcX* and *otcD* loci

4.1 Introduction

In the complex life cycle of *Streptomyces* spp., the variety of morphological and biochemical differentiation processes must be co-ordinated. This can be achieved at a number of molecular levels, the primary one of which is transcription. Efficient assembly of complete transcripts occurs with the transcribing enzyme (RNA polymerase) forming a stable complex with the DNA template until it reaches termination. Regulation of this transcription can allow the selective expression of the encoded genetic information. Early studies by Burgess *et al.*, (1969), found that a stimulatory factor, known as sigma (σ), was also required to initiate transcription leading to the possibility that σ factors perform a crucial function in regulation.

Sigma factors can be classified broadly into two families of bacterial proteins: those related in protein sequence to σ^{70}, the primary sigma factor in *E. coli*, and those related in protein sequence to σ^{54}, an alternative *E. coli* sigma involved in transcribing nitrogen-related genes. The existence of alternative sigma factors appears to provide a way of co-ordinately regulating gene expression. Sigma functions by binding to the core RNA polymerase to form a RNA polymerase holoenzyme which is capable of initating transcription. Other auxilliary proteins such as NusA, an elongation factor, play a role in modulation of the efficiency of initiation by binding competitively to the core polymerase. The sigma appears to recognise the DNA only after it has undergone conformational changes brought about by binding to the core polymerase. There is little information about the role of sigma in opening of strands of DNA, binding of the initiating nucleotide, and dissociation during promoter clearance, although previous studies (Burgess *et al.* 1969), show that sigma is essential because core RNA polymerase alone is unable either to locate promoters or to catalyse strand opening on covalently-closed circular DNA. A complete review of bacterial sigma factors can be found in Gross and Lonetto, 1992. The role of sigma in promoter recognition has been studied in more detail and will be discussed below.

The primary sigma factor in *E. coli*, σ^{70}, $E\sigma^{70}$, has been sequenced and comparisons with sigma factors in other organisms, both Gram-negative and Gram-positive, have produced a family of σ^{70}-related sigma's with similar structure and function (reviewed in Gross *et al.*, 1992). RNA holoenzyme containing σ^{70}-type sigma factors recognises promoters which have two

conserved blocks of sequence about 10 bp and 35 bp upstream of the transcriptional start site and are separated by 15-20 bp of non-conserved sequence. In a genetic approach to determine which regions might interact with the promoter sequence, mutant σ^{70}-type sigmas were constructed. Single base changes in the helix-turn-helix motif found in region 4 of the sigma sequence were found to be involved in recognising the -35 bp region (Gardella *et al.,* 1989) and a putative α-helix located in region 2.4 of sigma was found to be involved in recognising the -10 region (Zuber *et al.,* 1989).

DNA sequences involved in regulating transcription are fairly well understood for some bacterial systems such as *E.coli* and *B. subtilis* but little is known about such processes in *Streptomyces* spp. In *E. coli,* extensive DNA manipulation of regions surrounding promoters has been undertaken by a number of groups. Stefano *et al.,* (1982) investigated the spacing between the -35 and -10 regions of published *E. coli* promoter sequences and found that the majority were separated by 16-18 nucleotides. Altering the spacing between the -35 and -10 regions was shown subsequently to decrease the activity of the promoter. Jaurin *et al.,* (1984) showed that increasing the spacing from 16 to 17 in the *E. coli ampC* promoter also increased the promoter activity 16-fold. This, along with work of Siebenlist *et al.,* (1980), helped to confirm the premise of Losick *et al.,* (1981), that sigma factors generally contact nucleotides in both the -10 and -35 hexamers. The conservation of sequences at these two sites also appears to be essential in *E. coli* promoters; at the -10 region, TATAAT and at the -35 region, TTGACA (Hawley *et al.,* 1983). The compilation of promoter sequences of streptomycetes by Strohl (1992), has shown that although some consensus sequences in specific regions can be identified, streptomycete promoters are much more diverse in sequence and regulation of transcription may, therefore, be more complex. Of the 139 DNA sequences associated with apparent transcriptional start sites in streptomycetes, 29 promoters appeared to be similar to those recognised by the eubacterial RNA polymerases containing σ^{70}-like subunits. Consensus sequences at the -10 region [TAg(Pu)(Pu)T] and -35 region [TTGAC(Pu)] were observed. Mutational analysis of the 3' T in the -10 region showed this base to be significant (Janssen *et al.,* 1989), as it is in the consensus sequence in *E. coli* (Gardella *et al.,* 1989). Interestingly enough, mutational analysis of the -35 region did not affect significantly the promoter activity suggesting that this region plays only a minor role in the interaction of the DNA with RNA polymerase (Janssen *et al.,* 1989). Earlier transcription work of Thamchaipenet (1994), in the *OtcY* region of the *otc* cluster in *S. rimosus,*

showed divergent promoters separated by 134 nt. The putative promoter regions (-10 and -35) have been identified with a longer spacing of 19-20 nucleotides between them, different from those *E. coli* promoters that are recognised by $E\sigma^{70}$ RNA polymerase. The -10 and -35 regions of P*act*I and P*act*III of actinorhodin also show a longer spacer distance (21 nucleotides) (Parro *et al.*, 1991) which may explain why neither promoter is active in *E. coli* (Parro *et al.*, 1991). Although the promoters from both *E. coli* and streptomycetes are recognised by RNA polymerase containing σ^{70}-like subunits, the specific sequences recognised in each organism are unique, and in the case of *E. coli,* rigid. It is understandable, therefore, that not all streptomycetes promoters are recognised in *E. coli*. A good example of this is the gentamicin gene employed for gene disruption experiments described in Chapter 6.

A second category of promoter sequences are recognised by the RNA holoenzyme containing $E\sigma^{54}$-type sigma factors, similar to an alternative *E. coli* sigma involved in nitrogen regulation. Promoters which fall into this category consist of two blocks of conserved sequence located 12 bp and 24 bp upstream of the transcriptional start site and are separated by only 5 bp (for review, see Gross *et al.*, 1992). Alternative sigma factors do not, however, share the same high level of similarity as the σ^{70}-type sigma factors. $E\sigma^{54}$ interacts with cognate promoters to form a complex that is easily detected *in vivo* and *in vitro*. The complex formation requires the presence of an appropriate activator and ATP (Sasse-Dwight and Gralla; 1988). The mechanism of activation of the σ^{54}-type sigma factors involves contact between an activator protein and the RNA polymerase rather than a conformational change in the DNA brought about by binding the activator protein, as with the primary σ^{70}-type sigma factors. $E\sigma^{54}$ transcription is completely dependent on the presence of an activator, where the location of the activator binding is less critical and can, in fact, be remote from the RNA polymerase binding site (Gralla; 1991). At least eight different σ^{54}-type RNA polymerase holoenzymes have been found in *Bacillus subtilis*, several of which have active roles in the regulation of morphological development (Errington; 1991). Similarly, seven different alternative RNA polymerase holoenzymes have been found in *Streptomyces*, some involved in global regulation and others in regulating more specific functions (Buttner; 1989). Described below is an example of the complex transcriptional regulation found in *Streptomyces*.

In *S. coelicolor*, the *dagA* gene encodes an extracellular agarase gene. Trancription analysis of the *dagA* gene has revealed four separate promoters,

each of which is recognised by a separate sigma factor (Buttner; 1989). The most upstream promoter *dagA*P4 is recognised by a holoenzyme containing the vegatative σ factor, σ^{hrdB}, *dagA*P3 is recognised by $E\sigma^{52}$, *dagA*P2 is recognised by $E\sigma^{E}$, (formally $E\sigma^{28}$), and *dagA*P1, although not classified, appears to be recognised by a different and distinct σ factor (Buttner *et al.*, 1990) Recent studies with $E\sigma^{E}$, by Lonetto *et al.*, (1994), have revealed a distinctive subfamily of σ^{70}-type σ factors, all of which function as effector molecules responding to extracytoplasmic stimuli and thus have been designated as the ECF (extracytoplasmic function) subfamily. Other sigma factors in this subfamily include CarQ from *Myxococcus xanthus*, AlgU from *Pseudomoas aeruginosa* and SigX from *Bacillus subtilis* (Lonetto *et al.*, 1994 and references therein.). Sequence similarity is conserved across three of the four conserved regions of the σ^{70} family. What makes this family distinct is that the ECF σ factors regulate and respond to extracytoplasmic conditions and with many of these genes, the regulation is controlled negatively by gene products within their own operons.

Nucleotide sequencing of numerous antibiotic biosynthetic gene clusters has lead investigators to consider pathway-specific regulation at the DNA level. Divergent promoters have been found at a number of loci in *Streptomyces* antibiotic-producing strains, for example in the *redD* and *actI* and*III* loci in *S. coelicolor* (Narva *et al.*, 1990; Parro *et al.*, 1991) and the *aph* locus in *S. fradiae* (Janssen *et al.*, 1989). Recent transcriptional analysis work on the *otc* cluster by McDowall (1991) and Thamchaipenet (1994) have uncovered two sets of divergent promoters at different locations in the cluster. Back-to-back promoters were found for the *otcC* and *otcX* genes (McDowall, 1991). The *otc*Cp1 directs transcription of the *otcC-otcZ-otrA* polycistronic message whereas the *otc*Xp1 directs transcription of *otc*X and perhaps genes further downstream (McDowall; 1991). The *otcY*2p1 and *otcY*2p2 promoters the *OtcY* locus, are also arranged back-to-back. Tandem repeat regions have been found in the upstream -35 region at both loci. These repeated patterns strikingly resemble the common phosphate box sequences identified in *E. coli* (Makino *et al.*, 1986). Physiological evidence previously showed that anhydrotetracycline (ATC) oxygenase (encoded by *otcC)* is repressed by phosphate (Behal *et al.*, 1979; MJ. Butler, pers. comm.), suggesting that these repeats may act as regions subject to regulation by phosphate (McDowall, 1991).

Termination of transcription can also be important for the regulation of gene expression. Many genes are regulated by attenuation (Jensen *et al.*, 1986),

while the presence of weak terminators may allow differential expression of genes within a given operon. In spite of the high stability of the transcription complex, there are points where the stability drops sufficiently for the transcript to dissociate spontaneously. These are the transcriptional stop points of intrinsic terminators (reviewed in Richardson, 1993). Once the RNA polymerase has dissociated from the complex it can not reattach in a way which allows continued transcription. The best characterised intrinsic terminators are those found in *E. coli*. A second mechanism of termination in *E. coli* requires the binding of specific protein factors to terminate RNA synthesis efficiently. One such factor is the rho factor (reviewed in Pratt, 1994). Therefore, terminators in *E. coli* can be classified as rho-dependent or rho-independent. Rho-independent termination often occurs at a series of uridine residues after a hairpin structure has been formed by the nascent RNA molecule. One theory of intrinsic termination proposed by von Hippel and Yager (1991) is that the formation of the stem-loop causes the RNA polymerase to pause and the less stable pairing of the A to U residues contributes to destabilisation of the transcription complex, resulting in dissociation of the RNA polymerase. However, not all rho-independent terminators fit this generalisation. This will be discussed further, later in the chapter. In contrast, from the known sequences compiled to date, the rho-dependent terminators do not appear to conform to any consensus structure or sequence. The best-characterised termination factor is the rho protein from *E. coli,* whose discovery was reported more than two decades ago (Roberts; 1969). Rho is an essential cellular protein that is necessary for transcriptional termination at particular sites. In brief, it acts by binding to the nascent RNA chains upstream of the termination sites, in regions that are relatively unstructured and usually rich in cytosine residues. The binding of rho to RNA activates a nucleoside triphosphatase (NTPase) activity essential for rho-dependent termination. The NTPase activity, primarily ATPase, is thought to provide energy for translocation of rho along the RNA toward the polymerase and for unwinding the RNA-DNA duplex in the transcription bubble to facilitate release of RNA from the transcription complex (reviewed in Richardson; 1993).

Little is known thus far about transcription termination in *Streptomyces*. Glycerol catabolism in *S. coelicolor* is determined by two structural genes *gylA* , glycerol kinase, and *gylB,* glycerol-3-phosphate dehydrogenase. The two genes are induced coordinately by growth on glycerol and repressed by growth on glucose. Genetic analysis confirmed that they constitute an operon and transcriptional ananlysis of the operon was carried out (Smith; 1985). A *gyl* '0.9'

knt transcript was found, by low-resolution S1 mapping, to end shortly after an inverted repeat. The termination point of the transcript for the *hyg* (hygromycin phosphotransferase) gene of *S. hygroscopicus* was also determined and the sequences of the *Streptomyces aph, tsr* and *vph* genes also contain inverted repeats which might be transcriptional terminators (Thompson *et al.*, 1983; Bibb *et al.*, 1985). In all the above mentioned sequences, none of these inverted repeats was followed by a run of thymidines, characteristic of the rho-independent terminators in *E. coli*.

A number of potential stem-loop structures have been located in the *otc* cluster just after putative translational stop codons (McDowall, 1991; Thamchaipenet, 1994). One has been located following each of the two resistance genes,*OtrA*, and *OtrB,* and one after *OtcY1*-ORF3 in the *OtcY* locus. In this chapter, preliminary transcriptional analysis of*OtcX* and *OtcD* was undertaken.

4.2 Results

4.2.1 Total Preparation of RNA from *S. rimosus* M4018

S. rimosus M4018, which is a former OTC production strain (obtained from Pfizer), was used to prepare RNA for the transcription analysis described in this chapter. Thamchaipenet (1994), had previously determined the optimal conditions for growth to produce extracts with good quality RNA. These are summarised briefly below. OTC is produced at a high level in cultures grown in TS1 (production medium developed by Pfizer). However, attempts to prepare intact RNA from such cultures were not successful. Alternatively, the total RNA was prepared from cultures grown on tetracycline soluble medium (TSM6; Section 2.2.1). RNA was isolated more easily from this medium than TS1, which formed colloidal-like compounds, although the amount of mycelium and production of OTC were considerably lower than in the TS1 production medium (I.S. Hunter; pers. comm.). McDowall (1991), analysed the RNA isolated from cultures of *S. rimosus* M15883, a production strain descended in the Pfizer lineage from M4018, at different stages during the production phase of OTC by high resolution S1 nuclease protection assays and found that the level of *otcC-otcZ-otrA* and *otcX* transcripts increased by a factor of 4 between 22-46 hours and steadily declined over the next two time points of 69 and 93 hours. Similar results were seen with *S. rimosus* M4018 (Thamchaipenet; 1994). Therefore, the samples for isolation of RNA from *S. rimosus* M4018 were harvested at 40 hrs

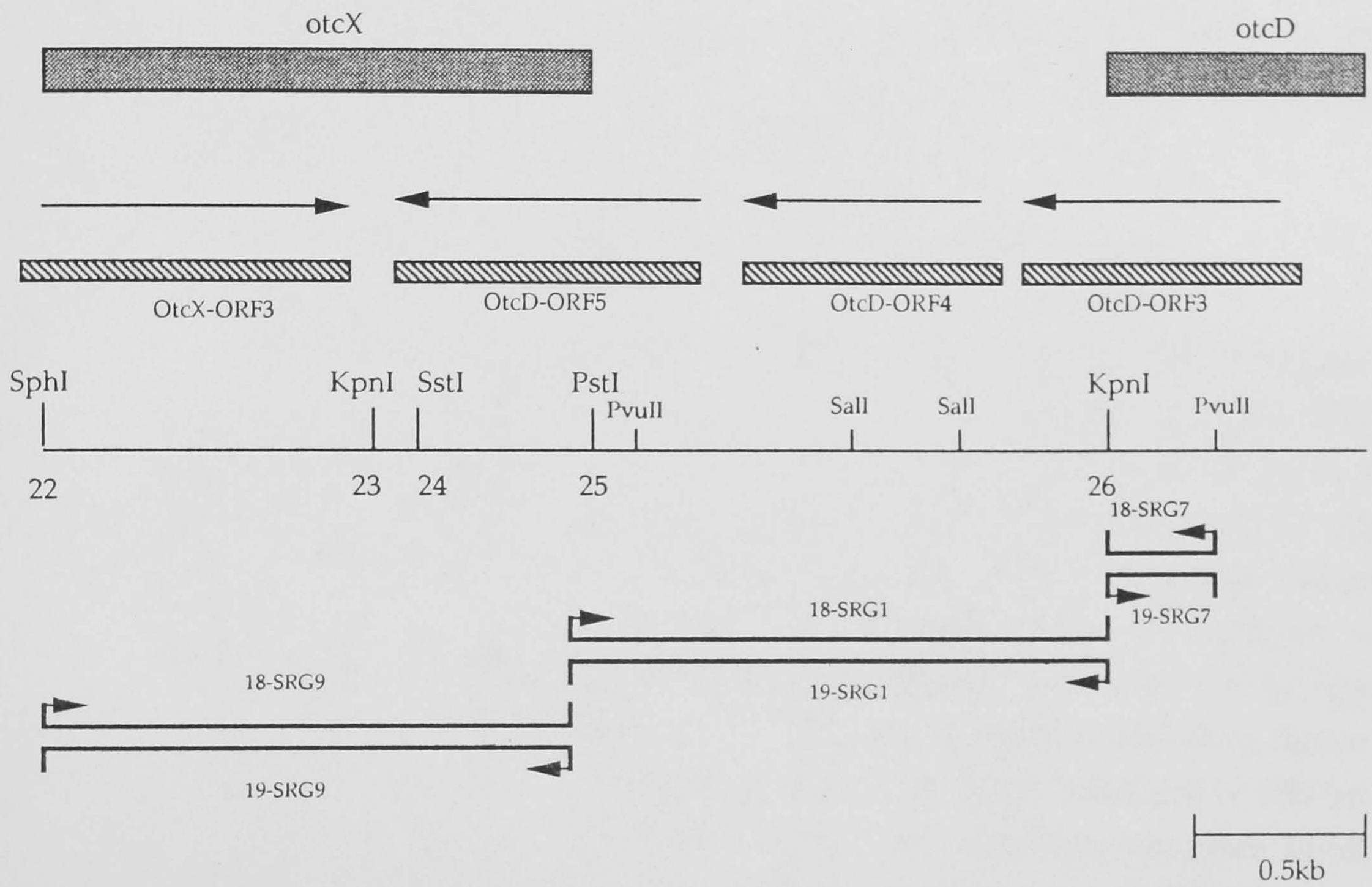


Figure 4.1
A schematic representation of the single-stranded templates used to transcriptionally map the *Sph*I$_{22}$ to *Pvu*II$_{26a}$ region of the *otc* cluster. Above the restriction map are the ( ) genetic loci, as described by Butler *et al.* (1989), and the deduced open reading frames ( ) resulting from sequence analysis. The long arrows above the deduced ORFs show the direction of probable transcription/translation. The arrowheads in show the direction of M13 DNA probes in which the transcripts are protected.

fermentation in TSM6 media. Total RNA was prepared and used in low- and high-resolution S_1 nuclease mapping.

4.2.2 Low-Resolution Mapping

Transcriptional analysis of the region, *Sph*I_{22} to right of *Kpn*I_{26} in the *otc* cluster (for sequence analysis see chapter 3), was performed by low-resolution mapping experiments using a series of single-stranded DNA templates. Figure 4.1 shows the region mapped and the templates used. Mapping was carried out in three separate experiments as described below.

4.2.2a Low-Resolution Mapping of the *Kpn*I_{26}-*Pvu*II_{26a} region

The single-stranded M13 templates 18-SRG7 and 19-SRG7 (figure 4.1) were prepared under RNAase-free conditions. The length of insert (350 bp) contains within it part of open reading frame *OtcD*-ORF3. One μg of each template was hybridised to 20μg of total RNA. Templates hybridised to 10μg of carrier tRNA and to each other served as controls. After S1 nuclease treatment, as described in Materials and Methods section 2.2.12, all samples were electrophoresed. The blotted gel was probed with a DIG labelled 350 bp *Kpn*I_{26}-*Pvu*II_{26a} fragment isolated from pSRG7. The resulting blot is shown in figure 4.2. The positive control lane shows the expected protected fragment of 350 bp. A protected fragment can also be seen in the lane containing template 18-SRG7. Even after prolonged exposure, not bands were visible in the 19-SRG7 lane.

4.2.2b Low-Resolution Mapping of the *Pst*I_{25} -*Kpn*I_{26} region

The templates employed in the second low resolution experiment were 18-SRG1 and 19-SRG1 (figure 4.1). The insert length is 1.5 kb and contains within it the C-terminus of *OtcD*-ORF3, the entire DNA sequence of *OtcD*-ORF4 and the N-terminus of *OtcD*-ORF5. The S1 digestion was done under RNAase-free conditions, electrophoresed and the samples blotted onto membrane. The DIG-labelled probe, a 1.5 kb *Pst*I_{25}-*Kpn*I_{26} fragment from pSRG1, hybridised to the expected 1.5 kb fragment in the positive control lane (figure 4.3). The probe also hybridised to two bands in the 19-SRG1 containing lane. The size of the lower band was calculated to be 1.1 kb and the upper band 1.6 kb. A band size of 1.6 kb does not seem plausible. Unexplained higher molecular weight artefacts have

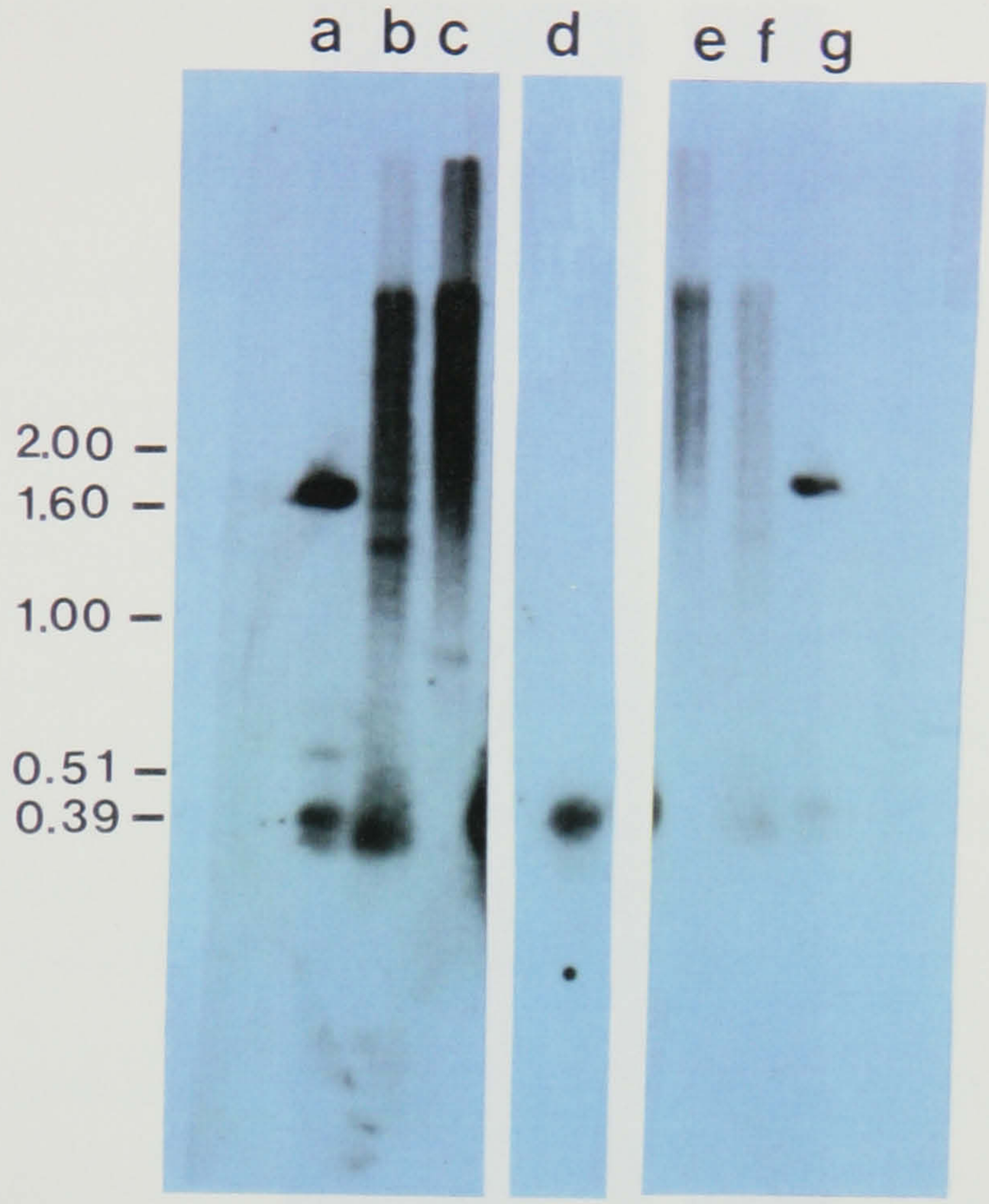


Figure 4.2
Low-resolution S1 mapping of transcripts in the *Kpn*1$_{26}$ to *Pvu*II$_{26a}$ region of the *otc* cluster using total RNA isolated from *S. rimosus* M4018 grown on Tetracycline Soluble Medium (TSM6). 20 μg of RNA were hybridised with single-stranded templates 18-SRG7 (lane b) and 19-SRG7 (lane c). 20 μg of tRNA were used as a control to hybridise with 18-SRG7 (lane e) and 19-SRG7 (lane f). A fully-protected fragment, 18-SRG7 hybridised to 19-SRG7 is shown in lane d. The size markers, 1 Kb ladder (BRL) occupy the outside lanes (a & g). The exposure time for lane d was 15 mins, all others were 2 hrs.

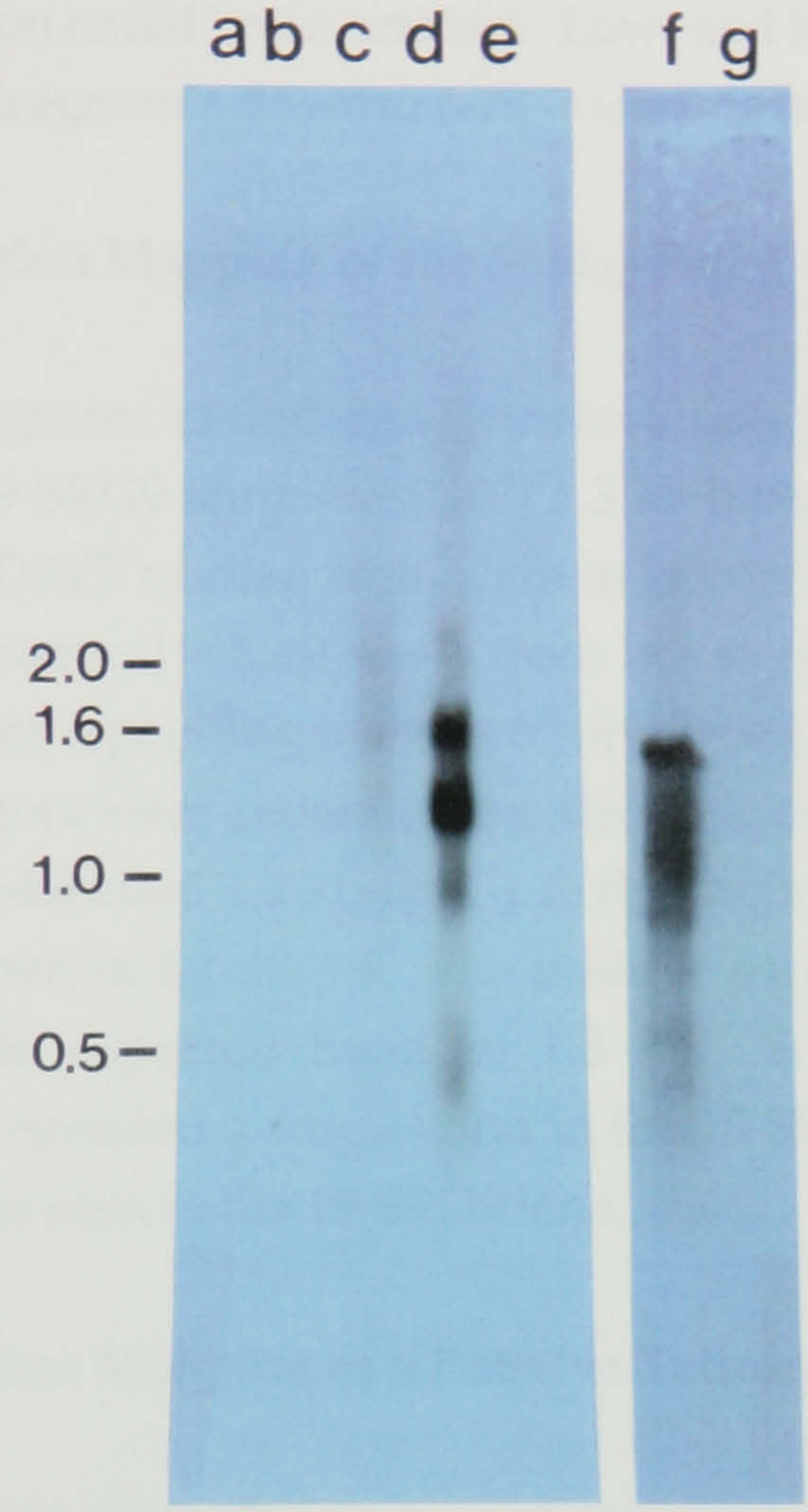


Figure 4.3
Low-resolution S1 mapping of transcripts in the *Pst*I$_{25}$ to *Kpn*1$_{26}$ region of the *otc* cluster using total RNA isolated from *S. rimosus* M4018 grown on Tetracycline Soluble Medium (TSM6). 20 μg of RNA were hybridised with single-stranded templates 18-SRG1 (lane c) and 19-SRG1 (lane d). 20 μg of tRNA were used as a control to hybridise with 18-SRG1 (lane b) and 19-SRG1 (lane e). A fully-protected fragment, 18-SRG1 hybridised to 19-SRG1, is shown in lane f. The size markers, 1 Kb ladder (BRL) occupy the outside lanes (a & g). The exposure time for lane f was 15 mins, all others were 2 hrs.

been seen previously in S1 nuclease protection assays (C. Ingham, pers. comm.). It may be a fully protected fragment, indicating readthrough of termination from the smaller termination product seen by the detection of the smaller 1.1 kb band. A second explanation could be an artefact. Low- and high-resolution mapping of the region and the fragments downstream aided interpretation of the data.

4.2.2c Low-Resolution Mapping of the *Sph*I$_{22}$-*Pst*I$_{25}$ region

The final fragment to undergo low-resolution mapping was contained in the 18-SRG9 and 19-SRG9 templates with 1.3 kb inserts (figure 4.1). Contained within it are *OtcD*-ORF5 reading from right to left and a putative acyltransferase (I.S. Hunter; pers. comm.) which reads from left to right. There is, therefore, a presumed terminator separating these two converging open reading frames. The S1-digested fragments were separated by electrophoresis. The blotted gel was probed with a DIG-labelled 1.3 kb *Sph*I$_{22}$-*Pst*I$_{25}$ fragment isolated from pSRG9. The results are shown in figure 4.4. The positive control of the two hybridised templates gave the expected band of 1.3 kb. A longer exposure of the experimental lanes revealed a single band in the 18-SRG9 lane of approximately 1.3 kb. No band was seen in the 19-SRG9 lane.

4.2.3 High-Resolution Mapping of a Putative Termination Region

From figure 4.3, a smaller protected fragment was observed for 19-SRG1 containing the *Pst*I-*Kpn*I region of the *otc* cluster (see figure 4.7). Estimated sizes indicated that the protected fragment was 1.1 kb, 0.4 kb smaller than the full length protected fragment size. To pinpoint the termination point, a higher resolution mapping experiment was undertaken. Two oligonucleotides, previously used for sequence analysis, were employed to produce a PCR product of the region encoding the putative termination site. Oligos P1 and P9 (See figure 3.1 for location of the primers) have been designed to bind to opposite strands of the DNA, 281 bp apart. A PCR fragment was generated using 18-SRG1 as the template and adding P1 and P9 as the primers. The resulting PCR product was then ligated to a linearised pT7Blue(R) vector (Novagen). pT7Blue has been engineered to simplify cloning of PCR products generated with *Taq* DNA polymerase. *Taq* DNA polymerase leaves a single 3' A-nucleotide which overhangs the reaction products (Clark; 1988). pT7Blue(R) has been prepared for T-cloning by digestion with *Eco*RV, followed by the addition of single 3' dT residues at each end. A blue/white selection process allows easy identification

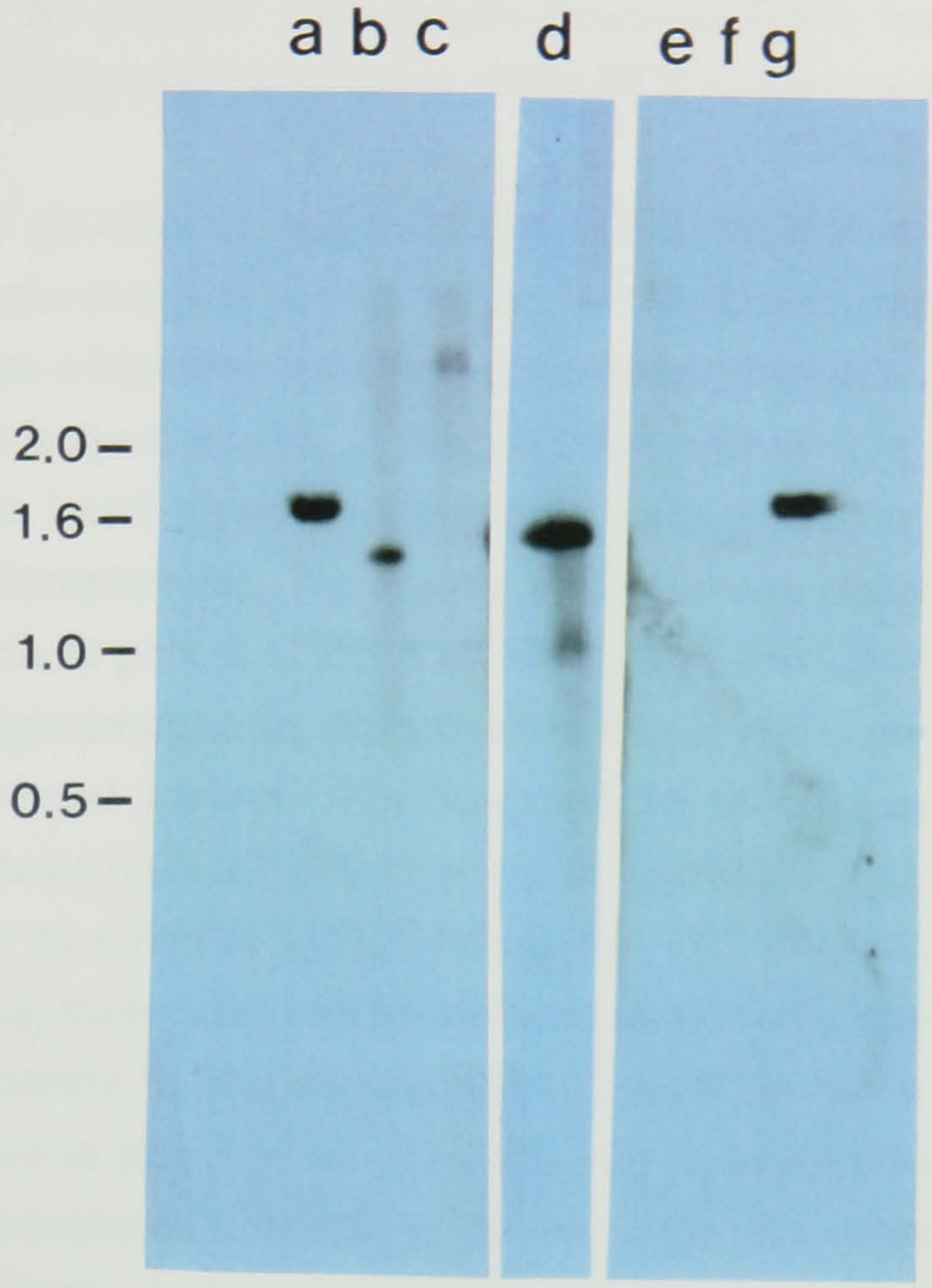


Figure 4.4
Low-resolution S1 mapping of transcripts in the *Sph*1$_{22}$ to *Pst*I$_{25}$ region of the *otc* cluster using total RNA isolated from *S. rimosus* M4018 grown on Tetracycline Soluble Medium (TSM6). 20 μg of RNA were hybridized with single-stranded templates 18-SRG9 (lane b) and 19-SRG9 (lane c). 20 μg of tRNA were used as a control to hybridize with 18-SRG9 (lane e) and 19-SRG9 (lane f). A fully-protected fragment, 18-SRG9 hybridised to 19-SRG9, is shown in lane d. The size markers, 1 Kb ladder (BRL) occupy the outside lanes (a & g). The exposure time for lane d was 15 mins, all others were 2 hrs.

of insert-containing products on indicator plates. A full description of the blue/white selection process is found in Sambrook *et al.*, (1989). Of the 10 colonies picked, 6 clones contained the correct size of insert. One clone was chosen, digested with *Pst*I and *Sst*I to excise the fragment, and cloned into M13mp18 and M13mp19. The single-stranded DNA was then sequenced to confirm the integrity and orientation of the PCR product within the vector.

The PCR product cloned in M13mp19 (19-PCR1) was used as template to generate a single-stranded DNA probe for high-resolution S1 mapping. The primer used was a biotinylated universal primer. Biotin is attached to the end of the primer as it is synthesised. A biotinylated primer was required to utilise the Dynal magnetic bead separation system. This system uses dynabeads M-280 conjugated to Streptavidin, for simple and efficient separation of biotinylated compounds such as proteins, DNA or RNA. Dynabeads M-280 Streptavidin are uniform superparamagnetic, polystyrene beads with streptavidin covalently attached to the bead surface. The Streptavidin protein, made up of 4 identical subunits, contains a high-affinity binding site for biotin and when the Dynabeads are placed next to a Dynal MPC® (Magnetic Particle Concentrator) the target biotin-containing molecules can be isolated in a highly specific manner because they bind efficiently to the streptavidin carried on the magnetic beads. The primer, therefore is attached to the magnetic dynabead and immobilised by a magnet field therefore eliminating the need for gel electrophoresis to purify radioactive products.

After digestion with *Pst*I, the beads (containing the attached probe), were separated away from the unincorporated label and excess nucleotides. The beads were washed thoroughly with RNAase-free solutions before hybridisation with total RNA. The labelled template was divided into two aliquots, one for the experimental sample and one for the control sample. The negative control used throughout this experiment was carrier tRNA. Total RNA isolated from an *otc*$^{-}$ mutant would have been an ideal negative control in this experiment, but was not available. S1 nuclease treatment digested the single-stranded DNA probe at the attachment site of the bead and the protected fragment was separated away from the beads by the introduction of a magnetic field. The resulting supernatant was precipitated and resuspended in loading buffer before being separated on a 6% (w/v) polyacylamide gel. A sequencing ladder of 18-PCR1 primed with oligo P9 was used as a control. The gel was run for two hours and the resulting overnight film exposure is seen in Figure 4.5. Different exposure times were

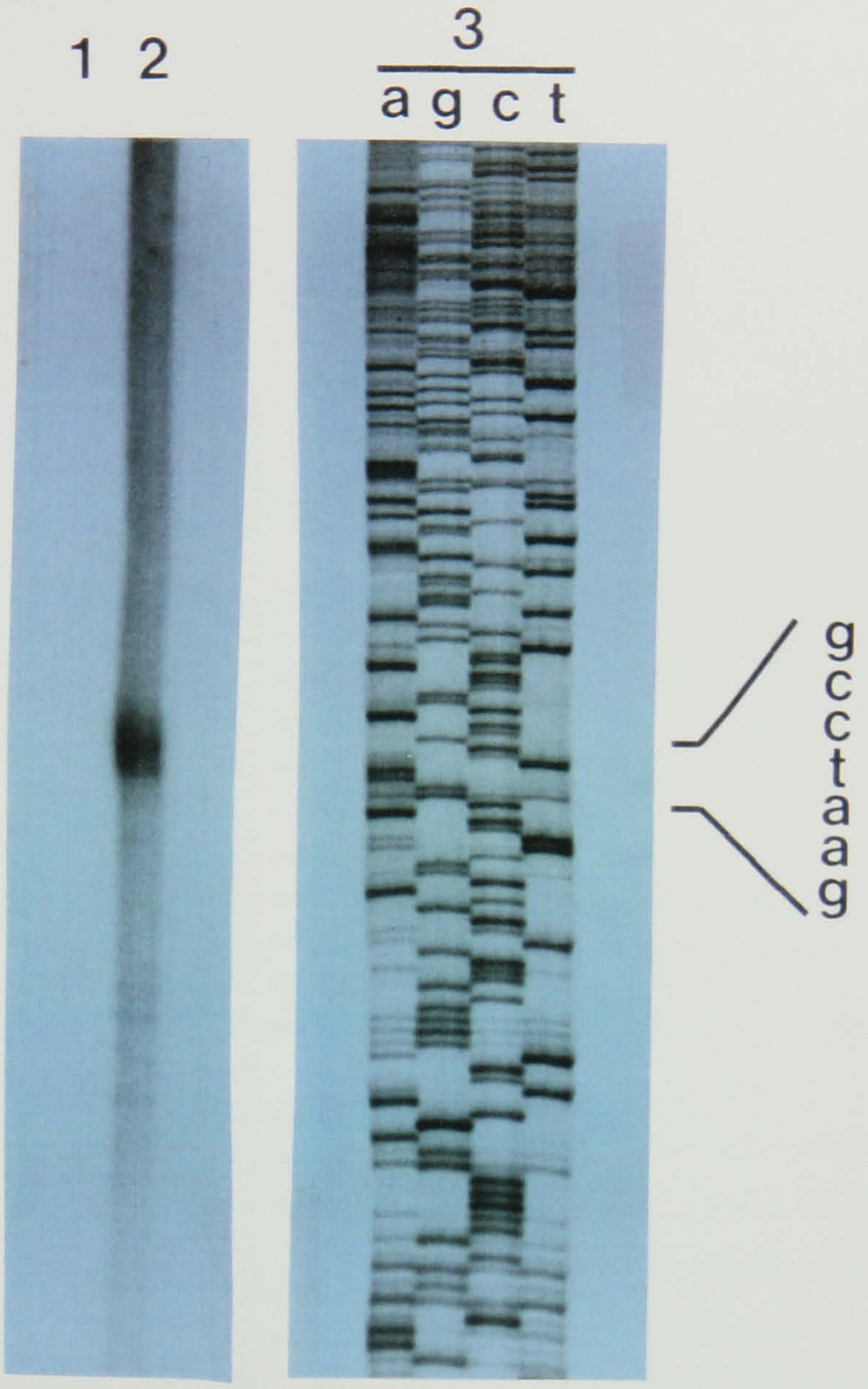


Figure 4.5
High-resolution mapping of the 3' end of the RNA transcript in the region 3' of *OtcD*-ORF4 in the *otc* cluster using total RNA isolated from *S. rimosus* M4018 grown on Tetracycline Soluble Medium (TSM6). 20 μg of RNA were hybridized with a radiolabelled single-stranded template, 19-PCR1 (lane 2). 20 μg of tRNA were used as a control to hybridize to the radiolabelled template (lane 1). A sequencing ladder from the strand complementary to the probe DNA using the same oligo primer is found in lane 3. Lanes 1&2 were exposed for 10 days at -70^{0}C and lane 3 was exposed overnight at RT.

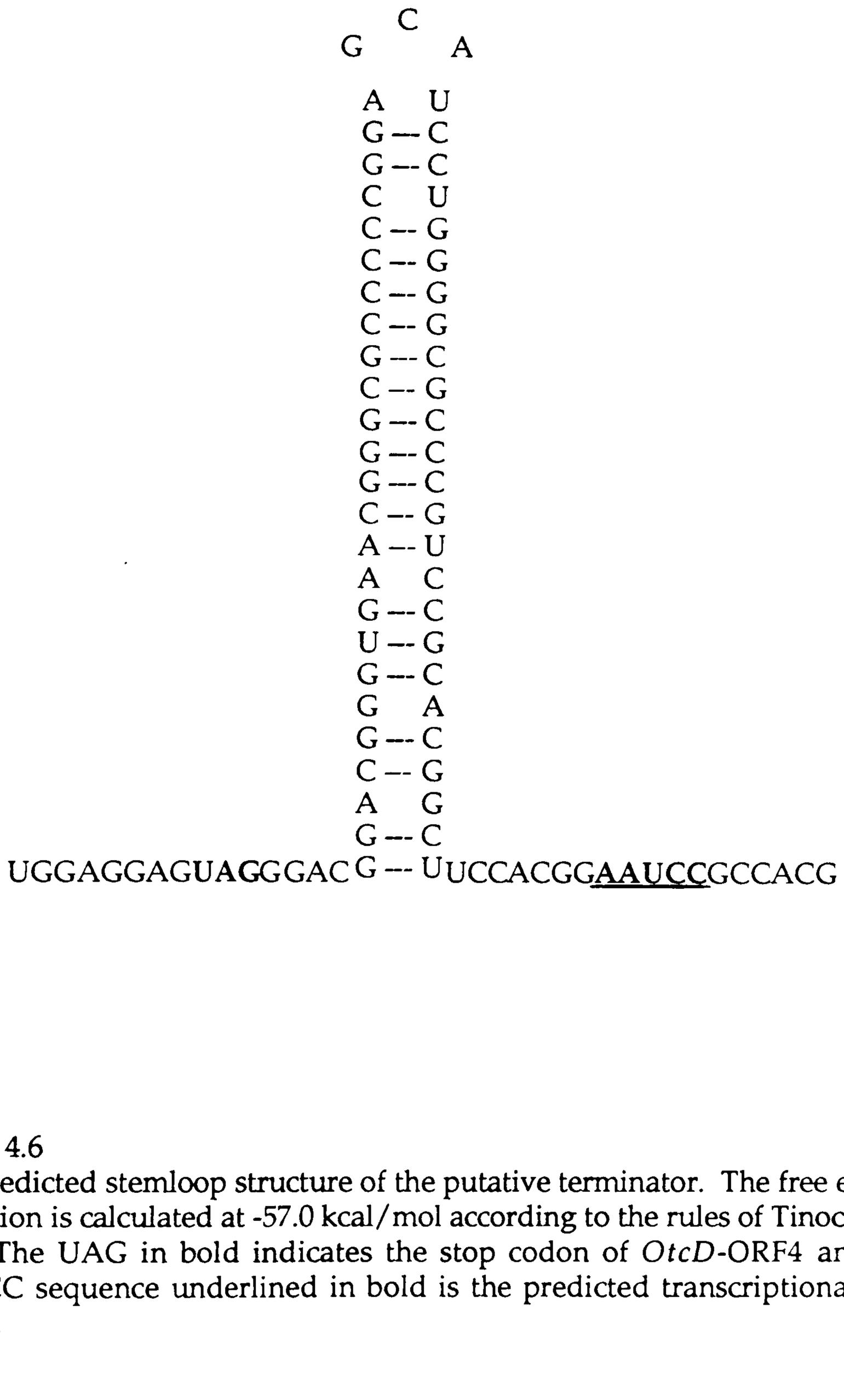


Figure 4.6
The predicted stemloop structure of the putative terminator. The free energy formation is calculated at -57.0 kcal/mol according to the rules of Tinoco *et al.* 1973. The UAG in bold indicates the stop codon of *OtcD*-ORF4 and the AAUCC sequence underlined in bold is the predicted transcriptional stop region.

required to visualise the sequencing ladder and experimental samples. Overnight exposure was sufficient for the ladder, 10 days was required to see the protected fragment. Multiple bands were observed corresponding to the sequence CGGAATCCGCCA. Those bases underlined indicate the approximate site of termination. Direct comparison between the sequencing ladder and the sample could be made because the primer used to generate the PCR product was also used to prime the sequencing reaction. FOLD (Devereux *et al.,*; 1984), a program that finds an optimum secondary structure for RNA molecules using the method of Zucker (1981) showed a strong inverted repeat immediately preceding this stretch of sequence. Figure 4.6 gives the stem-loop structure of the putative terminator showing both the translational and transcriptional stops. The terminator has an estimated free energy formation at 25^0C of -57.0 kcal/mol, according to the rules of Tinoco *et al.,* (1973) and Borer *et al.,* (1974).

4.4 Discussion

Low-resolution protection assays using S1 nuclease described in this chapter suggest that both the *OtcD*-ORF3 and *OtcD*-ORF4 gene products are transcribed from a single message (figure 4.7). These protection assays are in agreement with the predicted direction of translation using the CODONPREFERENCE program (Devereux *et al.,* 1984). The promoter *otcY2*p1 is located immediately upstream of the *OtcY2*-ORF1 gene. It is located upstream and reads in the same direction as *OtcD*-ORF3 and *OtcD*-ORF4. Immediately downstream of the *OtcY2*-ORF1 are the ORFs, *OtcD*-ORF1 and *OtcD*-ORF2. These ORFs are translationally coupled and do not appear to contain any potential stemloop structures or putative promoter sequences (Thamchaipenet; 1994). Therefore, it is presumed that the following ORFs are read from the same 6.2knt transcript, reading from right to left on *otc* cluster map (figure 7.1); *OtcY2*-ORF1, *OtcD*-ORF1, *OtcD*-ORF2, *OtcD*-ORF3 and *OtcD*-ORF4. Immediately downstream of the *OtcD*-ORF4 is a stable stemloop-forming structure.

The region surrounding the translational stop of *OtcD*-ORF4 was examined further using high-resolution mapping. The low-resolution S1 mapping showed a smaller protected fragment than the expected full length control sample. Initial computer analysis of the region using the FOLD and STEMLOOP programs (Devereux *et al.,* 1984), suggested that there may, indeed, be a terminator present. Caution must always be taken when analysing possible

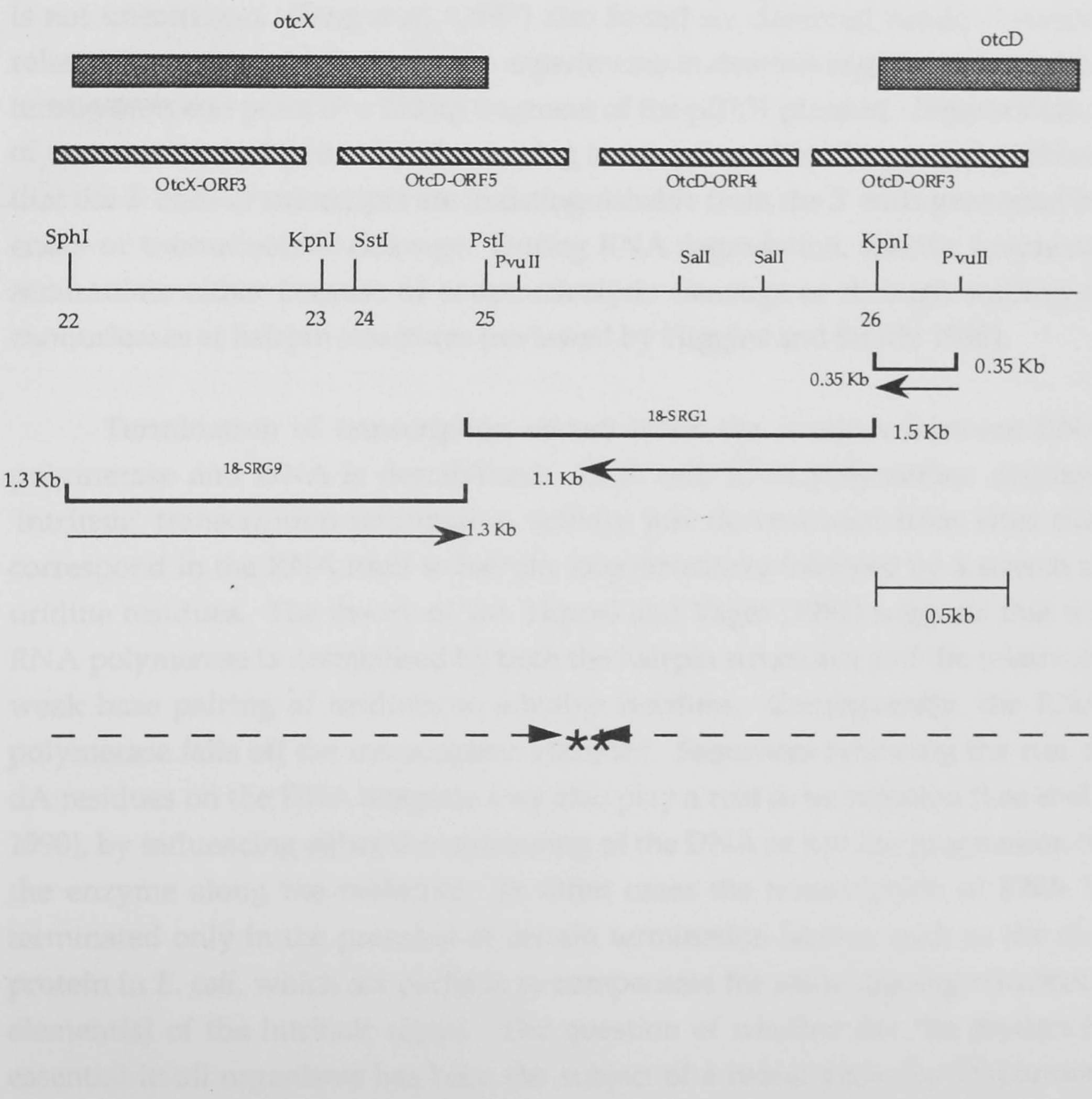


Figure 4.7
A partial transcription map of the *otcX* and *otcD* region of the *otc* cluster in *S. rimosus*. Fully-protected fragments are shown in bold, the S1 nuclease maping results are given below, the direction and length are also indicated. Dashed lines show putative transcripts, the * indicates termination of transcription.

terminator structures in *Streptomyces* mainly due to the high G+C content which gives an increased probability of stem-loop predictions. The high-resolution study revealed a protected fragment within the trancriptional termination point predicted by the FOLD program (Devereux *et al.,* 1984). The transcription seems to end at a series of positions lying within a few residues of each other, rather than at a single nucleotide. The presence of multiple S1 nuclease-generated bands is not uncommon. Deng *et al.,* (1987) also found six clustered bands of similar relative intensity in high-resolution experiments in determining the transcription termination end point of a 250 bp fragment of the pIJ101 plasmid. Determination of termination end-points by S1 mapping is complicated by the general problem that the 3' ends of transcripts are indistinguishable from the 3' ends generated by endo- or exonucleolytic cleavage. During RNA degradation, specific fragments accumulate either because of endonucleolytic cleavage or through stalling of exonucleases at hairpin structures (reviewed by Higgins and Smith; 1986).

Termination of transcription occurs when the complex between RNA polymerase and DNA is destabilised. In *E. coli,* RNA polymerase displays 'intrinsic' transcription termination activity just downstream from sites that correspond in the RNA itself to hairpin loop structures followed by a stretch of uridine residues. The theory of von Hippel and Yager (1991) suggests that the RNA polymerase is destabilised by both the hairpin structures and the relatively weak base pairing of uridines to adenine residues. Consequently, the RNA polymerase falls off the transcription complex. Sequences following the run of dA residues on the DNA template may also play a role in termination (Lee *et al.,* 1990), by influencing either the unwinding of the DNA or just the progression of the enzyme along the molecule. In other cases the transcription of RNA is terminated only in the presence of certain termination factors, such as the rho protein in *E. coli,* which act perhaps to compensate for some missing structural element(s) of the intrinsic signal. The question of whether the rho protein is essential in all organisms has been the subject of a recent paper by Opperman and Richardson (1994). Because transcription can be terminated intrinsically some organisms may not require a rho-type mechanism. A compilation of sequences from diverse bacteria such as *Pseudomonas fluorescens, Chromatium vinosum, Micrococcus luteus* and *Bacillus subtilis* (Quirk *et al.,* 1994) has revealed that Rho homologues are highly conserved (Opperman and Richardson, 1994). PCR primers were designed against conserved protein motifs of *rho* gene from *Micrococcus luteus,* an organism with a high G+C DNA content, similar to

Streptomyces to clone and sequence the rho homologue in *Streptomyces* (C. Ingham, unpublished results).

Until the recent discovery of a rho-like protein in *Streptomyces*, transcription termination had not been extensively studied in *Streptomyces*. Early high-resolution S1 mapping studies on various resistance genes such as *aph* and *tsr* have revealed that termination end points were located just distal to inverted repeats, which may be terminators (Thompson and Gray, 1983; Bibb*et al.*, 1985). None of these inverted repeats in the DNA sequence is followed by a run of thymidines (or uridines in RNA). The lack of extensive runs of uridine was also observed in three terminators located in the intergenic regions from the early region of actinophage ϕC31 (Ingham *et al.*, 1994). It appears to be more of a rule than an exception in the case of Streptomycete terminators. Deng *et al.*, (1987) have demonstrated that a 205 bp fragment from the plasmid pIJ101 has *in vivo* terminator activity in both *S. lividans* and *E. coli*. In both organisms, the termination end point occurs precisely at the same nucleotide, as detected by high-resolution mapping. This suggests that despite the absence of a poly-U tail, the *E. coli* polymerase can recognise the terminator from a G-C rich organism suggesting a common mechanism in both organisms.

Regulation of transcription of antibiotic biosynthetic gene clusters is a complicated process with numerous promoter and terminator sequences. Within the *otc* cluster, a number of putative terminator sequences have been identified. Terminators have been located at either end of the cluster immediately following the antibiotic resistance genes. A third putative terminator has also been found at the end of a polycistronic message within the *otcY* locus. The *otrA* terminator has been analysed by high-resolution S1 mapping and a termination end point was detected immediately after a large stem-loop structure (Doyle; 1987). However, only computer analysis of a potential stem-loop structure was used as the basis for determining the *otcY* (Thamchaipenet; 1994) and *otcB* terminator sequences (Reynes *et al.*, 1988). A fourth potential termination end point was located by high-resolution mapping of the region following *OtcD*-ORF4. As with the work of Deng *et al.*, (1987) a precise end point could not be determined, however a protected fragment was observed corresponding to a termination end-point just distal to a very stable stem-loop structure. Comparison of the four OTC terminators analysed to date is found in figure 4.8. All have high free energy of formation; OtrA

a)

otrA .....CCGG**C**ACCG**GA**CC**G**CG**C**C**CC**GCGCGG.. ...ATTTC ..CG**C**CC**GG**GCG**C**GGTCCGGTGGTTTCAGGTGC

otcD .....GGA**C**GGGT**GA**AC**G**GG**C**G**CC**CCCGG... ..AGCAT.....C**C**TG**GG**CGC**C**CGTCCGCACGGCTTCCACGG

otcY GTCTTGATCCT**C**GTAC**GA**CG**G**GA**C**C**CC**CGGCC... TAGCGTTTC...GG**C**CG**GG**GGTC**C**CGTCGTACGTGTGTGGCGAC

otrB .....ACC**C**GTAC**GA**.CGCG**C**T**CC**GGCCGCGGG GAACGTCCGA TTC**C**CC**GTG**CCGGAACGC.GCGGTTCCC

b)

tcmA GGTGAATGC**C**GGCGGGGTCCGG. .TCCGG ...**C**GGACCCCGCC**C**GGC**A**C**C**GCTGTTT

tcmN .....GGCG**C**C**G**CGGGGTCCGG**C** ACCCGG GC**C**GGACCCCGCC**C**GGAACC.....

Figure 4.8
Alignment of the putative terminators of the (a)*otc* cluster and the(b) tetracenomycin cluster. The left-hand panel shows the left arm of the inverted repeat, the middle panel, the 'loop' and the right-hand panel, the right arm of the inverted repeat. Those bases in bold are conserved in all sequences within the particular group.

-62kcal/mol; OtcY -65 kcal/mol; OtcB -54 kcal/mol and otcD -57 kcal/mol. The striking similarities found in all four tempts one to suggest that the mechanisms involved in termination may be specific to the antibiotics they produce. This is encouraged by the alignment of two tetracenomycin terminators, also shown in figure 4.8. The dissimilarity between the *tcm* and *otc* cluster terminators may also be significant. A closer look at downstream sequence did not reveal any obvious similarities.

Preliminary data from low-resolution S1 mapping suggested a read-through product of the *Pst*I-*Kpn*I region of pSRG1 (figure 4.3), and one might expect to see it on the high-resolution S1 mapping, however no band was detected. This may be due to the conditions used to run the gel. A third low-resolution experiment to examine the region immediately downstream, the *Sph*I-*Pst*I fragment from pSRG9, also showed no transcript reading in the same direction as the read-through product. The size of the read-through product was slightly higher than the fully protected fragment and could easily have been an artefact. Within this region is also the *OtcD*-ORF5 gene. No transcript was detected in RNA isolated from 40 hr cultures. A closer look at the sequence reveals a significant ribosome binding site and a -10 and -35 region with good similarity to the consensus sequences for *Streptomyces*. Strohl (1992), produced a consensus sequence of the -10 region [TAg(Pu)(Pu)T], the -10 sequence of *OtcD*-ORF5 is TGCGAT. The -35 region (TTCCAC), is also in good agreement with the consensus -35 region [TTGAC(Pu)] defined by Strohl, (1992). The spacing between the two regions is 20 bp, also well within the distance of the previously observed promoter regions in *Streptomyces*. Further computer analysis of the *OtcD*-ORF5 sequence has revealed a number of interesting features. The very hydrophobic repetitive region of predicted leucine residues in the C-terminal half of the protein has some resemblance to leucine zippers, which are involved in DNA binding (Csank *et al.*, 1992). The spacing of leucines every fifth amino acid does not rigidly conform to the leucine zipper motif of every seventh amino acid. However secondary structure analysis does reveal a helix-turn-helix in this region. Leucine zippers have been associated with regulatory functions which may also be the role of the *OtcD*-ORF5 and thus would explain why low-resolution S1 mapping showed no transcript. Transcription of regulatory genes is generally at a low level and for only a short period of time. The fact that this gene is located in the middle of the cluster may suggest that it would be transcribed in association with production of OTC. McDowall (1991) showed that transcripts originating from *Otr*Ap1 accounted for greater than 95% of the

total *otrA* transcription at 24 hours but by 32 hours the level had dropped to 80% and by 56 hours had dropped to 51%. The transcriptional analysis was further complicated however when *otrA* was shown to transcribed from two mRNA's (McDowall, 1991). The production of OTC is not detected until late in the growth phase, approximately 48 hours after cultures are inoculatd with spores. Further S1 mapping studies using RNA isolated from earlier time points may show a transcript in the *OtcD*-ORF5 region. It is not likely that this is a completely inactive gene,

A protected fragment was detected in the low-resolution S1 mapping experiment with the opposite strand template. Preliminary sequence analysis of *OtcX*-ORF3 shows good similarity with the acyltransferase of the daunorubicin-producing C5 strain (Ye *et al.*, 1994). It contains the 'serine' active site motif LGHSIGEM. Since this is a converging transcript, a transcriptional terminator is expected. A good stem-loop structure has not been located but preliminary sequencing in one direction has proved very difficult and could indicate strong secondary structures. Sequencing in the opposite direction usually solves this problem but has not been carried out as yet. The fully-protected fragment of 18-SRG9 would indicate that the transcriptional stop-point is to the right of the *Pst*I$_{25}$ site. No potential stemloops are seen in this vicinity, however a strong probable RNA stem-loop (with a free energy formation of -54.0 kcal/mol) is seen approximately 200 bp further downstream. If this were the case, one would expect a protected fragment of 200 bp with the 18-SRG1 template (figure 4.3). Unfortunately the gel running conditions used would not allow such a small fragment to be seen.

If the above speculations are accurate, then the transcript encoding 18-SRG9 template would overlap almost the entire *OtcD*-ORF5 gene. The translational stop codons for both the *OtcD*-ORF5 and putative acyltransferase genes terminate 21 bp apart. A potential stem-loop within the putative acyltransferase (with a free energy formation of -62kcal/mol) is a possible termination end-point for the *OtcD*-ORF5 gene product but again there would be transcriptional overlap. Figure 4.7 shows the overall transcriptional results of this chapter. The analysis here has shown just how complicated these systems can be. A number of experiments are warranted to help decipher the trancriptional control. Pinpointing the transcriptional stop of 18-SRG9 would initially require a low-resolution mapping experiment with a higher percentage gel

electrophoresis followed by high-resolution S1 mapping. The *OtcD*-ORF5 transcript will have to be found using samples of total RNA isolated at various time points. It is possible that the transcript will only be seen very briefly and at a low level. To show that all the potential terminators described here are active, a series of *in vivo* experiments using variations of promoter probe vectors are required. Ingham *et al.,* (1994) have produced a vector which will evaluate the level of termination by preventing the read through of a *tipA* promoter. These experiments may help to clear up the transcriptional regulation of the *otc* cluster.

A repressor-like gene product has been located next to the *otrB* gene (S. McGregor; pers. comm.). It shows good similarity with *E. coli* multi-drug resistance repressors (IS Hunter, pers. comm.). Sequencing the region between the two resistance markers is almost complete and no 'typical' activator has been found. The antibiotic clusters studied to date all contain regulatory genes, generally located in the middle of the cluster, which appear to be pathway-specific activators. An example of an activator is the *actII*-ORF4 gene of the *act* cluster, described in Chapter 1.6. The only exception is the negatively-regulated methylenomycin (mmy) cluster, also noted in Chapter 1.6. More than one activator has also been located within the chlorotetracycline (CTC) cluster (Ryan *et al.,* 1994). The location of the *OtcD*-ORF5 gives support to the speculation that it may be a variant on the typical activator structure. A simple experiment would be to introduce multicopies of the gene on a plasmid and look for an increase in OTC levels. Gene knock out experiments could also be done because preliminary transcriptional analysis indicates it is a monocistronic message. In either case, an answer can be sought to test this theory.

Chapter 5

Improving the Methods of Introducing DNA in the *S. rimosus* host

5.1 Introduction

Butler *et al.,* (1989) reported the cloning and characterisation of the DNA linking the two *otc* resistance determinants from *S. rimosus*. However the approach was very labour intensive. A cosmid library of total genomic DNA was constructed in *E. coli* with the intention of shunting clones containing *otc* DNA into non-producing *S. rimosus* mutants to assess the ability of the clones to complement and produce oxytetracycline. This proved to be unsuccessful as no transformants were obtained, presumably because the *E. coli* derived DNA was restricted by the host strain. Consequently, attempts to employ *E. coli*-type vectors were abandoned in favour of *Streptomyces* vectors. These *Streptomyces*-based subclones, carrying overlapping fragments of *otc* DNA, were passaged through *S. lividans* to avoid the highly-active host restriction system in *S. rimosus*. DNA from *S. lividans* was then transformed into a restriction-minus strain of *S. rimosus* (Hunter *et al.,* 1984) and the resulting transformants were then used to complement the non-producing mutant strains. The inability to transform DNA originating from *E. coli* greatly hampers further characterisation of the cluster. By understanding more about the mechanisms of the *S. rimosus* restriction system, we aim to deliver *E. coli* derived DNA into the *S. rimosus* host M4018.

A host restriction/modification system, by which exogenous DNA is degraded efficiently, has been seen in a number of *Streptomyces* species. (Bailey *et al.,* 1986; Qin *et al.,* 1994). Most restriction-modification (R-M) systems are composed of a methylase and an endonuclease. The methylases catalyse the addition of a methyl group to a specific sequence of four or more bases along the host DNA molecule. The accompanying endonuclease will introduce nicks in both DNA strands, causing subsequent DNA degradation, to any unmodified foreign DNA at or near this specific sequence (Bickle *et al.,* 1993). There are two different types of restriction systems: (1) the classical R-M systems, in which cellular DNA is protected from restriction by modification through methylation of adenosyl or cytosyl residues within the sequences recognized by the restriction enzymes, and (2) systems which cleave only DNA that carries specific modifications. In the latter case, host cell DNA is not restricted because it is not modified (Bickle *et al.,* 1993). The classical R-M systems of *E. coli* are organised into type I, II and III R-M systems (for a comprehensive review, see Bickle *et al.,*;

1993). Type I R-M systems have only been found in members of the *Enterobacteriaceae* family and are grouped in genetically-related families (Bickle *et al.*, 1993). The main enzyme is a multifunctional, multisubunit protein encoding three structural genes, which cuts randomly far from asymmetrical recognition sites. Type II R-M systems are the simplest, with two separate enzymes for restriction and modification. These enzymes cut at fixed positions within symmetrical recognition sites and have made them indispensible in many branches of molecular biology. Type III R-M systems are the smallest group of restriction enzymes which contain two subunits, one of which, the *mod* gene product, can function alone as a modification methylase. The second, the *res* gene product, is only active when complexed with Mod. The endonuclease activity of each of these R-M systems require Mg^{2+} as a cofactor.

A second group of restriction systems, distinct from the classical type I, II, and III R-M, have been found in *E. coli*. The classical restriction systems operate by recognising and destroying DNA that does not carry strain-specific modification. The second set of restriction systems operate specifically on modified DNA. These enzymes recognise and cut DNA carrying the DNA modification signature of other strains and, as a consequence, are not accompanied by the equivalent of a classical modification enzyme. Three such systems in *E. coli* are: McrA, McrBC and Mrr. The McrBC (for methyl-cytosine restricting) is the most extensively studied system. It contains two genes, *mcrB* and *mcrC* organised in a operon. The products of both genes are necessary for restriction. McrBC restricts DNA containing 5-methyl cytosine (5-meC), N^4 methyl cytosine (N^4meC), or hydroxymethylcytosine (hmC) in specific sequence contexts. The*McrBC* locus is located next to the *hsd* genes that code for the classical type I R-M systems, and another methyl-dependent restriction system, mrr, (restricting both 5-meC and N^6meA containing DNA), is located on the other side of the *hsd* genes. The genes for these three independent restriction systems are clustered in this small region of the genome and have thus been dubbed the immigration control region (Kelleher *et al.*, 1991). The recognition sequence for *mcrBC* has not yet been defined, but it seems clear that the nature of the modification is rather flexible, targets include R-mC, where "m" is one of the three cytosine modifications and R is either adenine or cytosine (Raleigh *et al.*, 1986). The second locus, mcrA, is located far away from the clustered restriction-modification genes. McrA codes for a single polypeptide and restricts only DNA which is methylated by the *Hpa* II and *Sss* I cytosine methylases. The Mrr system degrades DNA containing methylated by the *Hha*II or *Pst*I modification enzyme

(Heitman and Model; 1987). Both enzymes methylate adenosyl residues at the N^6 methyl adenosine (N^6meA).

Restriction systems are common in *Streptomyces*, (Matsushima *et al.*, 1987; Sànchez *et al.*, 1985) In a number of *Streptomyces* species, the restriction system is specific for modified DNA. This unmodified DNA transformed into these strains will be tolerated but DNA containing either 5-methylcytosine or N^6-methyladenine will be restricted. MacNeil *et al.*, (1988), investigated the restriction systems of a number of *Streptomyces* strains, with particular emphasis on the *S. avermetilis* strain. The aim was to characterise the methyl-specific restriction system of *S. avermetilis* and subsequently develop an *E. coli* host which produced compatible unmethylated DNA which could readily be transformed into the *Streptomyces* host. The slow growth rates of most *Streptomyces* strains have made *E. coli* shuttle vectors an essential tool in recombinant DNA technology. Of the nine *Streptomyces* spp. tested, seven strains had methyl-dependent restriction systems, including *S. coelicolor* and *S. lividans*. Only *S. avermetilis* restricted DNA modified at either N^6meA or 5meC. The other methyl-dependent strains restricted DNA only at one site, some proving to be more restrictive than others. *S. lividans*, although restrictive to N^6meA methylated DNA, has a very weak restriction barrier and can accept *E.coli* DNA with little effect on transformation efficiencies and therefore has become a popular *Streptomyces* host.

Transformation protocols have been optimised for the genetically well-characterised strains like *S. lividans* and *S. coelicolor*. However, they are not readily applicable to many other *Streptomyces* strains. The transformation procedure of Thompson *et al.*, (1982) is used for *S. lividans* and is also applicable to *S. rimosus*. The varying levels of transformation efficiency between hosts (>10^7 for *S. lividans* and 10^6 for *S. rimosus*) has forced researchers to look for other more efficient means of genetic transfer.

Conjugation is a highly-specific process whereby DNA is transferred unidirectionally from a donor to a recipient bacterium by a mechanism involving cell to cell contact. In *E.coli*, the donor cell usually carries a self-transmissible plasmid which encodes the capacity to carry out the transfer functions. The *tra* region encodes cell surface components as well as mobilisation proteins (*mob* gene products) which promote DNA transfer by interaction with a *cis*-acting sequence known as the origin of transfer (*oriT*) (Willetts *et al.*, 1984). Conjugation between

Gram-negative and Gram-positive bacteria was first demonstrated by Trieu-Cuot *et al.*, (1987). Gene transfer between distantly-related bacteria has been widely demonstrated (Heinemann *et al.*, 1989; Kreps *et al.*, 1990). Of particular interest is the demonstration of conjugation between *E. coli* and *Streptomyces* (Mazodier *et al.*, 1989 and Bierman *et al.*, 1992). The transfer depended on the presence of a 760-base-pair, *cis*-acting, *oriT*-containing fragment and on RP4 (IncP) functions supplied *in trans*. The recipient strain in the Mazodier *et al.*, (1989) study was *S. lividans*. Bierman *et al.*, (1992), employed *S. fradiae*, a strain that is notoriously difficult to transform by PEG-mediated protoplast transformation.

Mazodier *et al.*, (1989) utilised a specially-designed *E. coli* donor to transfer the plasmid pPM803 to a *S. lividans* host. The *E. coli* donor, S17-1 (Simon *et al.*, 1983) was created initially to demonstrate conjugation between *E. coli* and other Gram-negative bacteria. It carries the transfer functions of the broad host range plasmid RP4 to mobilise plasmids. To prevent self-transfer of RP4 plasmid into the recipient, it was immobilised in the donor strain by integration into the *E. coli* chromosome. This feature is not essential when conjugating *E. coli* and *Streptomyces* because the RP4 plasmid has no origin of replication for *Streptomyces*. However, Mazodier *et al.*, (1989) used this donor and mobilised the pPM803 into freshly-germinating spores of*S. lividans* TK24. Exconjugates were selected by resistance to neomycin and nalidixic acid (Mazodier *et al.*, 1989). The neomycin resistance gene was carried by pPM803 and the natural resistance of many *Streptomyces* to nalidixic acid was used to counterselect the sensitive *E. coli* donor. Bierman *et al.*, (1992) used the same donor and mobilised plasmids containing the 760bp *oriT* fragment from the IncP plasmid, RK2. Whereas Mazodier *et al.*, (1989) used freshly-germinating spores of the streptomycete, Bierman's conjugational recipient, *S. fradiae*, was prepared from vegetative mycelia. Thus, the system works well with those *Streptomyces* strains which do not sporulate well.

5.2 Objectives

The objective of this chapter was to characterise the restriction system of *S. rimosus* and find the most efficient means of introducing DNA into the *S. rimosus* host.

The immediate aims of the experiments discussed in this chapter were as follows:

i) to assess three different methods of introducing DNA into M4018: conjugation, transformation and electrotransformation.
ii) to determine the characteristics of the restriction barrier of *S. rimosus* strain M4018.

5.3 Methodology

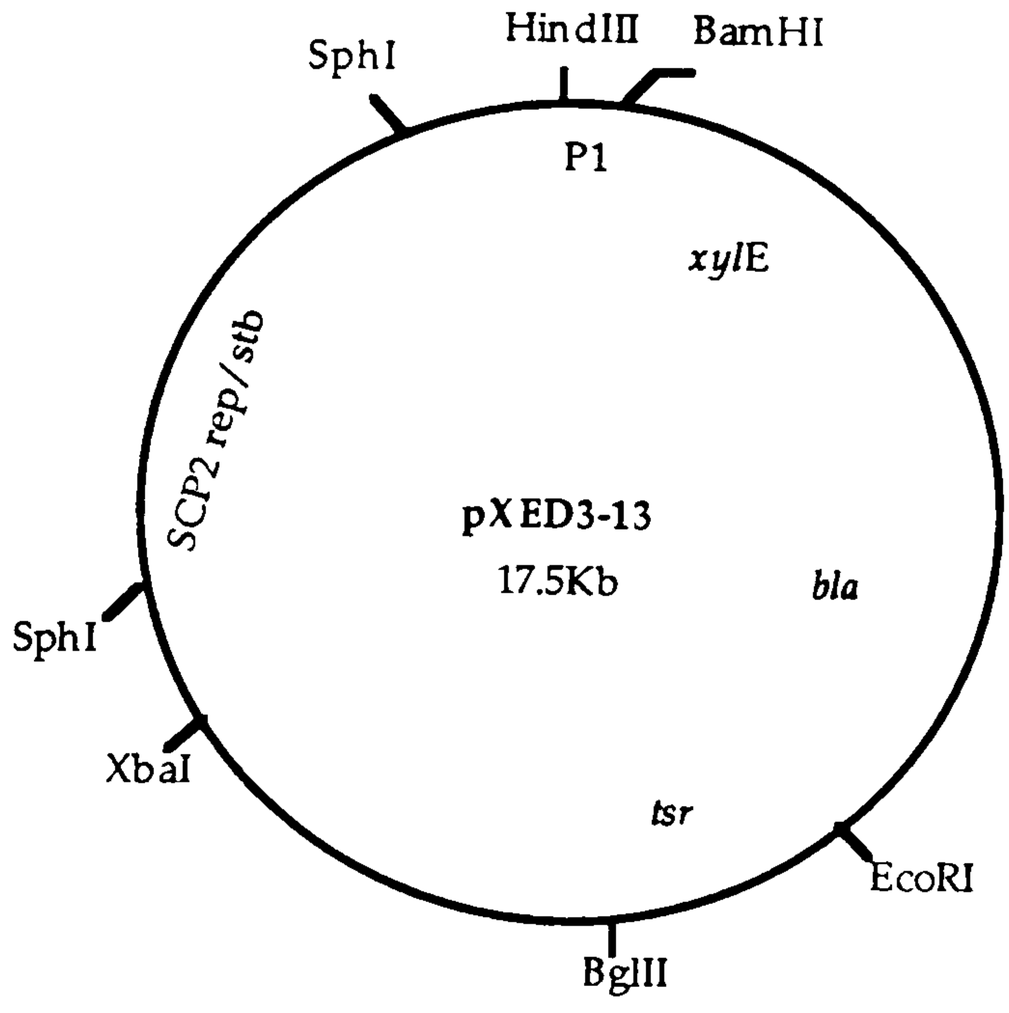
Two plasmids, pXED3-13 and pIJ699, were employed throughout the investigation of the host restriction system in *S. rimosus*. Both are bifunctional plasmids, carrying an origin of replication for both *Streptomyces* and *E. coli*.

pXED3-13 (figure 5.1a) is a SCP2*-based plasmid with a copy number of one to two per cell. A 5.9 kb fragment of DNA containing the *E. coli* replicon, a *gal*PI promoter fused to a *xyl*E reporter gene and an ampicillin resistance marker, was inserted between the *Eco*RI & *Hin*dIII sites of the *Streptomyces*-based plasmid. The resulting 17.5 kb plasmid, contained the SCP2 replication and stability functions along with the thiostrepton resistance gene (Ingram *et al.*, 1989). pIJ699 (figure 5.1b) is a high-copy number plasmid (40-300 copies per chromosome) developed by Kieser and Melton (1988). The vector comprises both an *E. coli* (P15A) and streptomycete (pIJ101) origin of replication, separated by transcriptional terminators from phage *fd*. The plasmid is 9.6 kb in size and selectable in *Streptomyces* by resistance to thiostrepton and in *E. coli* with kanamycin by virtue of the kanamycin/neomycin phosphotransferase gene. The neomycin resistance gene is promoterless and is therefore not functional.

At the outset of this investigation, two *E. coli* hosts were available, DS941 and ET12567. DS941 is a host capable of producing fully-methylated DNA. ET12567 is a methylation-deficient host; it is *dam*::Tn9,*dcm*$^-$,*hsd*M::Tn10, created by MacNeil *et al.*, (1988) to overcome the restriction barrier of *S. avermetilis*. ET12567 is poorly transformed by methylated DNA because hemimethylated DNA is replicated very poorly by DNA polymerase. Thus, some transformants will select for *dam*$^+$ revertants, selection on CAM ensured the presence of *dam*::Tn9. (MacNeil; pers. comm.).

A protoplast transformation procedure developed for *S. rimosus* by Butler *et al.*, (1989) routinely yielded $5X10^6$ transformants/μg of DNA. The electrotransformation procedure developed by J. Pigac (Pigac and Schrempf; 1994) yielded transformation efficiencies of the order of $5X10^6$/μg of DNA. In all

a)



b)

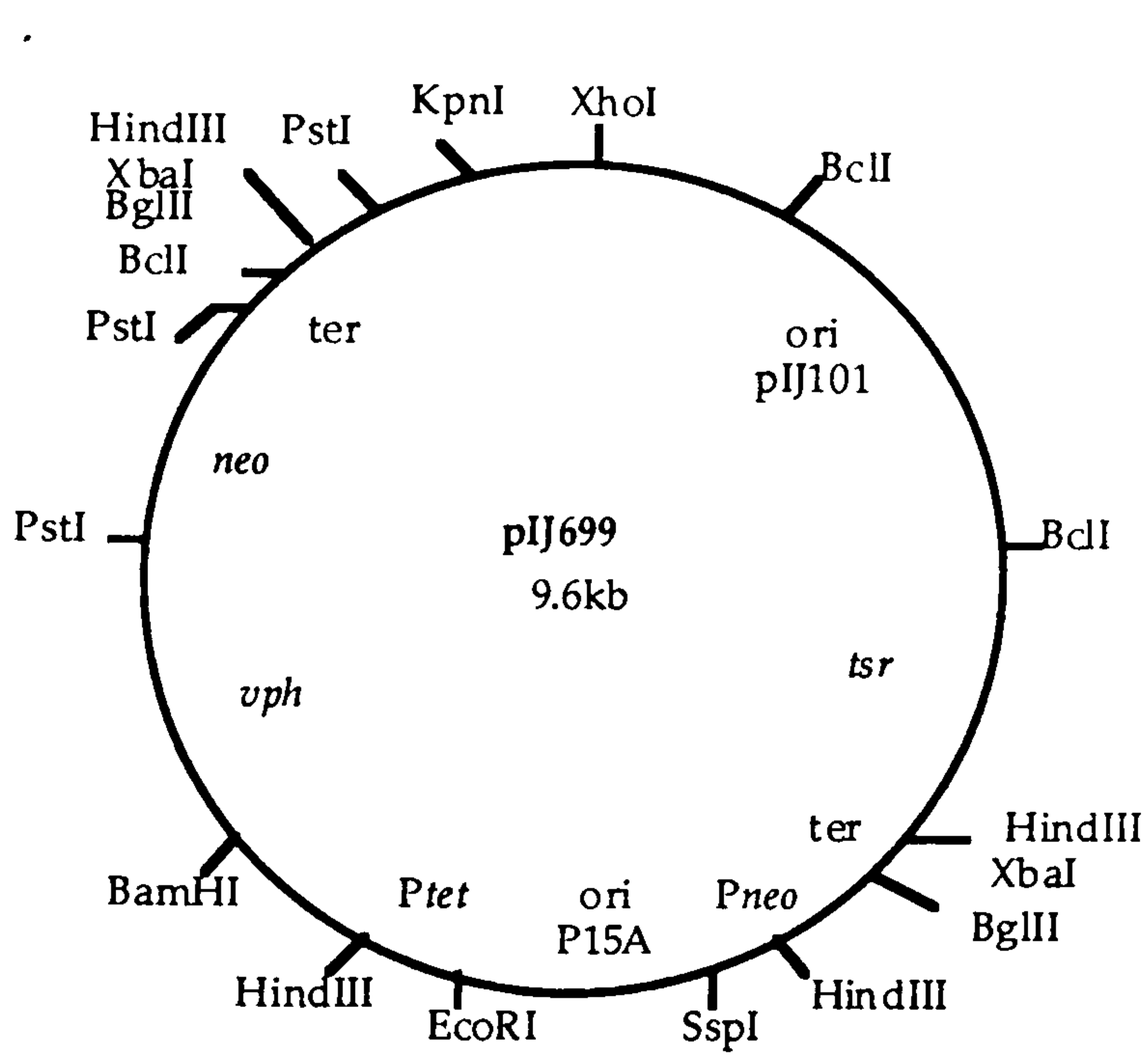


Figure 5.1
Plasmids pXED3-13 (a) [Ingram *et al.* 1989] and pIJ699 (b) [Kieser and Melton; 1988] used in the investigation of the *S. rimosus* restriction system.

transformation experiments, a duplicate dilution series was plated out from 10^0 through 10^{-3}. The average colony counts were based on at least 200 colonies per plate.

5.4 Results

5.4.1 Characterisation of *S. rimosus* Restriction System

Many *Streptomyces* spp. restrict methylated DNA originating from *E. coli* hosts (MacNeil *et al.*, 1988). In fully-methylated strains of *E. coli*, the *dam* methylase converts adenine to 6-methyladenine in the sequence GATC and the *dcm* methylase converts the second cytosine residue in the sequence, CC($^{A}/_{T}$)GG to 5-methylcytosine(Sambrook *et al.*, 1989). Therefore, DNA prepared in *E. coli*, with various levels of methylation, were tested for their relative ability to transform *S. rimosus* protoplasts.

The efficiency of transformation of *S. rimosus* strain M4018 protoplasts was tested with the high-copy number plasmid pIJ699 which had been isolated from *S. lividans* TK24. In two separate experiments the average transformation efficiency was 5.8×10^5 transformants/μg DNA.

DNA was isolated initially from two different *E. coli* hosts, DS941 and ET12567. Colonies from the *E. coli* strain ET12567 were tested on CAM plates to ensure the retention of the *dam*::Tn9 transposon. The transformation efficiencies from *S. rimosus*, of the high- and low-copy number plasmids, pIJ699 and pXED3-13, were also tested. Previous experiments had shown that there was a $>10^3$ fold difference in transformation efficiency between high- and low-copy number plasmids for strain M4018 whereas no difference was shown for the *S. lividans* strain TK24. (IS Hunter; pers. comm.).

The transformation procedure was carried out as described in the Materials and Methods (section 2.2.7). Ten ng of DNA was transformed into 100μl of protoplasts. The experiment was done in duplicate and the average numbers of transformants obtained were recorded in each case (Table 5.1). The pIJ699 plasmid isolated from ET12567 transformed with an efficiency approximately equal to the control plasmid pIJ699 derived from TK24. No transformants were obtained with the low copy number plasmid. A transformation efficiency of less than $10^2.\mu g^{-1}$ would not be picked up because

Plasmid	Transformants.μg^{-1} DNA ($X10^5$)
pXED3-13 (DS941)	< 1
pXED3-13 (ET12567)	< 1
pIJ699 (DS941)	< 1
pIJ699 (ET12567)	9
pIJ699 (*S. lividans*)	58

Table 5.1
Transformation efficiency of low-copy number plasmids (pXED3-13) and high-copy number plasmids (pIJ699) in *S. rimosus* strain M4018. The DNA originated from two different *E.coli* hosts; DS941 which produces fully-methylated DNA and ET12567 which produces methylation-deficient DNA.

only 10ng of DNA was used in the original transformation and since Butler *et al.*, (1989) found the transformation efficiencies of SCP2* based plasmids were at least 100-fold less than multi-copy plasmids, this is not surprising.

Confirmation that an intact *Streptomyces* origin of replication (ori) was present in the plasmids, came when DNA's from all 5 plasmids in the initial experiment were transformed into *S. lividans* TK24. This strain was chosen as the host because it was the most permissive host available. Table 5.2 shows that the *Streptomyces* ori is functional and the number of transformants is not affected by copy number. The activity of the *E. coli* ori was confirmed by transforming the original DNA back into the *E. coli* host TGI. (Table 5.3)

The transformation efficiency of pIJ699 (ET12567) into M4018 was slightly lower than the control pIJ699 (TK24) [Table 5.1]. To assess if the restriction barrier has been completely removed, a colony resulting from the transformation of pIJ699 (ET12567) into M4018, was picked and grown up in 50ml of YEME to purify the DNA. Ten ng of this DNA was retransformed back into the host M4018. The results of duplicate transformation experiments revealed an efficiency of $5.75X10^4$ transformants.μg^{-1} DNA. This is comparable to the transformation efficiency of the original *E.coli* DNA of $9X10^4$ transformants.μg^{-1} DNA (Table 5.1).

MacNeil *et al.*, (1988), have manipulated the *E. coli* host strain ET12567 extensively to remove all it's methylation capacity. This was required to overcome the restriction barrier of *S. avermetilis*. The loss of methylation appears to carry with it some yet further undefined mutations, as it is a poor growing strain with a doubling time ca. 35 mins, compared to the average *E.coli* host doubling time of 20 mins (Sambrook *et al.*, 1989).

Two *E. coli* strains, GM31 and CB51, were subsequently obtained to test whether the *S. rimosus* host strain M4018 requires completely unmethylated DNA before being accepted into the cell. GM31 is a *dcm*$^-$ host carrying a mutation in the gene which methylates C5-cytosine, the internal cytosine residue is non-methylated in either of the following two sequences CCAGG and CCTGG. CB51 (a *dam*$^-$ host) is also methylation-deficient as no methyl group is added to the adenine residue in the 5' GATC 3' sequence. Both strains are much healthier than the ET12567 strain and exhibit a more typical growth rate, with a doubling time of ca. 20 mins. pIJ699, pXED3-13 and pUC18 DNA were transformed into

Plasmid	Transformants.μg^{-1} DNA ($X10^5$)
pXED3-13 (DS941)	0.95
pXED3-13 (ET12567)	0.35
pIJ699 (DS941)	2.1
pIJ699 (ET12567)	1.6
pIJ699 (*S. lividans*)	5.0

Table 5.2
Transformation efficiency of low-copy number plasmids (pXED3-13) and high-copy number plasmids (pIJ699) into *S. lividans* strain TK24. The DNA originated from two different *E. coli* hosts; DS941 produces DNA which is fully-methylated and ET12567 which produces DNA which methylation-deficient.

Plasmid	Transformants.μg^{-1} DNA ($X10^5$)
pXED3-13 (DS941)	10
pXED3-13 (ET12567)	29
pIJ699 (DS941)	1.2
pIJ699 (ET12567)	3.6
pIBI24 (TG1)	4.8

Table 5.3
Transformation efficiency of low copy-number plasmids (pXED3-13) and high copy-number plasmids (pIJ699) in *E. coli* strain TG1. The DNA originated from two different *E.coli* hosts DS941 which produces fully-methylated DNA and ET12567 which produces DNA that is unmethylated.

the *E. coli* hosts CB51 and GM31. Because there are no nutritional requirements to distinguish the two strains, the DNA was digested with restriction enzymes which will only cleave non-methylated DNA. The restriction enzyme *Mbo*I will only cut DNA which is not methylated in the 5' GATC 3' sequence, characteristic of a *dam*$^-$ host. *Eco*RII restriction enzyme is specific for DNA in which the internal cytosine residue of the dcm recognition sites are non-methylated. Figure 5.2 illustrates the ability of the *Eco*RII and *Mbo*I restriction enzymes to cut the pUC18 DNA isolated from the CB51 and GM31 hosts. The 'GATC' sequence is commonly found in pUC18, *Mbo*I cuts this DNA 15 times (Sambrook *et al.,* 1989). pUC18 DNA is cut by the enzyme *Eco*RII 5 times. By comparison with the uncut control pUC18 DNA, it is clear that pUC18 DNA isolated from strain GM31 is cut by *Eco*RII but not *Mbo*I. Similarly, pUC18 DNA isolated from strain CB51 is digested by the *Mbo*I enzyme but remains uncut by *Eco*RII.

The high copy-number plasmid (pIJ699) and low copy-number plasmid (pXED3-13) were transformed into the two *E. coli* CB51, *E. coli* GM31. The DNA was isolated by the Birnboim Doly protocol and 25 ng of DNA was transformed into *S. rimosus* M4018. Two hundred µl of the transformation mix was plated. Transformation efficiencies are given in Table 5.4a, along with the transformation efficiency of the control plasmid pIJ486 which was isolated from a related *S. rimosus,* strain R6. This was kindly provided by J. Pigac. The ability of both the high- and low-copy number plasmids originating from the *dcm*$^-$ strain, GM31, to transform M4018 suggests that the methyl-dependent restriction occurs at cytosine residues. Removal of the methyl groups from the N^6-methyladenines, acheived by isolating DNA from the *dam*$^-$ strain, had no affect on transformation efficiency.

5.4.2 Electrotransformation

Attempts to make the transfer of DNA into *Streptomyces* more efficient and reliable has lead to the development, by J. Pigac, of a transformation protocol using electroporation technology. The cells are made permeable with the use of electric shocks. The rationale of the technique is as follows: since all membranes are charged (negatively on the outside, positively on the inside), application of an electric field, by displacing these charges, leads to an increase in the transmembrane potential. The amplitude of the potential modulation can be monitored by assessing the intensity, voltage, and duration of the electric field as well as taking into account the diameter of the cells. Membranes can be

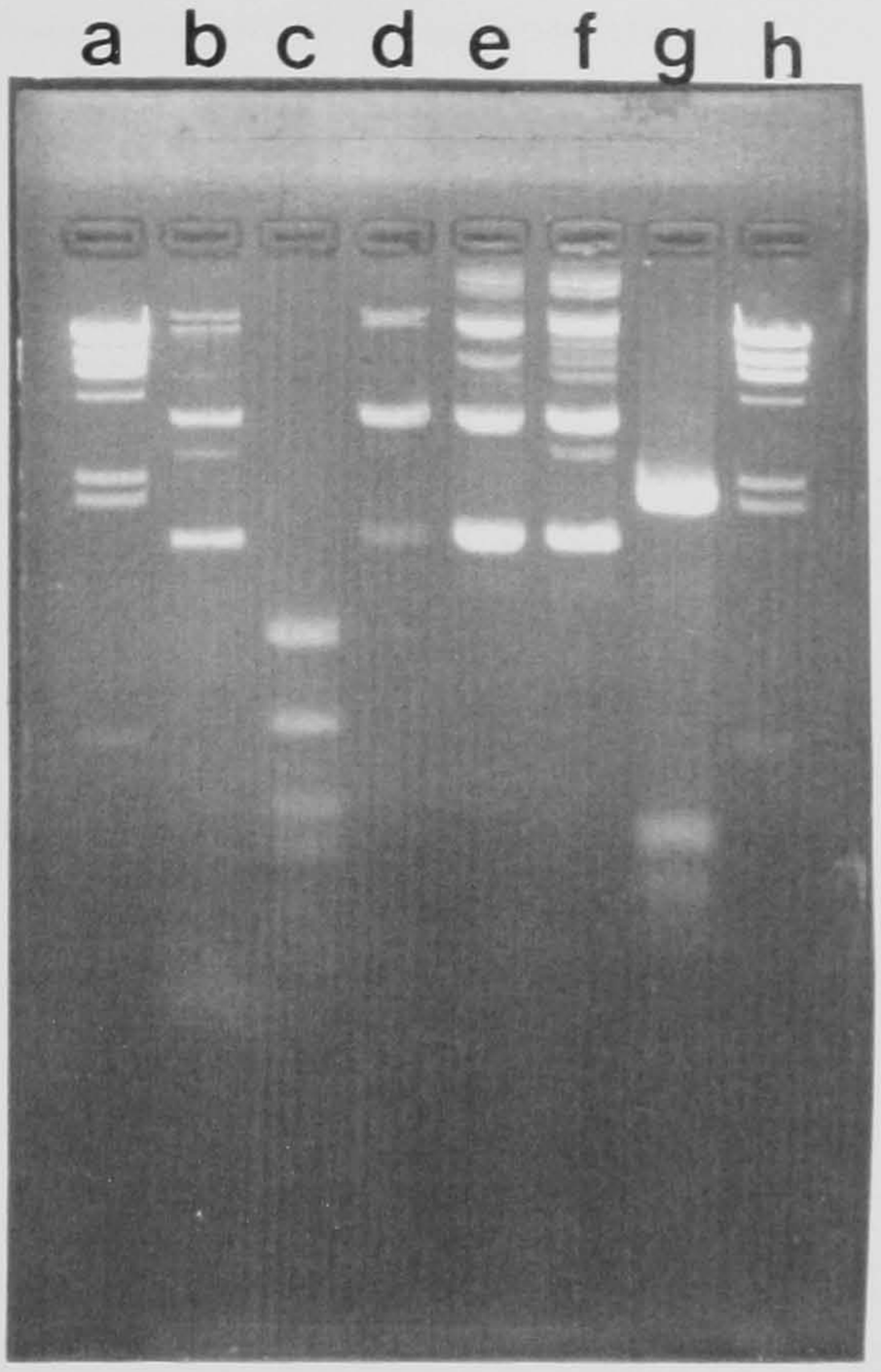


Figure 5.2 Comparison of Restriction patterns of DNA isolated from CB51 and GM31.
pUC18 DNA isolated from the *dam*⁻ host CB51 and *dcm*⁻ host GM31 verified with two restriction enzymes *Eco*RII and *Mbo*I.

a) λ*Hin*dIII marker
b) pUC18 in CB51 uncut
c) pUC18 in CB51 digested with *Eco*RII
d) pUC18 in CB51 digested with *Mbo*I
e) pUC18 in GM31 uncut
f) pUC18 in GM31 digested with *Eco*RII
g) pUC18 in GM31 digested with *Mbo*I
h) λ*Hin*dIII marker

a)

Host	DNA	Transformants.μg^{-1} DNA ($X10^2$)
E. coli CB51	pIJ699	< 1
	pXED3-13	< 1
E. coli GM31	pIJ699	470
	pXED3-13	7
S. rimosus MV9	pIJ486	1570

b)

Host	DNA	Transformants.μg^{-1} DNA ($X10^2$)
E. coli CB51	pIJ699	< 1
	pXED3-13	< 1
E. coli GM31	pIJ699	2200
	pXED3-13	320
S. rimosus MV9	pIJ486	10900

Table 5.4 Transformation efficiencies of high- and low- copy number DNA's from *dam*$^-$ host CB51 and *dcm*$^-$ host GM31 by

a) transformation of protoplasts and
b) transformation of electrocompetent cells.

reversibly reorganised, with some regions becoming thinner and pores being formed. The diameter of the pores is large enough to allow the passive transfer of macromolecules such as DNA. Other factors which lead to higher efficiencies of electrotransformation are the composition of the medium, the physiological state and concentration of the cells and the structure of the cell envelope. Pre-treatment with an enzyme such as lysozyme, can also increase the transformation efficiency through chemical breakdown of the cell wall. The damage created in the cytoplasmic membrane results in possible death of the cell. A delicate balance between minimal cell death and high yields of transformants is achieved by optimising the conditions for each organism (Joset and Guespin-Michel; 1993). J. Pigac optimised the conditions for *S. rimosus* R6, as previously described in Materials and Methods. Transformants grew up in 3 days rather than 7-10 days, as seen with protoplast transformations.

Plasmids pIJ699 and pXED3-13 prepared from the *dcm⁻ E. coli* host, GM31, along with the pIJ486 DNA from *S. rimosus* strain MV9, were employed to test the electroporation technology. Twenty-five ng of DNA from all three samples were mixed with 50µl of electrocompetent M4018 cells and electrotransformed under the conditions described previously. Table 5.4b shows that the overall transformation efficiency was approximately 1 order of magnitude higher for electrotransformed cells than transformation using protoplasts (Table 5.4a). This technique has the double advantage of 10-fold higher efficiency of transformation, and ability to pick transformants sooner, in 25% the incubation time of protoplast regeneration. This technique was used routinely in further *Streptomyces* transformations described in Chapter 6.

5.4.3 Conjugation

The ability of *E. coli* to conjugate with *Streptomyces* resulting in the unidirectional DNA transfer, was demonstrated by Mazodier *et al.,* (1989). The donor, *E. coli* S17-1 with the integrated RP4 *tra* functions was used to mobilise the pPM803 plasmid, with the *oriT,* into *S lividans* TK24. Exconjugates were selected using neomycin and nalidixic acid. The neomycin resistance gene was carried by the pPM803 plasmid and to counterselect the *E. coli* donor after transfer, nalidixic acid was used because many of the *Streptomyces* species have a natural resistance. The 1,200 exconjugates generated were resistant to neomycin and nalidixic acid. The same experiment was performed in this laboratory, with similar results. The experiment was done in duplicate and the presence of donor

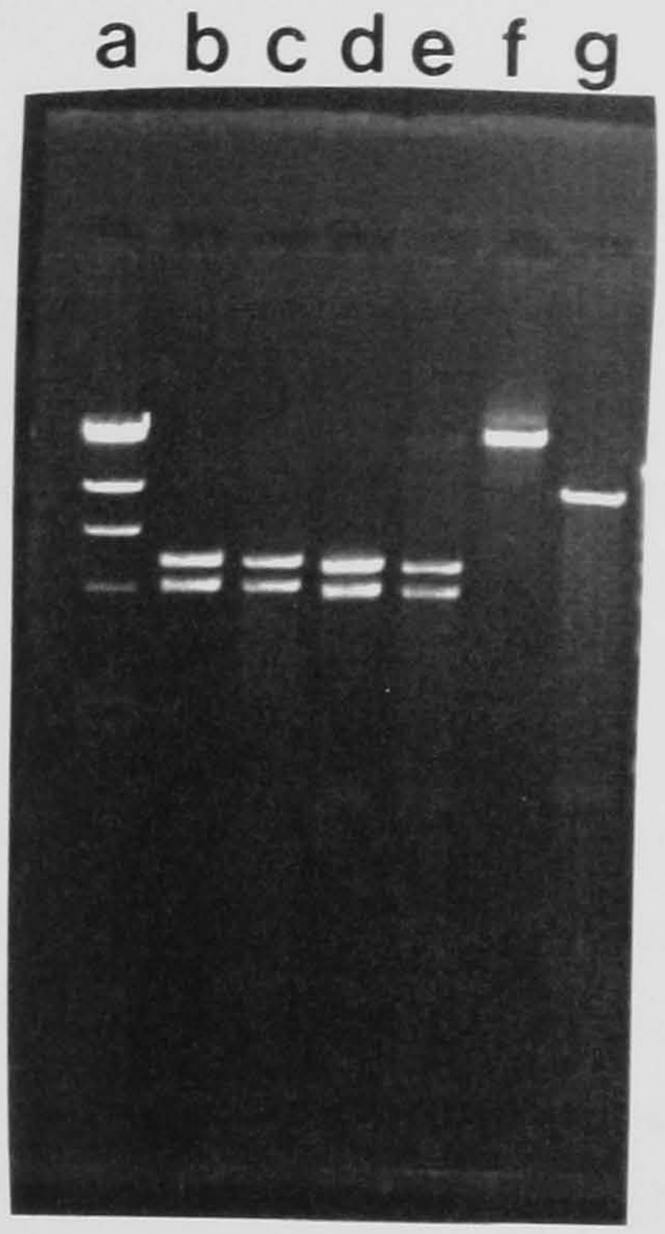


Figure 5.3
Restriction digest analysis of pPM803 DNA isolated from *E. coli* S17-1 and *S. lividans* TK24.

a) λ*Hin*dIII marker
b) pPM803/S17-1 digested with *Pst*I
c) pPM803/TK24 digested with *Pst*I
d) pPM803/S17-1 digested with *Hin*dIII
e) pPM803/TK24 digested with *Hin*dIII
f) pPM803/S17-1 digested with *Bcl*I
g) pPM803/TK24 digested with *Bcl*I

plasmid in the recipient exconjugants was verified by DNA analysis. Restriction patterns of the pPM803/S17-1 and pPM803/TK24 were identical for most enzymes. For*Bcl* I, which is sensitive to *dam* methylation, pPM803/S17-1 did not cut, whereas pPM803/TK24 did (figure 5.3). This proves that the plasmid had been transferred from the methylated *E. coli* host to the unmethylated *S. lividans* host.

Attempts to transfer DNA from *E. coli* to *S. rimosus* by conjugation initially employed many of the procedures of Mazodier *et. al.,* (1989). The pPM803 plasmid was be used as the mobilising plasmid because it carries the *oriT* functions and is functional in both *E. coli* and *Streptomyces*.. A tetracycline resistance marker was also incorporated into the system to counteract the small amount of tetracycline which could be produced by the *S. rimosus* host. This was provided by the SC101 plasmid (Cohen and Chang 1973; 1977). The RP4 functions were supplied by the *E. coli* host strain S17-1. As a control pPM803/S17-1 was also prepared for conjugational transfer into TK24. Spores of the *Streptomyces* hosts 4018 and TK24 were first induced to germinate by heat shock (see Materials and Methods, section 2.2.3) and then conjugated with the *E. coli* strain. The S17-1/TK24 plates were incubated overnight at 30^0C, the S17-1/4018 plates incubated for 6 hours at 30^0C. The plates were washed in LB and gently scraped to remove the *E. coli* layer with the *Streptomyces* layer still attached to the plate. Preliminary experiments showed that the optimal incubation time before removing the *E. coli* from the *S. rimosus* plates was in the range of 6-8 hours. If plates were left to incubate overnight (as is the case with *S. lividans*), flooding the plate to remove the *E. coli* resulted in the entire culture of *S. rimosus* detaching from the plate. A *S. rimosus* culture incubated for less time was still seen attached to the plates after the removal of *E. coli*.. Whereas 850 exconjugants per plate were obtained for *S. lividans*, none were obtained for *S. rimosus*.

A second procedure to conjugate from *E. coli* to *Streptomyces* was attempted which involved plating the donor strain with mycelial fragments of the recipient strain (as described in Materials and Methods section 2.2.9.2, Bierman *et al.,* 1992). The results showed that exconjugants were obtained from the *E. coli*/*S. lividans* matings but no exconjugants from matings between *E. coli* and *S. rimosus*. A control plate of *S. rimosus* with no drug added showed that *S. rimosus* was able to grow through the soft nutrient agar. Therefore, with two separate strategies (spore and mycelial recipients, the *E. coli* host was capable of conjugating with *S. lividans* but not with *S. rimosus*. This may however, be

explained by the restrictive nature of M4018 towards hemimethylated DNA. In the process of conjugal transfer, a single strand of DNA is transferred from donor to recipient. In both procedures mentioned above, the *E. coli* donor carries modified DNA and when it enters the *S. rimosus* host, the host replicates a second unmethylated strand. The resulting hemimethylated DNA is then recognised by the host and subsequently restricted by some *Streptomyces* hosts.

A third conjugational approach taken was to employ a donor which produced methylation-deficient DNA to attempt to circumvent the restriction system of *S. rimosus*. The *E. coli* host S17-1 was not a suitable donor because it produced methylated DNA. Therefore, a different source of transfer functions needed to be supplied. These need not be integrated into the chromosome of the *E. coli* host as long as the plasmid carrying the *tra* genes did not also carry a *Streptomyces* origin of replication. A number of *E. coli* strains carrying *tra* functions were kindly supplied by D. Sherratt, Genetics Department, Glasgow University. pDS620, a broad host range plasmid based on RP4, seemed most suitable. However resistance markers were incompatible with the pPM803 plasmid which carried the *oriT* functions. pPM803 carries only the aminoglycoside phoshotransferase gene conferring resistance to neomycin. The pDS620 plasmid carries the kanamycin resistance marker, which also confers resistance to neomycin, and therefore the selection of the two plasmids in the same host was not possible. The second choice was a IncP plasmid with a resistance marker requiring trimethoprim. Trimethoprim depresses the growth of Thy^+ cells, in the presence of high amount of thymine or thymidine.

A series of *E. coli* transformations was carried out to produce a suitable donor. The fully methylation deficient *E.coli* host ET12567 (*dam*$^-$, *dcm*$^-$), was the only methylation deficient host available at the time of the initial experiments. ET12567 was therefore the basis of the donor strain. The *tra* functions were supplied by the plasmid DS649, kindly provided to us by D. Sherratt. DS649 carries the broad host range IncP plasmid encoding the *tra* and *mob* functions. This plasmid also carries a number of resistance markers to facilitate selection. DS649 was transformed into ET12567 and selected by trimethoprim on special medium, isosensitest, containing a high concentration of thymine. A tetracycline-resistance marker was supplied by transforming the SC101 plasmid (Cohen and Chang; 1973, 1977) into the ET12567 host which also carried the

DS649 plasmid. The final requirement was to transform the pPM803 plasmid into the SC101/DS649/ET12567 host. The resulting *E. coli* donor, was an extremely slow growing strain, with doubling times in excess of 35 minutes.

Conjugation experiments were set up in parallel. As a control, the pPM803 plasmid was conjugated to *S. lividans* TK24 and to an *E. coli* recipient strain, DS801. The recently-constructed p501 strain was mated with *S. rimosus* 4018, *S. lividans* TK24 (using mycelial fragments of *Streptomyces*) and with the *E. coli* recipient strain DS801 (See Materials and Methods, section 2.2.9, for *E. coli* conjugation procedure). No exconjugants were found on either of the *Streptomyces* recipient plates employing the newly-constructed donor (Table 5.5). When the S17-1 donor was used as the donor, the number of exconjugants from DS801 was greater than 6X higher than when the newly constructed strain was the donor. Therefore, if the *E. coli*/*E. coli* conjugation, with the newly constructed strain was that poor, it was not unexpected that the *Streptomyces* conjugations showed no result.

A second *E. coli* host, GM31, was used for the conjugation experiments. However, this host was no more efficient as a donor, than ET12567 (Table 5.5). Although the reason was not readily apparent, a closer look at the compatibility of the *tra* and *oriT* functions suggested that this may be the problem. All previous studies of successful conjugations with *E. coli* and *Streptomyces* has employed the S17-1 host donor with the integrated RP4 *oriT* functions. This was not an option in this case because of the restrictive properties of the *S. rimosus* host.

A third approach was taken to try to employ the RP4 plasmid containing the *oriT* in the strain DS620 (D. Sherratt). For this approach, a second resistance marker had to be engineered into the pPM803 plasmid so it could be selected independently from the RP4 plasmid. There is a unique *Kpn*I site in the stability region of the pIJ101 replicion of the pPM803 plasmid. The gentamicin-resistance gene was removed from the pGM160 plasmid (Muth *et al.*, 1989) and cloned into pGEM7. The resulting clone, pGEMgmr, produced a 1.7Kb *Kpn*I cassette which could be inserted into the unique site of pPM803. The resulting clone pSG803 proved to be structurally unstable and therefore this approach was abandoned.

E.coli conjugations

Donor	Host Strain	Recipient	Conjugation Efficiency exconj./ml	Standard Deviation
pPM803	S17	DS801	$1.6X10^8$	6.7
p501	ET12567	DS801	$4.0X10^2$	14.1
p501	GM31	DS801	$6.3X10^2$	9.3

Streptomyces conjugations

Donor	Host Strain	Recipient	Conjugation Efficiency exconj./ml
p501	ET12567	TK24	$<1X10^2$
p501	GM31	TK24	$<1X10^2$
p501	ET12567	M4018	$<1X10^2$
p501	GM31	M4018	$<1X10^2$

Table 5.5
E. coli conjugations (top panel) using pPM803 and p501 from different host strains conjugated to the recipient strain DS801.
Streptomyces conjugations (bottom panel) using the same donor plasmid, p501 from different host strains conjugated to either TK24 or M4018.

5.5 Discussion

The advent of recombinant DNA technology has utilised *E.coli* hosts as the primary tool for the manipulation of DNA. *E. coli* is routinely the host of choice because of rapid growth rates and simple DNA isolation techniques. Unfortunately it is not possible to transfer DNA isolated in *E. coli* into most *Streptomyces* species because of the restrictive nature of the recipients. Methyl-specific restriction has been found in a number of *Streptomyces* species (Table 5.6).

Three other bacteria, *E. coli* (Raleigh *et al.*, 1989) *Acholeplasma laidlawii* (Sladek *et al.*, 1986) and *Bacillus thuringiensis* (Macaluso *et al.*, 1991) are known to restrict DNA containing 5-methylcytosine with little or no sequence specificity. In *E. coli*, two distinct genetic loci have been identified which encode genes controlling the restriction of DNA containing 5-methylcytosine, *mcrBC* and *mcrA*. In *A. laidlawii*, DNA modified with any of five 5-methylcytosine-specific methylases was restricted, but no sequence specificity was found (Sladek *et al.*, 1986). *Bacillus thuringiensis* is restrictive to DNA containing either N^6meA or 5-meC (Macaluso *et al.*, 1991). This is very similar to the *Streptomyces* spp., *S. avermetilis* (MacNeil *et al.*, 1988).

For direct introduction of shuttle vectors from *E. coli* hosts into *Streptomyces*, the DNA must be isolated from a non-methylating host. With *B. thuringiensis* poor transformation efficiencies lead to the investigation into the importance of DNA modifications. Results confirmed that DNA isolated from a *dam*$^-$ *dcm*$^-$ *E. coli* host gave a 500 to 1,000-fold increase in transformation efficiency over fully methylated DNA. The source of the plasmid DNA was especially important for larger plasmids as the larger plasmids might have more sites subject to modification and therefore more susceptible to restriction by *B. thuringiensis*. In *S. avermetilis*, plasmids with only 1-3 methylated bases per molecule were restricted only about 10-fold, however those plasmids with more than 10 sites were restricted greater than 1000-fold (MacNeil *et al.*, 1988). When modified DNA was introduced into *S. avermetilis* no transformants were obtained. This indicates that the decrease in transformation efficiency is greater than 10^6. Similarly, when fully modified DNA was introduced into *S. rimosus* 4018 no transformants were obtained (Table 5.1). This indicates a drop of transformation efficiency of greater than 10^5. The restriction barrier of *S. rimosus* was circumvented by a number of hosts. DNA isolated from a relatively permissive *Streptomyces* host like *S. lividans* or an *E. coli* host which lacks the *dcm*

Streptomyces sp.	*dam* (GmATC)	*dcm* ($CC^{A}/_{T}GG$)
S. avermitilis	++++	+++
S. bikiniensis		
S. coelicolor	++	
S. fervens		
S. lividans	+	
S. parvulus		+++
S. rimosus		+++

Table 5.6
Commonly used *Streptomyces* species and their methyl-specific restrictive nature. The number of (+) indicate the relative restrictive nature. The list was compiled from data taken from MacNeil *et al.* (1988).

methylase gene gave very similar efficiencies of transformation to that of DNA isolated from *S. rimosus*. The *E. coli* hosts capable of circumventing the restriction barrier of *S. rimosus* all carried a mutation in the *dcm* methylase gene.

E. coli has two methylase systems which do not carry an accompanying restriction endonuclease, and therefore do not belong, *sensu stricto*, to the restriction-modification processes, these are the DNA adenine-methylation (dam) and the DNA cytosine-methylation (dcm) systems. Mutations in these two genes have allowed the study of the role of these enzymes in the biology of *E. coli*. Dam methylation has been shown to play a role in the repair of DNA mismatches (Modrich; 1989). *Dam* methylates the adenine in the N^6 position of the sequence 'GATC'. This methylation occurs shortly after the DNA strand is synthesised. As a result, the DNA close to the replication fork remains hemi-methylated for a short time. This allows the mismatch repair system to discriminate between the template and daughter strand. The dam methylation proceeds shortly after the mismatch repair system. In strains with a mutation in the *dam* gene, both strands are unmethylated and the repair system can not distinguish between the nascent and daughter strands. The repair system still functions, however repair occurs on either strand and as a result *dam* strains have an increased spontaneous mutation rate (Marinus *et al.*, 1975). Other functions of the *dam* gene involve the regulation and co-ordination of several other cellular processes, along with providing a mechanism for coupling gene expression to the cell cycle.

The biological functions of the *dcm* gene are slightly more elusive. Originally, the only phenotype observed in *dcm* mutants was the susceptibility of IncN plasmids to restriction by *Eco*RII (Hattman; 1977). There is evidence that *Eco*RII and dcm methylases may have been common ancestors as there is good sequence similarity (Lauster *et al.*, 1989) and they are functionally very similar. The *dcm* gene methylates the inner cytosine residue of the sequence $CC^{A}/_{T}GG$. 5-methyl cytosine is more unstable than cytosine and the deamination results in the formation of thymine/guanine mismatches. If not repaired correctly before replication, C-to-T transitions occur. The repair system which recognises and repairs T:G mismatches has been termed very short patch (vsp) repair (Lieb; 1987). The *dcm* gene was originally thought to be involved directly along with the *mutL* and *mutS* gene products (Lieb *et al.*, 1988). Recent studies of the cloned *dcm* gene and surrounding region have identified that the gene product of *vsr* is responsible for the Vsp repair and not the overlapping *dcm* gene (Sohail *et al.*, 1990.) It therefore remains unclear what the function of the *dcm* gene is. It is

possible that recombination events between the IncN plasmid and the chromosome may have occurred and the *dcm* gene has no real purpose at all. However, there is evidence that a dcm recognition sequence within a promoter region may be under the regulation of *dcm* methylation, as in the case of the *lexA* gene (Miki *et al.*, 1981). Ringquist and Smith (1992) also showed that some dcm sites remain unmethylated in *dcm*$^+$ strains suggesting that these sites are protected by DNA binding molecules or changes in DNA topology.

DNA prepared from the *E. coli* host GM31, which is deficient in cytosine methylation, is able to overcome the restriction barrier of *S. rimosus* M4018. The status of the methylation at the dam sites has no influence on transformation efficiency of DNA into *S. rimosus*. This was outlined by the similar transformation efficiencies of both the ET12567 (*dam* $^-$, *dcm* $^-$) strain and the GM31 (*dcm* $^-$) strain (Tables 5.1 and 5.4a). The evolutionary differences in methylation deficient DNA requirements for various *Streptomyces* hosts may be the result of their natural competitive soil envirnoment and that the expression of varied and multiple restriction systems might give an advantage in survival This is exemplified by the *S. fradiae* strain. This strain contains at least five restriction systems that protect it from bacteriophage infections (Matsushima *et al.*, 1987). The total number of restriction systems has not been explored with *S. rimosus*. This could easily be tested by transforming DNA into *S. rimosus* which had been modified *in vitro* with various methylases. This experiment was not performed because for our purposes this is not significant as the *dcm*$^-$ strain GM31 is capable of circumventing the restriction of *S. rimosus*.

Transformation of protoplasts using polyethyleneglycol (PEG-mediated) protoplasting has been the established way of introducing DNA into the some of the best-characterised *Streptomyces* strains (Hopwood *et al.*, 1985). Although this method of transformation is generally applicable, it is necessary to establish the optimal conditions of growth, of formation of protoplasts and for regeneration of protoplasts. The transformation of the fragile protoplasts is tedious, time-consuming and frequently not reproducible.

Transformation by electroporation has been rendered possible for several Gram-negative (i.e. *E. coli* , Dower *et al.*, 1988; *P aeruginosa*, Diver *et al.*, 1990) and Gram-positive bacteria (i.e. *B. lactofermtum* Bonnassie *et al.*, 1990; *Streptococcus*, Powell *et al.*, 1988). The first attempt to electroporate *S. lividans* was made by MacNeil (1987). However it was less efficient than PEG-mediated

transformation. The conditions for the *S. rimosus* strain R6 has been optimised by Pigac and Schrempf (1994). These conditions are readily applicable to *S. rimosus* 4018 except that a greater concentration of lysozyme is required for 4018 (100μg/ml for R6 vs. 400μg/ml for 4018). By comparison to the transformation efficiency of pIJ486 into *S. rimosus* 4018 in Table 5.4 (a&b), there is a ten-fold increase using the electrotransformation procedure over transformation of protoplasts. The consistent drop in transformation efficiency of low-copy number plasmids over high-copy number plasmids is intriguing as this is not the case with the *S. rimosus* R6 strain (J. Pigac; pers. comm.). This however, may well be just one of the properties which differentiates the two strains. For *S. rimosus* this new technology has proved to be considerably more rapid, reproducible and approximately 10 times more efficient than PEG-mediated transformation. Although conditions for all of the commonly used *Streptomyces* strains have not been set, it is a good advance in recombinant technology.

Conjugation systems have been, in general, used less than other methods of gene transfer in recombinant DNA studies. However, there are practical advantages to using conjugation systems, (i.e. to overcome restriction barriers in gene transfer and allow for genetic manipulation of microbial species which are considered to be intractable to genetic manipulation). Until recently, it had been assumed that conjugal transfer was restricted to closely-related species. Conjugation has been very successful in promoting diverse gene exchanges (i.e. *E. coli* to *Entercoccus, Bacillus, Listeria Streptococcus,* and *Staphlococcus*, Trieu-Cuot *et al.*, 1987). An important factor in this success is likely to be the replicative transfer of single-stranded DNA during conjugation. Single-stranded DNA is known to be refractory to most restriction systems, thus the transfer of single-stranded DNA during bacterial conjugation may escape digestion the restriction systems of the recipient. The conjugal transfer of DNA between *E. coli* and *S. lividans* was successful for a number of reasons. As Streptomycetes germinate they extend by hyphal growth. The newly-formed cell wall may be more competent to take up DNA. The restriction system of *S. lividans* is a relatively weak one and can accept methylated DNA with only a slight decrease in efficiency of transformation. Successful conjugation may also be attributed to this relatively weak restriction system. This is not the case with*S. rimosus*. An initial experiment to test the pPM803 plasmid in the S17-1 *E. coli* host (with a tetracycline resistance marker to protect against the production of both OTC and TC by the M4018 strain) did not result in any exconjugants. This may be explained by the relatively strong restriction system of *S. rimosus*. Initial hopes

that the replicative transfer of single-stranded DNA would escape restriction were dashed. C. Smith (1992), demonstrated that, although fully methylated single-stranded DNA could be transformed into *S. coelicolor*, it was at a much reduced effiency, a $2X10^5$ reduction compared with a double-stranded *S. coelicolor*-derived plasmid, pIJ101. Unmethylated single-stranded DNA, on the other hand, produced a transformation efficiency only 10^3 fold lower the double-stranded DNA control, pIJ101. When single-stranded DNA enters the cell it is replicated into double-stranded hemi-methylated DNA and, based on Smith's results, is then subject to restriction.

Plasmids pPM803, SC101 and pDS649 (p501) were transformed into ET12567 (*dam*$^-$, *dcm*$^-$) in an attempt to resolve the possible hemimethylation problem. The single-stranded DNA to be transferred would be unmethylated and as a result, the double-stranded DNA produced by host replication, would not be susceptible to restriction in *S. rimosus*. This strain was a very slow growing strain, with doubling time of ca. 35 mins. This combination of plasmids in the ET12567 *E. coli* host has proved not to be a successful combination for conjugation into *Streptomyces*. The host of choice has to be the highly efficient S17-1 strain with the integrated RP4 plasmid. Unfortunately the RP4 plasmid, available to the lab, carried similar antibiotic markers to the pPM803 plasmid which made selection of both plasmids in the same host inconvenient. An attempt to insert a selectable antibiotic marker in a unique *Kpn*I site in the pPM803 plasmid resulted in a structurally unstable vector (data not shown). Data from Mazodier *et al.*, (1989) and Bierman *et al.*, (1992) on different protocols for conjugal transfer between *E. coli* and *Streptomyces* have both employed the S17-1 host, with the integrated transfer functions, suggesting it was an important factor in the process. However this is may not be the case as recent publication by Matsushima *et al.*, (1994) successfully employed a RP4 (RK2) plasmid, carrying the transfer functions, to conjugate DNA from *E. coli* to *Saccharopolyspora spinosa*. Buchanan-Wollaston *et al.*, (1987), showed that the *mob* and *oriT* functions of the broad-host range IncQ plasmid RSF1010 mediate the transfer of plasmids from *Agrobacterium* spp. into plant cells. The broad-host range plasmid pRK2 (IncP) was unsuccessful however, in similar tests of transfer. Trieu-Cuot *et al.*, (1991) also demonstrated that the transfer of IncP plasmids and some shuttle vectors were acutely sensitive to restriction by some type II R-M systems. Conversely, effective transmission of F-related plasmids is relatively resistant to type I restriction (Boyer; 1971). To determine the specific broad-host range plasmids which allow the transfer of DNA to *S. rimosus* may be a question of

trial and error as it appears that not all broad-host range plasmids are compatible with the recipient's restriction systems. The use of heat or UV irradiation on the recipient *S. rimosus* may destroy the restriction system and may then promote conjugation. Heat was used successfully to increase the viability of conjugation of corynebacteria with *E.coli* (Kreps *et al.*, 1990).

Therefore, at this time, conjugation does not appear to be a viable option for the genetic transfer of DNA from *E. coli* to *S. rimosus* due mainly to the restriction barrier of the *Streptomyces* host. The improved transformation efficiencies of DNA isolated directly from an *E. coli* host, which carries a mutation in the *dcm* gene, are a result of overcoming the methyl-specific restriction barrier of the *S. rimosus* host. This was further advanced with electrotransformation, a new technique in the *Streptomyces* field. This will prove to be a valuable asset in further genetic characterisation of the *S. rimosus* chromosome.

Chapter 6

Disruption of a Putative Ketoreductase Gene in the *otc*Cluster

6.1 Introduction

DNA sequencing of the *otc* cluster has revealed a number of open reading frames (ORFs). By comparing the deduced peptide sequences of these ORFs with a database (e.g. Genbank), functions of some genes can be assigned based on a substantial degree of similarity with other gene products of known function. Whether these gene products are catalytically active however, has not been addressed formally. Therefore, most of the ORFs, of known function, in the *otc* cluster have been assigned in this way. The most notable exception is the gene encoding ATC oxygenase, which through enzyme assays, was assigned to the *otcC* gene by *Binnie et al.,* (1989), using 'reverse genetics'. They first purified the ATC oxygenase protein from which N-terminal amino acid sequence was obtained. Oligonucleotide probes were constructed based on the protein sequence and used to detect the corresponding fragment of chromosomal DNA from *S. rimosus,* which was then cloned. Hence, the gene function assigned is known to be active, as it has been transcribed into protein. Another way to deduce the function of various genes in the *otc* pathway, using a molecular biological approach, is through gene disruption and replacement. A schematic representation of gene disruption is outlined in figure 6.1. The gene to be targeted is initially disrupted by inserting a selectable antibiotic resistance marker within the sequence of the gene. The disrupted gene can be introduced into a wild type cell through one (or two) 'Campbell-type' homologous recombination events (Chater and Bruton; 1983). Replacement of the wild-type sequence with the disrupted gene and selectable marker requires two crossover events. A single crossover event results in the incorporation of the the disrupted gene, selectable marker and shuttle vector sequence into the streptomycete chromosome. Results of both single and double crossover events (figure 6.6) disturb all downstream transcription. A second replacement event (a double crossover), is required to introduce a copy of the target gene carrying an in-frame mutation so that all downstream transcription is restored. At this point, the build up of metabolites produced by the mutant strain can be analysed chemically and the role of the target gene deduced.

There are many approaches to targeted gene disruption. MacNeil *et al.,* (1989) constructed a non-replicating suicide vector which carries the origin of replication in *E. coli* but not in *Streptomyces*. The aim is to force the plasmid to

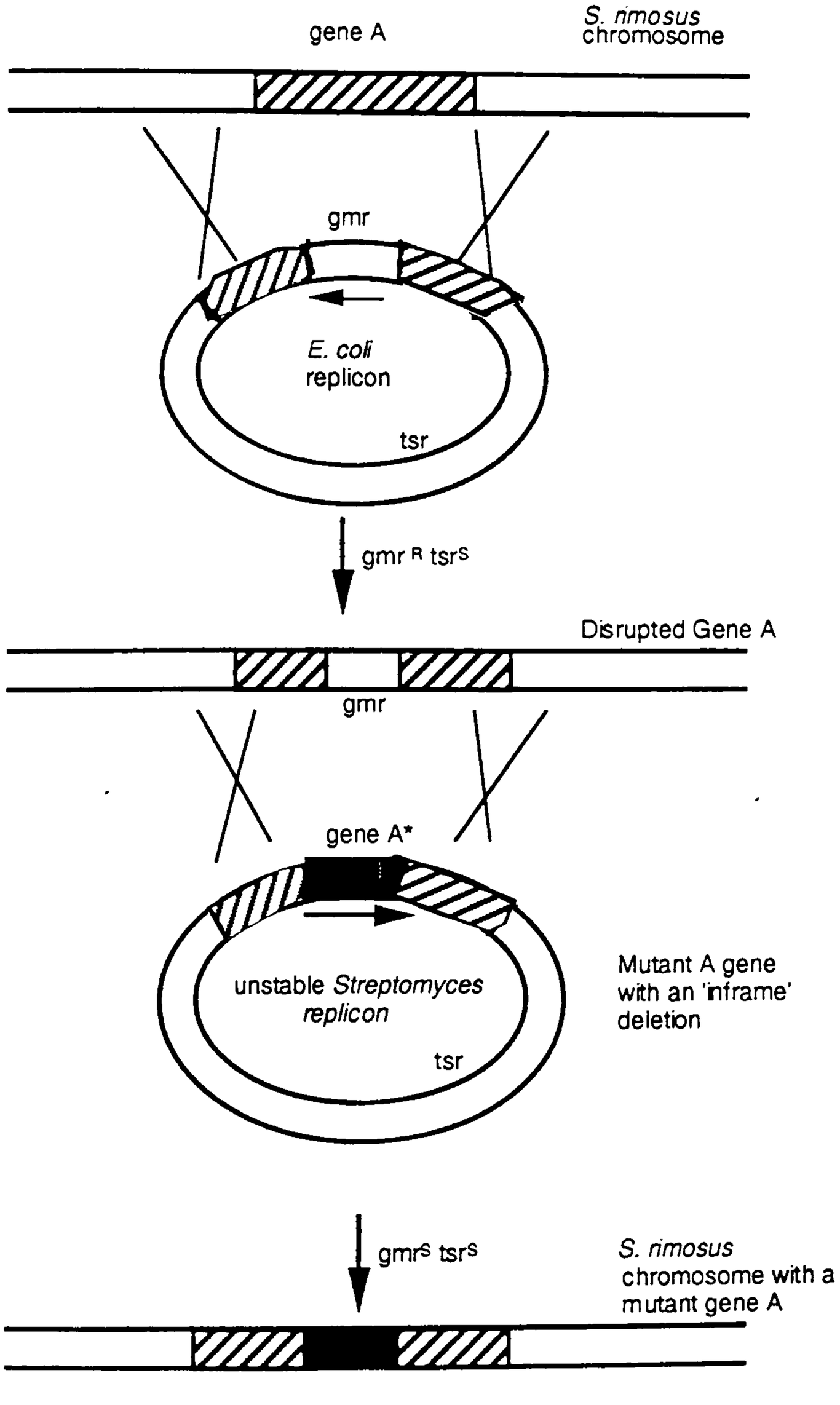


Figure 6.1
A schematic representation of the strategy for integration into *S. rimosus* chromosome employing the gentamicin resistance (*gmr*) gene.

integrate into the streptomycete chromosome in order to survive. Homologous regions flanking the disrupted target gene are first cloned out of the cluster into the plasmid. It is the homologous flanks of wild-type DNA that allow the 'Campbell-type' recombination event to occur. Suicide vectors have had great success with numerous *Streptomyces* strains (Khosla *et al.*, 1992; Richardson *et al.*, 1990; MacNeil *et al.*, 1989). Variations on this theme include the use of integration vectors in the single stranded form. Hillemann *et al.*, (1991) found that although there was very little difference in transformation frequency between double-stranded and single-stranded DNA, when a plasmid capable of replicating in *Streptomyces* was used, the integrative transformation was 10-100 times more efficient when the vector (carrying no *Streptomyces* origin of replication) was in the single-stranded DNA form. It has been observed in various bacteria, that integration *via* homologous recombination is greatly facilitated by the presence of single-stranded DNA (Lewin; 1994).

A second commonly-used approach to gene disruption is by introducing a replicating plasmid into the host chromosome. This plasmid will carry the *Streptomyces* origin of replication, which, when introduced into the host, confers the ability to replicate. After recombination events between the plasmid and chromosome occur, the plasmid must be removed from the cell before DNA analysis, since non-integrated plasmids contain the intact mutant gene. Various methods for segregating the replicating plasmid from the cell have been cited. Anzai *et al.*, (1988) cured *S. hygroscopicus* cells of the plasmid pIJ702 by regeneration of protoplasts in the absence of drug. Weber *et al.*, (1988) employed the plasmid pIJ702 and found that, although it was maintained stably in *S. lividans* and *S. coelicolor*, it was unstable in *Sac. erythraea*, being maintained poorly in mycelia and excluded almost totally from the spores. A third example of plasmid displacement employed the temperature-sensitive replicating plasmid pSG5 (Muth *et al.*, 1989). The replicon is inherited stably at temperatures below 34^{0}C, but is lost at incubation temperatures above this. Blanco *et al.*, (1992) were able to integrate exogenous DNA into the chromosome of *S. halstedii* using the temperature-sensitive replicon.

It is clear that gene-disruption experiments have been carried out on many *Streptomyces* species and each is unique to a specific strain. A number of strategies have been considered with respect to *S. rimosus*. A low copy-number vector was not feasible because of the inexplicable drop in transformation efficiency with low-copy number vectors compared with high-copy number

vectors (See Chapter 5). Two genes in the *otcY* locus were disrupted successfully in the *S. lividans* TK24 host which carried the entire *otc* cluster (plasmid pGLW101). However, attempts using a high-copy number, *E. coli* replicating vector pIBI24, as a suicide integration vector in *S. rimosus* were unsuccessful mainly because the recombination efficiency was not high enough to produce any transformants (Thamchaipenet; 1994). A reasonable frequency of double-crossover recombinants produced in the *S. lividans* host was obtained, all of which were unable to produce OTC. Despite the convenience and ease of integration in *S. lividans*, this species produces other antibiotics such as actinorhodin, and thus must express similar gene products to those disrupted in the *otc* cluster, namely the keto-reductase and cyclase genes. Interpretations of these integration experiments were difficult as any intermediates and/or shunt products produced from these disruptants could be the result of 'crossfeeding' between the *otc* and actinorhodin pathways. Temperature-sensitive plasmids were also not appropriate because when *S. rimosus* is subjected to temperatures above 34^0C, the organism does not produce oxytetracycline (Hunter pers. comm.). Therefore the data from disruptants made in this way would be difficult to interpret. The more plausible strategies involve the use of replicating plasmids and subsequent plasmid loss after integration. These techniques were looked at more closely and are the subject of this chapter.

The stability of the *Streptomyces* pIJ101-based plasmids was investigated based on the strategy of the gene disruption experiments carried out by Weber *et al.*, (1988). pIJ702, in particular, appears to show signs of instability when inherited in *Sac. erythaea,* but not in other strains. The ability of pIJ702 and two other pIJ101-based plasmids to be maintained stably in *S. rimosus* was explored as a possible basis for constructing an integration vector for the disruption of *OtcD*-ORF3, a putative 'reductase' gene, in the *otc* cluster. The sequence and deduced function of *OtcD*-ORF3 were described in Chapter 3.

pIJ101 is the parent of many of the vectors used in *Streptomyces*. Its derivatives range in copy number from 40-300 and the nucleotide sequence has been determined. Currently more is known about the biology of pIJ101 than any other *Streptomyces* plasmid (Kendall *et al.*, 1988). Replication of pIJ101 is thought to occur by a rolling-circle mechanism *via* a single-stranded DNA intermediate (Pigac *et al.*, 1988; Gruss *et al.*, 1989). The essential features of this mode of replication are (i) an origin of plus-strand synthesis, (ii) a replication protein (rep) which interacts with the plus origin to generate a nick which allows

displacement synthesis of the plus strand to occur, and (iii) a signal for efficient conversion of the single-stranded to the double-stranded form (Devine *et al.*, 1989). The sequence of the active site of the Rep protein of pIJ101 is similar to those found in several other plasmid-encoded Rep proteins which bind to pC194-like plus origins (Zaman *et al.*, 1993b).

Deng *et al.*, (1988) identified a non-coding region of pIJ101 DNA, called *sti*, which causes strong incompatibility when present in its natural orientation with respect to the basic replicon. A pair of plasmids can co-exist in the same host if they both possess *sti* in the correct orientation (Sti^+), or both possess *sti* in the reverse orientation (Sti^{rev}) or both lack *sti* (Sti^-). Sti^+ and Sti^- plasmids cannot co-exist in the same cell: if they occur together, the Sti^+ plasmid is retained while the Sti^- plasmid is eventually lost. Sti^- plasmids accumulate more ssDNA, the presumed intermediates in plasmid replication, than Sti^+ plasmids, implicating *sti* as the site where synthesis of the second strand is initiated. The primary site for second-strand initiation has been identified as a 0.53 kb *Spe*I-*Sst*II fragment in the pIJ101 plasmid (Zaman *et al.*, 1993). It is contained within a 1.2 kb *Bcl*I-*Bcl*I region of the pIJ101 plasmid which encodes *sti: cop/korB*. Deng *et al.*, (1988) suggested that Cop, a protein which is encoded by a gene adjacent to *sti* is a *trans*-acting negative regulator which decreases the copy number of Sti^+ plasmids by inhibiting the initiation of second-strand synthesis at *sti*. Mapping studies indicate that *cop* occupies the same region on pIJ101 as another characterised gene called *korB* (Kendall *et al.*, 1988). KorB represses the transcription of *kilB*, a gene necessary for normal intra-mycelial plasmid transfer (Kendall *et al.*, 1987; Stein *et al.*, 1989; Stein *et al.*, 1990; Zaman *et al.*, 1992).

The three plasmids in this study, pIJ702, pIJ486 and pIJ680 (figure 6.2), all contain the 2.0 kb *Bal*I-*Sst*II region of pIJ101 found to be essential for maintenance and replication in *Streptomyces* (Zaman *et al.*, 1993a). However, other important regions of the pIJ101 plasmid are missing (figure 6.2). These include the region encoding *sti: cop/korB.*, the *kilB* region and another region located between *sti: cop/korB.* and the essential region. Kieser *et al.*, (1982), have found that this 0.5 kb region is necessary for stable inheritance. Plasmids with deletions in this region are markedly unstable (ranging from 13%-55% loss after one round of growth on antibiotic-free media). A FRAME plot (described in Chapter 3), in this region does not indicate a defined ORF, although there are a number of direct repeats and two very strong stem-loop structures. Although not strictly defined, it is

structurally very similar to the *par* locus in the *E. coli* plasmid pSC101 (Meacock *et al.*, 1980; Miller *et al.*, 1983).

The *par* locus in the plasmid pSC101 has been identified and characterised genetically and physically. It does not encode a gene product or specify a function responsible for incompatibility but appears to be a *cis*-acting element that probably functions as the site that interacts with other cellular components during partitioning of plasmids. The precise mechanisms that govern partitioning of plasmids are not yet well understood. In *Bacillus subtilis,* the structure of the *par* locus that regulates plasmid partition is located very near the site for second-strand synthesis and has been shown to be orientation-specific (Chang *et al.*, 1987). Again, it is not encoded by a protein but has a series of inverted repeats and two sets of direct repeats. Direct sequence comparisons showed no significant degrees of similarity between the *B. subtilis* and pIJ101 sequences. However the series of direct and indirect repeats along with the loss of plasmid stability when the region is disrupted makes a strong case for similar roles. Further work by Devine *et al.*, (1989) showed that the accumulation of single-stranded DNA does not equate to segregational stability in *Bacillus subtilis.*

Segregational instability of plasmids in unicellular bacteria is usually monitored by plating samples on agar to reveal the proportion of colony-forming units that exhibit a plasmid-borne phenotype such as antibiotic resistance. A *Streptomyces* colony on a plate is not derived from a single cell, but from a clump in which only a small proportion of the mycelial filaments need to have a plasmid for a colony to be formed on selective agar. Therefore a simple plating method will not give a true estimate of the plasmid loss. *S. rimosus* grows in liquid as a very dispersed culture and therefore decreases the size of the microscopic clumps of mycelia to a level where truer estimates of plasmid loss can be analysed. A better method devised by Wrigley-Jones *et al.*, (1992), to monitor plasmid stability throughout growth is to measure plasmid copy number since segregational stability is often preceded by a decrease in plasmid copy number. This method was considered. However for our purposes, the liquid assay proved to be sufficient to deduce the percentage of cells which retained the plasmid.

6.2 Objectives

The objective of the work described in this chapter was to disrupt and replace a gene in the *otc* pathway without disturbing any downstream transcription.

The immediate aims of the experiments discussed in this chapter were as follows:

i) to characterise the stability of three *Streptomyces* vectors
ii) to develop a series of integration and replacement constructs
iii) to carry out gene disruption and replacement experiments on the *OtcD*-ORF3, a putative 'reductase' gene from *S. rimosus* 4018

6.3 Results

6.3.1 Choice of Vector

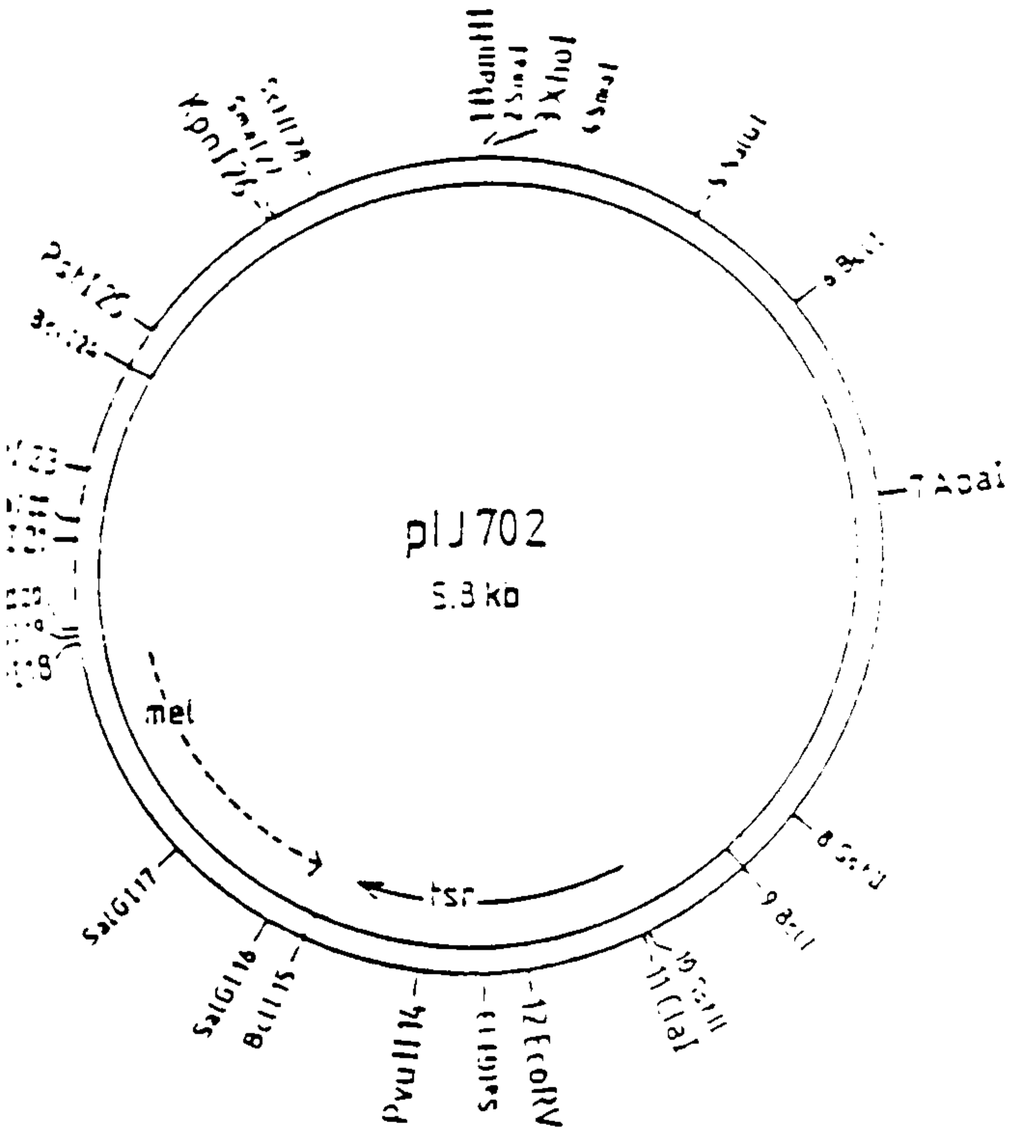
The stability of three *Streptomyces* plasmids was tested by introducing the plasmids into *S. rimosus* 4018 by PEG-mediated transformation, plating the resulting spores on non-selective media and then monitoring the loss of plasmid by colony counts on selective versus non-selective media. The unique characteristics of the three plasmids are outlined below. All three are non-conjugative derivatives of the natural pIJ101 replicon, and as such are high-copy number vectors (40-300 copies per cell - Kieser *et al.*, 1982). Each carries the *tsr* gene (from *S. azureus*), which confers resistance to thiostrepton by methylation of the rRNA (Thompson *et al.*, 1982), providing a selectable marker for propagation in *Streptomyces*.

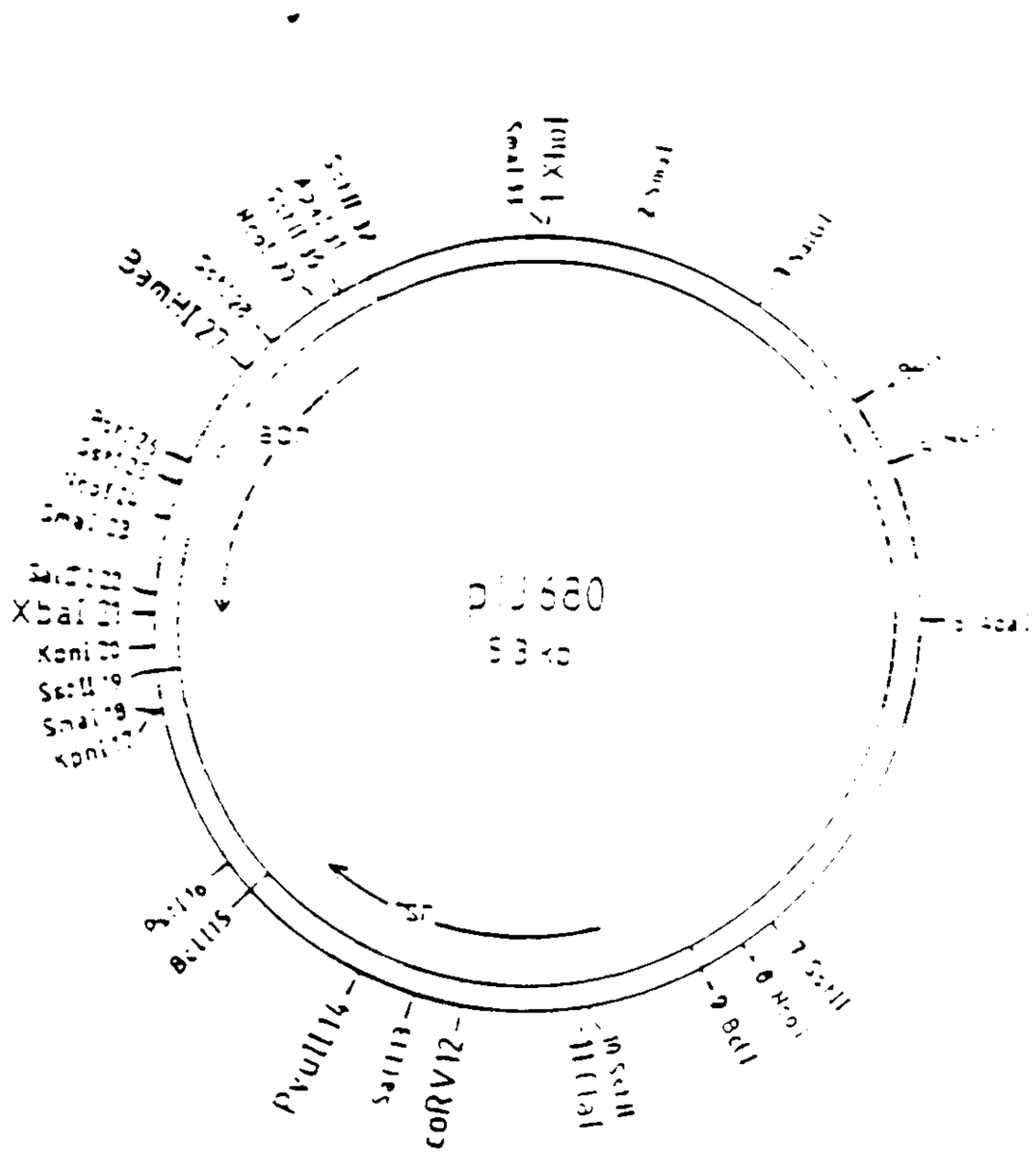
pIJ702

A 5.8 kb plasmid (Figure 6.2a), which carries 2.0 kb of pIJ101 found to be essential for the maintenance and replication (rep) in *S. lividans* . The rep region is fused with the thiostrepton resistance marker (*tsr*) and the *mel* operon (consisting of the two genes *mel* and *ORF438* from *S. antibioticus*), (Katz *et al.*, 1983).

pIJ680

A 5.3 kb plasmid (see Figure 6.2b) which, similar to pIJ702, carries the 2.0 kb essential region for maintenance and replication in *Streptomyces* along with the





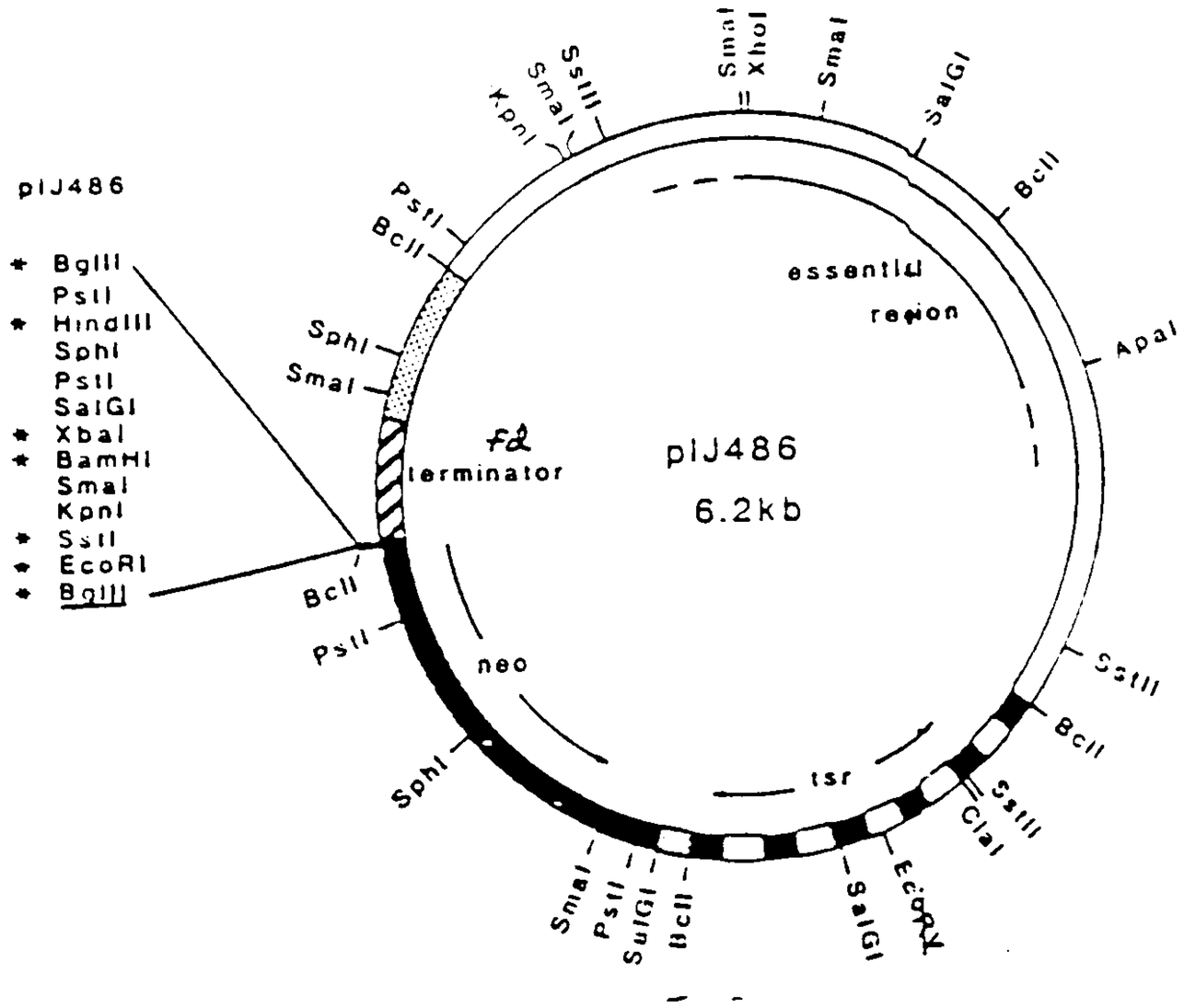


Figure 6.2
Restriction maps of three *Streptomyces* based plasmids employed for stability testing. References; pIJ702 and pIJ680 taken from Hopwood *et al.* 1985 and pIJ486 taken from Ward *et al.* 1986.

fused *tsr* gene. It also carries the neomycin phosphotransferase gene, *aph* I, from *S. fradiae* (Thompson *et al.*, 1982). The removal of the *Bam*HI site from the pIJ101 portion of the plasmid (T. Eckhardt, cited in Hopwood *et al.*, 1985), allowed for the unique *Bam*HI site in the *aph* gene to be utilised for insertional inactivation.

pIJ486

A 6.2 kb plasmid (see Figure 6.2c), again carries the essential region for maintenance and replication from the pIJ101 plasmid which is fused to the *tsr* gene. This plasmid uses the *aph*II gene of Tn5 (resistance to kanamycin or neomycin) as the indicator for promoter activity. A ribosome binding site and an in-frame stop codon lie between the *Bgl*II site (underlined in figure 6.2c) and the start codon for the *aph*II gene. The major terminator of phage fd prevents read-through from vector promoters into the *aph*II gene. pIJ486 differs from pIJ424 with the elimination of the *Bam*HI site from the pIJ101 region of the plasmid and the insertion of a polylinker in the *Bgl*II site.

All three plasmids (previously isolated from *S. lividans* TK24) were transferred into *S. rimosus* 4018 and recombinants selected by resistance to thiostrepton. A single colony was picked and streaked onto a Soya mannitol agar plate supplemented with 50μg.mL^{-1} thiostrepton to ensure a true transformant was selected. Spores were collected after three days and replated onto a Soya mannitol plate lacking thiostrepton. Figure 6.3 gives a schematic representation of the procedure followed for a single round of selection. This was carried out over several cycles for all three vectors. Spores collected from the Soya mannitol plate were divided into four aliquots. Aliquot A was replated onto Soya mannitol plate incubated for 3 days and used as the stock for the next round. 500μl of Aliquot B was used to inoculate 50 mL of TSB media. The liquid culture was incubated on an orbital shaker at 30^{0}C for 48 hrs. 500μl was then removed and plated onto a non-selective Soya mannitol plate. After 3 days at 30^{0}C, spores were collected and a series of serial dilutions made. In duplicate, 10^{-5}, 10^{-6} and 10^{-7} dilution cultures were plated on LB with and without 50μg.mL^{-1} thiostrepton. LB media was preferred because non-sporulating colonies could be seen easily after a 2-day incubation at 30^{0}C. Aliquot C was used to make a similar dilution series from the spores harvested directly from the Soya plate, as opposed to aliquot B whose spores came from a liquid culture. Aliquot D was stored frozen in 20% glycerol at -20^{0}C. The stability of the plasmids in both liquid and solid media is outlined in figure 6.4. Figure 6.4(a) shows the results for an agar based non-selection. Figure 6.4b shows the results

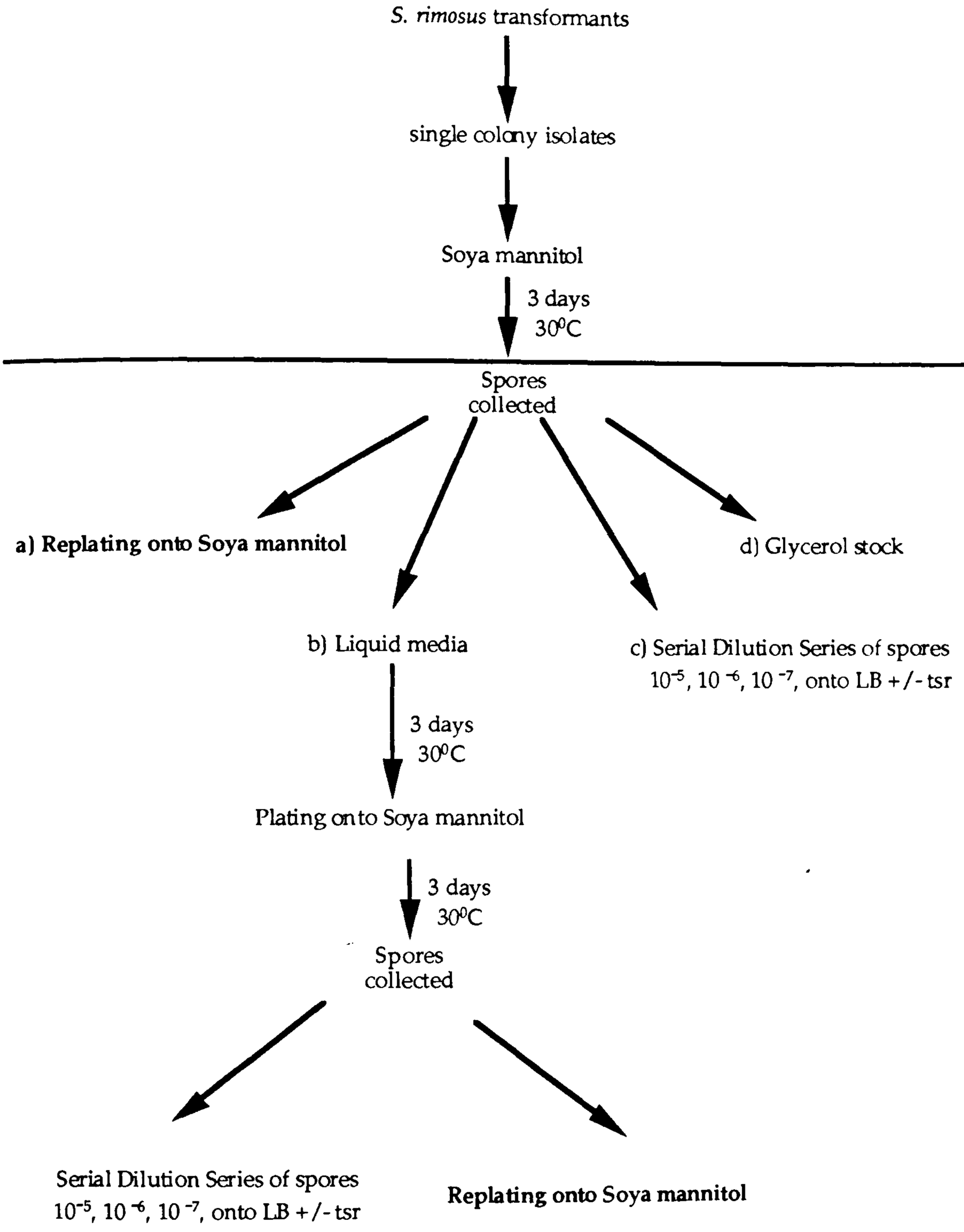
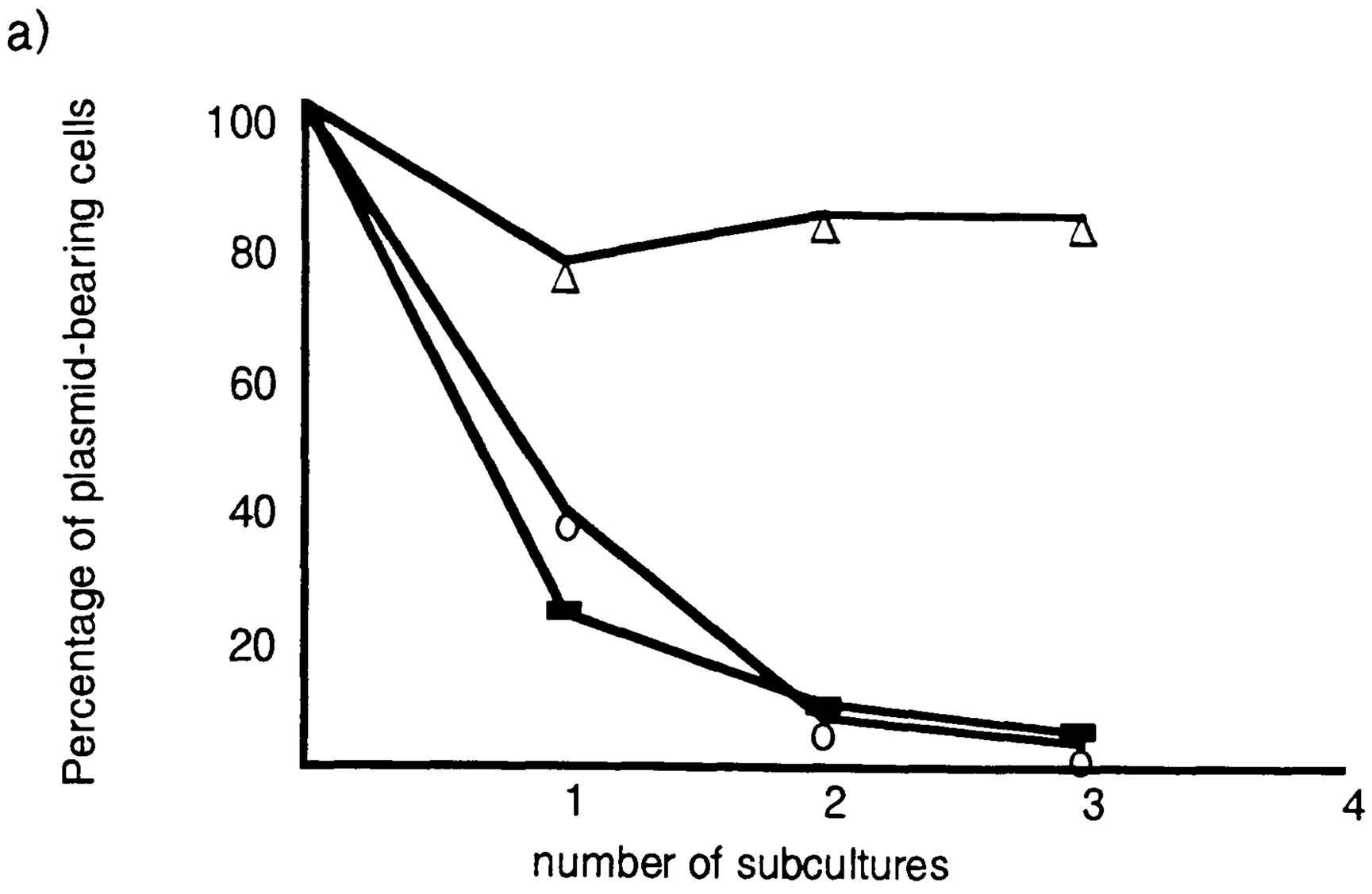


Figure 6.3
A schematic representation of a single cycle in both liquid and solid non-selection of 3 plasmids, pIJ702, pIJ486 and pIJ680 in *S. rimosus*.



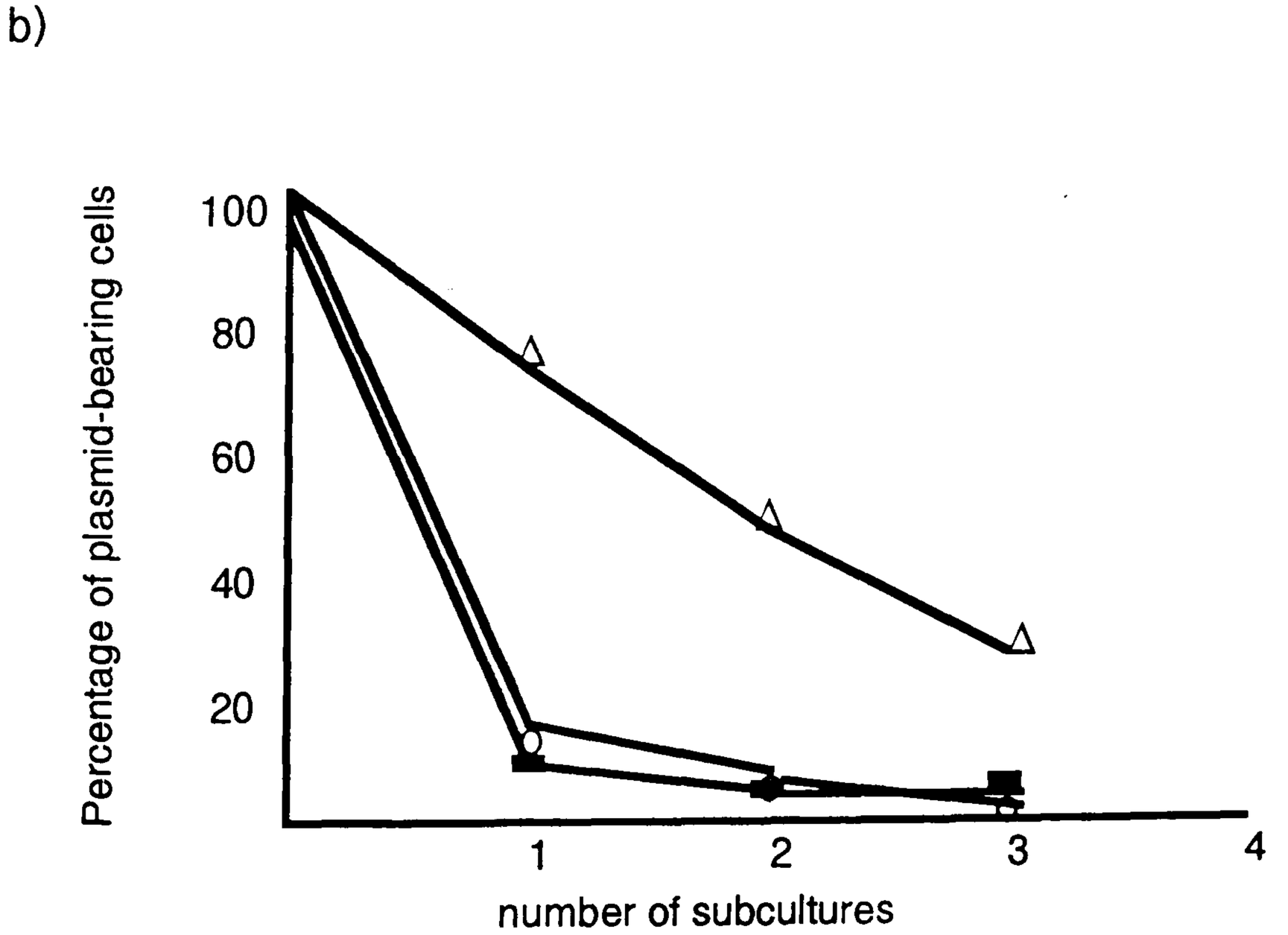


Figure 6.4
Plasmid loss in *S. rimosus* of pIJ702 (Δ), pIJ680 (▬) and pIJ486 (o) in the absence of thiostrepton. Plot (a) gives the results of agar plate selection, plot (b) gives the results of liquid selection. See text for further details. The plasmid-bearing cells were initially selected on thiostrepton.

of a liquid non-selection. Average colony counts were taken for duplicate plates. To determine the percentage of the spores retaining the plasmid, the average number of colonies (taking into account the dilution factor) on the LB+ tsr plates was divided by the average number of colonies on the LB plates and expressed as a percentage. The plasmid pIJ702 was quite stable in *S. rimosus* both in liquid and solid non-selection. Although there is a decline in the percentage of colonies retaining the plasmid, with both pIJ680 and pIJ486, the decline is significant and more rapid. There is very little difference between both pIJ680 and pIJ486. pIJ486 was selected as the unstable delivery vector for subsequent integration experiments primarily due to the greater choice of single restriction sites.

6.3.2 Construction of Integration Vectors

To deduce the function of various genes in the *otc* cluster, gene disruption experiments were set up. The gene to be disrupted, the *OtcD*-ORF3 gene, is a putative reductase (see Chapter 3). The ORF spans the *Kpn*I$_{26}$ site and this was chosen as the insertion site for the antibiotic-resistance marker (figure 6.5). The antibiotic-resistance gene chosen conferred resistance to gentamicin. There are very few appropriate antibiotics that inhibit the growth of *S. rimosus*. Initial tests on the suitability of gentamicin as a selectable marker were carried out by C. Ives (Glasgow University). A plasmid carrying both the *tsr* and *gmr* genes was introduced into *S. rimosus* and selected initially with gentamicin, resulting in confluent growth. These colonies were patched onto selective media containing thiostrepton and after three days at 30^{0}C only 30% of the colonies were found to be true transformants (C. Ives; pers. comm.). However, a plasmid carrying both *gmr* and *tsr* genes, for which the *tsr* gene is used for the primary selection, can be plated subsequently on media containing gentamicin and be found to be truly resistant to gentamicin as well. It was concluded that gentamicin could be used only as a secondary marker.

Gentamicin is an aminoglycoside which, together with neomycin, kanamycin and hygromycin, normally causes mistranslation in prokaryotes and inhibits protein synthesis (Holmes *et al.*, 1991). *Streptomyces tenebrarius,* which produces the aminoglycoside, contains a gene that encodes an S-adenosyl methionine-dependent methylase resistance gene that modifies the 16S ribosomal RNA. The resistance is due to methylation of the rRNA at a single site, which is characteristic for a given phenotype. For example, the gene encoding the *kgmB*

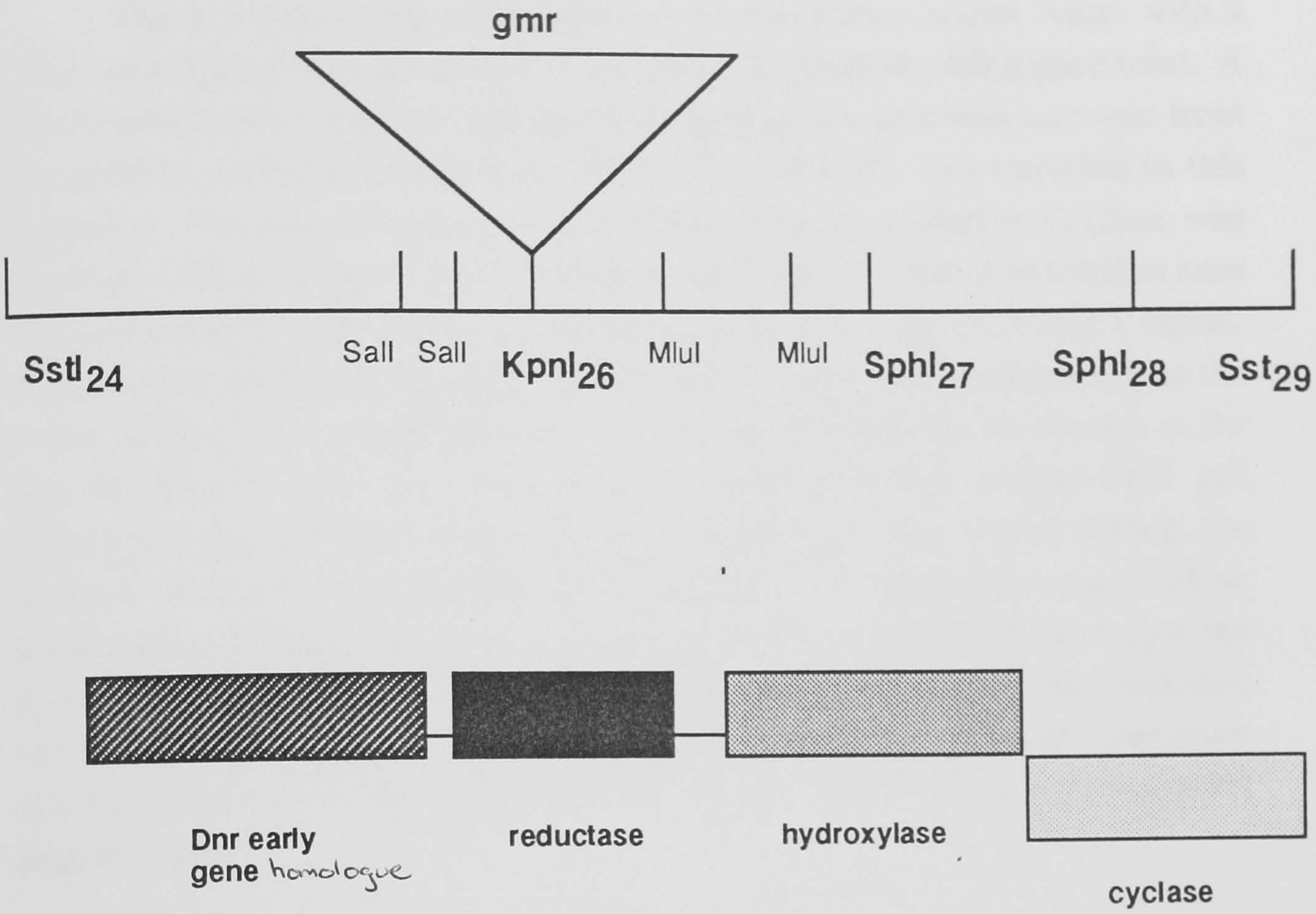


Figure 6.5
A partial restriction map of the *otc* cluster of *S. rimosus* showing the position of insertion of the gentamicin resistance marker into the reductase gene.

(for *k*anamycin-*g*entamicin resistance *m*ethylation) methylates residue G-1405 (Holmes *et al.*, 1991).

The first step in the construction of the integration vector began with a three-way ligation into the shuttle vector pIBI24 to produce p503 (figure 6.6a). A 1.5 kb *Bam*HI-*Sph* I fragment containing the gentamicin gene was removed from the pLST14 plasmid (Skeggs *et al.*, 1987). The promoter was included in this fragment. The second fragment from pLUS801, kindly supplied by C. Chen, was digested with *Bam*HI and *Xba*I. This plasmid carries a series of restriction sites that cut rarely in *Streptomyces* (containing a high percentage of A and T bases). This series of rare restriction sites (rrs), 57 bp in length, was incorporated into the vector so that upon integration, the rare sites would mark the *otc* cluster, at the altered *Kpn*I$_{26}$ site, for later genomic digestion and pulsed-field gel electrophoresis analysis to determine the location of the cluster within the genome. The three-way ligation with the *Bam*HI-*Sph* I fragment from pLST14, the *Bam*HI-*Xba*I fragment from pLUS801 and pIBI24 (*Sph*I-*Xba*I) produced the plasmid p503 (see Figure 6.6a). By utilising the restriction sites in the polylinker of the plasmid pGEM7, a *Xba*I cassette containing the gentamicin-rare restriction site (gmr-rrs) insert, was produced. The resulting plasmid was labelled p504 (Figure 6.6a).

A 3.4 kb fragment of the *otc* cluster from *Sst*I$_{24}$-*Sph*I$_{27}$ (see figure 3.2) was inserted onto the shuttle vector, pIBI24 to produce the plasmid, pSRG6. This plasmid contains the putative reductase gene that was chosen for disruption. The plasmid was linearised by digestion at the *Kpn*I$_{26}$ site and the overlapping ends filled in with T4 DNA polymerase. *Xba*I linkers were ligated to the blunt-ended fragment to produce the plasmid pSRG6X. The gmr-rrs *XbaI* cassette from plasmid p504 was then ligated to the *Xba*I site of pSRG6X to produce the plasmid p505. The orientation of the gentamicin gene is opposite to that of the *otc* genes to avoid a possible lethality caused by overexpression of any downstream *otc* genes. p505 now contained the gmr-rrs resistance marker flanked by 2.0 kb to the left and 1.4 kb to the right of the *Kpn*I$_{26}$ of the *otc* cluster in a pIBI24 shuttle vector (Figure 6.6b).

The final component in the construction of the integration vector was the unstable delivery vector containing the *Streptomyces* origin of replication. The plasmid p505 contains a unique *Sst*I site, which was cleaved to linearise the plasmid. The linear fragment was then ligated to pIJ486 restricted at the unique

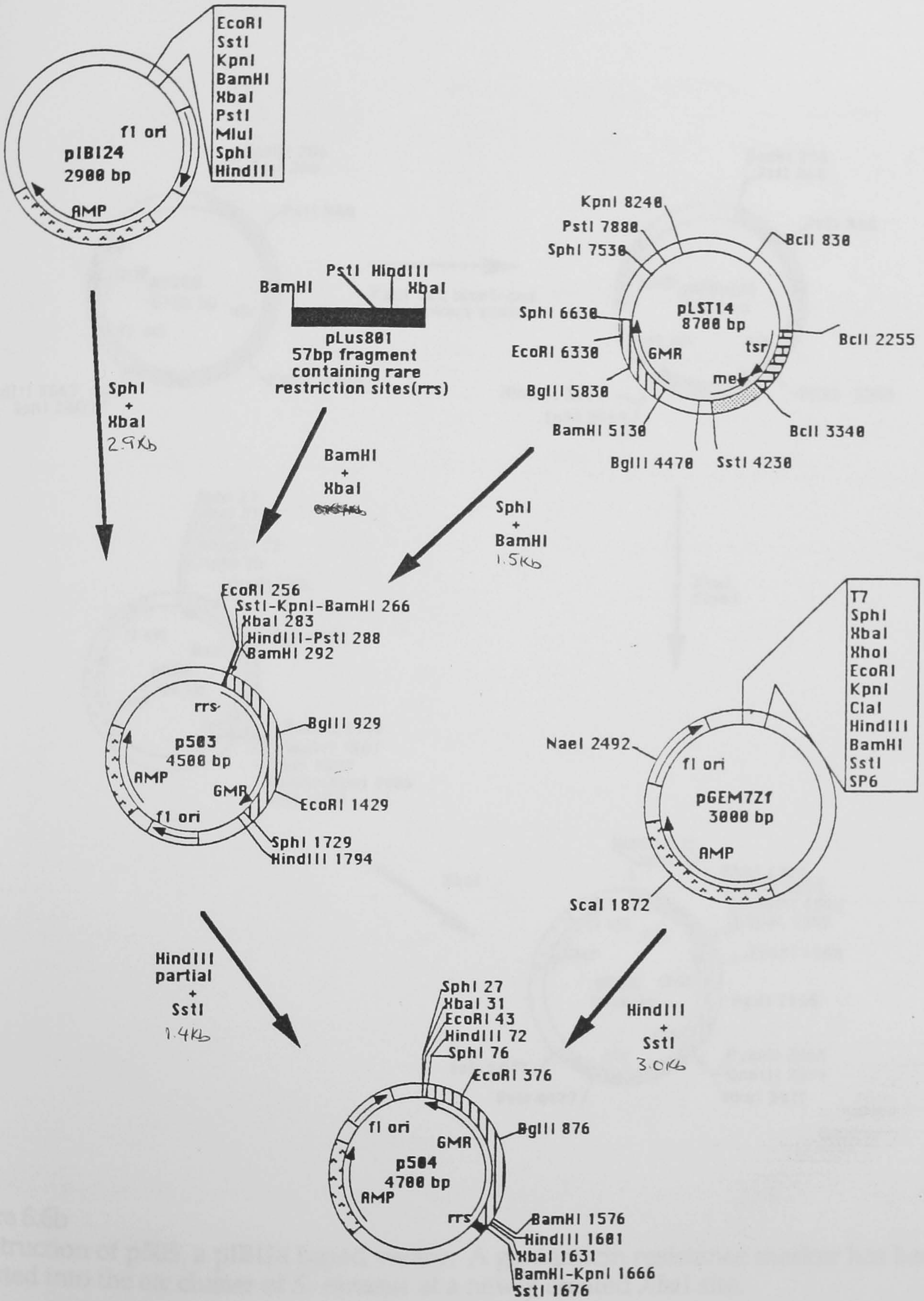


Figure 6.6a
Construction of p504, a *Xba*I cassette containing a gentamicin resistance gene.

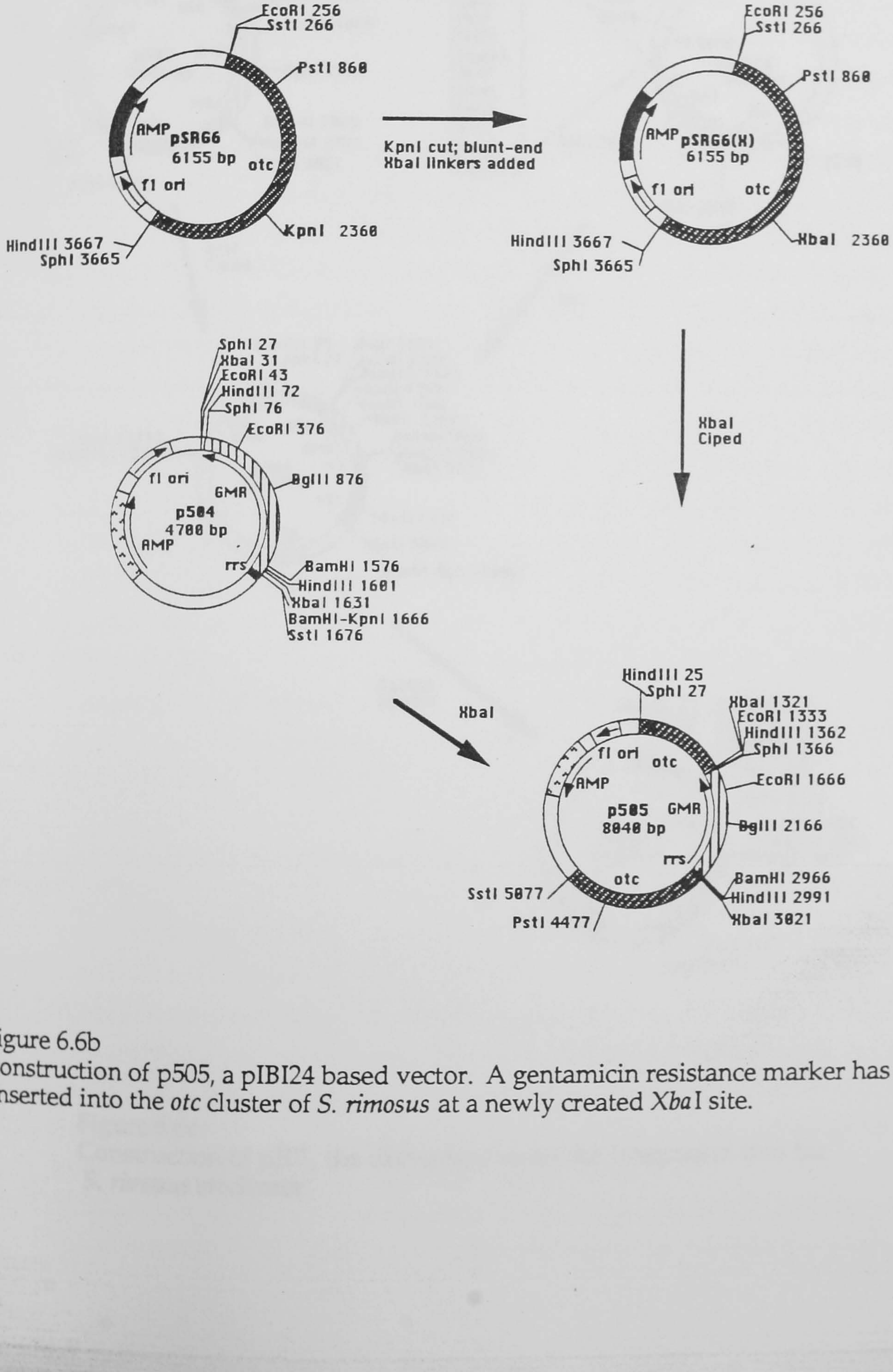


Figure 6.6b
Construction of p505, a pIBI24 based vector. A gentamicin resistance marker has been inserted into the *otc* cluster of *S. rimosus* at a newly created *Xba*I site.

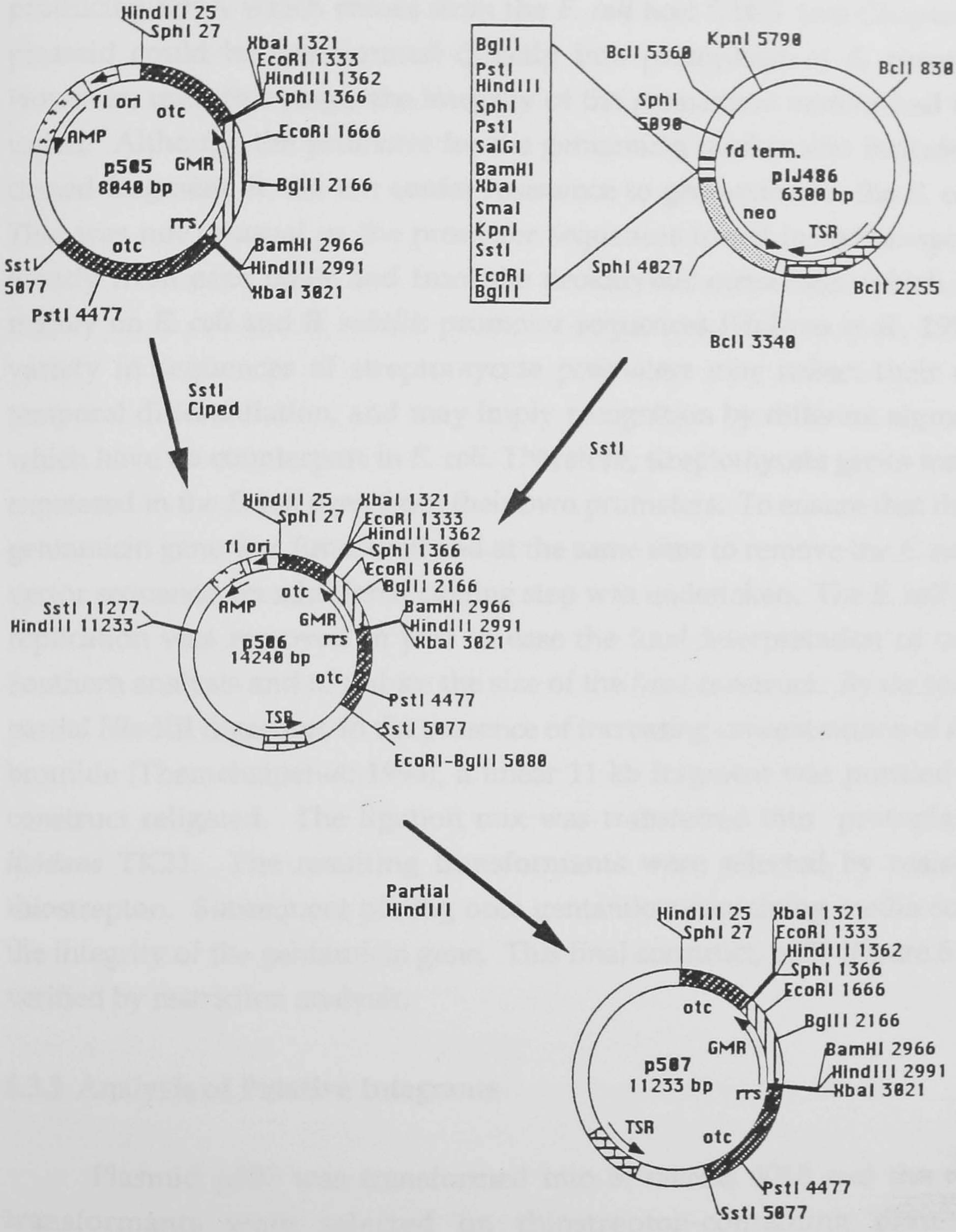


Figure 6.6c
Construction of p507, the disruption vector for integration into the *S. rimosus otc*cluster.

*Sst*I site. To facilitate cloning, the plasmid carrying the *E. coli* origin of replication was treated with calf intestinal phosphatase, so that the vector could not self-ligate. The resulting bifunctional plasmid was named p506 (figure 6.6c). By producing DNA which comes from the *E. coli* host GM31 (see Chapter 5), this plasmid could be transformed directly into protoplasts of *S. rimosus* 4018. However, until this stage, the integrity of the gentamicin marker had not been tested. Although the promoter for the gentamicin marker was included in the cloned fragment, it did not confer resistance to gentamicin in the *E. coli* host. This was not unusual as the promoter sequences found in *Actinomyces* differ greatly from each other and from the prokaryotic consensus, which is based mainly on *E. coli* and *B. subtilis* promoter sequences (Holmes *et al.*, 1991). The variety in sequences of streptomycete promoters may reflect their complex temporal differentiation, and may imply recognition by different sigma factors which have no counterpart in *E. coli*. Therefore, streptomycete genes may not be expressed in the *E. coli* even from their own promoters. To ensure that the cloned gentamicin gene was functional and at the same time to remove the *E. coli* pIBI24 vector sequence, an additional cloning step was undertaken. The *E. coli* origin of replication was removed in part to ease the final interpretation of data after Southern analysis and to reduce the size of the final construct. By carrying out a partial *Hin*dIII digestion, in the presence of increasing concentrations of ethidium bromide (Thamchaipenet; 1994), a linear 11 kb fragment was purified and the construct religated. The ligation mix was transferred into protoplasts of *S. lividans* TK24. The resulting transformants were selected by resistance to thiostrepton. Subsequent plating onto gentamicin-containing media confirmed the integrity of the gentamicin gene. This final construct, p507 (figure 6.6c), was verified by restriction analysis.

6.3.3 Analysis of Putative Integrants

Plasmid p507 was transformed into *S. rimosus* 4018 and the resulting transformants were selected on thiostrepton-containing plates. The transformation efficiency was $4.0X10^5$ transformants.μg^{-1} DNA. Ten colonies were selected and spore suspensions made. From the spore suspensions, 30ml of liquid TSB media was inoculated. No antibiotic was added to the media, therefore relaxing the selective pressure on the strain to retain the plasmid. After 3 days, the culture was streaked for single colonies onto a Soya mannitol plate containing no antibiotic. A 1 mL aliquot was also stored in 20% glycerol at -20^0C. The appearance of the individual colonies appeared to indicate that some sort of

recombination event was taking place. A control plasmid with no homologous *S. rimosus* chromosomal DNA gave recombinants with a uniform brown pigment indicative of an *otc*- producing strain. The colonies on the experimental plates were of mixed pigmentation, with various shades of brown and a number were completely white in colour. Fifty colonies from each of the 10 plates were patched simultaneously onto Soya mannitol, Soya mannitol plus 50μg.mL^{-1} thiostrepton, and Soya mannitol plus 200μg.mL^{-1} gentamicin. An attempt to replica plate the colonies proved to be unsuccessful and therefore individual patching of colonies was required. The Soya mannitol plate was used as a master plate for further plating of individual colonies. The control strain 4018 and a recombinant containing plasmid pIJ486 were also plated onto the Soya mannitol plates. Putative positives were streaked onto a higher concentration of gentamicin (500μg.mL^{-1}), as there was a slight growth by recombinants containing the control pIJ486. Of the 500 colonies originally picked, 106 were gentamicin-resistant and thiostrepton-sensitive, indicative of the double crossover event (as illustrated in figure 6.7). Ten colonies, producing various shades of brown pigment, were selected for further analysis. Chromosomal DNA was purified and digested with *Sst*I. A single *Sst*I site in the integration vector facilitates the interpretation of hybridisation results. Wild-type genomic DNA cut with *Sst*I gives rise to a 5.4 kb fragment encompassing the region surrounding the putative reductase gene (see figure 6.7). A single cross-over will produce two *Sst*I fragments, 13.3 kb and 5.4 kb. A double cross-over event will produce a single *Sst*I fragment of 7.4 kb. A Southern blot was hybridised with the random-primed, DIG labelled, 800 bp *Mlu*I-*Sal*I fragment containing part of the wild-type reductase gene. The positive control used was the *S. rimosus* strain G7 which also produces OTC. The chromosomal DNA was digested with *Sst*I and blotted. When probed, a number of clones contained the expected 7.4 kb fragment, indicative of a double crossover event. However, most of these also appeared to have an amplified number of the copies of the gmr::reductase gene (See figure 6.8) There was one colony which had only a single copy, similar to the G7 strain. This was colony 16, and was labelled Int16, (figure 6.8, lane g). The integrity of the Int16 DNA was checked by probing with both the 1.6 kb*gmr* and 1.1 kb*tsr* genes. As shown in figure 6.8, the *gmr* gene was present, by interpretion of the band size, and in figure 6.9 (lane b), the *tsr* gene was not. In figure 6.9, upper band observed in the pIJ486 control lane (f) represents supercoiled undigested plasmid DNA. The Int16 colony was pale in colour and did not produce OTC when checked in plate assays but was resistant to OTC when plated on media containing 50μg.mL^{-1} OTC.

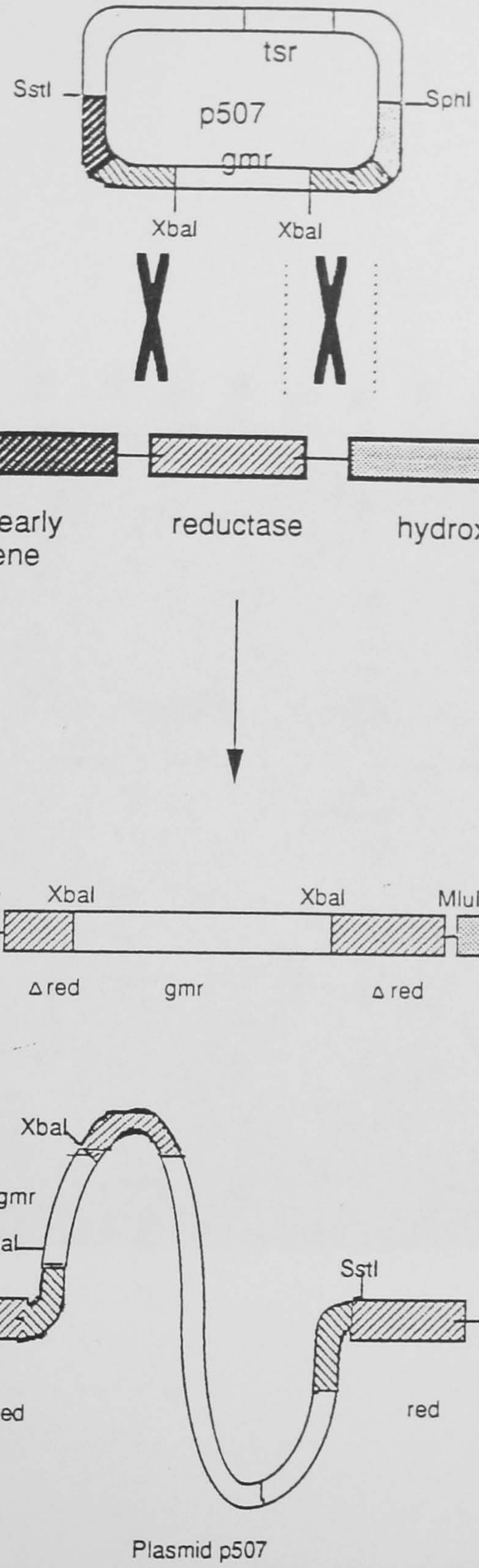


Figure 6.7
A schematic diagram of the possible outcomes of homologous recombination events with p507 and the *S. rimosus* chromosome. A double cross-over event (a) results in a 7.4 kb fragment when digested with *Sst*I. A single cross-over event (b) results in two fragments, 13.3kb and 5.4kb, when digested with *Sst*I.

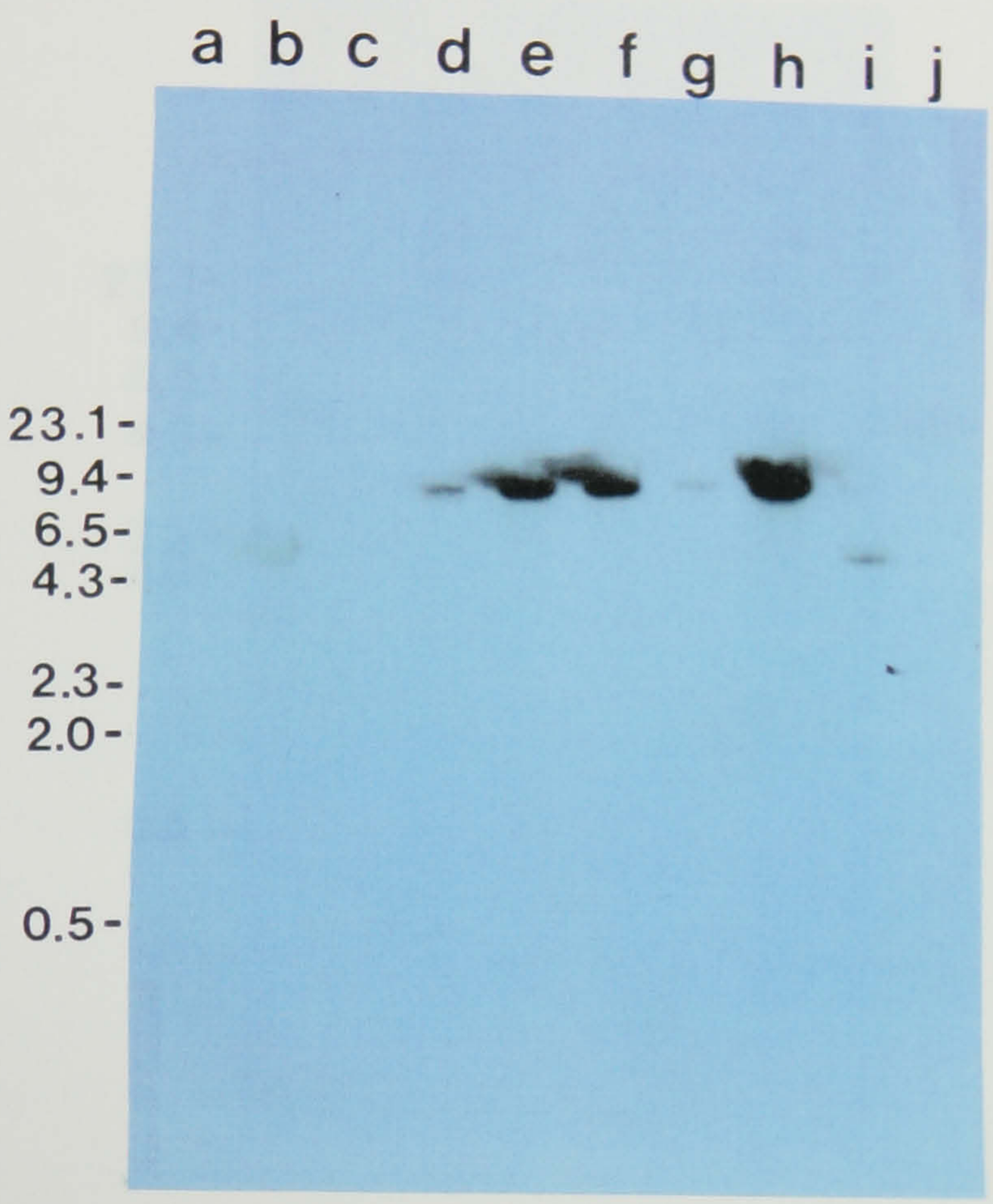


Figure 6.8
DNA from putative recombinants with *gmr* integrated in the chromosome were digested with *Sst*I and probed with an 800 bp *Mlu*I-*Sal*I *otc* fragment from pSRG1 (lanes b-h). The control lane (i) contains DNA from wild-type S. rimosus, digested with *Sst*I showing hybridisation to a 5.4 kb fragment. λ*Hin*dIII markers are marked in Kb in lanes a & j. The ethidium bromide gel (not shown) showed equal loading in all tracks. Int16 is found in lane g, Int17 is found in lane h.

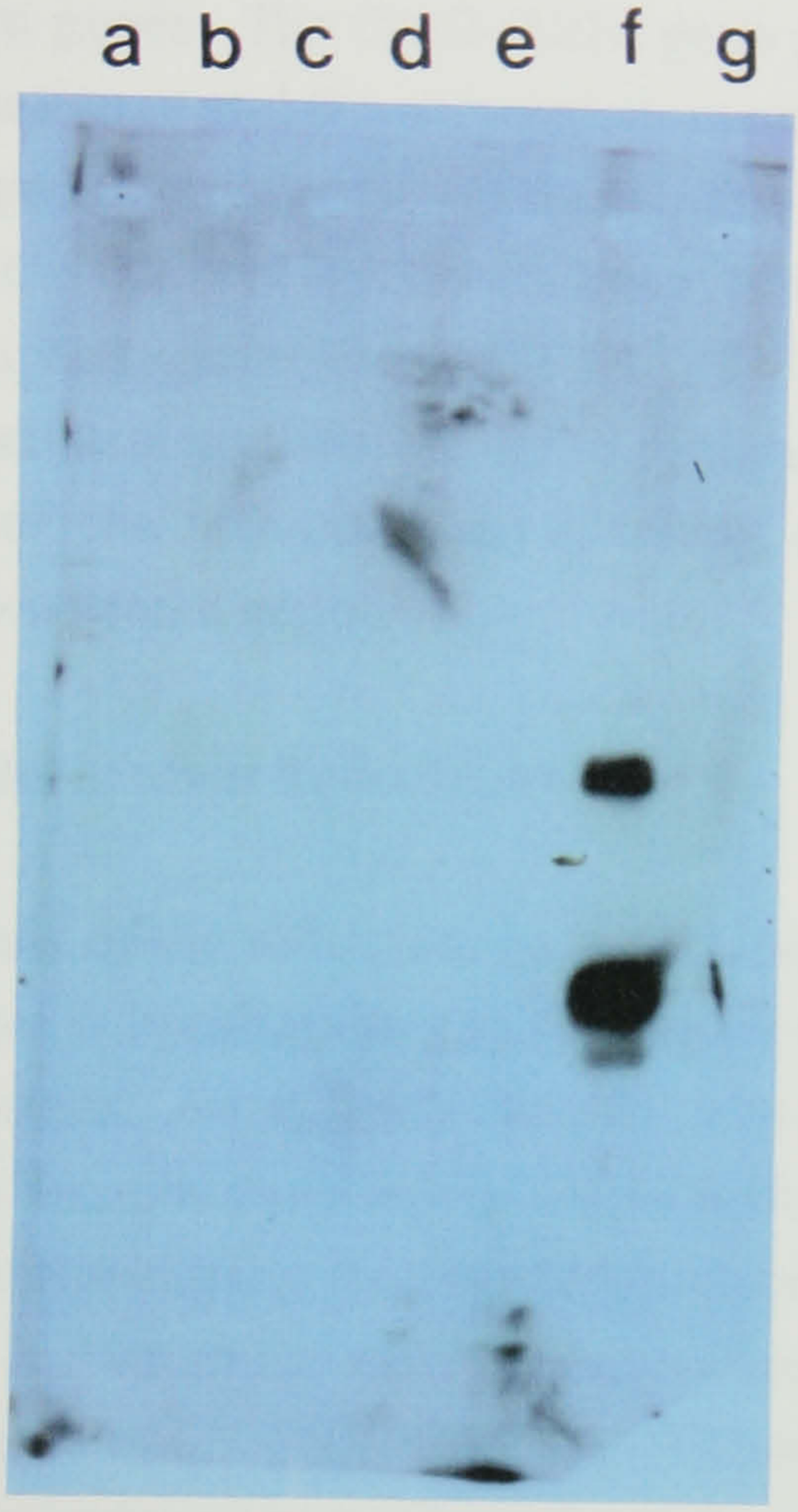


Figure 6.9
DNA from putative recombinants with *gmr* integrated in the chromosome probed with 1.1 kb *Bcl*I *tsr* gene, (lanes b-e). The 1.1 kb *Bcl*I fragment of the control plasmid, pIJ486 (lane f) is hybridised to the *tsr* probe. The flanking lanes contain λ *Hin*dIII markers (kb). Int16 (lane b) corresponds to lane (g) in figure 6.8. Lanes c, d and e represent the same samples as lanes b, c and d repectively, in figure 6.8.

Strain Int16 contains a gentamicin gene that disrupts the putative reductase gene and any genes downstream which run off the same transcript. Preliminary sequence analysis shows that the open reading frames of the *OtcD*-ORF4 and *OtcD*-ORF5 also run in the same direction (see Chapter 3). It was assumed at this point that the polarity effects might prevent or reduce expression of the two downstream genes. The *OtcD*-ORF4 gene product is thought to be involved early in the pathway and the only pale mutants known so far (Butler *et al.*, 1989) are those of genes early in the pathway. The pale colour observed for the Int16 strain maybe the result of lost downstream transcription at *OtcD*-ORF4, rather then the disrupted gene, *OtcD*-ORF3. To facilitate experimental interpretation, a second strategy was set up to replace the gentamicin marker with a mutant form of the reductase gene, taking care not to disturb the transcription of the downstream genes.

6.3.4 Construction of the Mutant Reductase Gene

By careful analysis of the reductase gene and the homologous sequences of other reductase genes it became clear that a highly-conserved region was a good target for mutagenesis. An in-frame deletion was favoured over any point mutations or insertions because there is less risk of restoring the wild-type gene. A schematic diagram of the cloning strategy to produce the mutant reductase is shown in figure 6.10. A five amino acid sequence was to be deleted from the highly conserved sequence of GXXX**GXGXAX**A. (Those in bold were removed). PCR was the technique employed to produce the mutant gene. From sequencing analysis, single-stranded templates were available in both orientations (See figure 3.1). This allowed the use of the universal primer, which recognises and binds to vector sequence, as one of the primers needed to generate the PCR product. In fact two fragments would be generated, one to the left of the missing bases and one to the right. The two oligos designed to 'loop-out' the conserved region also incorporated a new *Bam*HI site. This allowed the mutant gene to be located by simple restriction analysis. After the PCR was complete, a second series of constructions was made to produce a vector that could replace the gentamicin gene in Int16 by two 'Campbell-type' recombination events. As shown in figure 6.10, the constructions ensured that the only DNA generated by PCR in the final construct was that which was contained within the reductase gene.

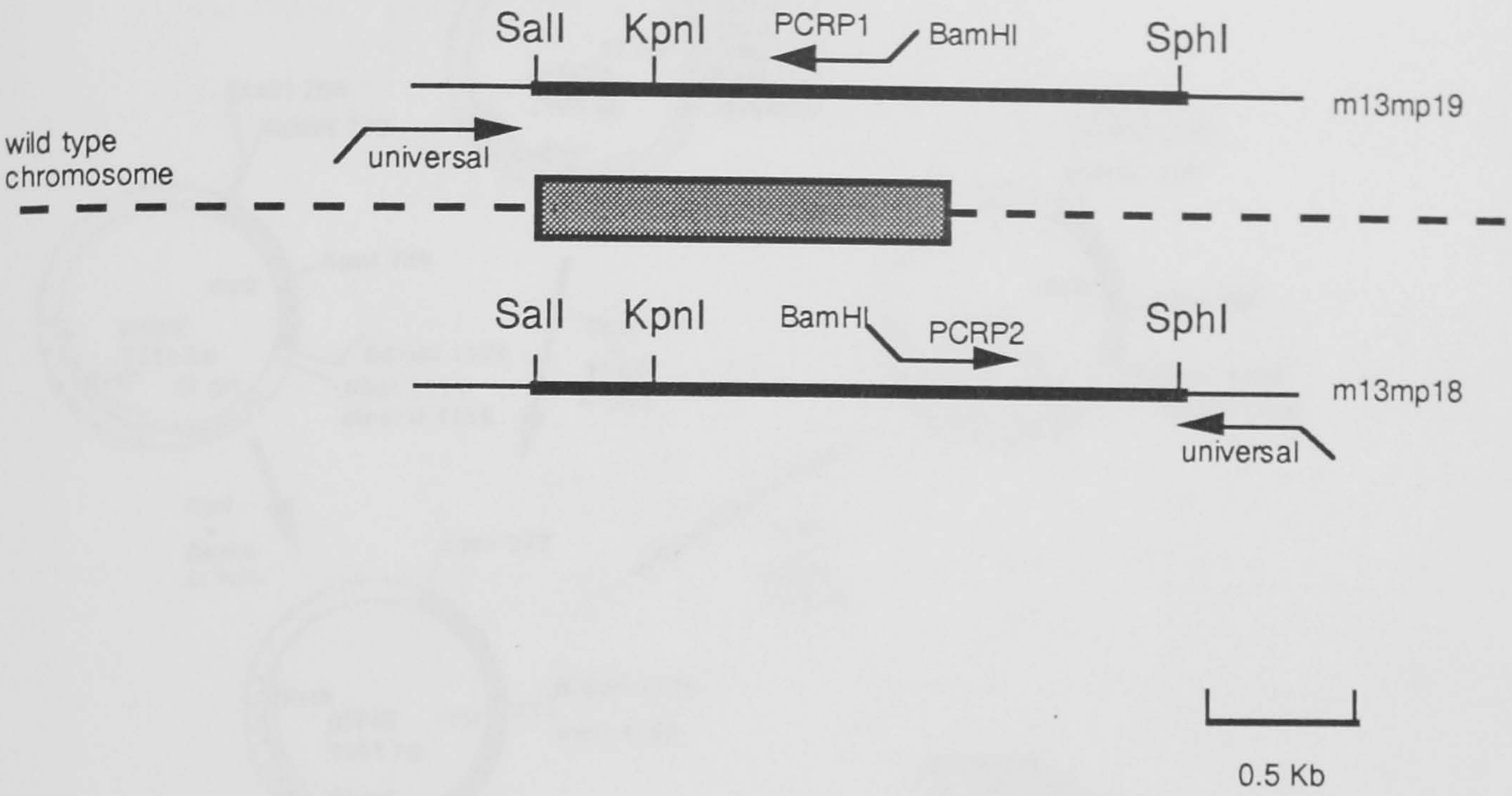


Figure 6.10
The mutation strategy using PCR to create an 'in-frame' deletion of the reductase gene in the *otc* cluster. Thin lines indicate vector sequence, thick lines indicate chromosomal sequence. The boxed region shows the location of the reductase gene. PCRP1 and PCRP2 are the mutagenic oligos which create a *Bam*HI site. Universal denotes the universal primer.

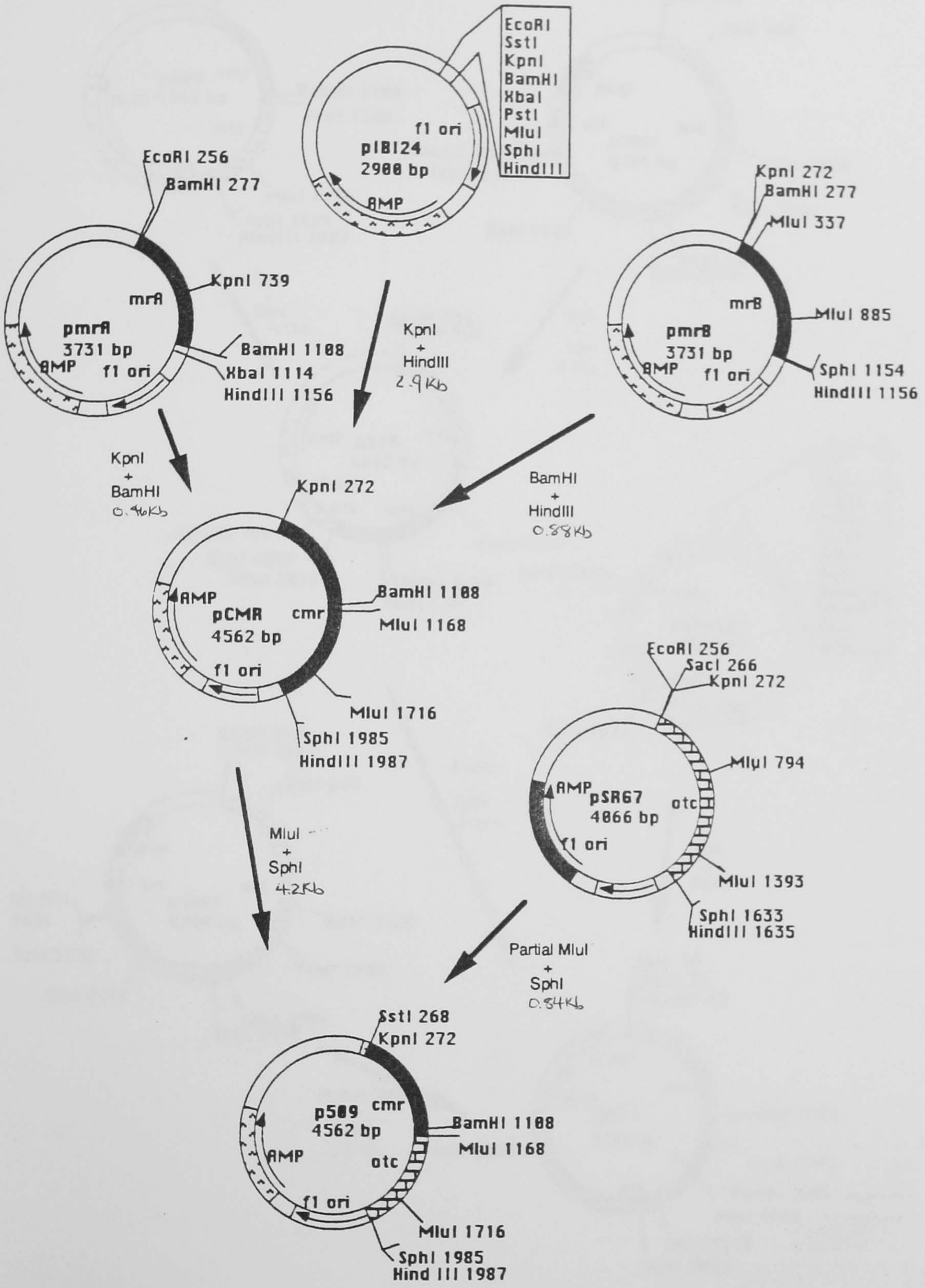


Figure 6.11a
Construction of p509 from two PCR-generated fragments, mrA and mrB. 'otc' denotes wild-type *otc* DNA, 'cmr' denotes PCR-generated *otc* DNA.

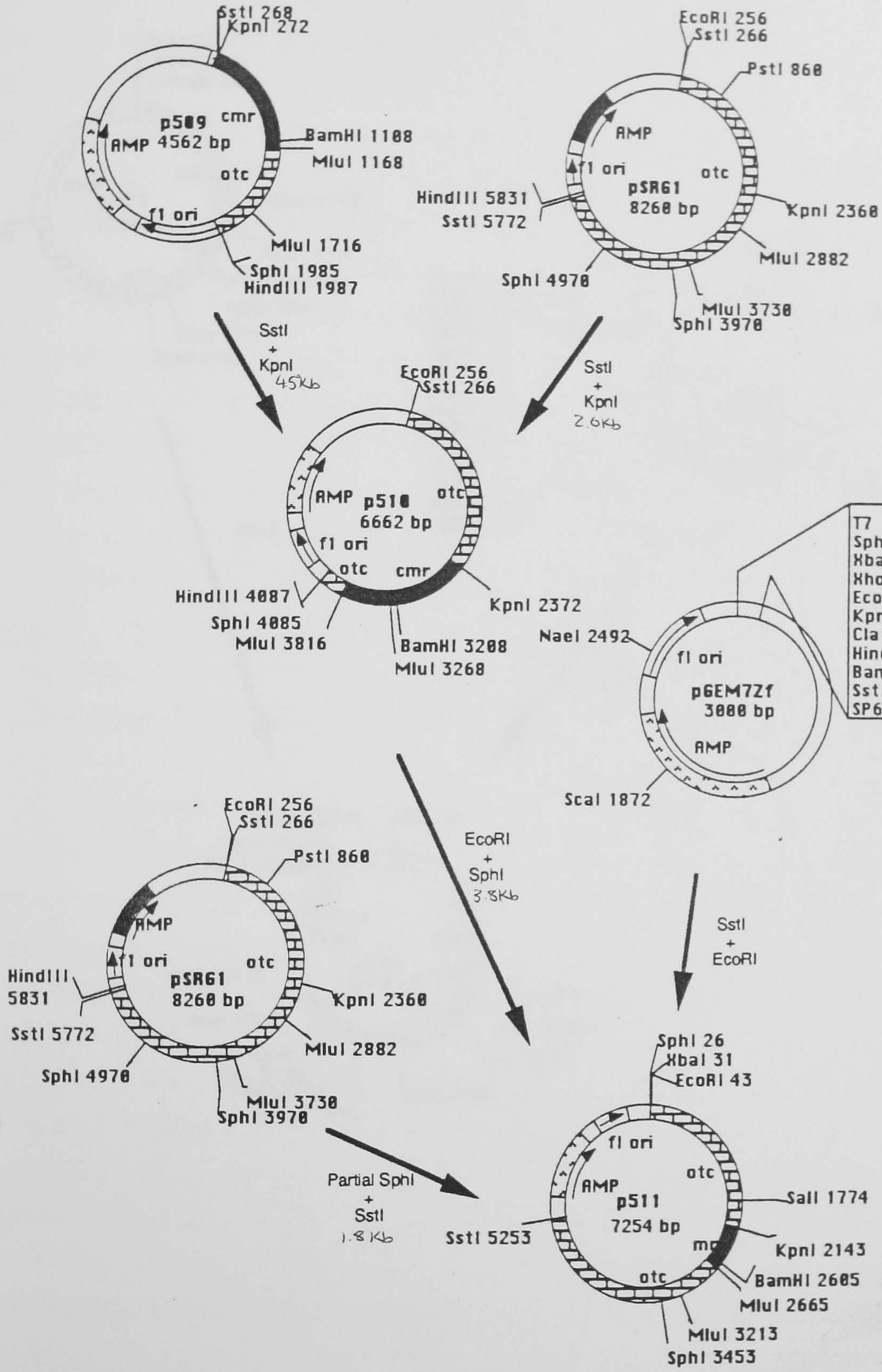
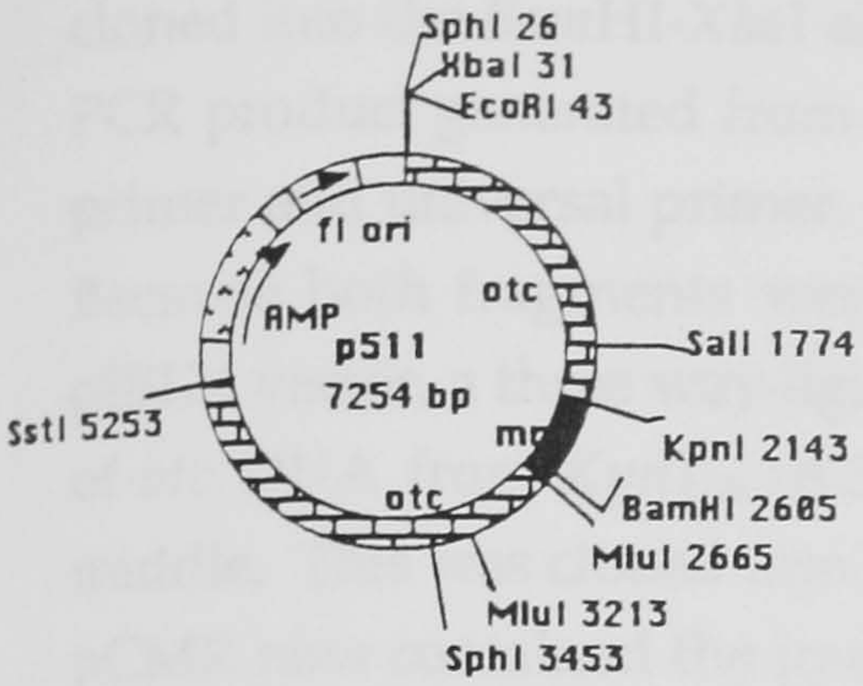
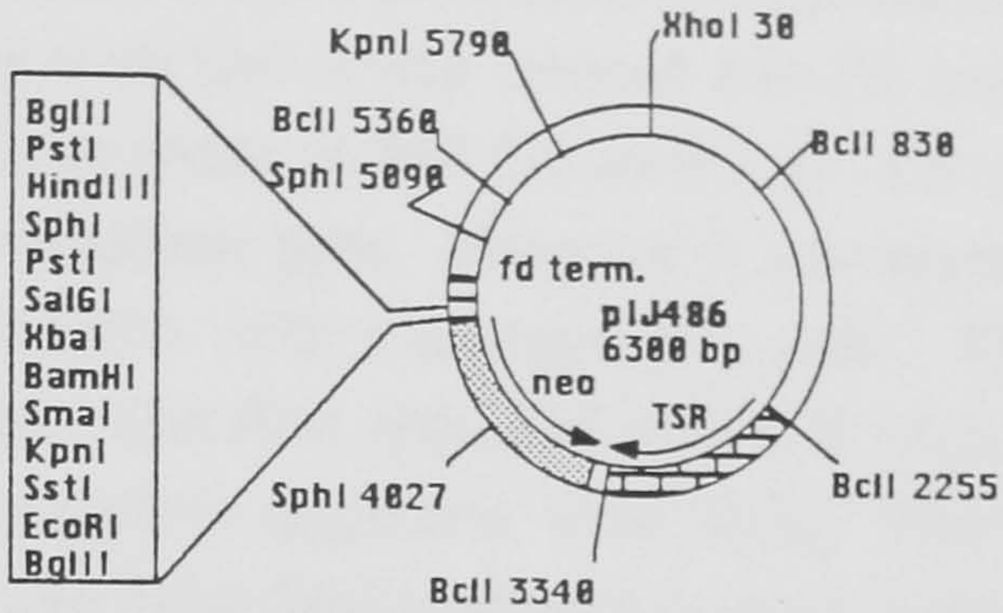


Figure 6.11b
Construction of p511, which contains a mutated 'reductase' gene from the *otc* cluster of *S. rimosus*.





Xba I

XbaI
Ciped

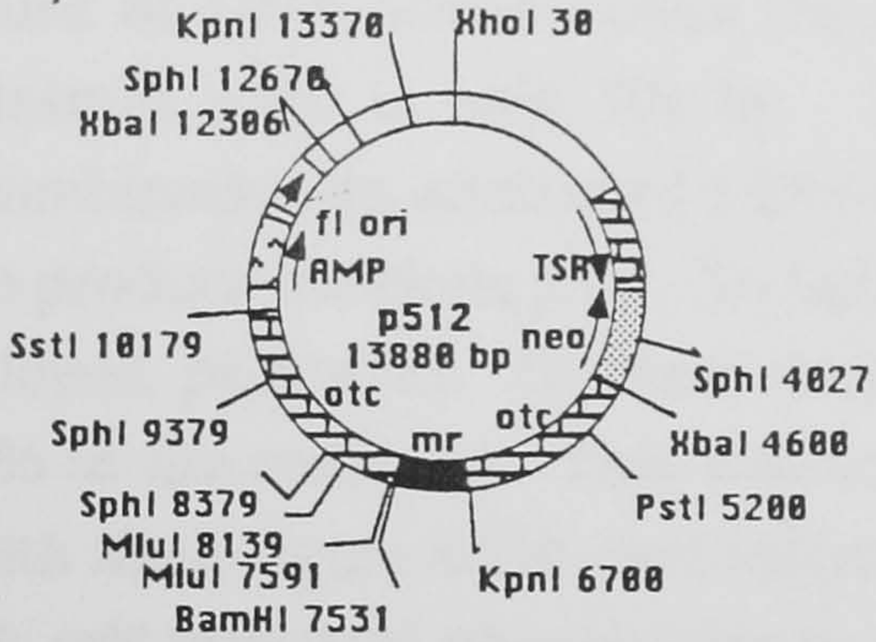


Figure 6.11c
Construction of p512, which carries the mutant 'reductase' gene of the *otc* cluster in *S. rimosus*.

Plasmid pmrA contains the PCR product generated from left of the missing amino acids using the PCRP1 primer and the universal primer. It was cloned into the *Bam*HI-*Xba*I sites of the vector, pIBI24. Clone pmrB contains the PCR product generated from right of the deleted amino acids using the PCRP2 primer and universal primer. It was cloned into the *Bam*HI-*Sph*I sites of pIBI24. Because both fragments were originally cloned in the same direction in the pIBI24 vector, a three way ligation was set up to produce a contiguous fragment of *otc* DNA from *Kpn*I$_{26}$ to *Sph*I$_{27}$ with the newly created *Bam*HI site in the middle. This was cloned into pIBI24 to create pCMR (figure 6.11a). The plasmid pCMR now contained the mutant reductase gene. However it was necessary to replace all of the surrounding *otc* DNA with wild-type sequence. This was accomplished in two cloning steps. The first required a partial digestion of pSRG7 with *Mlu*I followed by complete digestion with *Sph*I. This 838 bp fragment was then ligated to the large *Mlu*I-*Sph*I fragment created from pCMR. The resulting plasmid, p509, (figure 6.11a), now contained wild-type *otc* DNA to the right of the *Mlu*I site, very close to the newly-created *Bam*HI site. The second cloning step added wild-type *otc* DNA (*Sst*I$_{24}$ to *Kpn*I$_{26}$) to the left of the reductase gene from plasmid pSRG1 to produce the plasmid p510 (figure 6.11b).

The amount of DNA which flanks the right-hand side of the mutant reductase in plasmid p510 is only 500 bp. To increase the probability of homologous recombination an additional 2 kb fragment of *otc* DNA was ligated to the plasmid to produce the clone p511. To facilitate cloning, a different shuttle vector was employed, pGEM7Zf. The final cloning requirement was to include the vector, pIJ486 in the construct. This was accomplished by opening up the p511 plasmid with *Xba*I, figure 6.11b, and inserting *Xba*I-digested pIJ486. p511 was treated with calf intestinal phosphatase to avoid self-religation. Selection was carried out in the *E. coli* host GM31 so that the resulting recombinant plasmid, p512, (figure 6.11) could be transformed directly into *S. rimosus* Int16. The transformation was done using electrocompetent Int16 cells (Chapter 2.2.8), with selection on TSB media containing 50μg.mL^{-1} thiostrepton.

6.3.5 Analysis of Intergrants with Mutant Reductase

A second series of transformants was analysed. Ten transformants were picked from the transformation plate. For simplicity, these were labelled mr 1-10. TSB liquid media containing no antibiotic was used to segregate the unstable

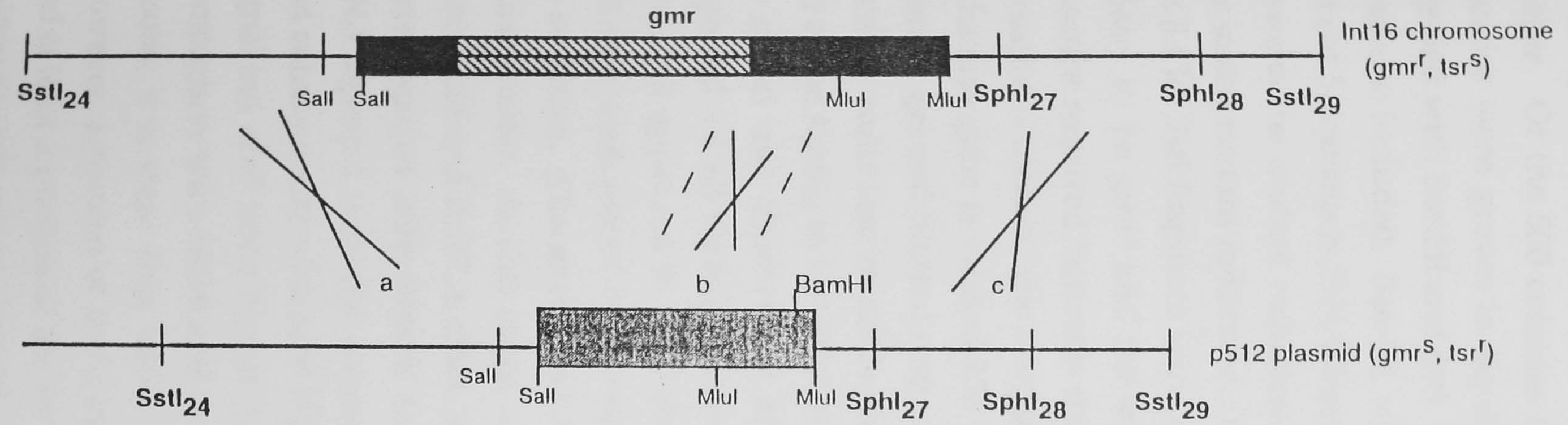


Figure 6.12
Two possible outcomes of double cross-over events with p512. The wild-type 'reductase gene (■) and mutant reductase (▒) are indicated. A *Bam*HI-*Sst*I digest will produce a single band of 5.4 kb with cross-overs at a and b and two fragments (2.5 and 2.9 kb) if crossovers occur at a and c, when probed with a *Sal*I-*Sph*I$_{27}$ fragment of wild-type *otc* reductase DNA.

delivery vector from the cells, similar to the procedure outlined above, section 6.3.3, for the p507 plasmid. Single colonies were patched onto selective and non-selective media as above (Section 6.3.3). Only those colonies which were thiostrepton-sensitive (tsr^s) and gentamicin-sensitive (gmr^s), were analysed further. Of the 500 colonies tested, 61 were both tsr^s and gmr^s. Ten of these colonies were grown in liquid culture, chromosomal DNA was purified and digested with *Bam*HI and *Sst*I. Positive controls of wild type G7 and Int16 DNA's were also included. *Bam*HI, which has the recognition sequence GGATCC, does not cut *Streptomyces* DNA frequently. A *Sst*I digest alone would not distinguish between the mutant reductase and the wild type. With the engineered *Bam*HI site in the mutant reductase, the *Bam*HI and *Sst*I double digestion would divide the 5.4 kb *Sst*I fragment into two fragments of 2.5 and 2.9 kb. It is possible for a colony to be gmr^s and tsr^s without mutating the reductase gene due to the distance required between the crossover points, as shown in figure 6.12. The gentamicin marker was inserted at the *Kpn*I_{26} site and the deleted portion of the reductase gene is located 456 bp distal. The digested DNA was run on a 0.8% agarose gel and blotted onto nylon membrane. The probe used to identify the putative reductase mutants was a 800 bp fragment stretching from the *Sal*I site left of the *Kpn*I_{26} to the *Mlu*I site to the right of *Kpn*I_{26}. As shown in Figure 6.13, no signal was observed in any of the clones. The G7 sample contained the expected 5.4 kb signal, lane m. Int16 contained the expected 7.4 kb fragment, lane l. It appeared as if all the mr clones lacked the *otc* DNA (lanes b-k). The colonies were plated onto media containing 50 $\mu g.mL^{-1}$ OTC and were found to be sensitive. The attempted replacement event induced the cells to delete the entire cluster. At this stage, a closer look at the Int16 clone was taken. DNA's from Int16 and Int17, a clone which showed an amplified signal, (see figure 6.8), were digested with *Bam*HI and *Sst*I and run along side a sample of wild-type DNA digested with the same enzymes. The blot was probed initially with *otrA* and subsequently with *otrB* (figure 6.14). It was evident that *otrA* (lane a) gave a signal but *otrB* (lane b) did not. The signal of the Int17 was also amplified in comparison with Int16 and wild type DNA (lane c in both blots). From these results, it is clear that Int16 clone did indeed undergo an integration event. However, a portion of the cluster was lost as well. In retrospect, the loss of *otrB* and at least a portion of the *otcY* locus, might explain the pale colour of the Int16 colony as *otcY* genes, close to *otrB*, are involved in polyketide biosynthesis. The second event augmented the first deletion event to remove the entire cluster.

Figure 6.13
Putative recombinants likely to contain an integrated mutant reductase gene (lanes b-k), digested with *Sst*I-*Bam*HI and probed with a *Mlu*I-*Sal*I fragment from pSRG1. Int 16 (lane l) and wild-type *S. rimosus* (lane m) showed the expected hybridisation bands. The ethidium bromide gel (not shown) showed equal loading in all tracks. λHindIII markers occupy the outside lanes.

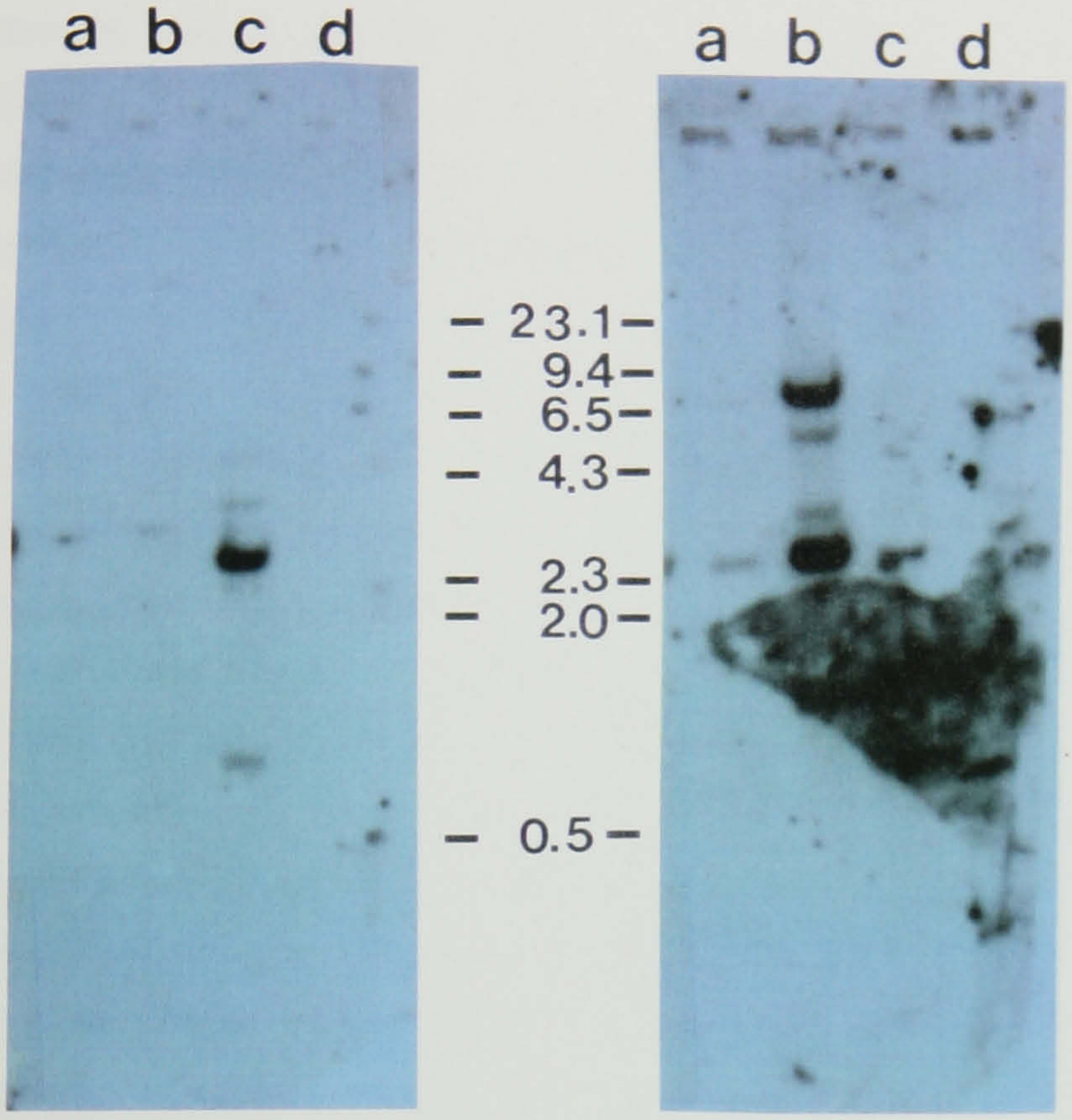


Figure 6.14
Int16 (lane a)and Int17 (lane b) chromosomal DNA was digested with *Bam*HI and *Sst*I and probed with a) *otrB* and b) *otrA*. The positive control DNA used was *S. rimosus* G7, lane c, λ HindIII, lane d.

With the advent of electrotransformation and the consequent increase in the transformation efficiency of *S. rimosus* M4018, integration using a suicide vector system was attempted. This would eliminate any deleterious effects the unstable *Streptomyces* replicon might have had on the integration events. Preliminary tests were carried out using different selective media to monitor the growth of electrotransformants. Previous work (C. Ives; pers. comm.), showed that gentamicin could not be used as the primary selectable marker for transformation of protoplasts, see Section 3.3.2. The plasmid, p507, which carries both the tsr and gmr antibiotic resistance markers was transformed into electrocompetent *S. rimosus* M4018 cells. Following electrotransformation, the cells were plated onto either Soya mannitol or Tryptic Soya broth (TSB) media containing varying concentrations of gentamicin. The concentrations ranged from 200μg.mL^{-1} to 600μg.mL^{-1}. The positive control plate contained 50μg.mL^{-1} thiostrepton. The plates were incubated at 30^{0}C for three days. In the case of Soya mannitol, there was confluent growth at all concentrations of gentamicin. However, the TSB plates contained individual colonies, some of which sporulated after 3 days and some that did not. The transformation efficiency calculated from the control plates containing thiostrepton was 9.0X10^{5}. The sporulating and non-sporulating colonies on the gentamicin-containing TSB plates were picked onto a separate Soya mannitol plate containing 50μg.mL^{-1} thiostrepton. After three days at 30^{0}C, all of the sporulating colonies grew and all of the non-sporulating colonies did not. This provided a good visual screen for the true transformants. This was an encouraging result, and 50ng of the plasmid p505 isolated from the GM31 *E. coli* host, was transformed into *S. rimosus* M4018 by electroporation, and plated onto TSB plates containing 400μg.mL^{-1} gentamicin. Plasmid p505 (see figure 6.6) carries only an *E. coli* origin of replication, and therefore integration is required for survival. The entire one mL was plated and after 3 days, there were 23 sporulating colonies. Twelve of these colonies were grown up in TSB for analysis. Gentamicin was added to the media to maintain the selective pressure on the *gmr* gene should a single cross-over event have occurred. The chromosomal DNA preps were digested with *Sst*I, blotted and probed with the 1.6 kb *gmr* gene. The resulting blot showed that all of the clones carried the integrated *gmr* gene, but again the DNA was amplified (blot not shown). The number of copies of the amplified region was well in excess of 500. No further analysis was undertaken with these clones.

6.4 Discussion

Gene disruption and replacement is a valuable tool in determining the function of specific genes. The primary approach taken here employed an unstable *Streptomyces* replicon as the delivery vehicle. All three plasmids considered, pIJ702, pIJ680 and pIJ486, are based on the pIJ101 replicon.

pIJ702 is maintained quite stably in some *Streptomyces* hosts such as *S. lividans*, *S. coelicolor* (Katz *et al.*, 1983) as well as *S. rimosus* (this work), and poorly maintained in others (*Sac. erythraea*, Weber *et al.*, 1988). The variability between strains also extends to variability of different plasmids within a given host. For example, pIJ680 and pIJ486 are unstable in *S. rimosus* although there is very little difference in plasmid size or content when compared to the more stable pIJ702. Many naturally-occurring plasmids, irrespective of their copy number, are maintained stably in a growing bacterial population. Chimeric plasmids generated by *in vitro* recombination methods, however, are frequently deficient in partition because they usually consist of only the replication origin, the associated replication functions, and one or two additional selectable markers such as antibiotic-resistance genes. This is the case with all three plasmids, and until the mechanisms that govern partitioning are fully understood there is bound to be plasmid variability within a given *Streptomyces* host.

Using resistance to the antibiotic marker gentamicin as a screen to identify integration into the *S. rimosus* "reductase" gene *via* the unstable delivery vector, pIJ486, has lead to chromosomal amplifications and deletions. The strategy employing the unstable vector was sound as the gentamicin gene was incorporated into the gene which was targeted (see figure 6.5). Unfortunately, the organism elicited its own repair mechanisms to counteract the integration event. The genetic instability of *Streptomyces* is well documented (Leblond *et al.*, 1991; Birch *et al.*, 1990; Dyson *et al.*, 1987; Betzler *et al.*, 1987) with many strains exhibiting extraordinarily high [illegible] rates in certain genes. Unstable phenotypes are lost spontaneously at frequencies as high as 0.1% of plated spores (Birch *et al.*, 1990). Also associated with this genetic instability, which often correlates with deletion of unstable genes, is the formation of very high-copy number tandemly-amplified DNA sequences (ADS) that takes place in the absence of any obvious selective pressure. A range of different sequences can be amplified, and to varying extents (5 to 10% of the total DNA), in different mutants derived from the same progenitor (Hornemann *et al.*, 1986). The

deletions can extend to in excess of 800 kb (Birch *et al.*, 1987) and the physically-amplified arrays can extend to 3,000 kb (Häusler *et al.*, 1989) constituting 18-45% of the chromosome, respectively (Birch *et al.*, 1990).

Amplifications can be classified into two main types. Type I amplifications are non-reproducible and, when characterised, have been shown to originate from overlapping regions of the chromosome. The majority of amplifiable units of DNA (AUDs) in *S. glaucescens* have been mapped to a single chromosomal locus of 100 kb (Häusler *et al.*, 1989). There is no consensus core sequence involved and the AUDs are present only in a single copy in the progenitor chromosome, flanked by short imperfect direct repeats (Häusler *et al.*, 1989). Type II amplifications, in contrast to type I amplifications, are the result of events which are reproducible, and the AUDs are either flanked by extensive (1.0 to 2.2 kb) direct repeats (Fishman *et al.*, 1985) or are already present in the chromosome in duplicate (Altenbuchner *et al.*, 1985). In *S. fradiae*, the flanking repeat structure is present in three copies in the chromosome, and in *S. lividans*, the repeat structure consists of three copies of a 1.0 kb repeat with two copies of a 4.7 kb sequence sandwiched between them. At first glance one may see these repeats as insertion sequences. However, the restriction patterns of the different repeats in *S. lividans* clearly show divergence (Altenbuchner *et al.*, 1985). This suggests that the direct repeat itself and not any other properties which it may possess, is the critical factor. This was confirmed when Altenbuchner *et al.*, (1985) found that *S. lividans*, with only two copies of the 1.0 kb repeat flanking the 4.7 kb sequence, was amplified at a much lower frequency than the larger duplicated repeat.

Amplifications have been closely linked to large chromosomal deletion events. The integration of the *gmr* gene into the *otc* cluster of *S. rimosus* has lead to the amplification and subsequent deletion of approximately one half of the cluster. The chromosomal DNA of Int16 appears to be in single-copy form. It appears that during the liquid non-selection stage it may have been in a higher copy number and when allowed to sporulate on the Soya mannitol plate, the organism attempted to stabilise the chromosome by deletion of the amplified sequence, in this case, the *otrB* end of the cluster. Several instability systems in *Streptomyces* spp. display a characteristic stepwise loss of markers (Dyson *et al.*, 1987), in which primary high-frequency mutant types give rise to secondary mutant types at even higher frequencies. This pattern of appearance of mutants is highly suggestive of a two-step mechanism whereby a primary deletion

facilitates further secondary events. This theory was supported in this work when, with the second integration event, the *S. rimosus* organism created further deletions to remove the entire cluster. This was verified by the loss of *otc* resistance on a Soya mannitol plate containing 50μg.mL^{-1} OTC. This step-wise loss of markers has been also demonstrated in *S. glaucescens* (Hintermann *et al.*, 1984) with the sequential loss of the unstable genes coding for phosphotransferase (*strS*) and subsequently tyrosinase (*melC*) proteins. In *S. glaucescens* , the loss of 800 kb of genetic material certainly impairs the viability of the cells. However, they are still able to grow on minimal media, indicating that they have lost no essential genes. This is also the case with *S. rimosus*, indicating that the deleted region apparently contains relatively dispensable DNA which may be rich in genes encoding the biosynthesis of secondary metabolites. This work is further corroborated by that of Gravius *et al.*, (1993), who during screening programs to improve oxytetracycline production in *S. rimosus* R6, showed that some high-yielding isolates showed high levels of genetic instability manifested by the production of morphological variants. Five phenotypic traits were affected: sporulation, pigmentation, colony morphology, oxytetracycline production and oxytetracycline resistance. The variants were classified into three classes based on oxytetracycline resistance. Class I variants (99%) had parental levels of OTC resistance. They showed a variety of phenotypes affecting colony morphology, pigmentation, degrees of sporulation and *otc* production. These variants were very unstable and gave rise to further variants and some reversions to the parental phenotypes. There were no DNA rearrangements detected in the primary Class I variants. Class II variants were sensitive to OTC, but showed uniform phenotypic properties. Hybridisation analysis showed the deletion of the *otrB* resistance gene and indeed all of the cluster. Restriction analysis of these Class II variants using pulsed-field gel electrophoresis confirmed a deletion of 455 kb, all with identical restriction patterns suggesting that there are specific deletion end-points. Class III variants were the rarest observed (0.1%), exhibited uniform phenotypes and showed a 30-40% overproduction of OTC and a corresponding increased level of OTC resistance. There were extensive, reproducible, large-scale DNA rearrangements of about 200 kb, including the *otrB* gene.

It is presumed that the second replacement event created a series of Class II variants which have lost a total of 455 kb of DNA, including the *otc* cluster and flanking sequence. Pulsed-field gel electrophoresis would have to be undertaken to confirm that these variants contain the same deletion end-point as those cited

in Gravius *et al.*, 1993.. It can also be presumed that within this 455 kb there are non-essential genes due to the uniform phenotypes observed.

The physical map of *S. rimosus* is currently being constructed by J. Cullum (pers. comm.) with the use of pulsed-field gel electrophoresis. Because of the high G+C content of the *Streptomyces* DNA (70 to 74%), restriction enzymes with recognition sites containing only A and T nucleotides, such as *Ase*I (ATTAAT) and *Dra*I (TTTAAA), will generate relatively few fragments. *Ase*I and *Dra*I linking cosmids (i.e. recombinant cosmids including either *Ase*I or *Dra*I sites) were isolated from a gene bank and used as hybridisation probes against Southern transfers of the pulsed-field gel electrophoresis restriction patterns. The precise mapping with both single and double digests supports the interpretation that the chromosome is a linear structure. A preliminary map is shown in figure 6.15.

Bacterial linear plasmids were first described in *Streptomyces rochei* (Hayakawa *et al.*, 1979) and have now been detected in at least 10 other *Streptomyces* spp. (Hinnebusch *et al.*, 1993). The linear plasmid DNA has a terminal protein attached covalently to the 5' end of each DNA strand, as well as inverted terminal repeats, dubbed 'invertrons' (Sakaguchi; 1990). *Streptomyces* linear plasmids as well as other invertrons probably replicate by a protein-primed DNA replication mechanism that has been well characterised for *B. subtilis* phage ϕ29 and Adenovirus (Salas, 1991). The telomere is the origin of replication and is recognised by specific initiation proteins that promote unwinding. The 5'-terminal protein (TP) serves as the primer for a specific DNA polymerase. In the priming reaction, a free TP first complexes with DNA polymerase at the origins of the telomere. The DNA polymerase then catalyses the formation of a covalent bond between the free TP and a dNTP. This TP-linked dNTP becomes the 5' terminal nucleotide of the nascent strand, which can then grow to the end of its template.

Linear chromosomes have been found recently in a number of *Streptomyces* spp including *S. coelicolor* and *S. lividans* (Chen; 1993). This linear structure is similar to the linear plasmids with two identical free ends each carrying a covalently bound protein and a 25 kb inverted repeat. The linear chromosome contains, in the middle of it, a functional *oriC* (Zakrzewska-Czerwinska *et al.*, 1992) and has been shown to replicate by this mechanism following forced circularisation (Lin *et al.*, 1993) and therefore is able to exist, in a

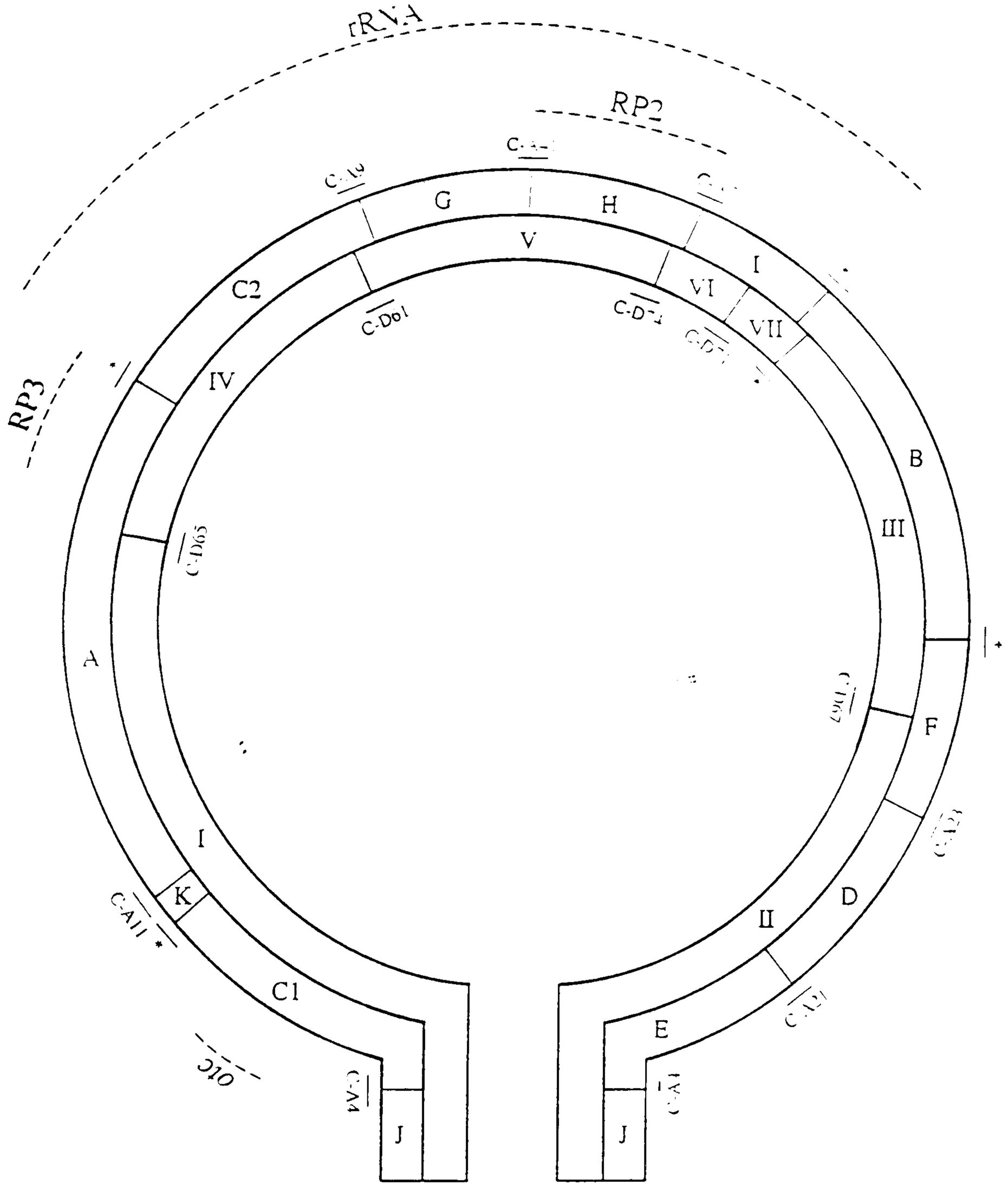


Figure 6.15
A schematic representation of the linear chromosome of *S. rimosus* R6. The construction of the physical map employed restriction digests of *Ase*I and *Dra*I and pulse-field electrophoresis (Hranueli *et al.*, unpublished).

viable cell, as linear or a circular molecule. Physical mapping of the *S. lividans* 66 chromosome located both of the telomeres in the 'unstable region' where large deletions (up to 2 Mb) and extensive amplifications frequently occur (Redenbach *et al.*, 1993). Lin *et al.*, (1993) speculated that large deletions may remove the DNA together with the terminal proteins from both telomeres, thereby forcing the organism to fuse the linear chromosomal DNA into a circle and this was confirmed by hybridisation analysis. To carry this further, the location of the *otc* cluster has been provisionally mapped very close to one end of the chromosome (see figure 6.15; D. Hranueli; pers. comm.). A more definite location might have been assigned employing the Int16 clone. It carries, within the *gmr* cassette, a series of AT-rich restriction sites. However with the loss of the *otrB* end of the cluster, comparative restriction mapping is not possible. The orientation of the cluster can however be assigned, with the *otrB* gene closest to the telomere. Further investigation is needed to confirm whether the genome of Int16 clone is linear and the replacement clones, which are sensitive to OTC, are circular. The organism, in its struggle to survive may have deleted the dispensable DNA along with the terminal proteins and re-circularised to form a stable chromosome with the minimal DNA requirements for survival.

The location and orientation of the *otc* cluster is a stroke of misfortune. It appears that the cluster is located between two large repeat sequences that has attached itself to the end of the genome, not unlike a transposon (J. Cullum; pers. comm.). The consistent genetic instability makes it very difficult to integrate into the cluster without subsequent loss of some of the genes. Of over 1500 colonies patched, and more than 50 chromosomal DNA preparations, only Int16 appeared not to be amplified. If the orientation of the cluster had been reversed and the *otrA* gene closest to the telomere, the antibiotic minocycline could have been added to the growing cultures to prevent the loss of the cluster. Minocycline acts on the ribosomes. The *otrA* resistance gene mediates resistance *via* non-covalent ribosomal modification (Doyle *et al.*, 1991). The *otrB* resistance gene, on the other hand, acts as an exporter of OTC (McGregor; pers. comm.). Therefore, minocyline could have been incorporated into the media to add selective pressure to retain the *otrA* resistance gene, had the *otrA* gene been closest to the telomere. The organism would have been forced into retaining the entire cluster, but as it is, the addition of minocycline would have no affect on retaining the *otrB* end of the cluster.

It appears that integration into the *S. rimosus* chromosome to study the specific functions of genes in the *otc* cluster is not a viable option. The cluster has been cloned into the SCP2* plasmid, GLW101 (IS Hunter; pers. comm.). When transformed into *S. lividans* integration using suicide vectors is possible (Thamchaipenet; 1994). This is not an ideal environment to test possible shunt metabolites because metabolites produced could be either the result of disruption of the *otc* cluster or a reponse of *S. lividans* to the presence of a disrupted cluster. The next best situation would be to relocate the cluster within the *S. rimosus* genome using either transposons, or phage attachment sites. The attachment sites for RP2 and RP3 phage have been located on the genome and they are much closer to the centre of the chromosome (J. Cullum; pers. comm.)[figure 6.15]. Vectors with compatible phage attachment sites, which enable the plasmid carrying them to integrate into the chromosome, could be employed. Another possibility is to transform the plasmid containing the entire *otc* cluster into an OTC-sensitive mutant. This has been attempted but has been unsuccessful so far. The plasmid containing the cluster is very large and therefore difficult both to isolate and to transform. If this approach does prove to be successful, suicide plasmids could then be used to disrupt the cluster in the plasmid. The resulting shunt metabolites could then be analysed and identified chemically and gene functions assigned.

Chapter 7

Concluding Remarks

Concluding Remarks

The primary objective of the work described in this thesis was to determine the functions of genes encoded by part of the *otc* cluster from *S. rimosus*. A strategy for the integration and disruption of various gene products was also investigated. A 3.0 kb *Kpn*I-*Mlu*I region of the cluster was sequenced and its transcription analysed. The deduced function of one of the ORFs, *OtcD*-ORF3 (a putative reductase), was determined by similarity to other polyketide oxidoreductases. However, the precise function within the pathway was only speculative. Gene disruption and replacement experiments were therefore undertaken to help deduce it's function. These experiments were facilitated by development of a strategy for the introduction of DNA into *S. rimosus*.

Much of the *otc* cluster has been sequenced (figure 7.1). The functions of the ORFs have been deduced by sequence similarity with other gene products in the databases. A 3.0 kb region between the *otcX* and *otcD* loci was sequenced in this work. The overall G+C composition was 72.8% and three open reading frames were uncovered. FRAME (Bibb; 1984) and CODONPERENCE (Devereux *et al.*, 1984) programs revealed that all three ORFs were transcribed in the same direction. The first ORF, labelled *OtcD*-ORF3, encodes a protein 252 amino acids in length with a predicted molecular weight of 25,274 D. The *OtcD*-ORF3 gene product has an average 21.8% amino acid sequence identity and 43.6% amino acid sequence similarity with FAS-type ketoreductases, such as, NodG (*nod*G) from *Rhizobium meliloti* (Fisher *et al.*, 1987) and *E. coli* FabG(β-ketoacyl acyl carrier protein reductase) (Cronon *et al.*, 1992). A much greater similarity to type II PKS ketoreductases was seen, including *OtcY*-ORF1 (Thamchaipenet; 1994), DauB (Ye *et al.*, 1994), ActIII (Hallam *et al.*, 1988), Gra-ORF5 and Gra-ORF6 (Sherman *et al.*, 1989) as shown in figure 3.5. The function of the ActIII gene product has been determined genetically, (Bartel *et al.*, 1990), and shown to reduce a keto group at position C-9 from the carboxy terminus of the actinorhodin molecule. The sequence similarity suggests that the function of *OtcY*-ORF1 is similar to that of actIII, performing the reduction at position C-9 in the OTC molecule. In these amino acid sequences there is a consensus NADP(H) binding motif, Gly-X-Gly-X-X-Ala (where X is any amino acid) (Scrutton *et al.*, 1990). This was proposed as a consensus NADP(H) binding motif in putative ketoreductases (Hopwood and Sherman; 1990). However, this motif is not conserved at the same position in other reductases. It was proposed that another conserved motif, Gly-3(X)-Gly-X-

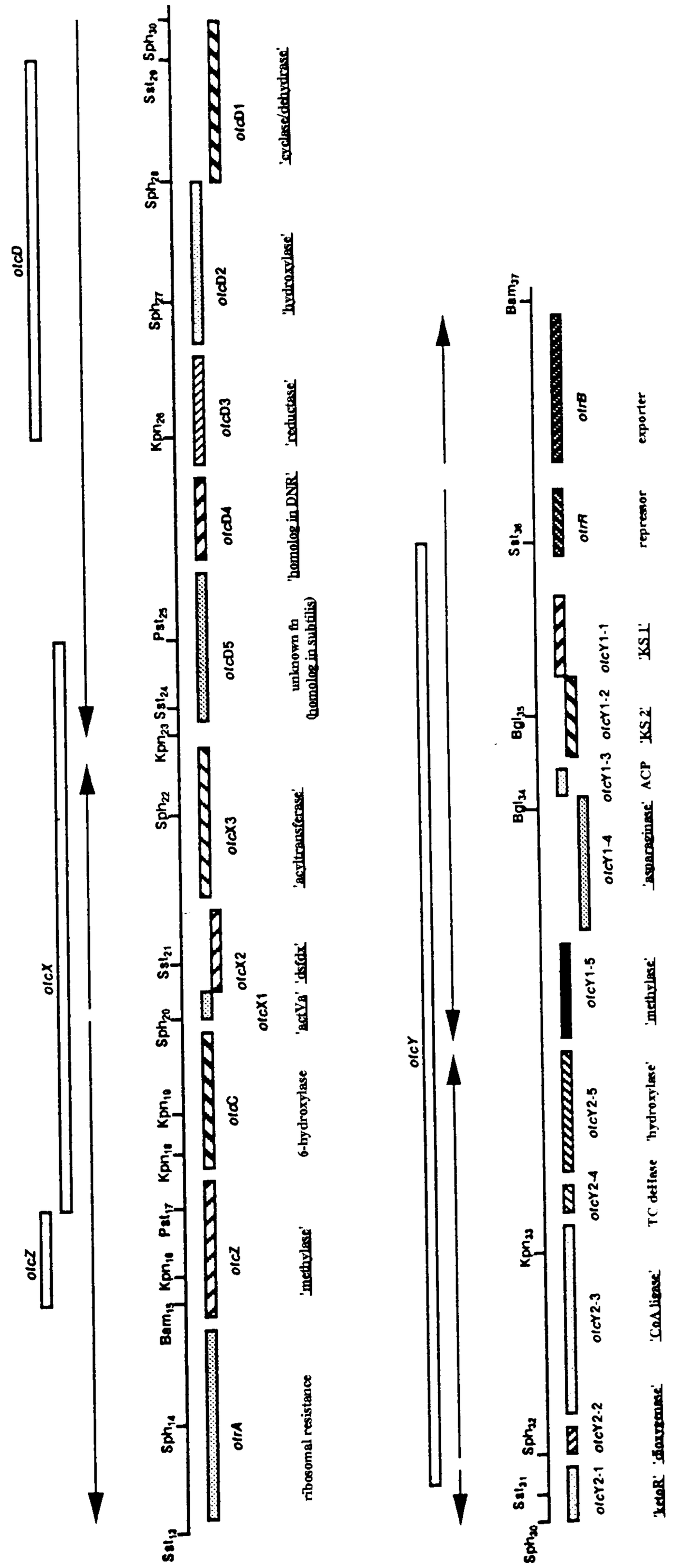


Figure 7.1
Overall architecture and 'putative' gene functions of the oxytetracycline cluster in *S. rimosus*.

Gly-3(X)-Ala-6(X)-Gly, which was conserved near the N-terminus of the putative protein was likely to be an NADPH binding motif (Thamchaipenet; 1994). This consensus sequence has been found in all the aligned sequences (figure 3.6). In the biosynthesis of OTC, three oxidoreductase steps are required. One, already mentioned above, at position C-9, one in the series of steps which ultimately convert 4-keto-ATC to ATC and one at the last step of the pathway, OTC dehydrogenase. The *OtcD*-ORF3 gene product has tentatively been assigned to the reduction step at position C4, see figure 1.10, that is involved in the conversion of 4-keto-ATC to ATC however, gene disruption experiments would help to confirm this function.

The second ORF, *OtcD*-ORF4 encodes a protein of 258 amino acids with a predicted molecular weight of 28,571 D. It shows good end-to-end similarity, (79.7%), with ORF1 in the *lkmB* region of the gene cluster for biosynthesis of daunorubicin from *S. griseus* (Krügel *et al.*, 1993). Mutations in this region of the *S. griseus* cluster were confirmed by complementation analysis of a gene product found early in the biosynthetic pathway. However no matches with genes of known function were found in database searches and therefore the function of the gene product could not be deduced. The final ORF, *OtcD*-ORF5, encodes a protein of 353 amino acids with a predicted molecular weight of 38,113 D. The composition of amino acids that make up this gene product is unique. It is a very hydrophobic molecule and carries a series of repeat sequences, with a consensus of LRGAD, close to the C-terminus. The significance of this region is not readily apparent but the evenly-spaced leucine residues and a secondary structure of a helix-turn-helix in the protein suggests that it may be involved in bending of DNA or binding of protein to DNA (Csank *et. al.*, 1992). Further searches of the databases revealed a strong similarity to a *B. subtilis* gene product of unknown function. The region of similarity was confined solely to this LRGAD consensus region, adding credibility to the premise that the motif is important. Hybridisation studies using *OtcD*-ORF5 as a probe to DNA of *S. peucetius* and *S. galilaeus* produced positive results. The antibiotic clusters of both strains are currently being studied and as a result of the hybridisation results with *OtcD*-ORF5, comparative studies should reveal ORFs with substantial similarity.

Preliminary transcriptional analysis of the recently-sequenced region between the *Kpn*I$_{23}$ to *Pvu*II$_{26a}$ (figure 6.1) was undertaken with the premise that all three ORFs were part of the same polycistronic message originating from *OtcY2*-ORF1 (Thamchaipenet; 1994). Previous work by Thamchaipenet (1994)

showed that *OtcY2*-ORF1, *OtcD*-ORF1 and *OtcD*-ORF2 were all contained on the same message reading in the same direction and upstream of the three ORFs sequenced here. The 3.0 kb region was divided into three sections and mapped using low-resolution S1 protection assays. *OtcD*-ORF3 and *OtcD*-ORF4 were found to be transcribed in the same direction as the above mentioned ORFs and are transcribed from the same 6.2 knt polycistronic message. A very stable stemloop structure was found immediately downstream of the *OtcD*-ORF4 gene product and high resolution mapping confirmed that a mRNA does end in this region. *In vivo* assays are warranted to confirm that the terminator is active. This can easily be done using promoter probe-type constructs. The potential terminator is placed between a promoter and reporter gene and the level of termination is measured by the relative decrease in gene expression (Ingham *et al.*, 1994). The third ORF, *OtcD*-ORF5 did not produce a transcript under the conditions during which the RNA was isolated, 40 hours after inoculation. The *OtcD*-ORF5 region has proved to be very interesting. Sequence analysis shows a potential promoter region upstream with good agreement to the consenus sequences of the -10 and -35 regions of *Streptomyces*. There is no sequence similarity with any other *Streptomyces* gene products in the database. However, a potential variation on a leucine zipper motif cannot be ruled out. Leucine zippers have been associated with regulatory-type functions of proteins. An activator has not been located within the otc cluster to date, and following the general layout of other streptomyces antibiotic clusters, such as actII of the actinorhodin cluster (Fernández-Moreno *et al.*, 1991), one would expect to find one. A repressor that shows good similarity to multidrug resistance repressors in *E. coli,* has been located next to the *OtrB* resistance gene and the cluster could therefore be negatively-regulated. Most clusters contain more than one regulatory mechanism so it possible that there will be an activator as well. Further time-point analysis of the RNA will be required to determine if this gene is transcribed and then potential knock-out experiments can be performed to help determine its function.

Recombinant DNA technology has used *E.coli* as the primary host of manipulation of DNA. The reasons for the use of *E. coli* as the primary host arise from its rapid growth rate and simple techniques for isolation of DNA. *Streptomyces* molecular biology cannot benefit directly from these advances because of the restrictive nature of the hosts. Most *E. coli* strains tend to methylate adenine at the N^6 position in the sequence 5' GATC 3' and cytosine at the C5 or N^4 position in the sequence $CC^{A}/_{T}GG$ (Sambrook *et al.*, 1989). A

number of *Streptomyces* strains encode a restriction system which will digest DNA containing either 5-methylcytosine or N^6-methyladenine. MacNeil *et al.*, (1988), investigated the restriction systems of a number of *Streptomyces* strains, with particular emphasis on *S. avermetilis*. Of the nine *Streptomyces* spp. tested, seven strains had methyl-dependent restriction systems, including *S. coelicolor* and *S. lividans*. Only *S. avermetilis* restricted DNA modified at either N^6meA or 5meC. The other methyl-dependent strains restricted DNA methylated only at one site. *S. lividans*, although restrictive to DNA methylated at N^6meA, has a very weak restriction barrier and can accept DNA derived from *E.coli* with little affect on transformation efficiencies and therefore has become a popular *Streptomyces* host. Three *E.coli* hosts were used to investigate the *S. rimosus* restriction system; ET12567, which produces unmethylated DNA, CB51, which produces DNA unmethylated at N^6-methyladenine sites and GM31, which produces unmethylated DNA at 5-methylcytosine sites. The efficiencies of transformation of *S. rimosus* with the DNA isolated from these three *E.coli* hosts were compared with DNA isolated from a *Streptomyces* host, *S. lividans*. DNA iolated from both ET12567 and GM31 showed comparable levels of transformants per μg DNA as DNA isolated from *Streptomyces* . This, along with the low transformation efficiencies of DNA isolated from CB51, indicated that *S. rimosus* contains a methylcytosine-specific restriction system. High- and low-copy number plasmids tested revealed a consistent 10^3 drop in efficiency for the low-copy number plasmids. This is not observed with *S. lividans* or another *S. rimosus* strain R6 (J. Pigac; pers. comm.). No explanation is readily available but this precludes the use of low-copy number plasmids in any integration strategies.

An increased efficiency in the transfer of DNA into the host *S. rimosus*, was desirable before any attempts at chromosomal integration could be made. Two different strategies were employed. The first involved conjugal transfer. A series of constructs were made which contained all the requirements for successful conjugation. The *tra* locus which encodes the cell surface components as well as mobilisation proteins (*mob* gene products), promotes DNA transfer by interaction with a *cis*-acting sequence known as the origin of transfer (*oriT*) (Willetts *et al.*, 1984). Conjugation has been very successful in promoting gene exchanges between distantly-related species (i.e. *E. coli* to *Entercoccus, Bacillus, Listeria Streptococcus,* and *Staphlococcus,* Trieu-Cuot *et al.*, 1987). Of particular interest was the demonstration of conjugation between *E. coli* and *Streptomyces* (Mazodier *et al.*, 1989; Bierman *et al.*, 1992). An important factor in this success is likely to be the transfer of single-stranded DNA during conjugation. Single-stranded DNA is

known to be refractory to most restriction systems (Hillemann *et al.*, 1991), thus the transfer of single-stranded DNA during bacterial conjugation may escape digestion by restriction systems of recipients. Successful conjugal transfer from *E. coli* to a relatively weakly-restricting *Streptomyces* host has been demonstrated with *S. lividans*. The more restricting nature of *S. rimosus*, which actively degrades hemi-methylated DNA, proved that conjugation was not a viable option. The second strategy involved improving the existing transformation procedure. The advent of the electrotransformation technology has provided a good option for increased efficiencies of transformation. The protocol was developed by J. Pigac and H. Schrempf (1995), and provided a much more efficient method of obtaining transformants. The increased transformation efficiency in itself was not the most important advance, but the shortened incubation times for regeneration of transformed cells has improved recombinogenic methods.

Many methods for gene disruption have been employed for the numerous *Streptomyces* spp. It appears that each has variations to accomodate the specific organism in question. The same proved to be true with *S. rimosus*. A series of constructs was made to introduce a 1.7 kb gentamicin-resistance marker into the $KpnI_{26}$ site of the *otc* cluster. This disrupted the *OtcD*-ORF3 gene product, a putative reductase gene. The construct was introduced on an unstable replicon, pIJ486. Preliminary evidence indicated that the plasmid pIJ486 could be lost through a single round of liquid non-selection at 89%. The resulting integrants were analysed by hybridisation and found in many cases to contain amplified copies of the *otc* cluster. Further analysis revealed that the amplifications were also accompanied by deletions of one end of the cluster. *OtrB* used as a probe did not produce a signal whereas *otrA* used as a probe did. The integration event did take place as the *gmr* gene was found in the *S. rimosus* chromosome. However, the loss of a section of the cluster was also found. The *otc* cluster has recently been located on the linear chromosome of *S. rimosus*. It is located close to one end and is situated between large direct repeats. Amplified sequences have been observed in other *Streptomyces* strains including *S. lividans*. These unstable amplified sequences are frequently accompanied by large deletions. The deletions do not disrupt the viability of the cell and are therefore thought to encompass predominately genes for secondary metabolism. This is in good agreement with the observations made for the chromosome*S. rimosus*. The integration strategy in itself is viable. However the *otc* cluster must be relocated in the chromosome before specific gene disruptants can be obtained. The 34 kb

otc cluster has been cloned into the SCP2* plasmid and this could be introduced into a OTC$^-$ *S. rimosus* strain. Attempts have been made to isolate and transform this plasmid, but its size has proved to be very cumbersome. A second option is to introduce the cluster into the chromosome on a plasmid containing phage attachment sites. Two phage attachment sites have been located towards the middle of the *S. rimosus* chromosome and this may increase the chances of integrated single copies.

The resulting disruptions will then be replaced with a copy of the gene product containing a specific 'in-frame' mutation which renders the product non-functional and at the same time does not disturb transcription of any downstream genes. The shunt metabolites produced can then be chemically analysed to establish definitive gene functions.

The number of natural antibiotics used in human therapy has decreased steadily as the number of resistant bacteria has grown. Antibiotics are a fundamental tool in the fight against disease and new strategies in this battle are needed. One approach is to produce hybrid antibiotics. A number of antibiotic clusters have already been elucidated and as this number grows and gene functions defined, a concerted effort can be made to 'mix and match' gene functions of a number of related antibiotics to produce a whole new range of genetically-engineered products. The work presented here is another step towards this goal, using *otc* as the model biosynthetic pathway, to help in the battle against bacterial infection.

Bibliography

Altenbuchner J and Cullum J. (1984). DNA amplifications and an unstable arginine gene in *Streptomyces lividans* 66. Mol. Gen. Genet. **210:** 192-197.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. (1990). Basic local alignment search tool. J. Mol. Biol. **215:** 403-410.

Anzai H, Murakami T, Imai S, Satoh A, Nagaoka K and Thompson CJ. (1987). Transcriptional regulation of bialaphos biosynthesis in *Streptomyces hygroscopicus*. J. Bacteriol. **169:** 3482-3488.

Aretz W, Koller K-P and Riess G. (1989). Proteolytic enzymes from recombinant *Streptomyces lividans* TK24. FEMS Microbiol. **65:** 31-36.

Bailey CR and Winstanley DJ. (1986). Inhibition of restriction in *Streptomyces clavuligerus.* by heat treatment. J. Gen. Microbiol. **132:** 2945-2947.

Bartel PL, Conners NC and Strohl W.R. (1990a). Biosynthesis of anthracyclines: analysis of mutants of *Streptomyces* sp. strain C5 blocked in daunomycin biosynthesis. J. Gen. Microbiol. **136:** 1877-1886.

Bartel PL, Zhu C-B, Lampel JS, Dosch DC, Conners NC, Strohl WR, Beale JM,Jr. and Floss HG. (1990b). Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions. J. Bacteriol. **172:** 4816-4826.

Behal V andHunter IS. (1994). Tetracyclines. In: *Genetics and Biochemistry of Antibiotic Production*. Edited by Vining, L. Acadamic Press. (in press).

Behal V, Hostalek Z and Vanek Z. (1979). Anhydrotetracycline oxygenase activity and biosynthesis of tetracycline in *Streptomyces aureofaciens*. Biotech. Lett. **1:** 177-193.

Betzler M, Dyson P and Schrempf H. (1987). Relationship of an unstable *argG* gene to a 5.7-Kilobase amplifiable DNA sequence in *Streptomyces lividans* 66. J. Bacteriol. **169:** 4804-4810.

Bibb MJ, Bibb MJ, Ward JM and Cohen SN. (1985). Nucleotide sequences encoding and promoting expression of three antibiotic resistance genes indigenous to *Streptomyces*. Mol. Gen. Genet. **199:** 26-36.

Bibb MJ, Biro S, Motamedi H, Collins JF and Hutchinson CR. (1989). Analysis of the nucleotide sequence of the *Streptomyces glaucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. EMBO J. **8:** 2727-2736.

Bibb MJ, Findlay PR, Johnson MW. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-codoning sequences. Gene. **30:** 157-166.

Bibb MJ, Janssen GR and Ward JM. (1986). Cloning and analysis of the promoter region of the erythromycin-resistance gene (*ermE*) of *Streptomyces erythraeus*. Gene **41:** E357-E368.

Bickle TA and Kruger DH. (1993). Biology of DNA restriction. Microbiol. Reviews. **57:** 434-450.

Bierman M, Logan R, O'Brien ET, Seno R, Nagaraja R and Schoner BE. (1992). Plasmid cloning vectors for the transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene **116:** 43-49.

Binnie C, Warren M and Butler MJ. (1989) Cloning and heterologous expression in *Streptomyces lividans* of *Streptomyces rimosus* genes invloved in oxytetracycline biosynthesis J. Bacteriol. **171:** 887-895.

Birch A, Häusler A and Hütter R. (1990). Genome instability in *Streptomyces* spp. J. Bacteriol. **172:** 4138-4142.

Birch AJ. (1967). Biosynthesis of polyketides and related compounds. Science **156:** 202-206.

Birnboim HC and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nuc. Acid Res. **7:** 1513-1523.

Blanco G, Brian P, Pereda A, Méndez C, Salas JA and Chater KF. (1993). Hybridization and DNA sequence analyses suggest an early evolutionary divergence of related biosynthetic gene sets for polyketide antibiotics and spore pigments in *Streptomyces* spp. Gene **130:** 107-116.

Blanco G, Pereda A, Méndez C and Salas JA (1992). Cloning and disruption of a fragment of *Streptomyces halstedii* DNA involved in the biosynthesis of spore pigment. Gene **112:** 59-65.

Bonnassie S, Burini J-F, Oreglia J, Trautwetter A, Patte J-C and Sicard AM. (1990). Transfer of plasmid DNA to *Brevibacterium lactofermentum* by electrotransformation. J. Gen. Microbiol. **136:** 2107-2112.

Borer PN, Dengler B, Tinoco I. Jr. and Uhlenbeck OC. (1974). Stability of ribonucleic acid double stranded helices. J. Mol. Biol. **86:** 843-853.

Boyer H. (1971). DNA restriction and modification mechanisms in bacteria. Annu. Rev. Microbiol. **25:** 153-176.

Brana AF, Méndez C, Diaz LA, Manzanal MB and Hardisson C. (1986). Glycogen and trehalose accumulation during colony development in *Streptomyces antibioticus*. J. Gen. Microbiol. **132:** 1319-1326.

BuchananWollaston V, Passiatore JE and Cannon F. (1987). The *mob* and *oriT* mobilizaton functions of a bacterial plasmid promote its transfer to plants. Nature **328:** 172-175.

Burgess RR, Travers AA, Dunn JJ and Baultz EKF. (1969). Factor stimulating transcription by RNA polymerase. Nature **221:** 43-46.

Butler MJ, Binnie C, Hunter IS, Sugden DA and Warren M. (1990). Genetic manipulation of the oxytetracycline biosynthetic pathway genes. In: *Developments in Industrial Microbiology,* pp 41-50 31. Edited by Pierce, G. E. Society for Industrial Microbiology.

Butler MJ, Friend EJ, Hunter IS, Kaczmarck FS, Sugden DA and Warren M. (1989). Molecular cloning of resistance genes and architecture of a linked gene cluster involved in the biosynthesis of oxytetraclycline by *Streptomyces rimosus.* Mol. Gen. Genet. **215:** 231-238.

Buttner MJ, Chater KF and Bibb MJ. (1990). Cloning, disruption and transcriptional analysis of 3 RNA-polymerase sigma-factor genes of *Streptomyces coelicolor* A3(2). J. Bacteriol. **172:** 3367-3378.

Buttner MJ. (1989). RNA polymerase heterogeneity in *Streptomyces ceolicolor* A3(2). Mol. Microbiol. **3:** 1653-1659.

Caballero J. Martinez E, Malpartida F and Hopwood DA. (1991). Organization and functions of the *actVA* region of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor*. Mol. Gen. Genet. **230**: 401-412.

Calzone FJ, Britton RJ and Davidson EH. (1987). Mapping of gene transcripts by nuclease protection assays and cDNA primer extension. In: *Guide to Molecular Cloning Techniques*, pp. 611-632. Edited by Berger SL, and Kimmel AR. Academic Press, London.

Chang S, Chang SY and Gray O. (1987). Structural and genetic analyses of a *par* locus that regulates plasmid partition in *Bacillus subtilus.* J. Bacteriol. **169:** 3952-3962.

Chater KF and Bruton CJ. (1983). Mutational cloning in *Streptomyces* and the isolation of antibiotic production genes. Gene **26:** 67-78.

Chater KF and Bruton CJ. (1985). Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. EMBO J. **4:** 1893-1897.

Chater KF and Hopwood DA. (1989a). Diversity of bacterial genetics. In: *Genetics of Bacterial Diversity*, p. 24-52. Edited by DA Hopwood and KF Chater, Academic Press, Inc. (London), Ltd., London.

Chater KF, Bruton CJ, Plaskitt KA, Buttner MJ, Mendez C and Helmann JD. (1989b). The developmental fate of *Streptomyces ceolicolor* hyphae depends

crucially on a gene product homologous with the motility factor of *Bacillus subtilus*. Cell **59:** 133-143.

Chater KF, Bruton CJ, Plaskitt KA, Buttner MJ, Mendez C and Helmann JD. (1989c). The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility sigma factor of *B. subtilis*. Cell **59:** 133-143.

Chater KF. (1972). A morphological and genetic mapping study of white colony mutants of *Streptomyces ceolicolor* . J. Gen. Microbiol. **72:** 9-28.

Chater KF. (1989a). Multilevel regulation of *Streptomyces* differentiation. Trends in Genet. **59:** 372-377.

Chater KF. (1989b). In: *Regulation of Procaryotic Development*, p. 277-299. Edited by I. Smith, R. Slepecky and P. Setlow. American Society for Microbiology.

Chater KF. (1990). The improving prospects for yield increase by genetic engineering in antibiotic-producing *Streptomyces*. Bio/Technol. **8:** 115-121.

Chater KF. (1993). Genetics of differentiation in *Streptomyces* Annu. Rev. Microbiol. **47:** 685-713.

Chen C, Lin Y-S, Kieser HM and Hopwood DA. (1993). The chromosomal DNA of *Streptomyces lividans* 66 appears to be linear. J. Cell Biochem Suppl **17E:** 291.

Chopra I, Hawkey PM and Hinton M. (1992). Tetracyclines, molecular and clinical aspects. J. Antimicrob. Chem. **29:** 245-277.

Chopra I. (1985). Mode of action of the tetracyclines and the nature of bacterial resistance to them. In: *The Tetracyclines*, pp. 317-392. Edited by Hlavka JJ and Boothe, JH. Springer-Verlag, Berlin .

Clark J. (1988). Novel non-templated nucleotide addition reactions catalysed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. **16:** 9677-9686.

Cohen SN and Chang ACY. (1973). Recircularization and autonomous replication of a sheared R-factor DNA segment in *E. coli* transformants. Proc. Nat. Acad. Sci. USA. **70:** 1293-1297.

Cohen SN and Chang ACY. (1977). Revised interpretation of the origin of the pSC101 plasmid. J. Bacteriol. **132:** 734-737.

Cole SP, Rudd BAM, Hopwood DA, Chang CJ and Floss HG. (1987). Biosynthesis of the antibiotic actinorhodin. Analysis of blocked mutants of *Streptomyces coelicolor*. J. Antibiot. **40:** 340-347.

Cortes J, Haydock SF, Roberts GA, Bevitt DJ and Leadley PF. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature. **348**: 176-178.

Coulson A. (1994). High-performance searching of biosequence databases. Tibtech **12:** 76-80.

Cronon JE. Jr. and Rawlings M. (1992). The gene encoding *Escherichia coli* acyl carrier protein lies within a cluster of fatty acid biosynthetic genes. J. Biol. Chem. **267:** 5751-5754.

Crueger W and Crueger A. (1989). In: Biotechnology, A textbook of Industrial Microbiology. Second edition. Science Technology Publishers, Madison Ws.

Csank C and Martindale DW. (1992). Isoleucyl-transfer RNA-synthetase from the ciliated protozoan *Tetrahymena-thermophila* - DNA sequence, gene-regulation, and leucine zipper motifs. Mol. Cell. Biol **267:** 1719-1729.

Daniels L, Bakheit N and Herman K. (1985). Systematic and Applied Microbiology **6:** 12-17.

Deng Z, Kieser T and Hopwood DA. (1987). Activity of *Streptomyces* transcriptional terminator in *Escherichia coli.* Nucl. Acids Res. **15:** 2665-2675.

Deng Z, Kieser T and Hopwood DA. (1988). "Strong incompatibility" between derivatives of the *Streptomyces* multi-copy plasmid pIJ101. Mol. Gen. Genet. **214:** 286-294.

Dente L, Cesareni G and Cortese R. (1983). pEMBL- A new family of single stranded plasmids. Nuc. Acid Res. **11:** 1645-1655.

Devereux J, Haeberli P and Smithies O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nuc. Acid Res. **12:** 387-395.

Diver JM, Bryan LE and Sokol PA. (1990). Transformation of *Pseudomonas aeruginosa* by electroporation. Anal. Biochem. **189:** 75-79.

Donadio S, Staver MJ, McAlpine JB, Swanson SJ and Katz L. (1991). Modular organisation of genes required for complex polyketide biosynthesis. Science **252:** 675-679.

Donadio S, Staver MJ, McAlpine JB, Swanson SJ and Katz L. (1992). Biosynthesis of the erythromycin macrolactone and a rational approach for producing hybrid macrolides. Gene **115:** 97-103.

Dower WJ, Miller JF and Ragsdale CW. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16:** 6127-6145.

Doyle D, McDowell KJ, Butler MJ and Hunter IS. (1991). Characterization of an oxytetracycline-resistance gene, *otrA* of *Streptomyces rimosus* . Mol. Microbiol. 5: 2923-2933.

Dyson P and Schrempf H. (1987). Genetic instability and DNA amplifications in *Streptomyces lividans* 66. J. Bacteriol. **169:** 4796-4803.

Ekhardt K and Wagner C. (1988). Biosynthesis of anthracyclinones. J. Basic Microbiol. **28:** 137-144.

Enquist LW and Bradley SG. (1971). Characterisation of deoxyribonucleic acid from *Streptomyces venezulae* species. Dev. Ind. Microbiol. **12:** 2431-2441.

Errington J. (1991). Possible intermediate steps in the evolution of a prokaryotic developmental system. Proc. R. Soc. Lond. **B 224:** 117-121.

Fernández-Moreno MA, Caballero JL, Hopwood DA and Malpartida F. (1991). The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. Cell **66:** 769-780.

Fernández-Moreno MA, Martinez E, Boto L, Hopwood DA and Malpartida F. (1992). Nucleotide sequence and deduced functions of a set of co-transcribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. J. Biol. Chem. **174:** 2958-2967.

Fishman SE and Hershberger CL. (1983). Amplified DNA in *Streptomyces fradiae* J. Bacteriol. **155:** 459-466.

Gardella T, Moyle H and Susskind MM. (1989). A mutant *Escherichia coli* σ^{70} subunit of RNA polymerase with altered promoter activity. J. Mol. Biol. **206:** 579-590.

Gatenbeck S. (1961). The biosynthesis of oxytetracycline. Biochem. Biophys. Res. Comm. **6:** 422-426.

Gibson TJ. (1984). "Studies on the Ebstein-Barr virus genome". PhD Thesis, Cambridge University, UK.

Gold L, Pribnow D, Schneider T, Shinelding S, Swebilium BS and Stormo G. (1981). Translational initiation in prokaryotes. Ann. Rev. Microbiol. **35:** 365-403.

Gralla JD. (1991). Transcriptional Control- Lessons from an *Escherichia coli* promoter data-base. Cell **66:** 415-418.

Granozzi C, Billeta R, Passantino R, Sollazzo M and Puglia AM. (1990). A breakdown in macromolecular synthesis preceding differentiation in *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. **136:** 713-718.

Gravius B, Bezmalinovic T, Hranueli D and Cullum J. (1993). Genetic instability and strain degeneration in *Streptomyces rimosus*. Appl. Environ. Microbiol. 59: 220-2228.

Gross CA, Lonetto MA and Losick R. (1992). Bacterial Sigma Factors. In: *Transcriptional Regulation* , pp. 129-176. Edited by Yamamoto, K. and McKnight, S. (Cold Spring Harbor Lab. Press, Plainview, NY.

Gruss AD and Ehlrich SD. (1989). The family of highly interrelated single-stranded deoxyriboncleic acid plamids. Microbiol. Reviews **53:** 231-241.

Guijarro J, Santamaria R, Schauer A and Losick R. (1988). Promoter determining the timing and spatial localization of transcription of a cloned *Streptomycces coelicolor* gene encoding a spore-associated polypeptide. J. Bacteriol. **170:** 1895-1901.

Hallam SE, Malpartida F and Hopwood DA. (1988). Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor* . Gene **74:** 305-320.

Hanahan D. (1983). Studies on transformation of *E. coli* with plasmids. J. Mol. Biol. **166:** 557-580.

Hattman S. (1977). Partial purification of the *Escherichia coli mec+* deoxyribonucleic acid-cytosine methylase: *in vitro* methylation completely protects bacteriophage lambda deoxyribonucleic acid against cleavage by *Eco*RII. J. Bacteriol. **129:** 1330-1334.

Häusler A, Birch A, Krek W, Piret J and Hütter R. (1989). Heterogenous genomc amplificaton in *Streptomyces glaucescens* tend to carry amplifications and deletions. Mol. Gen. Genet. **217:** 437-446.

Hawley DK and McClure WR. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nuc. Acid. Res. **11**: 2237-2255.

Hawley DK and McClure WR. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucl. Acids Res. **11:** 2237-2255.

Hayakawa T, Tanaka T, Sakaguchi K, Otake N and Yonehara H. (1979). A linear-like plasmid DNA in *Streptomyces* sp. producing lankasidin group antibiotics. J. Gen. Appl. Microbiol. **25:** 255-260.

Heinemann JA and Sprague GF Jr. (1989). Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature **340:** 205-209.

Heitman J and Model P. (1987). Site-specific methylases induce the SOS DNA repair response in *Escherichia coli.* J. Bacteriol. **169:** 3243-3250.

Helmann JD and Chamberlin MJ. (1988). Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **42:** 839-872.

Hershberger CL, Arnold B, Larsen J, Skatrud P, Reynolds P, Szoke P, Rostreck PR, Swartling J and McGilvray D. (1989). Role of giant linear plasmids in the biosynthesis of macrolide and polyketide antibiotics. In: *Genetics and Molecular Biology of Industrial Microorganism* p. 147-155. Edited by CL Hershberger, SW Queener and G. Hegeman. American Society for Microbiology, Washington D.C.

Hillemann D, Pülher A and Wohlleben (1991). Gene disruption and gene replacement in *Streptomyces* via single stranded DNA transformation of integration vectors. Nucleic Acid Res. **19:** 727-731.

Hinnebusch J and Tilly K. (1993). Linear plasmids and chromosmes in bacteria. Mol. Micro. **10:** 917-922.

Hintermann G, Crameri R, Vögtli M and Hütter R. (1984). Streptomycin sensitivity in *Streptomyces glaucescens* is due to deletions comprising the structural gene coding for a specific phosphotransferase. Mol. Gen. Genet. **196:** 513-520.

Holmes DJ and Cundliffe E. (1991). Analysis of a ribosomal RNA methylase gene from *Streptomyces tenebrarius* which confers resistance to gentamicin. Mol. Gen. Genet. **229:** 229-237.

Hong S-K, Kito M, Beppu T and Horinouchi S. (1991). Phosphorylation of the AfsR product, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). J. Bacteriol. **173:** 2311-2318.

Hopwood DA and Sherman DH. (1990). Molecular genetics of polyketides and its comparision to fatty acid biosynthesis. Annu. Rev. Genet. **24**: 37-66.

Hopwood DA, Bibb MJ, Chater KF, Janssen GR, Malpartida F and Smith CP. (1986). Regulation of gene expression in antibiotic-producing *Streptomyces,* In: *Regulation of Gene Expression - 25 years on*, pp. 251-276. Edited by I.R. Booth and C. F. Higgins. Cambridge University Press.

Hopwood DA, Bibb MJ, Chater KF, Kieser T, Bruton CJ, Kieser HM, Lydiate DJ, Smith CP, Ward JM and Schrempf H. (1985). Genetic manipulation of *Streptomyces*: A laboratory manual. Norwich. John Innes Foundation.

Hopwood DA, Wildermuth H and Palmer HM. (1970). Mutants of *Streptomyces coelicolor* defective in sporulation. J. Gen. Microbiol. **61:** 397-408.

Horinouchi S and Beppu T. (1984). Production in large quantities of actinorhodin and undecylprodigiosin induced by *afsB* in *Streptomyces lividans*. Agric. Biol. Chem. **48:** 2131-2133.

Horinouchi S and Beppu T. (1992). Regulation of secondary metabolism and cell differentiation in *Streptomyces*: A-factor as a microbial hormone and the afsR protein as a component of a two-component regulatory system. Gene (Amst) **115:** 167-172.

Horinouchi S, Kito M, Nishiyama M, Furuya K, Hong S-K, Miyake K and Beppu T. (1990). Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2) and "*Streptomyces lividans* ". J. Bacteriol. **168:** 257-269.

Hornemann U, Otto CJ and Hoffmann GG. (1986). Abstracts from 5th International Symposium of the Genetics of Industrial Microorgisms. 1986 S4-P2, p.29.

Hunter IS and Friend EJ. (1984) Restriction deficient mutants of *Streptomyces rimosus* . Biochem. Soc. Trans. **12:** 643-644.

Ingham CJ, Hunter IS and Smith MCM. (1994). Rho-independent termination without 3' poly-U tails from the early region of actinophage ϕC31. Nucl. Acids Res. (in press)

Ingram C, Brawner M, Youngman P and Westpheling J. (1989). *xylE* functions as an efficient reporter gene in *Streptomyces* spp.: Use for the study of *galP*1, a catabolite-controlled promoter. J. Bacteriol. **171:** 6617-6624.

Itoh Z, Suzuki T Nakaya M, Inoue M and Mitsuhashi S. (1984). Gastrointestinal motor-stimulating activity of macrolide antibiotics and analysis of their side effects on the canine gut. Antimicrob. Agents Chemother. **26:** 863-869.

Janssen GR, Ward JM and Bibb MJ. (1989). Unusual transcriptional and translational features of the aminoglycoside phosphotransferase gene (*aph*) from *Streptomyces fradiae*. Genes Dev. **3:** 415-429.

Jaurin B and Cohen SN. (1984). *Streptomyces lividans* RNA polymerase recognises and uses*Escherichia coli* transcription signals. Gene **28:** 83-91.

Jensen KF, Bonekamp F and Poulsen P. (1986). Attenuation at nucleotide biosynthetic genes and amino acid biosynthetic operons of *Escherichia coli*. TIB 11: 362-365.

Joset F and Guespin-Michel J. (1993). Modification of the genetic information in: *Prokaryotic genetics genome organization, transfer and plasticity*, Blackwell Scientific Publications.

Kaiser D. (1989). Multicellular development in *Myxobacteria*. In: *Genetics of Bacterial Diversity*, pp. 243-263. Edited by DA Hopwood KF Chater, London: Academic.

Katz E, Thompson CJ and Hopwood DA. (1983). Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. J. Gen. Microbiol. **129:** 2703-2714.

Katz L and Donadio S. (1993). Polyketide Synthesis: Prospects for hybrid antibiotics. Annu. Rev. Micro. **47:** 875-912.

Kelleher JE and Raleigh EA. (1991). A novel activity in *Escherichia coli* K-12 that directs restriction of DNA modified at CG dinucleotides. J Bacteriol. **173:** 5220-5223.

Kendall KJ and Cohen SN. (1987). Plasmid transfer in *Streptomyces lividans*: identification of *kil-kor* system associated with transfer region of pIJ101. J. Bacteriol. **169:** 4177-4183.

Kendall KL and Cohen SN. (1988). Complete nucleotide sequence of the *Streptomyces lividans* plasmid pIJ101 and correlation of the sequence with genetic properties. J. Bacteriol. **170:** 4634-4651.

Khosla C, Ebert-Khosla S, and Hopwood DA. (1992). Targeted gene replacements in a *Streptomyces* polyketide synthase gene cluster: role for the acyl carrier protein. Mol. Microbiol. **6**(21): 3237-3249.

Khosla C, McDaniel R, Ebert-Khosla S, Torres R, Sherman DH, Bibb MJ and Hopwood DA. (1993). Genetic construction and functional analysis of hybrid polyketide synthases containing heterologous acyl carrier proteins. J. Bacteriol. **175:** 2197-2204.

Kieser T, Hopwood DA, Wright HM and Thompson CJ. (1982). pIJ101, a multicopy Broad Host range *Streptomyces* plasmid: Functional analysis and delevopment of DNA cloning vectors. Mol. Gen. Genet. **185:** 223-238.

Kieser T. and Melton RE (1988). Plasmid pIJ699, a multicopy, positive-selection vector for *Streptomyces* . Gene **65:** 83-91.

Kim E-S, Bibb MJ, Butler J, Hopwood DA and Sherman DH. (1994). Sequences of the oxytetracycline polyketide synthase-encoding *otc* genes from *Streptomyces rimosus* . Gene **141:** 141-142.

Kirby D and Hopwood DA. (1977). Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. 98: 239-252.

Kirby KS, Fox-Carter E and Guest M. (1967). Isolation of deoxyribonucleic acid and ribosomal ribonucleic acid from bacteria. Biochem. J. **104:** 258-262.

Kitamura I, Tobe H, Yoshimoto A, Oki T, Nagahawa H, Takeuchi T and Umezawa H. (1981). Biosynthesis of aklavinone and aclacinomycins. J. Antibiot. 34: 959-964.

Kreps S, Ferino F, Mosin C, Gerits J and Mergeay M. (1990). Conjugative transfer and autonomous replication of a promiscuous IncQ plasmid in the cyanobacterium *Synechocystis* PCC 6803. Mol. Gen. Genet. **221:** 129-133.

Krügel H, Schumann G, Hänel F and Fiedler G. (1993). Nucleotide sequence analysis of five putative *Streptomyces griseus* genes, one of which complements an early function in dauorubicin biosynthesis that is linked to a putative gene cluster involved in TDP-daunosamine formation. Mol. Gen. Genet. **241:** 193-202.

Lauster R, Trautner TA and Noyer-Weidner M. (1989). Cytosine specific type II DNA methyltransferases. A conserved enzyme core with variable target recognition domains. J. Mol. Biol. **206:** 305-312.

Lawlor EJ, Baylis HA and Chater KF. (1987). Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor*. Genes and Development **1:** 1305-1310.

Leblond P, Demuyter P, Simonet J-M and Decaris B. (1991). Genetic Instability and Associated Genome Plasticity in *Streptomyces ambofaciens*: Pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region. J. Bateriol. **173:** 4229-4233.

Leskiw BK, Bibb MJ and Chater KF. (1991). The use of a rare codon specifically during development? Mol. Microbiol. **5:** 2861-2867.

Levy SB. (1984). Resistance to tetracyclines. In: *Antimicrobial Drug Resistance,* pp. 191-204. Edited by Bryan, LE. Academic Press, Orlando FL.

Lewin B. (1994). In: Genes V. pp. 241-243. Oxford University Press.

Lieb M and Bhagwat AS. (1988). Very short patch mismatch repair activity assaociated with gene *dcm* is not conferred by a plasmid coding for *Eco*RII methylase. J. Bacteriol. **170:** 4967-4968.

Lieb M. (1987). Bacterial genes *mutL*, *mutS* and *dcm* Participate in repair of mismatches at 5-methylcytosine sites. J. Bacteriol. **169:** 5241-5246.

Lin Y-S, Kieser HM,Hopwood DA and Chen C. (1993). The Chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol. Micro. **10:** 923-933.

Lonetto MA, Brown KL, Rudd KE and Buttner MJ. (1994). Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. Proc. Natl. Acad. Sci. **91:** 7573-7577.

Losick R and Pero J. (1981). Cascades of sigma factors. Cell **25:** 582-584.

Macaluso A and Mettus A-M. (1991). Efficient transformation of *Bacillus subtilus* requires nonmethylated plasmid DNA. J. Bacteriol. **173:** 1353-1356.

MacNeil DJ, Gewain KM, Ruby CL, Dazeny G, Gibbons PH and MacNeil T. (1992). Analysis of *Streptomyces avermetilus*. genes required for avermectin biosynthesis utilizing a novel integration vector. Gene **111:** 61-68.

MacNeil DJ. (1988). Characterization of a Unique methyl-specific restriction system in *Streptomyces avermetilus*. J. Bacteriol. **170:** 5607-5612.

Makino K, Shinagawa H, Amemura M and Nakata A. (1986). Nucleotide sequence of the *phoB* gene, the positive regulatory gene for the phosphate receptor of *Escherichia coli* K-12. J. Mol. Biol. **190**: 37-44.

Makino K, Shinagawa H, Amemura M, Kimura S, Nakata A and Ishihawa A. (1988). Regulation of the phosphate regulon of *Escherichia coli*: activation of *pstS* transcription by PhoB protein *in vitro*. J. Mol. Biol. **203**: 85-95.

Malek LT, Soostmeyer G, Davey CC, Krygsman P, Compton J, Gray J, Zimny T and Stewart D. (1990). Secretion of granulocyte macrophage colony stimulating factor (GM-CSF) in *Streptomyces lividans* . J. Cell. Biochem. Suppl. **14A:** 127.

Malpartida F and Hopwood DA. (1986). Physical and genetic characterisation od the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor*. Mol. Gen. Genet. **205:** 66-73.

Malpartida FM and Hopwood DA. (1984). Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotics and its expression in a heterologous host. Nature. **309**: 462-464.

Malpartida FM, Hallam SE, Kieser HM, Motamedi H, Hutchinson CR, Butler MJ, Sugden DA, Warren M, McKillop C, Bailey CR, Humphreys GO and Hopwood DA. (1987). Homology between *Streptomyces* genes coding for different polyketides used to clone antibiotic biosynthetic genes. Nature. **325**: 818-821.

Maniatis T, Fritsch EF, Sambrook J. (1989). In: *Molecular Cloning: "A Laboratory Manual"*. Vol II. Cold Spring Harbor Laboratory, New York.

Maplestone RA, Stone MJ and Williams DH. (1992). The evolutionary role of secondary metabolites-a review. Gene. **115**: 151-157.

Marinus MG and Morris NR. (1973). Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* . J. Bacteriol. **114**: 1143-1150.

Marinus MG and Morris NR. (1975). Pleiotopic effects of a DNA adenine methylation mutation (*dam* -3) in *Escherichia coli* K-12. Mutation Res. **28:** 15-26.

Marinus MG andPalmer BR. (1994). The *dam* and *dcm* strains of *Escherichia coli* - a review. Gene **143:** 1-12.

Matsushima P, Broughton MC, Turner JR and Baltz RH. (1994). Conjugal transfer of cosmid DNA from *Escherichia coli* to *Saccharopolyspora spinosa* : effects of chromosomal insertions on macrolide A83543 production. Gene (submitted).

Matsushima P. Cox KL and Baltz RH. (1987). Highly transformable mutants of *Streptomyces fradiae* defective in several restriction systems. Mol. Gen. Genet. 206: 393-400.

May MS and Hattman S. (1975). Analysis of Bacteriophage Deoxribonucliec Acid sequences Methylated by Host- and R-Factor -Controlled Enzymes. J. Bacteriol. 123: 768-770.

Mazodier P, Petter R and Thompson C. (1989). Intergeneric conjugation between *Escherichia coli* and *Streptomyces* Species. J. Bacteriol. 171: 3583-3585.

McCarthy AD and Hardie DG. (1984). Fatty acid synthase - an example of protein evolution by gene fusion. Trends Biochem. Sci. 9: 60-63.

McCormick JRD, Johnson S and Sjolander NO. (1963). Biosynthesis of the tetracyclines. V. Naphthacenic precursors. J. Am. Chem. Soc. 85: 1692-1694.

McCormick JRD, Miller PA, Johnson S, Arnold NH and Sjolander NO. (1962). Biosynthesis of the tetracyclines. IV. Biological rehydration of the 5a,6-anhydrotetracyclines. J. Am. Chem. Soc. 84: 3023-3024.

McCormick JRD. (1966). Biosynthesis of the tetracyclines: an integrated biosynthetic scheme. In *Antibiotics - Advances in Research, Production and Clinical Use,* pp. 556-573. Edited by M. Harold and Z. Gabriel. Prague: Czechoslovak Medical Press.

McDaniel R, Ebert-Khosla S, Hopwood DA and Khosla C. (1993). Engineering biosynthesis of novel polyketides: Manipulation and analysis of an aromatic polyketide synthase with unproven catalytic specificities. J. Am. Chem. Soc. 115: 11671-11675.

McDowall KJ. (1991). "An investigation of the molecular genetics of oxytetracycline biosynthesis and resistance in *Streptomyces rimosus*." PhD Thesis, University of Glasgow. UK.

McGuire JC, Thomas MC, Pandey R, Toussaint M and White RJ. (1980). Biosynthesis of daunorubicin glucosides: analysis with blocked mutants. Adv. Biotechnol. 3: 117-122.

Meacock PM and Cohen SN. (1980). Partitioning of bacterial plasmids during cell division: a *cis*-acting locus that accomplishes stable plasmid inheritance. Cell 20: 93-103.

Miki T, Ebina Y, Kishi F and Nakazawa A. (1981). Organization of the *lexA* gene of *Escherichia coli* and nucleotide sequences of the regulatory region. Nucleic Acids Res. **9:** 2145-2147.

Miller CA, Tucker WT, Meacock PM, Gustafsson P and Cohen SN. (1983). Nucleotide sequence of the partition locus of *Escherichia coli* plasmid pSC101. Gene **24:** 309-315.

Miller PA, Hash JH, Lincks, M and Bohonos N. (1965). Biosynthesis of 5-hydroxy-tetracycline. Biochem. Biophys. Res. Com. 18: 325-331.

Modrich P. (1989). Methyl-directed DNA mismatch correction. J. Biol. Chem. **264:** 6597-6600.

Molnár I. (1994). Secretory production of homologous and heterologous proteins by recombinant *Streptomyces*: What has been accomplished? In: *Recombinant Microbes for Industrial and Agricultural Applications*, p. 81-103. Edited by Y. Murook and Imanaka Marcel Dekker, Inc.

Muth G, Nuβbaumer B, Wohlleben W and Pühler (1989). A vector system with temperature-sensitive replication for gene disruption and mutational cloning is streptomycetes. Mol. Gen. Genet. **219:** 341-348.

Narva KE and Feitelson JS. (1990). Nucleotide sequence and transcriptional analysis of the *redD* Locus of *Streptomyces ceolicolor* A3(2). J. Bacteriol. **172:** 326-333.

Neal RJ and Chater KF. (1991). Bidirectional promoter and terminator region bracket *mmr*, a resistance gene embedded in the Streptomyces coelicolor A3(2) gene cluster encoding methylenomycin production. Gene **100:** 75-83.

Ogasawara N, Nakai S and Yoshikawa H. (1994). Systematic sequencing of the 180 kilobase region of the *Bacillus subtilus* chromosome containing the replication origin. DNA Res. **1:** 1-14.

Ohnuki T, Katoh T, Imanaka T and Aiba S. (1985). Molecular cloning of tetracycline resistance genes from *Streptomyces rimosus* in *Streptomyces griseus* and characterization of the cloned genes. J. Bacteriol. **161:** 1010-1016.

Ohnuki T, Kotoh T, Imanaka T and Aiba S. (1985). Molecular cloning of tetracycline resistance genes from *S. rimosus* and *S. griseus* and characterisation of the cloned genes. J. Bacteriol. **161:** 1010-1016.

Oki T, Kitamura I, Matsuzawa Y, Shibamoto N, Ogasawara T, Yoshimoto A and Inui T. (1979). Antitumor anthracycline antibiotics, aclacinomycin A and analogs. J. Antibiot. **42:** 1567-1577.

Omura S. (1992). The expanding horizon for macrobial metabolites - a review. Gene **115:** 141-149.

Opperman T and Richardson JP. (1994). Phylogenetic analysis of sequences from diverse bacteria with homology to the *Escherichia coli rho* Gene. J. Bacteriol. 176: 5033-5043.

Parro V, Hopwood DA, Malpartida F and Mellado P. (1991). Transcription of genes involved in the earliest steps of actinorhodin biosynthesis in *Streptomyces coelicolor*. Nuc. Acids Res. 19: 2623-2627.

Patel M, GulloVP, Hegde VR, Horan AC, Gentile F, Marquez JA, Miller GH, Puar MS and Waitz JA. (1987a). A novel tetracycline from *Actinomadura brunnea*: fermentation, isolation and structure elucidation. J. Antibiotics. 40: 1408-1413.

Patel M, GulloVP, Hegde VR, Horan AC, Marquez JA, Vaughan R, Puar MS and Miller GH. (1987b). A new tetracycline antibiotic from *Dactylosporangium sp.*: fermentation, isolation and structure elucidation. J. Antibiotics. 40: 1414-1418.

Peoples OP and Sinskey AJ. (1989). Fine structure analysis of the *Zoogloea ramigera phbA-phbB* locus encoding β-ketothiolase and acetoacetyl-CoA reductase: nucleotide sequence of *phbB*. Mol. Microbiol. 3: 349-357.

Piepersberg W, Distler J, Ebert A, Hienzel P, Mansouri K, Mayer G and Pissowotzki K. (1988). Expression of genes for streptomycin biosynthesis. In: *Biology of Actinomycetes '88* p. 86-91. Edited by Y. Okami, T. Beppu and H. Ogawara, Japan Scientific Societies Press, Tokyo.

Pigac J, Vujaklua D, Toman Z, Gamulin V and Schrempf H. (1988). Structural instability of a bifunctional plasmid pZG1 and single-stranded DNA formation of *Streptomyces* . Plasmid **19:** 222-230.

Piret JM and Chater KF. (1985). Phage-mediated cloning of *bldA*, a region involved in *Streptomyces coelicolor* morphological development and its analysis by genetic complementation. J. Bacteriol. **163:** 965-972.

Powell IB, Achen MG, Hillier AJ and Davidson BE. (1988). A simple and rapid method for genetic transformation of lactic streptococci by electroporation. Appl. Environ. Microbiol. **54:** 665-660.

Pratt, T. (1994). Rho and RNA: models for recognition and response. Mol. Micro **11:** 983-990.

Qin Z, Peng K, Zhou X, Liang R, Zhou Q, Chen H, Hopwood DA, Kieser T and Deng Z. (1994). Development of a Gene Cloning System for *S. hydroscopicus* subspecies *yingchengensis*, A producer of three useful antifungal compounds, by elimination of three barriers to DNA transfer. J. Bacteriol. **176:** 2090-2095.

Raleigh EA and Wilson G. (1986). *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA 83: 9070-9074.

Raleigh EA, Trimarchi R and Revel H. (1989). Genetic and physical mapping of the *mrcA* (*rglA*) and *mcrB* (*rglB*) loci of *Escherichia coli* K-12. Genetics 122: 279-296.

Rawlings M and Cronan JE.Jr. (1992). The gene encoding *Escherichia coli* acyl carrier protein lies within a cluster of fatty acid biosynthetic genes. J. Biol. Chem. **267**: 5751-5754.

Redenbach M, Flett F, Piendl W, Glocker I, Rauland U, Wafzig O, Kliem R, Leblond P and Cullum J. (1993). The *Streptomyces lividans* 66 chromosome contains a 1 MB deletogenic region flanked by two amplifiable regions. Mol. Gen. Genet. **241:** 255-262.

Reynes J-P, Calmeb T, Procourt D and Tiraby, G. (1988). Cloning and expression in *Escherichia coli* and the nucleotide sequence of a tetracycline resistance gene from *Streptomyces rimosus*. J. Gen. Microbiol. **134:** 585-598.

Rhodes PM, Hunter IS, Friend EJ and Warren M. (1984). Recombinant DNA methods for the oxytetracycline producing *S. rimosus*. Biochem. Soc. Transac. 12: 586-587.

Rhodes PM, Winskill N, Friend EJ and Warren M. (1981). Biochemical and genetic characterisation of *Streptomyces rimosus* mutants impaired in oxytetracycline biosynthesis. J. Gen. Microbiol. **124:** 329-338.

Richardson MA, Kuhstross S, Huber MLB, Ford L, Godfrey O, Turner JR and Rao RN. (1990. Cloning of spiramycin biosynthetic genes and their use in constructing *Streptomyces ambofaciens* mutants defective in spiramycin biosynthesis. J. Bacteriol. **172:** 3790-3798.

Ringquist S and Smith CL. (1990). The *Escherichia coli* chromosome contains specfic, unmethylated *dam* and *dcm* sites. Proc. Natl. Acad. Sci. USA **89:** 4539-4543.

Roberts J. (1969). Termination factor for RNA synthesis. Nature **224:** 1168-1174.

Rudd BAM and Hopwood DA. (1979). Genetics of actinorhodin biosynthesis by *Streptomyces ceolicolor* A3(2). J. Gen. Microbiol. **114:** 35-43.

Ryan MJ, Lotvin JA, Strathy NK, Rosenfeld SA, Tilley BC, Fantini SE and Smolin DE. (1994). Molecular cloning and characteristion of the chlorotetracycline pathway. In: *Seventh International Symposium on the Genetics of Industrial Microorganisms*. GIM 94 Montreal Canada.

Sakaguchi K (1990). Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of Adeno-Type viruses. Microbiol. Reviews **54:** 66-74.

Salas M. (1992). Protein binding of DNA replication. Annu. Rev. Biochem **60:** 39-71.

Sambrook J, Fritsch EF and Maniatus T. (1989) *Molecular Cloning: A Laboratory Manual* 2nd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sànchez J, Barbés C, Hernandez A, de los Reyes Gavilàn C-R and Hardisson C. (1985). Restriction-modification systems in *Streptomyces antibioticus* . Can. J. Microbiol. **31:** 942-946.

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

Shen B and Hutchinson CR. (1993). Enzymatic-synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. Science. **262:** 1534-1540.

Sherman DH, Kim ES, Bibb MJ, Kieser HM, Bibb MJ and Hopwood DA. (1992). Functional replacement of genes for individual polyketide synthase components in *Streptomyces coelicolor* 3A(2) by heterologous genes from a different polyketide pathway. J. Bacteriol. **174:** 6184-6190.

Sherman DH, Malpartida F, Bibb MJ, Kieser HM, Bibb MJ and Hopwood DA. (1989). Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tü22. EMBO J. **8:** 2717-2725.

Siebenlist U. Simpson RB and Gilbert W. (1980). *E. coli* RNA polymerase interacts homologously with two different promoters. Cell **20:** 269-281.

Simon R, Priefer U and Pühler A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in gram negative bacteria. Bio/Tech. **1:** 784-791.

Skeggs PA, Holmes DJ and Cundliffe E. (1987). Cloning of aminoglycoside-resistant determinants from *Streptomyces tenebrarius* and comparison with related genes for other Actinomycetes J. Gen. Microbiol. **133:** 915-923.

Sladek TL, Nowak JA and Maniloff J. (1986). Mycoplasma restriction: identification of a new type of restriction specificity for DNA containing 5-methylcytosine. J. Bacteriol. **165:** 219-225.

Smith C. (1992). Gene disruption and sequence replacements in the glycerol operon of *Streptomyces ceolicolor* . Streptomyces Gene Disruption Workshop. Norwich.

Smith CP. (1985). Molecular biology of the glycerol utilisation operon os *Streptomyces coelicolor* A3(2). PhD thesis. University of East Anglia, Norwich.

Sohail A, Lieb M, Dar M and Bhagwat AS. (1990). A gene required for very short patch repair in *Escherichia coli* is adjacent to the DNA cytosine methylase gene. J. Bacteriol. **172:** 4214-4221.

Southern EM. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503-517.

Stefano JE and Gralla JD. (1982). Spacer mutations in the *lac*pS promoter. Proc. Nat. Acad. Sci. **79:** 1069-1072.

Stein DA, and Cohen SN. (1990). Mutational and functional analysis of the *korA* and *korB* gene products of *Streptomyces* plasmid pIJ101. Mol. Gen. Genet. **222:** 337-334.

Stein DA, Kendall KJ and Cohen SN. (1989). Identification abd analysis of transcriptional regulatory signals for the *kil* and *kor* loci of *Streptomyces* plasmid pIJ101. J. Bacteriol. **171:** 5768-5775.

Stock JB, Ninfa AJ and Stock AM. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. **53:** 450-490.

Stone MJ and Williams DH. (1992). On the evolution of functional secondary metabolites (natural product). Mol. Microbiol. **6:** 29-34.

Strauss D and Fleck W. (1975). Leukaemomycin, an antibiotic with antitumor activity. Z. Allg. Mikrobiol. **15:** 65-623.

Strohl W.R. (1992). Complilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res. **20**: 961-974.

Strohl WR, Bartel PL, Conners NC, Zhu C-B, Dosch DC, Beale JM. Jr. Floss HG, Stutzman-Engwall K, Otten SL and Hutchinson CR. (1989). Biosynthesis of natural and hybrid antibiotics by anthracycline-producing streptomycetes. In: *Genetics and Molecular Biology of Industrial Microorganisms,* pp. 68-84. Edited by CL. Hershberger, SW. Queener and G. Hegeman. Washington DC: American Society of Microbiology.

Stromo GD, Schneider TD and Gold LM. (1982). Characterization of translation initiation sites in *E. coli*. Nuc. Acid Res. **10:** 2971-2996.

Stutzman-Engwall KJ, Otten S and Hutchinson CR. (1992). Regulation of secondary metabolism in *Streptomyces* spp. and overproduction of daunorubicin in *Streptomyces peucetius*. J. Bacteriol. **174:** 144-154.

Summers RG, Wendt-Pienkowski E, Motamedi H and Hutchinson CR. (1992). Nucleotide sequence of the *tcmI-tcmIV* region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* and evidence that the *tcmN* gene encoded a mutlifunctional cyclase-dehydrogenase-*O*-methyltransferase. J. Bacteriol. **174:** 1810-1820.

Thamchaipenet A. (1994). Characterisation of Some Genes Involved in the Early Steps of the Biosynthetic Pathway for Oxytetracycline in *Streptomyces rimosus* PhD Thesis. University of Glasgow. Glasgow, UK.

Thomas R and Williams DJ. (1983). Oxytetracycline biosynthesis: mode of incorporation of [1-^{13}C]- and [1,2-$^{13}C_2$] -acetate. J. Chem. Soc., Chem. Comm. 128-130.

Thompson CJ and Gray GS. (1983). Nucleotide sequence of a streptomycetes aminoglycoside phosphotransferase gene and its relationship to phosphotransferases encoded by resistance plasmids. Proc. Natl. Acad. Sci. **80:** 5190-5194.

Thompson CJ, Ward MJ and Hopwood DA. (1982). Cloning of antibiotic resistance and nutritional genes in streptomycetes. J. Bacteriol. **151:** 668-677.

Tinoco I. Jr, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM and Gralla J. (1973). Improved estimation of secondary structure in Ribonucleic acids. Nature New Biol. **246:** 40-41.

Trieu-Cuot P, Carlier C, Martin P and Couvalin P. (1987). Plasmid transfer by conjugation from *Escherichia coli* to gram-positive bacteria. FEMS Microbiol. Lett. **48:** 289-294.

Trieu-Cuot P, Carlier C, Poyart-Salmeron C and Couvalin P. (1991). Shuttle vectors containing a multiple cloning site and a *lacZ* gene for conjugal transfer of DNA from *Escherichia coli* to Gram positive bacteria. Gene **102:** 99-104.

Vanek Z, Cudlin J, Blumauerova M, Hostalek Z, Podojiil M, Rehacek Z and Krumphanzl V. (1981). Physiology and pathophysiology of the production of excessive metabolites. Institute of Microbiology. Czechoslovak Academy of Sciences, Prague.

Vögtli M and Hütter R. (1987). Characterisation of the hydroxystreptomycin phosphotransferase gene (*sph*) of *Streptomyces glaucescens* : nucleotide sequence and promoter analysis. Mol. Gen. Genet. **208:** 195-203.

von Hippel PH and Yager TD. (1991). Transcription elongation and termination are competitive kinetic processes. Proc. Natl. Acad. Sci. U.S.A. **88:** 2307-

Wagner C, Ekhardt K, Schumann G, Ihn W and Tresselt D. (1984). Microbial transformation of aklanonic acid, a potential early intermediate in the biosynthesis of anthracyclines. J. Antibiot. **37:** 691-692.

Walsh C. (1986). Naturally occurring 5-deazaflavin coenzymes: Biological redox roles. Accts. Chem. Res. **19:** 216-221.

Wang I-K, Reeves C and Gaucher GM. (1990). Isolation and sequencing of a genomic DNA clone containing the 3' terminus of the 6-methylsalicylic acid polyketide synthetase gene of *penicillum urticae*. Can. J. Microbiol. 37: 86-95.

Ward JM, Janssen GR, Kieser T, Bibb MJ, Buttner MJ and Bibb MJ. (1986). Construction and characterisation f a series of multicopy promoter probe plasmid vectors for*Streptomyces* using aminoglycoside phosphotransferase gene from Tn5 as indicator. Mol. Gen. Genet. **203:** 468-478.

Weber JM and Losick R. (1989) The use of a chromosome integration vector to map erythromycin resistance and production genes in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) Gene **68:** 173-180.

Wildermuth H. (1970). Development and organization of the aerial mycelium in *Streptomyces coelicolor* . J. Gen. Microbiol. **69:** 353-364.

Willetts N and Wilkins B. (1984). Processing of plasmid DNA during bacterial conjugation. Microbiol. Reviews. **48:** 24-41.

Willey J, Santamaria R, Guijarro J, Geistlich M and Losick R. (1991). Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by *Streptomyces coelicolor*. Cell **65:** 641-650.

Wrigey-Jones CA, Richards H, Thomas CR and Ward JM. (1992). A method for plasmid copy number determination in recombinant *Streptomyces*. J. Microbiol. Methods **16:** 69-80.

Yannish-Perron C, Vieira J and Messing J. (1985). Improved M13 phage cloning vectors and host strain nucleotide sequences of the M13mp18 and mp19 vectors. Gene. **33:** 103-119.

Ye J, Dickens ML, Plater R, Li Y, Lawrence J and Strohl WR. (1994) Isolation and sequence analysis of genes encoding daunomycin polyketide synthase of *Streptomyces* sp. strain C5. J. Bacteriol. (submitted).

Yu T-W, Bibb MJ, Revill P and Hopwood DA. (1994). Cloning, sequencing and analysis of the griseusin polyketide synthase gene cluster from *Streptomyces griseus*. J. Bacteriol. **176:** 2627-2634.

Zakrzewska-Czerwinska J and Schrempf H. (1992). Characterisation of an autonomously replicating region from the *Streptomyces lividans* chromosome. J. Bacteriol. **147:** 2688-2693.

Zaman S, Radnedge L, Richards H and Ward JM. (1993b). Analysis of the site for second-stand initiation during replication of the *Streptomyces* plasmid pIJ101. J. Gen. Microbiol. **139:** 669-676.

Zaman S, Richards H and Ward JM. (1993a). Identification of the minimal replicon of the *Streptomyces* plasmid pIJ101. Plasmid **29:** 57-62.

Zengel JM, Arche RH and Lindahl L. (1984). The nucleotide sequence of the *Escherichia coli fus* gene, coding for elongation factor G. Nucl. Acids Res. **12:** 2181-2192.

Zuber P, Healy J, Carter, III HL, Cutting S, Moran, Jr. CP and Losick R. (1989). Mutation changing the specificity of an RNA polymerase sigma factor. J. Mol. Biol. **206:** 605-614.