Theses Digitisation:
https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given
An Investigation of the Molecular Genetics of Oxytetracycline Biosynthesis and Resistance in *Streptomyces rimosus*.

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

by

Kenneth John McDowall

Institute of Genetics
The University
Church Street
Glasgow, G11 5JS

January 1991
The research reported in this thesis is my own original work except where otherwise stated and has not been submitted for any other degree.

Dedicated to mum and dad.
## CONTENTS

**ABBREVIATIONS**

**ACKNOWLEDGEMENTS**

**SUMMARY**

**CHAPTER 1 INTRODUCTION**

1.1 Introduction to *Streptomyces* 1

1.2 *Streptomyces* Life Cycle 3

1.3 Developmental Biology of *Streptomyces coelicolor* 5

1.4 Medical and Commercial Interest in *Streptomyces* 8

1.5 Organisation of the Genes for Antibiotic Biosynthesis 9

1.6 Pathway-specific Regulatory Genes 12

1.7 Catabolite Regulation of Secondary Metabolism 14

1.8 Role of "Signal" Molecules in Controlling Development 21

1.9 RNA Polymerase Heterogeneity in *Streptomyces* 23

1.10 Polyketide Biosynthesis 26

1.11 Actinorhodin Biosynthesis 32

1.12 Organisation of Polyketide Synthase Genes 34

1.13 Module Hypothesis 36

1.14 Biosynthesis of Oxytetracycline 38

1.15 Scope of Present work 46

**CHAPTER 2 MATERIALS AND METHODS**

2.1 Bacterial Strains and Vectors 47

2.2 Microbiological Techniques and Standard Media 50

2.2.1 Standard Media for the Propagation of *E.coli* 50

2.2.2 Standard Media for the Propagation of *Streptomyces* 50
2.2.3 Sterilization 52
2.2.4 Preparation of *Streptomyces* Spore Suspensions 52
2.2.5 *E. coli* Growth Conditions 53
2.2.6 Growth of *Streptomyces* Mycelium 53
2.2.7 Quantification of *Streptomyces* Growth 53
2.2.8 Quantification of Oxytetracycline in Culture Supernatants 54
2.2.9 Antibiotics and Indicators 54
2.2.10 Preservation of *Streptomyces* and *E. coli* Strains 55
2.2.11 Introduction of Plasmid DNA into *E. coli* 55
2.2.12 Transfection of *E. coli* TGI with Bacteriophage M13 56
2.2.13 Introduction of Plasmid DNA into *Streptomyces* 57

2.3 Nucleic Acid Isolation and Manipulation 59
2.3.1 Isolation of Plasmid DNA from *E. coli* and *Streptomyces* 59
2.3.2 Isolation of Total RNA 61
2.3.3 Quantification of Nucleic Acids 63
2.3.4 Precipitation of Nucleic Acid using Ethanol or Isopropanol 64
2.3.5 Digestion of DNA with Restriction Endonucleases 64
2.3.6 Ligation of DNA Fragments 64
2.3.7 Removal of the 5’ Phosphate from Linearised DNA 64
2.3.8 Removal of Protein from Nucleic Acid Solutions using Organic Solvents 65
2.3.9 Preparation of Radio-Labelled Probes 65
2.3.9.1 Random Primed DNA Labelling Method 65
2.3.9.2 5’-End Labelling of Oligonucleotides 66
2.3.9.3 Sephadex G50 Column Chromatograph 66
2.3.10 Agarose Gel Electrophoresis 66
2.3.10.1 Mini Gels 67
2.2.10.2 Large Gels 67
2.3.10.3 Photograph of Resolved Nucleic Acid 68
2.3.11 Denaturing Polyacrylamide Gel Electrophoresis 68
2.3.12 Recovery of DNA from Agarose Gels 69
2.3.12.1 GeneClean™ 69
2.3.12.2 Elecro-Elution 70
2.3.13 Isolation of DNA from Polyacrylamide Gels 70
2.3.14 E.coli Colony Transfer to Amersham Hybond-N™ Membrane 71
2.3.15 Southern Analysis 72
2.3.15.1 Capillary Transfer 72
2.3.15.2 Aqueous Hybridisation Conditions 72
2.3.15.2a Using Random Primed Probes 72
2.3.15.2b Using Oligonucleotides 73
2.3.16 Preparation of Single-Stranded M13 DNA 74
2.3.17 DNA Sequencing 75
2.3.18 Transcript Mapping using Complementary Single-Stranded DNA Probes 76
2.3.19 Primer Extention Mapping of Transcript 5' Ends 79
2.3.20 Autoradiography 80
2.3.21 Densitrometry Scanning 80

CHAPTER 3 ANALYSIS OF PROMOTER ACTIVITY UPSTREAM OF OTRA

3.1 Introduction 81
3.1.1 Objectives 82
3.1.2 Methology 83
3.2 Results
3.2.1 Construction of Recombinants for in vivo Promoter Analysis 85
3.2.2 Assay of Promoter Activity from pIJ486 Constructs 88
3.2.3 S1 Nuclease Mapping of the Putative otrAp2 Transcript 89
3.2.4 Transcription of otrA during OTC Production 93
3.3 Discussion 102

CHAPTER 4 THE OTCZ GENE: NUCLEOTIDE SEQUENCE, DEDUCED FUNCTION AND TRANSCRIPTION

4.1 Introduction 113
4.2 Results
4.2.1 Sequencing Method 115
4.2.2 Sequencing Strategy 118
4.2.3 Computer-Assisted Sequence Analysis 125
4.2.3.1 Potential Protein Coding Regions 126
4.2.3.2 Amino Acid Sequence Comparisons 131
4.2.4 Analysis of Transcription of otcZ 136
4.3 Discussion 139

CHAPTER 5 CHARACTERISATION OF THE OTCC/OTCX PROMOTER REGION

5.1 Introduction 142
5.1.1 Overall Strategy 145
5.2 Results and Discussion
5.2.1 Construction of a Promoter Library in pIJ2843 146
5.2.2 Assay of In Vivo Promoter Activity 150
5.2.3 Characterisation of the xylE* Transformants 150
5.2.4 Location and Characterisation of the Promoter Region Responsible for the Transcription of otrA During Antibiotic Production 155
5.2.4.1 Location of the 5' Termini of the Divergent Transcripts 157
5.2.4.2 Quantification of the 5' Termini of the Divergent Transcripts 163
5.2.5 Comparison of the otcCp1/otcXp1 Promoter Sequences 166

CHAPTER 6 THE SPHI12-SSTI13 REGION: NUCLEOTIDE SEQUENCE AND LOCATION OF DNA ELEMENTS DIRECTING THE EXPRESSION OF XYLE IN pIJ2843 CONSTRUCTS

6.1 Introduction 174
6.2 Results 177
6.2.1 DNA Sequencing of the SphI-SstI Region 177
6.2.2 Features of the Nucleotide Sequence 179
6.2.3 Analysis of the Deduced Amino Acid Sequence of PPCR1 and PPCR2 185
6.2.4 Location of the DNA Elements from the SphI-SstI Region which Direct Transcription of xyle in pIJ2843 Constructs 185
6.2.5 Analysis of Transcription within the Sau3AIc-e and Sau3AIf-g Regions 189
6.3 Discussion 189

CHAPTER 7 TRANSLATION INITIATION

7.1 Introduction 196
7.2 Results and Discussion 198
7.2.1 Comparison of 16S rRNA Sequences from Streptomyces and E. coli 198
7.2.2 Complementarity Between Downstream Sequences and the 5' End of 16S rRNA 199
7.2.3 Nucleotides of 16S rRNA Implicated in Binding to Downstream Regions 205
7.3 Scope of Future Studies 206

CHAPTER 8 CONCLUDING REMARKS 209

BIBLIOGRAPHY 215
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ap</td>
<td>ampicillin</td>
</tr>
<tr>
<td>APS</td>
<td>ammonium persulphate</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>Ci</td>
<td>curie</td>
</tr>
<tr>
<td>CIAP</td>
<td>calf intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>D</td>
<td>dalton</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulphoxide</td>
</tr>
<tr>
<td>DMF</td>
<td>dimethylformamide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide 5'-triphosphate</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine 5'-triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine 5'-triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymidine 5'-triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>dH2O</td>
<td>distilled water</td>
</tr>
<tr>
<td>E</td>
<td>core RNA polymerase</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diaminetetra-acetic acid (disodium salt)</td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
</tr>
<tr>
<td>EtOH</td>
<td>ethanol</td>
</tr>
<tr>
<td>g</td>
<td>centrifugal force equivalent to gravitational acceleration</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropylthio-β-D-galactoside</td>
</tr>
<tr>
<td>hrs</td>
<td>hours</td>
</tr>
<tr>
<td>kb</td>
<td>kilo bases</td>
</tr>
<tr>
<td>kM</td>
<td>kanamycin</td>
</tr>
<tr>
<td>l</td>
<td>litres</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>mins</td>
<td>minutes</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Mr</td>
<td>molecular weight</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>OTC</td>
<td>oxytetracycline</td>
</tr>
<tr>
<td>°C</td>
<td>degrees centigrade</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>P&lt;sub&gt;i&lt;/sub&gt;</td>
<td>orthophosphate</td>
</tr>
<tr>
<td>PIPES</td>
<td>piperazine-NN'-bis-2-ethane sulphonic acid</td>
</tr>
<tr>
<td>PPCR</td>
<td>potential protein coding region</td>
</tr>
<tr>
<td>RBS</td>
<td>ribosome binding site</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>sp. act.</td>
<td>specific activity</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecylsulphate</td>
</tr>
<tr>
<td>TC</td>
<td>tetracycline</td>
</tr>
<tr>
<td>TEMED</td>
<td>NNN-N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>TES</td>
<td>N-tris (hydroxymethyl) methyl-2-aminoethanesulphonic acid</td>
</tr>
<tr>
<td>Tris</td>
<td>tris (hydroxymethyl) methylamine</td>
</tr>
<tr>
<td>TSB</td>
<td>tryptone soya broth</td>
</tr>
<tr>
<td>tsr</td>
<td>thioestrepton</td>
</tr>
<tr>
<td>U</td>
<td>units</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>volts</td>
</tr>
<tr>
<td>v/v</td>
<td>volume by volume</td>
</tr>
<tr>
<td>w/v</td>
<td>weight by volume</td>
</tr>
<tr>
<td>W</td>
<td>watts</td>
</tr>
<tr>
<td>Xgal</td>
<td>5-bromo-4-chloro-3-indoyl-β-galactoside</td>
</tr>
<tr>
<td>YEME</td>
<td>Yeast extract-malt extract</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I express first my gratitude to Iain Hunter for his friendship, support and words of wisdom over the last three years.

Special thanks is also extended to "Auntie" Marion for keeping the laboratory running smoothly and providing me with blankets when I arrived in Glasgow to keep me warm during the cold winter nights. Needless to say the media ladies deserve a mention for preparing many batches of foul smelling media, while always remaining cheerful. Thank you Margaret and Linda for deciphering my writing and typing the reference section.

I also acknowledge my colleagues in the Streptomyces group for their friendship and invaluable discussion. Alison is due a special mention for proof-reading early drafts of this manuscript and reminding me of a few rules of my own language.

Thanks also to Mike Butler for arranging my visit to the Pfizer Laboratories, where I undertook several experiments described in this thesis. I am also grateful for the help and advice given by Craig Binnie and Mel Warren at Sandwich.

Last, but certainly not least, I thank my wife, Janine, for her love and understanding during this work, particularly whilst this thesis was being written.

This work was supported by a grant from the Medical Research Council.
SUMMARY

Prior to the work described in this thesis, an oxytetracycline (OTC) resistance gene, otrA, had been cloned from *S. rimosus* and sequenced. S1 nuclease mapping experiments suggested that the otrA gene was transcribed from two promoters, 129bp (otrAp1) and 339–344bp (otrAp2) upstream of the translation start codon.

Further transcriptional analyses described in this thesis confirmed the presence of the otrAp1 promoter, but showed that a technical artifact had led to the erroneous assignment of otrAp2. The transcription of the otrA resistance gene was examined at different stages in the fermentation of OTC. During the period of intense vegetative growth which precedes the production of OTC, the otrA resistance gene was transcribed predominantly from otrAp1. As the growth rate slowed and the onset of antibiotic production approached, the resistance gene was transcribed increasingly as part of an otcC-otcZ-otrA polycistronic message, while transcription from otrAp1 decreased. The inclusion of the otrA resistance gene in a transcription unit along with the biosynthetic genes otcZ and otcC probably ensures that the level of resistance increases in line with the biosynthesis of oxytetracycline. Furthermore, the delay in the production of OTC can be attributed, at least in part, to the temporal regulation of transcription.

The 1.3kb *SmaI*-BcII fragment containing the otcZ biosynthetic gene was sequenced. The amino acid sequence of the predicted gene product showed homology to a methyltransferase from bovine pineal glands and to a region of the tetracenomycin cluster, which is also implicated in methylation. The otcZ gene product is thought to catalyse the C-6 methylation of the tetracyclic nucleus to form 6-methylpretetramid.

Sequencing of a 0.96kb *SphI*-SstI fragment which was located immediately upstream of otcC identified two PPCRs, otcX PPCR1 and
PPCR2, that were encoded on the opposite strand to otcC, otcZ and otrA. No homology was found between the amino acid sequences of the predicted gene products and entries in the NBRF-protein and EMBL databases. The SphI-SstI fragment had been shown previously to hybridize to a region of the actinorhodin cluster of S. coelicolor, which was implicated in hydroxylation. It is possible that the polypeptides encoded by otcX PPCR1 and PPCR2 have roles in the C-4 hydroxylation of 6-methylpretetramid.

Primer extension analyses identified the 5' ends of both the mRNA of otcX PPCR1 and PPCR2 and the otcC-otcZ-otrA message. These divergent transcripts were non-overlapping. Alignment of the DNA sequences immediately upstream of the 5' ends of these transcripts revealed several conserved features, which may be involved in the initiation and regulation of transcription. The most striking of these was a tandem repeat, which had sequence similarity to the recognition sequence of the PhoB regulatory protein. Under conditions of low phosphate, PhoB activates transcription of the phosphate operon in E. coli. As oxytetracycline is only produced under conditions of low phosphate, it is possible that a functional analogue of PhoB regulates transcription of the otc cluster in S. rimosus.

An analysis of sequences downstream of the putative translation start codons of the genes described in this thesis and a selection from other Streptomyces spp. has identified at least three consecutive nucleotides in the first 10 codons of every gene, which are complementary to some part of the 5' end of 16S rRNA from Streptomyces. It is proposed that base pairing between the 5' end of 16S rRNA and complementary nucleotides at the 5' end of a gene maybe important in determining the efficiency of translation.
CHAPTER 1
GENERAL INTRODUCTION
General Introduction

The primary objective of this chapter is to provide the background to the experimental investigations, which are described in this thesis. Firstly, a review of the mechanisms, which control the global changes in gene expression that underlie the development of *Streptomyces*, will be presented. Particular emphasis will be given to the integration of secondary metabolism (generally referred to as antibiotic production) into the life cycle.

During the course of this work, conserved sequences were identified within promoters of the oxytetracycline cluster. These DNA determinants were similar to the *pho* boxes, which are found in *E. coli* promoters that are regulated by phosphate. Although a molecular genetic approach was used to analyse the production of oxytetracycline in this work, physiological factors that influence the onset of secondary metabolism and morphological differentiation, particularly phosphate limitation, will be surveyed briefly to provide the necessary background for discussion on the possible role of the DNA determinants revealed by molecular analysis.

Within this thesis, experimental evidence is presented, which suggests that the delay in the production of oxytetracycline can be attributed, at least in part, to the temporal regulation of transcription. Therefore, recent publications, which have contributed significantly to our understanding of *Streptomyces* gene expression will also be discussed.

Finally, recent advances in our understanding of the biosynthesis of oxytetracycline and related polyketide antibiotics will be reviewed.

1.1 Introduction to *Streptomyces*

Members of the genus *Streptomyces* are aerobic, Gram-positive soil bacteria that undergo complex morphological and physiological differentiation. *Streptomyces* genomes have been estimated by renaturation kinetics to be in the range 6 to 9 megabases (Antonov *et al.*, 1978), which is up to twice the values for *E. coli* K12 and *Bacillus*
Figure 1.1. Development of *Streptomyces coelicolor*. Italicised gene symbols indicate the involvement of particular gene products in the accomplishment of given stages of development. The inner cycle (---) occurs only with *bldA* and *bldD* mutants grown on medium containing glucose as the carbon source. During growth on mannitol phenotypic suppression of *bldA* and *bldD* mutants occurs and the outer cycle occurs. From Chater and Merrick, 1979.
*subtilis* (Hutter and Eckhardt, 1988). Recent analysis of the *S. coelicolor* chromosome, by pulse-field gel electrophoresis, does not contradict the size estimates based on kinetic analysis (cited in Hopwood and Kieser, 1990). A distinct feature of streptomycetes is the high proportion of guanine (G) and cytosine (C) in their genomes, averaging at 73 mol% G+C (Enquist and Bradley, 1971). The analysis of codon usage in *Streptomyces* shows strong bias in favour of codons with G or C in the third position (Bibb et al., 1984). From a file containing 63 *Streptomyces* genes (compiled by Mervyn Bibb, unpublished work), it was shown that G residues (42%) are favoured over C residues (27%) in the first codon position, C residues (29%) are preferred over G residues (16%) in the second position and C residues (55%) are preferred to C residues (36%) in the third codon position. Therefore, the average mol G+C composition in the first, second and third position of a codon is 69%, 45% and 91% respectively.

*Streptomyces* are of interest for two principal reasons. Firstly, their morphological differentiation from vegetative mycelium to chains of spores is a relatively simple system for studying the mechanisms, which regulate the expression of developmental genes. Secondly, this genus produces a plethora of complex chemical structures, many of which have important pharmaceutical properties. These compounds have been designated "secondary metabolites" as they are not involved in the vegetative growth of the microorganism and they are thought to be produced generally as an adjunct to the differentiation process.

1.2 *Streptomyces* life cycle.

The *Streptomyces* life cycle (Figure 1.1), can be considered to start with the germination of a spore (Chater, 1984). The germling grows apically and branches to form a mat of irregular septate filaments, known as the substrate mycelium. When nutrients become limiting, aerial mycelium are produced from the substrate mycelium. The tips of these mycelium coil and the hyphae septate to form spores. After maturation, which involves the thickening and rounding of the spore wall, the propagules are released.
Chapter 1 General Introduction 4

In soil and leaf litter, Streptomyces spp. are ecologically-important microorganisms for recycling dead plant material, which is composed largely of complex polysaccharides, lignin and cellulose. Similar to other soil microorganisms such as fungi and bacilli, the streptomycetes secrete a battery of extracellular enzymes to degrade these recalcitrant substrates.

The mycelial growth habit is well adapted to utilise these complex, insoluble polymers, which have to be degraded extracellularly before they can be taken up by the organism and used for growth. Significant factors limiting growth are, presumably, the rate of diffusion of enzymes away from the colony and the rate of diffusion of soluble degraded products back to the colony.

Mycelial outgrowth from an initial point of colonisation, as a dispersed mat probably minimises rapid nutrient exhaustion by ensuring that the growing parts of the colony are continually encountering new sources of nutrients. However, growth in a localised area does not favour dispersal of the species. Having exhausted a nutrient source, a species must be able to colonise new areas in order to have any chance of survival. In Streptomyces, this requirement is met by the formation of spores, which can be dispersed by air or water.

Historically, it was thought that antibiotics might play a regulatory role in differentiation. Recent evidence suggests that pamamycin stimulates the formation of aerial mycelium and may act as a signal for colonial development in the producing strain S. alboniger (Kondo et al., 1988). This is unlikely to be a general role of secondary metabolites, however, as mutants of many Streptomyces spp., which have lost the ability to produce antibiotics, are viable and can complete their life cycle under laboratory conditions. By a similar argument, secondary metabolites are unlikely to be shunt products, which alleviate the potential toxic accumulation of metabolic intermediates, as growth slows and cultures enter stationary phase (as proposed by Hunter and Baumberg, 1989).
Chapter 1 General Introduction 5

It is generally envisaged that the degradation of the substrate mycelium provides precursors and energy for the growth of the aerial mycelium. Antibiotic production is usually coincidental with the formation of the aerial mycelium. It has been argued eloquently, on the basis of teleological considerations, that the general role of secondary metabolites is to prevent invasion of the lysing substrate mycelium by motile, unicellular microorganisms, for example, bacilli (Chater, 1984). The timing of antibiotic production is clearly central to this hypothesis.

1.3 Developmental Biology of Streptomyces coelicolor

There has been a great deal of interest in the morphological development of Streptomyces and many mutants, blocked at different stages in the life cycle of S. coelicolor A3(2), have been isolated (Chater and Merrick, 1979). Two classes of mutants were identified; bald (bld) mutants, which do not form aerial mycelium or produce, in general, any of the antibiotics associated with this strain, and white (whl) mutants, which form aerial mycelium and produce antibiotics, but do not form pigmented, mature spores (Figure 1.1). With molecular genetic approaches now possible, the characterisation of these mutants is revealing interesting facets about the processes that are instrumental in Streptomyces development.

BldA mutants are unable to form aerial mycelium and produce antibiotics (Piret and Chater, 1985). Sequencing of a DNA segment which complements this mutation did not identify any potential protein coding regions (PPCR). Instead, it was deduced that this region of DNA could encode a RNA product with the structure of a leucine tRNA_{TTA} (Lawlor et al., 1987). The existing evidence suggests that TTA codons may be absent from the structural genes for morphological differentiation and antibiotic production, but present in a small subset of genes associated with the regulation of these processes. Recently, in S. coelicolor, it has been shown that the expression of five genes containing TTA codons is dependent on the bldA gene product (Leskiw et al., cited in Guthrie and
Chater, 1990). This dependency provided compelling evidence that \textit{bldA} specifies a leucine tRNA which is actively involved in protein synthesis. The rarely used TTA codon has been identified in the \textit{actII} gene, which regulates positively actinorhodin production (Section 1.6), and in a putative actinorhodin resistance gene (Hopwood pers. comm.). Very recently, the accumulation of an undescribed transcript encoding at least one gene from the undecylprodigiosin (red) cluster of \textit{S. coelicolor} A3(2) was shown to be dependent on the expression of \textit{bldA} (Guthrie and Chater, 1990). On media containing low levels of phosphate, however, the \textit{bldA} mutant was able to produce undecylprodigiosin. These experiments revealed that the inability of the \textit{bldA} mutant to produce this antibiotic cannot be accounted for by the presence of TTA codons in the \textit{red} structural genes, but rather, \textit{bldA} influences the transcription of \textit{red} genes. Under conditions of low phosphate, an alternative regulatory pathway (independent of \textit{bldA}) can lead to \textit{red} gene expression (Guthrie and Chater, 1990).

Sequencing of the DNA that complements the \textit{whiG} mutant, revealed an ORF with the potential to encode a polypeptide with striking sequence similarity to the $\sigma^{23}$ polypeptide from \textit{B. subtilis} (Chater \textit{et al.}, 1989). This suggested strongly that the promoters for a subset of genes, whose products were required for development of spores from aerial hyphae, were recognised by an alternative sigma factor, $\sigma^{\text{whiG}}$. In \textit{B. subtilis} and \textit{E. coli}, the $\sigma^{23}$-RNA polymerase holoenzyme transcribes genes associated with chemotaxis and motility (Chamberlain, pers. comm.), at the end of exponential growth when filamentous, multinucleate, non-motile cells divide to form uninucleate, motile cells. Mutants of \textit{B. subtilis} deficient in $\sigma^{23}$ remain filamentous and non-motile (Helman \textit{et al.}, 1988). It has been suggested that, apart from motility, the above phenotype is very similar to that of \textit{whiG} mutants, which are unable to develop beyond the extension of aerial hyphae (Chater, 1989). This family of $\sigma$ factors may have a general role in the dispersal of bacteria during nutrient limitation.

DNA complementing the \textit{whiE} mutant, which is deficient in spore pigmentation, has been sequenced and shown to contain ORFs that
correspond clearly to those of known polyketide synthase (PKS) genes (Section 1.12) (Davis and Chater, 1990). The biosynthesis of the grey-green spore pigment appears to be a specific feature of aerial hyphae (Davis and Chater, 1990), unlike other antibiotics which are presumed to be synthesized by the substrate mycelium. The biosynthesis of spore pigment in \textit{S. coelicolor} may be under novel temporal and spatial control. It would be very interesting, therefore, to identify the factors necessary for the transcription of the \textit{whiE} locus.

A model for the osmoregulation of differentiation has been proposed on the basis of the observed oscillations in the intracellular levels of glycogen (Chater, 1989). During the differentiation process, there is a cycle between the synthesis and degradation of carbon storage compounds. Glycogen granules are accumulated prior to the formation of aerial mycelium and then degraded during the extension of the aerial hyphae (Wildermuth, 1970). Glycogen is again synthesized, immediately prior to the formation of the uninucleate spores (Hardisson \textit{et al.}, 1986). The switch between synthesis and degradation is presumed to cause fluctuations in the osmotic potential of the mycelium. It is proposed that increased turgor pressure due to the degradation of glycogen granules provides the "driving force" for the extension of the aerial hyphae and the "rounding off" of the spore compartments. Similarly, the accumulation of glycogen at the end of hyphal extension is predicted to cause a reduction in turgor pressure, which permits the formation of septa. At present, it is open to debate whether the fluctuations in the abundance of glycogen has a real role in the osmoregulation of differentiation or is merely a consequence of changes in the supply and demand for cellular precursors. This model identifies the importance of the overall physiology of the mycelium and in particular, the osmotic balance. It was also proposed that the products of some \textit{bld} genes may regulate polymer synthesis and degradation. Depending on the growth substrate, turgor pressure may be generated by metabolic pathways not regulated by \textit{bld} genes and this could provide a possible explanation for the dependence of some \textit{bld} phenotypes on the type of growth medium (Chater, 1989).
1.4 Medical and Commercial Interests in *Streptomyces*

*Streptomyces* produce a multitude of secondary metabolites, which range from simple amino acid analogues to complex macrocyclic compounds (Goodfellow and O'Donnel, 1989). Of the six thousand or so bioactive compounds now characterised, 60% of these are produced by *Streptomyces* (Omura, 1986). However, only about 70 of the compounds derived from *Streptomyces* satisfy the high standards required clinically for efficacy and tolerability.

Many of the antibiotics produced by *Streptomyces* are of medical importance. Examples include; the β-lactams penicillin G and cephamycin, the tetracycline series of polyketide antibiotics, the aminoglycosides neomycin, spectinomycin and kanamycin and the peptide antibiotics actinomycin and vancomycin. Compounds produced by the streptomycetes are not only used as antibacterial agents in clinical infections. The polyethers monensin and lasalocid produced by *S. cinnamonensis* and *S. lasalocid*, respectively, are the most widely used anticoccidial agents for controlling parasitic protozoa in poultry. These compounds are used also as growth promotants in ruminants (Schumard and Callender, 1968; Mitrovic and Schildnecht, 1974). The antihelminthics avermectin and milbemycin produced by *S. avermitilus* and *Streptomyces* spp. NRRL 5370 respectively, have potent activity against nematode infections of farm animals (Burg *et al.*, 1979; Pat. No. DT2329485). The tripeptide antibiotic bialaphos, produced by *S. hygroscopicus*, has a potent herbicidal activity when cleaved by a peptidase in plants (Bayer *et al.*, 1972). The *Streptomyces* produce also a vast range of other pharmalogically-active compounds, including anti-inflammatory, vasodilatory and anti-cancer agents (Ninomija *et al.*, 1980; Blum and Carter, 1974; Yoshida *et al.*, 1982). The potential to discover secondary metabolites with novel pharmaceutical properties is an obvious reason for continuing the study of *Streptomyces*.

Recently, the ability of *Streptomyces* spp. to secrete a variety of extracellular enzymes has been the focus of considerable attention. The analysis of several genes encoding secreted polypeptides suggests
that, as in *E. coli*, a signal sequence at the amino terminus of the gene product contains the targeting information for secretion of the polypeptide. The promoter and signal sequence of a protein protease inhibitor (LTI) in *S. lividans* has been shown to direct expression and export of sCD4, a soluble derivative of the human T-cell receptor CD4 that acts as a receptor for the HIV virus (Brawner *et al*., 1990). The vast amount of knowledge that has been gained from the development of advanced fermentation systems for antibiotic production, could benefit the development of *Streptomyces* spp. as hosts for the production and secretion of small genetically-engineered polypeptides.

1.5 Organisation of Genes for Antibiotic Biosynthesis

The structural genes for the biosynthesis of all the antibiotics (used here as a general term for secondary metabolites), so far studied, are closely linked in single clusters (Hunter and Baumberg, 1989). Orthogonal-field gel electrophoresis has revealed that the genes for methylenomycin production are contained on the "giant linear plasmid", SCP1 (ca. 350-520 kb) (Kinashi *et al*., 1987) confirming the early genetic evidence (Kirby and Hopwood, 1977), which suggested that the genes had an extrachromosomal location. Apart from the gene cluster for methylenomycin A, all the other antibiotic clusters would appear to be located on the chromosome.

There have been few reports on the overall organisation of transcription within antibiotic clusters. This may reflect the difficulty experienced by many groups in identifying *Streptomyces* transcripts using Northern analysis. The overall transcriptional organisation of the actinorhodin and methylenomycin cluster has been deduced, however, by virtue of the mutational cloning strategy used to clone and analyse genes involved in antibiotic biosynthesis. Using this approach, DNA fragments inserted into an *att*-deleted derivative of actinophage *sC31* direct integration of the vector into the homologous region of the host genome, thereby potentially disrupting gene expression. It was predicted that integration would be mutagenic when the insert was completely internal to a transcription unit (Chater and Bruton, 1983).
**Chapter 1**

*General Introduction*

---

**Figure 1.2.** Preliminary transcription map of the methylenomycin production gene cluster. *PstI* sites are shown by a vertical bar. From Chater and Bruton, 1985.

---

**Figure 1.3.** Preliminary transcription map of the gene cluster for actinorhodin production. (a) Putative transcripts from the mutational cloning experiments. (b) Putative transcripts deduced from S1 nuclease mapping experiments. (c) Location of genes as deduced by complementation tests (Malpartida and Hopwood, 1986). Modified from Hopwood *et al.*, 1986.

---

**Figure 1.4.** Preliminary transcription map around the *aphD* gene, which is part of the gene cluster for streptomycin production. From Distler *et al.*, 1988.
Chapter 1  General Introduction

The analysis of $\phi$C31 constructs, which caused gene disruption, suggested that the genes for the biosynthesis of methylemomycin (the mmy genes) were organised into two transcription units, one of which also encoded the methylemomycin resistance (mmr) gene (Figure 1.2) (Chater and Bruton, 1985). Sequencing of a DNA segment encoding the mmr gene revealed, however, that this particular gene has to be transcribed in the opposite direction to that predicted by mutational cloning (Neal and Chater, 1987). It is not clear why an anomalous result with the mutational cloning was obtained, although the artifactual production of mmy-antisense RNA might be a plausible explanation (Hopwood et al., 1986). The methylemomycin cluster is organised, therefore, into at least three transcription units; the central mmr gene being transcribed from the opposite strand to that for the longer mRNAs of the biosynthetic genes.

Mutational cloning, using att-deleted derivatives of $\phi$C31 (as described by Chater and Bruton, 1983), and S1 nuclease mapping experiments revealed, that at least four transcription units were present in the actinorhodin cluster (Figure 1.3) (Hopwood et al., 1986). At least two of the transcripts (for mutant classes I,VII,IV and VI,V) would appear to be polycistronic, while another of less than 1 kb encode the actIII region. A fourth transcript was predicted for the actII region encoding the putative positive-regulatory gene (Section 1.6).

The large polycistronic messages in the actinorhodin cluster would appear to correspond to the subclustering of genes for "early" and "late" steps in the biosynthetic pathway. This feature has also been reported for the tylosin cluster (Seno and Baltz, 1988) and tetracenomycin (Motamedi and Hutchinson, 1987). These observations prompted the suggestion that subclustering may facilitate the ordered expression of genes for sequential stages in the production of the antibiotic (Motamedi and Hutchinson, 1987).

The transcriptional organisation of the streptomycin cluster around the aphD resistance gene has been deduced by Northern analysis (Distler et al., 1987; 1988). The current working model (Figure
1.4) includes three promoters for the \textit{strR}, \textit{aphP} and \textit{strB1} genes. At least seven different RNA species have been detected in this region. An anti-termination mechanism has been proposed to account for the differences in the 3' termini of transcripts originating from the same promoter (Section 1.6) (Distler \textit{et al}, 1988).

1.6 Pathway–specific Regulatory Genes

In addition to the global regulation of \textit{Streptomyces} development (Sections 1.3 and 1.8), the biosynthesis of a number of antibiotics has been shown to be controlled by pathway–specific regulatory elements, most of which remain to be characterised satisfactorily.

Genetic analysis of the actinorhodin pathway provided the first example of a positive–acting regulatory element. The \textit{actII} mutants were unable to produce actinorhodin in cosynthesis tests with any other classes of mutant and were presumed, therefore, to be either polar mutants within a large operon or deficient for a positive regulator (Rudd and Hopwood, 1979). The over–production of actinorhodin by the wild–type strain when DNA complementing the \textit{actII} mutants was introduced on a plasmid, gave credence to the latter hypothesis (Hopwood, Malpartida and Chater, cited in Hopwood \textit{et al}, 1986). Similarly, increased production of undecylprodigiosin (red) was achieved by introducing random fragments of the \textit{red} cluster into \textit{S. coelicolor} on a high–copy–number plasmid (Guthrie, 1989). Interestingly, the same recombinant plasmid, containing the putative positive–regulatory (\textit{redD}) gene, was also able to restore undecylprodigiosin production in a \textit{bldA} mutant, providing further support for the proposal that the structural genes for the antibiotics produced by \textit{S. coelicolor} do not contain any TTA codon (Section 1.3).

DNA segments have been isolated from \textit{S. hygroscopicus} and \textit{S. violaceoruber}, which complement \textit{actII} mutants of \textit{S. coelicolor} and are closely linked to the gene cluster for milbemycin and granaticin. These DNA segments are thought to encode positive–regulatory elements (Malpartida \textit{et al}, 1987). Recombinants of \textit{S. lividans} containing the milbemycin element, \textit{milbII}, over–produced actinorhodin (Humphreys \textit{et
al., 1987). These observations suggest that the actII, milbII and granaticin positive-regulatory elements belong to family of regulators which have similar functions and recognize similar targets. The bialophos cluster of S. hygroscopicus has also been shown to contain a positive regulator, brpA, which activates transcription of the biosynthetic genes (Anzai et al., 1987).

A possible negative effector was identified in the methylenomycin cluster (Chater and Bruton, 1985), as mutational cloning experiments (Section 1.5), identified a region at the "leftmost" end of the cluster which caused over-production of methylenomycin when disrupted by the integration of sC31. The simplest explanation was that a negatively-acting regulator was encoded by the disrupted locus. The region immediately to the right of this segment also has a regulatory role since a slight increase in production resulted from gene disruption in this region (Chater and Bruton, cited in Hopwood et al., 1986).

Evidence has been obtained for both positive and negative regulation of streptomycin biosynthesis in S. griseus (Distler et al., 1988). The predicted product of the strR regulatory gene, which is required for the maximum expression of the aphD resistance gene and the strB amidotransferase gene, has significant homology to the lambda Q protein involved in the antitermination of transcription. It has been proposed that the strR gene product facilitates transcription through a number of termination sites, which have been identified by Northern analysis (Section 1.5) (Distler et al., 1988) A DNA segment has also been identified in the streptomycin cluster between the aphD resistance gene and the strB aminotransferase gene, which controls negatively expression of the aphD gene (Tohyama et al., 1986).

In summary, the analysis of a number of different gene clusters has shown that resistance and regulatory elements are closely linked to the structural genes for the biosynthesis of antibiotics. It has also been observed that genes involved at similar stages in the biosynthesis of some antibiotics may be subclustered within the same transcription unit. Clustering probably reflects the need to coordinate the
expression of several related functions. At present, little is known about the promoters and RNA polymerase holoenzymes which are responsible for the transcription of antibiotic clusters. Although regulatory elements have been identified, the regulatory circuits of which they are a part have not yet been elucidated.

1.7 Catabolite Regulation of Secondary Metabolism

The biosynthesis of tetracyclines (including chloro- and oxytetracycline) and other antibiotics, including aminoglycosides, macrolides, polyenes, anthracyclines, ansamycins and polyethers is particularly sensitive to phosphate regulation (Martin et al., 1988; Martin and Demain, 1980). Tetracycline can only be produced abundantly in medium containing levels of phosphate lower than that required for maximal growth of the culture. Phosphate starvation is used to limit the growth of tetracycline producers, including the Pfizer strains M4018 and M15883, and induce the production of antibiotic. The amount of phosphate required for optimum production of tetracycline is dependent on the individual production strain (Behal, 1986; 1987). The activity of anhydrotetracycline (ATC) oxygenase (encoded by otcC in S. rimosus) is reduced in cultures of S. aureofaciens and S. rimosus, when grown on medium containing levels of phosphate supraoptimal for tetracycline production (Behal, 1979; Butler pers. comm.). Although it was stated in a recent review (Liras et al., 1990), no direct evidence has been published which supports the view that the repression of ATC oxygenase activity by phosphate is due to a reduction in synthesis of the enzyme. From the published data, it could be argued that phosphate regulates the catalytic efficiency of ATC oxygenase.

The synthesis of several enzymes associated with secondary metabolism (ESM) is regulated by phosphate, for example; p-aminobenzoate (PABA) synthase from the cãndidin producer S. griseus and dTDP-glucose-4,6-dehydratase, dTDP-mycarose synthetase and macrocin O-methyltransferase from the tylosin producer Streptomyces T59-235 (cited in review by Liras et al., 1990). In addition, the catalytic activity of some other ESM can be inhibited by growth on medium containing high levels of phosphate, for example; deacetoxycephalosporin
C synthetase in *Streptomyces clavuligerus*, *Nocardia lactamurans* and *Acremonium chrysogenum*. This indicates that there are two levels at which phosphate control can be exerted (Liras *et al.*, 1990).

The biosynthesis of candidicidin by *S. griseus* has been studied intensively as a model of phosphate control (Liras *et al.*, 1977; Martin and Demain 1977). PABA synthase activity can only be detected in *S. griseus* during the antibiotic production phase and is absent in a series of non-producing mutants (Gil *et al.*, 1985). Furthermore, this particular enzyme activity is not found during growth on medium containing high levels of phosphate. The structural gene (*pabS*) and promoter for PABA synthase have been cloned and characterised (Gil and Hopwood, 1983; Rebello *et al.*, 1989). Recently, the phosphate control of PABA synthase activity has been shown to be exerted at the level of transcription (Asturious *et al.*, 1990). The promoter for *pabS* has been reduced to a 35bp core sequence which is still sensitive to phosphate control. A hexamer, centred 10 bp upstream from the transcription start point of the *pabS* promoter, has good homology with the -10 region from the consensus sequence for the major class of eubacterial promoters. However, no homology to a canonical -35 region could be detected in the *pabS* promoter. Computer analysis of the *pabS* promoter revealed the presence of a hyphenated direct repeat on the sense strand, which overlaps with the tentative -10 region. This hyphenated repeat has significant sequence similarity to the consensus for *pho* boxes in *E. coli* (Liras *et al.*, 1990). The *pho* boxes are the DNA binding sites of the PhoB protein, which activates transcription of the phosphate regulon during phosphate limitation (Makino *et al.*, 1986, 1988). The primary *pho* boxes, in all the *E. coli* promoters regulated by phosphate, are located 10 bp upstream of the -10 regions on the same strand. It has been proposed that PhoB protein, bound to these *pho* boxes, substitutes for a typical -35 region in the phosphate-controlled promoters of *E. coli* (Makino *et al.*, 1986, 1988).

The PhoB protein belongs to a family of bacterial regulatory proteins that are an integral part of the sensory and regulatory mechanisms responsible for cellular responses to environmental changes.
Figure 1.5. The proposed system that regulates the uptake of phosphate in *E. coli*. Extracellular phosphate (P_i) crosses the outer membrane through the PhoE, OmpF and OmpC porins. Alternatively, phosphoesters in the periplasm are hydrolysed to P_i by alkaline phosphatase, PhoA. Under conditions of low phosphate, uptake into the cytoplasm is via the phosphate-binding protein (PstS) through the phosphate-specific transport (PST) system. The PST system in conjunction with PhoU regulates the activity of the transcriptional activator PhoB, which regulates the expression of all the operons in the Pho regulon, including *phoE, phoA, pstSCABphoU* and *phoBR*. The active form of PhoB is P-PhoB, produced by the action of the PhoR kinase. The PhoU protein together with PhoR is proposed to dephosphorylate P-PhoB. From Stock *et al.*, 1989.
(Ronson et al., 1987; Yamada et al., 1989). The following summary of phosphate uptake illustrates the salient points of these regulatory mechanisms. Most of the material for this brief discussion was taken from an in depth review by Stock et al. (1989).

E. coli has a very complex mechanism for regulating the expression of genes involved in the uptake and metabolism of phosphate (Figure 1.5). In addition to the regulatory protein PhoB, another central component of the system is a membrane receptor PhoR, which phosphorylates PhoB under conditions of low phosphate, thus allowing PhoB-P to activate transcription of the phosphate operon. Frequently membrane receptor proteins are kinases that respond to environmental signals and phosphorylate response regulators that control transcription.

Active transport of phosphate across the membrane is mediated by the phosphate-specific transport (PST) system and several sugar phosphate transport systems, including glycerol-3-phosphate (encoded by the ugp operon) and hexose phosphates (encoded by the uhp operon). An anion-specific porin, PhoE, facilitates the passage of phosphate across the outer membrane into the periplasmic space. Within the periplasmic space, the PhoA alkaline phosphatase (which hydrolyses polyphosphates) and binding proteins such as PstS and UgpB, assist the delivery of phosphate to the transport systems within the cytoplasmic membrane.

Extracellular phosphate levels regulate the expression of many of the genes involved in phosphate uptake. High levels of phosphate shut down expression of the phoA and phoE genes, as well as the pst and ugp operons, while sugar phosphates such as glucose-6-phosphate and hexose phosphates induce genes that encode their own respective transport systems. Interestingly, a number of different protein kinases and response regulators are involved in controlling the expression of these genes. The PhoR protein kinase and the PhoB regulator modulate the expression of phoA, phoE, the pst operon and the phoBR operon itself. This group of genes have been designated the Pho regulon, since, all of their promoters require phosphorylated PhoB for maximum
transcription. At present, it is not known precisely how low-phosphate concentrations activate PhoR. PhoR activity does not appear to be regulated directly by either phosphate in the periplasmic space or indirectly by an interaction with PstS. An intact PTS system is essential, however, for sensing high levels of phosphate and reducing the expression of the Pho regulon.

A component of the pst operon PhoU, which is not involved directly in phosphate uptake, is required to shut down expression of the Pho regulon. PhoU appears to promote the dephosphorylation of PhoB-P (the active form) in response to a signal that reflects the activity of the PST. PhoM, another typical membrane receptor-protein kinase, can also phosphorylate PhoB. The PhoM kinase appears to be part of a distinct regulatory system that has some cross-specificity for PhoB.

In conclusion, the PhoR/PST system exerts both positive and negative control on PhoB. Under conditions of low phosphate, PhoR and PhoM probably both function to activate Pho expression by catalysing the transfer of phosphoryl groups to PhoB (the pst operon does not appear to be involved in this process). Under conditions of high phosphate, PhoR along with the PST system act to shut down the expression of the Pho regulon; in particular PhoU, which is a component of the PST system, probably activates a PhoR-dependent phosphatase that dephosphorylates PhoB-P.

Analysis of the promoter sequences of the aphD gene, encoding a streptomycin phosphotransferase in the streptomycin gene cluster of S. griseus and an open reading frame, ORFX, which is divergent to the thiostrepton-inducible tipA gene in S. lividans, has revealed tentative pho-like sequences in both of these phosphate-regulated promoters (Liras et al., 1990). These sequence similarities prompted the suggestion that pho-like boxes in Streptomyces (also called phosphate-control (PC) sequences) may affect the binding kinetics of the RNA polymerase, possibly through a specific interaction with a putative protein bound to sequences within the vicinity of phosphate-controlled promoters.
Further to this, it was proposed that the intracellular effector which mediates phosphate control may be a highly phosphorylated nucleotide or a sugar–phosphate that acts as a sensor of the phosphate levels in the culture medium (Liras et al., 1990).

Pairs of protein kinases and response regulators are involved in numerous, dynamic responses to changing environmental conditions, for example; chemotaxis and repellent stimuli in enterobacteria (Hess et al., 1988), regulation of outer membrane porin expression in response to osmolarity and other conditions in E. coli (Igo et al., 1989a; 1989b) and regulation of gene expression in response to nitrogen deprivation (Nifa et al., 1988). Sporulation in B. subtilis also involves response regulators Spo0A and SpoOF, which are homologous to PhoB (Smith, 1989). Mutations in spo0A and spo0F have been identified which prevent initiation of the sporulation process at the onset of stationary phase when the growth of the cells becomes nutrient-limited. It is possible, therefore, that some of the bld mutants of S. coelicolor (described in Section 1.3) could be deficient in proteins related to the protein kinases and response regulators described above.

From in vitro assays as well as genetic analysis, it would appear that many protein kinases can phosphorylate response regulators from different systems. For example, PhoB can be phosphorylated by both PhoR and PhoM. The gene encoding PhoM is not part of the Pho operon. It has been suggested by Stock and other workers in this field that "cross-talk" between parallel phosphotransfer pathways may provide a global regulatory network for sensing the overall physiological state of the organism (Stock et al., 1989). It is conceivable that integrated phosphotransferase systems may be important in controlling the switch to morphological differentiation and secondary metabolism in Streptomyces. Should this be the case, it could provide an additional explanation for the growth-medium dependence of some bld phenotypes (Section 1.3) (Chater, 1989).

The "physiological" state of Streptomyces is presumed to be the key factor determining the switch to antibiotic production and
morphological differentiation. It is proposed that some bld mutations may disrupt phosphotransfer pathways, which sense the level of nutrients. Assuming that the phosphotransferase systems for different nutrients are integrated and have a cumulative effect on the decision to undergo differentiation, the result of a mutation, which disrupts a single phosphotransferase system, may depend on the growth medium. For example, a hypothetical mutant, which disrupts results in the organism always "believing" that it is supplied well with phosphate, may undergo the complete life cycle when it is grown on medium containing high levels of phosphate. The depletion of readily utilisable carbon and/or nitrogen source may be sufficient to "trigger" differentiation. It is possible, however, that the same mutant might not undergo differentiation when grown on a medium containing good carbon and nitrogen sources but, low levels of phosphate, as the organism does not "sense" the phosphate limitation, which would "trigger" differentiation in the wild-type strain. The above example is only intended to illustrate the possible role of putative sensing mechanisms in determining the switch to differentiation. It should be noted that there is no evidence to support this model.

Physiological analysis of antibiotic production has revealed that secondary metabolism can be regulated by the availability of carbon and nitrogen, as well as phosphate (Martin and Demain, 1980; Campbell, 1986). The expression of phenoxazinone synthase (PHS) of S. antibioticus that forms the phenoxazinone nucleus of actinomycin is regulated by glucose (Jones, 1984). Transcriptional analysis revealed that the increase in the specific activity of PHS, observed in medium with galactose as a carbon source was due, at least in part, to increased production of PHS mRNA. The repression of PHS activity caused by glucose was also shown to be exerted at the level of transcription (Jones, 1984). Interestingly, high levels of glucose do not repress the biosynthesis of PHS in S. lividans. It has also been suggested that glucose represses puromycin biosynthesis in Streptomyces alboniger (Sankaran et al., 1975) and represses transiently erythromycin production in Saccharopolyspora erythraea (Escalante et al., 1982). In the latter example, the repression of synthesis of ESM rather than the inhibition of the catalytic activities
of some or all of the ESM was inferred, as glucose added after the onset of antibiotic production had no effect. The repression of synthesis of deacetoxycephalosporin C synthase has been proposed to be the key factor responsible for the suppression of cephamycin biosynthesis by glucose (Cortes et al., 1986). Glucose is not the only carbon source which can have a repressive effect. For example, glycerol can repress cephalosporin production by \textit{S. clavuligerus} (Aharonwitz and Demain, 1978) and citrate can repress novobiocin production by \textit{S. niveus} (Kominek, 1972).

Readily utilisable nitrogen sources can also repress antibiotic production. For example, the biosynthesis of cephalosporin in \textit{S. clavuligerus} has been shown to be depressed by ammonium (Bascaran et al., 1989a; 1989b). The production of chloramphenicol by \textit{S. venezuelae} has been studied with regard to the influence of nutrient conditions prevailing at the onset of secondary metabolism. Under conditions of nitrogen limitation, the activity of ESM was shown to depend on the concentration of glucose (Bhatnager et al., 1988). Conversely, the extent to which glucose suppressed the production of cephamycin depended on the residual nitrogen source.

All of the above examples are consistent with the proposal that integrated sensory systems monitor the level of carbon, nitrogen and phosphate in the growth medium and may have a cumulative effect on the decision to produce antibiotics and undergo morphological differentiation. Although the molecular mechanism(s) which mediate the carbon and nitrogen regulation of secondary metabolism have not been described, it is presumed that the binding of ancillary proteins to DNA around the sites of transcription initiation will be important.

1.8 Role of "Signal" Molecules in Controlling Development

As described in the previous section, the principal factor controlling the onset of antibiotic biosynthesis is probably the deficiency of one or more nutrients required for growth. Recently, the possible role of an "endogenous" signal molecule, guanosine 3'–
diphosphate 5'-diphosphate (ppGpp), in mediating the switch to antibiotic production has been investigated. In *E. coli*, ppGpp mediates the stringent response, brought about by amino acid starvation (Cashel, 1969; 1975). The depletion of amino acids in the growth medium leads to a rise in the intracellular concentration of ppGpp. In turn, this results in a rapid decrease in the rates of accumulation of tRNA and rRNA, due to changes in the transcription rate and promoter specificity of RNA polymerase. A similar type of response has also been demonstrated in streptomycetes (Ochi, 1986). Relaxed mutants (unable to elicit the stringent response) of various *Streptomyces* spp., for example *S. coelicolor* and *S. griseus*, are also impaired in antibiotic production and have reduced levels of sporulation (Ochi, 1990a and 1990b). Recently, studies on *S. coelicolor* A3(2) have shown that ppGpp could not be detected in the early and mid exponential growth phases, but increased levels were observed at the end of exponential growth (Strauch *et al.*, 1990). The *actIII* gene, which is involved in actinorhodin production, was transcribed immediately after the culture entered stationary phase. After nutrient shiftdown, a switch to secondary metabolism was demonstrated by the appearance of the *actIII* transcript one hour later. It was proposed that ppGpp, along with other factors, plays a major role in determining the onset of secondary metabolism (Strauch *et al.*, 1990). It would be interesting to establish how the expression of *bldA* responds to changes in the guanosine nucleotide pools, since in contrast to the general decrease in the levels of other tRNA transcripts, an increase in the abundance of *bldA* transcripts would be expected prior to secondary metabolism (Chater, 1989). In *E. coli*, there is some evidence for the positive regulation of some genes during the stringent response (Cashel and Rudd, 1987).

Another interesting example of a "signal" molecule which controls secondary metabolism and morphological differentiation in some *Streptomyces* spp. is A-factor (2-(6'-methylheptanoyl)-3R-hydroxymethyl-4-butanolide). In *S. griseus*, A-factor induces sporulation, streptomycin biosynthesis and resistance, even at extremely low concentrations (10^-9 M) (Hara and Beppu, 1982a and 1982b). This factor may have a key role in synchronising antibiotic production and resistance by preventing
hyphal filaments, whose growth is not yet limited by nutrient depletion, being killed by streptomycin produced from other regions of the colony which have reached the end of vegetative growth. A protein has been isolated that binds A-factor in *S. griseus* (Mikaye et al., 1989). Mutants of *S. griseus* that are deficient in A-factor–binding protein do not require A-factor for streptomycin biosynthesis or sporulation. Indeed, the mutants undergo morphological and physiological development at an earlier stage of growth than the wild type. It was proposed that A-factor–binding protein is a repressor-type regulator which binds to regulatory regions within some or all of the genes for streptomycin biosynthesis, blocking transcription. A-factor is assumed to interact with the repressor protein, rendering it unable to bind DNA and thus, inducing transcription (Miyake et al., 1990).

A-factor analogues, with a butanolide ring as the nucleus of their structure, have also been reported, for example I-factor controlling anthracyclic biosynthesis in *S. viridochromogenes* (Gräfe et al., 1982), a factor controlling cytodifferentiation and anthracyclic production in *S. bikiniensis* and *S. cyaneofuscatus* (Gräfe et al., 1983) and "inducing factors" controlling virginiamycin production in *S. virginiae* (Nihara et al., 1988). Recently, a protein has been detected which binds the "inducing factor" of *S. virginiae* (cited in Miyake et al., 1990).

A-factor production was shown to be distributed widely among the streptomycetes (Hara and Beppu, 1982). The role of A-factor is not obvious for most of these species, other than the streptomycin-producing *S. griseus* and *S. bikiniensis*. A-factor-deficient mutants of *S. coelicolor A3(2)* and *S. lividans* showed no detectable phenotype (Horinouchi et al., 1986 and 1987). Interestingly, no binding protein specific for A-factor could be detected in either of these species (Miyake et al., 1990). It is unlikely, therefore, that A-factor has a role in controlling the development of these *Streptomyces* spp.

1.9 RNA Polymerase Heterogeneity in *Streptomyces*

The controlled expression of alternative sigma factors (σ), which direct the transcription of different sets of genes, is presumed to have
an integral role in coordinating both the morphological and physiological development of *Streptomyces*. Alternative sigma factors and thus, alternative RNA polymerase holoenzymes (Eo) have a key role in the control of sporulation in *B. subtilis* (Losick and Pero, 1981; Pero et al., 1982), T4 phage development in *E. coli* (Kassavetis and Geiduschek, 1984), nitrogen utilisation in enteric bacteria (Hirshman *et al.*, 1985; Hunt and Magasanik, 1985), the production of heat shock proteins in *E. coli* (Grossman *et al.*, 1984; Grossman *et al.*, 1986) and the expression of flagellar and chemotaxis genes in *E. coli*, *Salmonella typhimurium* and *B. subtilis* (Helmann and Chamberlain, 1987; Chamberlain, pers. comm.). Recently, the development of spores from aerial hyphae in *S. coelicolor* has been shown to be under the control of an alternative sigma factor, $\sigma^{\text{whiC}}$ (Chater *et al.*, 1989). Interestingly the amino acid sequence of $\sigma^{\text{whiC}}$ closely resembles $\sigma^{28}$ of *B. subtilis*, the sigma factor involved in directing expression of the chemotaxis and motility regulons. This suggested to Chater and colleagues that Eo$^{\text{whiC}}$ may make contact with promoters whose -35 and -10 regions conform to the consensus sequence derived from promoters recognised by Eo$^{28}$ from *B. subtilis*. In support of this hypothesis, the cloning of a DNA fragment containing a $\sigma^{28}$-dependent promoter from *B. subtilis*, on a high-copy-number plasmid, caused reduced sporulation in *S. coelicolor* (Chater *et al.*, 1989).

RNA polymerase heterogeneity was first demonstrated in *S. coelicolor* using the *veg* and *ctc* promoters from *B. subtilis* as substrates in *in vitro* transcription "run-off" experiments (Westpheling *et al.*, 1985). The *veg* promoter belongs to the major eubacterial class, which is recognised by Eo$^{43}$ in *B. subtilis*, while the *ctc* promoter is transcribed by a minor form of RNA polymerase holoenzyme, Eo$^{37}$. Fractionation of crude RNA polymerase preparations from *S. coelicolor* permitted the isolation of two distinct forms of RNA polymerase, Eo$^{35}$ and Eo$^{48}$, that recognised *veg* and *ctc* respectively. The *ctc*-transcribing holoenzyme, but not the *veg* form, was also shown to transcribe the *endoH* gene (encoding a secreted endonuclease) from *S. plicatus*. The *endoH* promoter resembles the consensus for *B. subtilis* promoters recognised by Eo$^{37}$. 
Chapter 1  General Introduction 25

The *dagA* gene (encoding an extracellular agarase gene) is transcribed by four promoters, whose sequences are markedly heterologous (Buttner *et al*., 1987). Using the *dagA* gene as a template in *in vitro* transcription assays, three different transcribing activities were resolved. Each RNA polymerase holoenzyme could transcribe only one of the four promoters (Buttner *et al*., 1988). The Ec49, which was identified previously and recognizes *ctc*, transcribes from the *dagAp3* promoter, whereas a novel species, Ec26, recognizes the *dagAp2* promoter. Circumstantial evidence suggested that the third holoenzyme, which recognizes the *dagAp4* promoter, is Ec35 (veg—transcribing holoenzyme) which was identified previously. Consistent with this proposal, the promoter sequence of *dagAp4* resembles closely the consensus for the major class of eubacterial promoters. The *dagAp1* promoter could not be transcribed by any of the purified holoenzymes and so, the authors suggested that another form of RNA polymerase must recognize this promoter (Buttner *et al*., 1988). It is also possible that ancillary proteins are required to activate transcription from *dagAp1*. The labile nature of the *dagAp1*-transcribing activity during purification is consistent with this proposal. The galactose operon of *S. coelicolor* is transcribed from two promoters (Westpheling and Brawner, 1989). Transcription from *galp1* is repressed by growth on glucose and is induced to high levels by growth on galactose. Constitutive transcription from *galp2* occurs at low levels. The *galp1* and *galp2* promoters are thought to be transcribed by Ec26 and Ec35 respectively.

Using an oligonucleotide probe, designed from a sequence of 10 amino acids that are conserved completely between σ33 of *B. subtilis* and σ70 of *E. coli* (the "housekeeping" sigma factors encoded by *rpoD*), four genes from *S. coelicolor* have been cloned and sequenced (Tanaka *et al*., 1988). Each gene encodes a predicted product with high sequence similarity in the regions conserved between σ70 of *E. coli* and σ33 of *B. subtilis* (*rpoD* boxes). These genes have been designated *hrdA*, *hrdB*, *hrdC* and *hrdD* (homologous to *rpoD*). Independently, the *rpoD* gene of *Myxococcus xanthus* was used also to clone *hrdB*, *hrdC* and *hrdD* (Buttner *et al*., 1990). The degree of similarity between the predicted gene
products of the \textit{hrd} genes and those of the \textit{rpoD} genes of \textit{E. coli} and \textit{B. subtilis}, far exceeds that observed between different $\sigma$ factors for minor classes of promoters (Tanaka \textit{et al.}, 1988).

Gene disruption experiments have shown that \textit{hrdA}, \textit{hrdC} and \textit{hrdD} mutants are viable and are unaffected apparently in differentiation, gross morphology and antibiotic production (Buttner \textit{et al.}, 1990; Takahashi \textit{et al.}, cited in Buttner \textit{et al.}, 1990). The inability to isolate \textit{hrdB} mutants using an integration system based on the actinophage $\phi$C31, suggested that the $\sigma$ factor encoded by this gene may be essential for the transcription of the general "housekeeping" genes. Furthermore, it was also noted that the putative $\sigma$ factors encoded by \textit{hrdA}, \textit{hrdC} and \textit{hrdD} appeared to be unable to complement the transcribing activity encoded by \textit{hrdB} (Buttner \textit{et al.}, 1990).

It has been proposed that closely-related families of $\sigma$ factors may recognise distinct subsets of promoters with only minor differences in the sequence or spacing of their "-35" and "-10" boxes (Westpheling and Brawner, 1989). Although RNA polymerase heterogeneity has been demonstrated in \textit{Streptomyces}, disappointingly little is known about the discrete gene sets which are recognised by the alternative sigma factors.

1.10 Polyketide biosynthesis

The diverse range of secondary metabolites produced by the \textit{Streptomyces} can be classified on the basis of their biosynthetic origins (Turner and Aldridge, 1983). The following products of primary metabolism were identified as being important precursors for secondary metabolites; (i) small carboxylic acids, (ii) amino acids and intermediates of the shikimate pathway, (iii) nucleosides and (iv) intermediates of the tricarboxylic acid cycle.

The tetracyclines, along with the anthracyclines, macrolides, polyethers, ansamycins, polyenes and isochromanequinones are members of the polyketide class of antibiotics, which are synthesized from small carboxylic acids, in a manner analogous to fatty acid biosynthesis. The
Figure 1.6. Schematic representation of the biosynthesis of fatty acids and polyketides. The circle labelled FAS or PKS represent the fatty acid synthases or polyketide synthases. These complexes carry two thiol groups, one on the $\beta$-ketoacyl synthase (□) and the other on the acyl carrier protein (O). Also labelled are the different functions required for each step; acetyl transferase (AT), acyl transfer reaction (TR, which is not unambiguously assigned to a specific component of the complex), malonyl transferase (MT), $\beta$-ketoacyl synthase (KS), ketoreductase (KR) dehydratase (DH), enoyl reductase (ER), palmityl transferase (PT) or thioesterase (TE) involved in the release of palmityl CoA (X=CoA) or free palmitic acid (X=OH) respectively. From Hopwood and Sherman, 1990.
similarity of these reactions was demonstrated convincingly by the
inhibition of both these pathways using the antibiotic cerulenin (Arison
and Omura, 1974). Recently, DNA regions known to be involved in the
earliest stages in polyketide biosynthesis have been sequenced. The
predicted protein products of these sequences have compelling homology
to the β-ketoacyl synthase, β-ketoacyl reductase and acyl carrier
protein (ACP) of fatty acid biosynthesis (for review, see Hopwood and

The synthesis of the fatty acid palmitate will be considered
(Figure 1.6) to provide a basis for the discussion of polyketide
assembly. Palmitic acid is an important precursor for the synthesis of
phosphoglycerides, which are the polar lipids found almost exclusively in
biological membranes. This fatty acid is a fully-saturated 16 carbon
polyketide which is synthesized from one acetyl-CoA unit and seven
malonyl-CoA units, by a complex with eight enzymatic activities, called
the fatty acid synthase (FAS). The malonyl-CoA extender unit is
formed from acetyl-CoA and CO₂ by the action of acetyl-CoA carboxylase.
The starter acetyl group is transferred, by the action of an acetyl
transferase, to the ACP which in turn transfers the acetyl group to the
active site of the β-ketoacyl synthase. Condensation then ensues
between the acetyl starter unit and a malonyl extender unit, which is
attached to the ACP vacated by the acetyl group, resulting in the loss
of CO₂ (the same CO₂ molecule used to produce malonyl-CoA from
acetyl-CoA). The ACP transfers the growing polyketide chain, in strict
sequence, to the active sites of ketoreductase, dehydratase and enoyl
reductase where the keto group from the malonyl CoA extender unit is
converted first to a hydroxyl group, then to an enoyl group and finally,
to an alkyl group. After this sequence of events the growing
polyketide chain is transferred back to the active site of β-ketoacyl
synthase. Only now can the second malonyl-CoA extender unit attach to
the ACP and participate in a further round of condensation-reduction-
dehydration-reduction. The remaining malonyl-CoA units are added in a
reiterant cycle of reactions to form palmitoyl-ACP. A palmityl transferase
then releases the palmitic acid from the FAS complex. Therefore, a
typical FAS is a multivalent system involving eight functional units;
acetyl, malonyl and palmityl transferases, ACP, ketoacyl synthase, keto reductase, dehydratase and enoyl reductase. The fatty acid synthase that catalyses these steps varies in different life forms. In most bacteria and plant plastids, the multienzyme complex is formed from eight separate polypeptides (Type II FAS). In contrast, the same eight functions are carried out by domains on a single, multifunctional polypeptide (Type I FAS) in invertebrates. The yeast, *Saccharomyces cerevisiae*, represents an intermediate case in which the FAS consists of two polypeptides, carrying three and five catalytic activities respectively.

Despite the obvious similarity in the chemistry of the biosynthesis of polyketide backbones and fatty acids, PKSs appear to be able to introduce complexity at five different levels. (i) The cycle of reduction-dehydration-reduction that follows the addition of two carbons in fatty acid biosynthesis is omitted or curtailed at some or all of the points in polyketide assembly. Thus, the PKS has to choose between four possibilities at each step in assembly and depending on the decision, leaves alkyl, enoyl, hydroxyl or keto groups at different carbon atoms (the occurrence of keto groups on the carbon backbone gives rise to the name "polyketide"). As in fatty acid biosynthesis, the modification of the keto groups probably takes place immediately after condensation, rather than after the complete backbone has been formed. This has been suggested by the incorporation of intermediates, predicted from a FAS-like progressive pathway, into the small polyketide aspyrrole in *Aspergillus*. Putative intermediates that contained functional groups, which could not be predicted from a FAS-type progression, were not incorporated (Staunton, pers. comm.). Furthermore, intermediates in the assembly of the polyketide backbone of the mycinamicins and tylosin have been isolated, which show the pattern of reduction, dehydration, reduction and stereochemistry found in the final product (Kinoshita *et al.*, 1988 and Huber *et al.*, 1990, cited in Hopwood and Sherman, 1990). (ii) Although polyketide chain initiation is primed commonly with acetyl-CoA, this is not an absolute requirement. For example, *Saccharopolyspora erythraea* uses propionate as a starter unit in the biosynthesis of erythromycin. In contrast to FAS, many PKS make a
choice of extender unit from a range of possibilities including acetate, propionate, butyrate and occasionally more complex alternatives. (iii) As a consequence of the condensation-reduction-dehydration-reduction cycle being broken during the assembly of many polyketides, the carbon chains often contain chiral centres, unlike the products of FAS which are achiral. (iv) The biosynthesis of many polyketides including oxytetracyclines, actinorhodin, granaticin, and tetracenomycin, involves cyclization or aromatisation of the polyketide backbone. The protein products of act ORF4 and gra ORF 4, are deduced to contain a β-ketoacyl cyclase activity as they can complement actVII mutants which accumulate an incorrectly cyclised shunt product, mutactin (Zhang et al., 1990). Due to significant homology to the deduced products of act and gra ORF4, the predicted products of whiE\textsc{v} and tcm ORF4 are also thought to have cyclase activity (Hopwood and Sherman, 1990). The later stages of monensin biosynthesis involve epoxidases which generate an intermediate that undergoes a cascade of ring closures, to form the heterocyclic rings seen in this antibiotic (Cane et al., 1983). The three oxygens at positions 7, 8 and 9 (Figure 1.7) of the cyclic ethers, are derived from molecular oxygen by epoxidation of C=C double bonds. The other oxygen atoms, including the two in the remaining cyclic ethers are derived from their respective propionate or acetate extender units. (v) The polyketide structure is modified usually by the addition of side groups to the polyketide nucleus. Unusual sugar motifs are attached often to macrolide antibiotics. For example, the deoxysugars mycarose and mycinose and the deoxy-amino sugar mycaminose, are added to the tylactone ring during the later stages of tylosin biosynthesis (Baltz et al., 1981). The biosynthesis of oxytetracycline (Section 1.14) provides an example of a polyketide nucleus which is modified by the addition of numerous side chains. Several antibiotics have bilateral symmetry and are thought to be produced by dimerisation of two identical polyketide nuclei. Examples include, actinorhodin, produced by \textit{S. coelicolor} (Section 1.11) and chlorothricin, produced by \textit{S. antibioticus} (Lee et al., 1986).

The biosynthesis of the polyether monensin A by \textit{S. cinnamonesis} provides an excellent example of the complex "programming" required to
Figure 1.7. The proposed biosynthetic route to monensin A. The hypothetical polyketide skeleton is derived by the condensation of five acetyl (A), seven propionyl (P) and one butryl (B) residues in the order shown. Also indicated are the polyketide group substitutions; alkyl (W), enoyl (X), hydroxyl (Y) and keto (Z) functionalities. Modified from Cane et al., 1983.
synthesize the polyketide backbone of some antibiotics. During polyketide assembly, the monensin PKS has to make 37 decisions: choices of 13 building units in a specific order from acetate, propionate and butyrate; a choice of four types of chemistry after each of the 12 condensations, to give either keto [Z], hydroxyl [Y], enoyl [X] or alkyl [W] and correct choice of stereochemistry at 12 points in the assembly of the carbon skeleton. The monensin PKS produces only one main polyketide product from the vast number of theoretical possibilities (Figure 1.7).

1.11 Actinorhodin Biosynthesis

The amenability of *S. coelicolor* to both *in vivo* and *in vitro* genetic manipulations and the relative ease with which actinorhodin producing colonies can be detected (on alkali medium they are deep blue) have contributed to the biosynthesis of actinorhodin being a paradigm for the chain assembly of other polyketides, including oxytetracycline. The primary carbon skeleton of actinorhodin is derived from 16 acetate units. A series of mutants blocked in actinorhodin biosynthesis (*act* mutants) was isolated and classified into seven phenotypic classes (Rudd and Hopwood, 1979). The *act* loci were shown to be linked on a short segment of the *S. coelicolor* chromosome, by analysing the frequencies of recombinants after conjugal matings between *act* mutants and wild-type cultures. Pairwise cosynthesis between *act* mutants allowed the mutant classes to be arranged in a linear sequence of biosynthetic blocks; I, III, VII, IV, VI and V (Rudd and Hopwood, 1979). The *actII* mutants did not produce actinorhodin in cosynthesis tests with any of the other mutants and were presumed to be deficient in a positive-regulatory factor. The *actI* and *actIII* mutants failed to 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 other mutant classes to actinorhodin. This cosynthesis data suggested that *actI/III* mutants were defective for enzymes involved in the earliest steps of biosynthesis. The *actIII* mutants accumulated a red pigment, whereas *actI* mutants did not and were deduced, therefore, to be blocked at a later stage in the construction of the carbon skeleton (Rudd and Hopwood, 1979).
Figure 1.8. Organisation of ORFs in the PKS gene cluster for actinorhodin in *S. coelicolor*, granaticin in *S. violaceoruber*, tetracenomycin in *S. glaucescens*, the putative spore pigment in *S. coelicolor* and oxytetracycline in *S. rimosus*. From Hopwood and Sherman, 1990.
Chapter 1

General Introduction

It was suggested that genes or gene clusters responsible for the synthesis of polyketide chains, might have a common evolutionary origin (Malpartida and Hopwood, 1984). Using probes derived from segments of DNA that complemented actI and actIII mutants respectively, Southern analysis identified homologous fragments in genomic-DNA digests of several other polyketide-producing streptomycetes (Malpartida et al., 1987). Fragments cross-hybridizing with the actI probe, were identified in the oxytetracycline producer S. rimosus, the tetracenomycin producer S. glaucescens and the granaticin producer S. violaceoruber. Analysis of the predicted amino acid sequence of the actIII gene product revealed that it encodes a reductase (Hallam et al., 1988). No fragment hybridized with the actIII probe in genomic DNA digests of S. glaucescens. However, this was not unexpected as the biosynthesis of tetacenomycin C does not require the corresponding reduction.

1.12 Organisation of Polyketide Synthase Genes

The actI and actIII regions have been sequenced (Fernandez and Malpartida, cited in Hopwood and Sherman, 1990) along with the actVII region, which is colinear with the actI region and is thought to have a role in the cyclization of the polyketide chain (Zang et al., 1990). A number of ORFs have been identified, which encode protein products with convincing homology to enzymes involved in fatty acid synthesis. These ORFs have been designated the actinorhodin polyketide synthase (act PKS) genes. Sequence data has also been obtained for the PKS genes of granaticin (Sherman et al., 1989), oxytetracycline (Sherman, pers. comm.) and tetracenomycin (Bibb et al., 1989), which cross-hybridized with the act probes described above. The sequencing of DNA, which complements mutants of S. coelicolor that are deficient in spore pigmentation (whiE), has also revealed a set of ORFs that clearly correspond to those of known PKS genes (Davis and Chater, 1990).

The organisation of the act, gra, otc, tcm and whiE PKS genes has been compared in a recent review (Hopwood and Sherman, 1990). Each PKS gene cluster contains three characteristic ORFs, designated ORF 1-2-3 (except for the whiE ORFs which are called III-IV-V) (Figure 1.8).
The predicted protein sequences of ORF 1-2 show convincing homology to the fatty acid β-ketoacyl synthase of E. coli, encoded by fabB. The predicted protein sequence of ORF 3 in each cluster resembles strongly acyl carrier proteins (ACPs) from FAS in a variety of life forms, including bacteria, plants and animals. Immediately downstream of ORF 1-2-3 there is a fourth ORF that shows relatedness between different PKS clusters.

In contrast with the predicted products of ORF 1-2-3, the presumed products of the fourth ORFs only resemble each other in a specific domain. The act and gra ORF4 sequences show homology throughout their entire length. These particular ORFs are both thought to encode a bifunctional cyclase/dehydratase, as they complement actVII mutants, which accumulate mutactin and are thought to be defective in cyclization and an accompanying dehydration. The predicted product of whiE ORF VI, resembles only the amino-terminal half of the act and gra ORF4 proteins. It was deduced, therefore, that the whiE ORF VI may encode a monofunctional polyketide cyclase (Hopwood and Sherman, 1990). The tcmla ORF4 protein was thought originally to be a bifunctional cyclase/O-methyltransferase. However, recent evidence (described in Chapter 4), suggests that there may be an error in the sequence and that the tcml PKS, similar to whiE, may actually contain a monofunctional cyclase (Hutchinson, pers. comm.). The DNA sequence of the ORF4 region of S. rimosus has not been completed (Sherman, pers. comm.). However, a region of DNA immediately downstream of otc ORF1-2-3 does hybridize with actVII DNA (Butler et al., 1990).

The actIII gene (act ORF5), which is encoded on the opposite strand to ORF 1-2-3, is predicted to encode a ketoreductase (Hallam et al., 1988). Recently, the introduction of a plasmid carrying only the intact actIII (act ORF5) gene into the 2-hydroxyalkavinone producer Streptomyces galilaeus ATCC 31671 was shown to result in the exclusive formation of the hybrid antibiotic alkavinone (Bartel et al., 1990). The chemical analysis of alkavinone and other hybrid polyketides produced by S. galilaeus, when transformed with DNA from S. coelicolor containing different combinations of the actI, III, IV and VII loci, suggested that
the function of the act ORF5 product was to reduce the keto group at C-9 from the carboxy terminus to a hydroxyl group (Bartel et al., 1990). This proposal is consistent with the finding that the predicted sequence of the act ORF5 product has significant homology to other reductases (Hallam et al., 1988). The predicted amino acid sequence of the gra ORF5 and ORF6 proteins show high similarity to the predicted actIII gene product and possibly encodes a heterodimeric ketoreductase (Hopwood and Sherman, 1990). Sequencing of a fragment within the otc cluster, which cross-hybridized with actIII DNA (Butler et al., 1990), has revealed a potential protein coding region that is predicted to encode a product with high amino acid sequence identity to the deduced actIII product (Thamchaipenet, pers. comm.). The regions in the otc cluster, which hybridized to actI and actIII DNA, are separated by approximately 8 kb (Butler et al., 1990) indicating that the genetic organisation of the otc PKS is topologically different to the other PKS gene clusters so far analysed.

1.13 The Module Hypothesis

The biosynthesis of the macrolide antibiotic erythromycin A (erA) by Saccharopolyspora erythrae, is a complex pathway that includes; (i) the formation of a 14-membered macrolactone ring from the condensation of propionyl-CoA and 6 methylmalonyl-CoA residues, (ii) the synthesis and attachment of sugar residues to form erythromycin D (erD) and (iii) the hydroxylation and O-methylation of erD to form erA. Genes involved in the synthesis of the macrolactone ring are designated eryA and are thought to be closely linked in a 30 kb region of DNA. Cross-hybridization between different fragments within the 30kb segment and limited sequencing suggested that the eryA region was organised into six repeated modules of approximately 5 kb, each containing an acetyl transferase, ACP, β-ketosynthase and β-ketoreductase (Katz et al., 1990). As the assembly of erythromycin requires six condensation, it was proposed that each module contains the enzymatic activities required for the condensation and processing of only one of the six methylmalonyl-CoA extender units (Katz et al., 1990).

Recently, a 10kb segment of DNA from the eryA region, which was
Figure 1.9. Potential catalytic sites in the deduced protein of ORFA in the eryA region of the *Saccharopolyspora erythraea*. ACP, acyl carrier protein; KS, \( \beta \)-ketoacyl-ACP synthase; AT, acyltransferase or thioesterase and KR, \( \beta \)-ketoacyl-ACP reductase. Abbreviations: S, *SacI*; Sm, *SmaI*. Modified from Cortes et al. (1990).
cloned by chromosome walking from the \textit{ermE} resistance determinant, has been sequenced completely (Cortes \textit{et al.}, 1990). Analysis of the sequence revealed a PCR extending over 9.5kb. The deduced protein product contains nine separate portions, whose amino acid sequences are very similar to the amino acid sequences for the catalytic domains of known fatty acid and polyketide synthases (Figure 1.9). The N-terminal and C-terminal halves of the deduced protein each contain a set of four predicted activities in the same order, except for the presence of an additional acyltransferase (or thioesterase) domain at the C-terminus. It would appear that a single, multifunctional peptide, which resembles type II FAS in invertebrates, might catalyse more than two of the six possible condensation cycles. In contrast, relatively simple polyketide structures such as actinorhodin, granaticin, tetracenomycin and oxytetracycline would appear to be synthesised by multienzyme complexes, which contain single polypeptides for each catalytic activity (Hopwood and Sherman, 1990).

It was suggested that the large number of genes involved in the assembly of the macrolactone ring may reflect the complexity of the structure (Katz \textit{et al.}, 1990). It is conceivable that other macrolide antibiotics, including tylosin, avermectin and perhaps polyethers such as monensin will also require a number of PKS modules for assembly of the carbon chain. The sequencing of many of these clusters is currently underway and should prove enlightening in the near future.

1.14 Biosynthesis of Oxytetracycline

Oxytetracycline (5-hydroxytetracycline; OTC) is one of the commercially-important broad-spectrum, tetracycline series of polyketide antibiotics produced in large-scale fermentations by \textit{S. rimosus}. Within the tetracycline group, deoxycycline (6-deoxy-hydroxytetracycline), which is made by a chemical modification of OTC and chlorotetracycline (N-demethyl-7-chlorotetracycline; CTC), produced by \textit{S. aureofaciens}, also have important pharmaceutical properties. A large number of tetracyclines have been reported, some of which are made by microorganisms other than streptomycetes, for example, \textit{Nocardia} (Sinclair \textit{et al.}, 1962).
Figure 1.10. Biosynthesis of oxytetracycline. Indicated are the proposed steps at which various OTC-negative mutants are blocked. From Rhodes et al., 1981.
The plausible biosynthetic pathway for the tetracyclines was elucidated at the Lederle Laboratories (McCormick, 1966; McCormick and Jensen, 1969) by the biochemical analysis of intermediates produced by mutants blocked in production. Confirmation of the polyketide nature of the carbon skeleton was obtained by NMR spectroscopy of OTC, labelled with $^{13}$C-labelled acetate and malonate (Thomas and Williams, 1983).

Chemical considerations predict at least nine stages leading to oxytetracycline (Figure 1.10) (Rhodes et al., 1981); (i) the sequential condensation of eight malonyl-CoA extender units onto a malonamyl-CoA primer unit to produce a nonaketide and then cyclization to form the tetracycline nucleus, (ii) C-6 methylation to form the first isolatable intermediate 6-methylpretetramid (6MPT), (iii) C-4 hydroxylation to form 4-hydroxy-6MPT, (iv) C-6 oxidation to form 4-keto-anhydrotetracycline (4-amino-ATC), (v) C-4 transamination to form 4-keto-ATC, (vi) double N-methylation to form ATC, (vii) C-6 hydroxylation to form dehydrotetracycline (DHTC), (viii) C-5 hydroxylation to form dehydroxytetracycline (DHOTC) and finally (ix) CSF1-dependent reduction to form OTC.

Genetic analysis of mutants blocked in OTC production indicated that the oxytetracycline (otc) genes were located on the chromosome of S. rimosus (Pigac and Alacevic, 1979). The first integrated study of the genetics and biochemistry of OTC production was undertaken at the Pfizer Laboratories (Rhodes et al., 1981). A series of mutants blocked in OTC production was isolated and classified primarily on the basis of cosynthesis tests in liquid culture, into nine groups. Single representatives from each group (otc-4, -19, -20, -25, -56, -75, -90, -151, and -155) were selected for further analysis. Three of these mutants, otc-19, -56 and -155, were able to produce OTC when supplied with culture filtrate of the Coenzyme F$_{420}$-producing strain S. aureofaciens ATCC 13190. This suggested that these otc mutants were blocked in the synthesis of this cosynthetic factor, which is essential for the final enzymatic step in OTC biosynthesis. The remaining six classes of otc mutants were presumed to be blocked at six different steps in the biosynthetic sequence from malonyl-CoA to oxytetracycline.
Chapter 1 General Introduction

The absolute order in which the \textit{otc} -4, -20, -25, -75, -90 and -151 mutants were blocked could not be deduced, as each representative failed to cosynthesize \textit{OTC} with the other \textit{otc} mutants in agar strip test in any of the possible pairwise combinations (as described by Delic \textit{et al.}, 1969). The feeding of known intermediates of \textit{OTC} did, to some extent, delineate the steps which could be blocked (Rhodes \textit{et al.}, 1981). All of the \textit{otc} mutants, except \textit{otc}-25, could convert a derivative of \textit{DHTC} to \textit{OTC}. It was argued that the \textit{otc}-25 mutant could not be blocked in C-5 hydroxylation, as dehydroxytetracycline would probably be converted to tetracycline, which is biologically active: the \textit{otc}-25 mutant has an antibiotic-negative phenotype. The \textit{otc}-25 mutant was presumed, therefore, to be blocked in the reduction of \textit{DHOTC} to form \textit{OTC}. The \textit{otc}-4 mutant was able to convert \textit{ATC}, but not 4-amino-\textit{ATC}, to \textit{OTC} and was deduced, therefore, to be blocked in N-methylation of 4-amino-\textit{ATC}. Similarly, the \textit{otc}-75 mutant could not convert \textit{ATC} to \textit{OTC} and was thought consequently to be blocked in the C-6 hydroxylation that forms \textit{DHTC}. Having assigned the \textit{otc}-25, -4 and -75 mutants to the steps between 4-amino-\textit{ATC}, which would produce antibiotic-negative phenotypes if disrupted, it was concluded that \textit{otc}-20, -90 and -151 were blocked before the formation of 4-amino-\textit{ATC}. Mutant in the class represented by \textit{otc}-90 were pale and nonpigmented. As many of the intermediates and shunt products of the pathway are highly coloured, it was considered probable that the \textit{otc}-90 mutant was blocked in the earliest biosynthetic steps (Hunter, pers. comm.). The mutant classes represented by \textit{otc}-4, -20, -25, -75, -90, and -151 were assigned to putative loci \textit{otcD}, \textit{otcX}, \textit{otcA}, \textit{otcC}, \textit{otcY} and \textit{otcZ} respectively (Rhodes \textit{et al.}, 1981). The conversion of \textit{DHTC} to \textit{DHOTC} was attributed to a putative locus called \textit{otcB}.

Genetic mapping, using the representatives from each mutant group, suggested that the \textit{otc} genes were located in two distinct clusters (Rhodes \textit{et al.}, 1981). The first (located at 4 o'clock on the circular map of the \textit{S. rimosus} chromosome), was responsible for the formation of the tetracycline nucleus and its conversion to anhydrotetracycline (ATC), while the second (located at 10 o'clock),
Figure 1.11. Restriction map of the otc cluster. Hatched boxes (■■■■) denote the location of the otr resistance genes. Closed boxes (■) show the minimum DNA segments that can complement the OTC-negative classes of mutants (Butler et al., 1989). Stippled boxes (■■■■) display homology to act DNA (Butler et al., 1990). The open box (■■■■) denotes the position of the otcC gene located by "reverse genetics" (Binnie et al., 1989). The line (■■■■) delineates the region of DNA in pPFZ163 which confers the ability to produce OTC when introduced in S. lividans and S. albus (Binnie et al., 1989). Abbreviations: B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; K, KpnI; P, PstI; R, EcoRV; S, SacI; Sp, SphI. Modified from Butler et al., 1990.
Chapter 1  General Introduction 43

contained two sets of genes whose products either had an enzymatic role in the conversion of ATC to OTC or were involved in the biosynthesis of Coenzyme F₄₂₀ (also called CSF1).

Initially, the early steps encoded in the 4 o'clock were the focus of attention at the Pfizer Laboratories. Genomic DNA fragments of *S. rimosus*, which conferred resistance to OTC and complemented all of the *otc* mutants (*otc*-4, -20, -90 and -151) blocked in the "early" biosynthetic sequence from malonyl CoA to anhydrotetracycline (ATC), were cloned. The cloned DNA was linked physically within a 35 kb segment of the chromosome. The gene cluster is flanked by two resistance genes, which can independently confer protection against the antibiotic (Figure 1.11) (Butler et al., 1989).

At the Pfizer laboratories, attention turned subsequently to the functions involved in the conversion of ATC to OTC. The product of the *otcC* gene, anhydrotetracycline (ATC) oxygenase, has an easily assayable enzymatic activity. A "reverse genetic" approach was taken to clone the *otcC* gene and linked DNA, that is the 10 o'clock cluster. The ATC oxygenase was purified and the N-terminal sequence determined. Oligonucleotide probes were designed and used to clone a corresponding fragment of chromosomal DNA from *S. rimosus*. The screening of a cosmid library, using a probe derived from the cloned DNA, located the *otcC* gene within the cluster for the early steps in OTC production and not as expected (Rhodes et al., 1981) at a separate locus on the opposite side of the chromosomal map (Binnie et al., 1989). The hybridization results contradicted the original genetic mapping data and it became a formal possibility that all the structural genes for OTC biosynthesis were located in the 4 o'clock cluster. The ability of a plasmid carrying this gene cluster (on a 34 kb EcoRI fragment) to confer OTC production when introduced into the heterologous hosts *S. albus* and *S. lividans*, demonstrated convincingly that all the structural genes for OTC were indeed linked on a single segment of the *S. rimosus* chromosome (Binnie et al., 1989).
Recently, it has been shown that the ATC oxygenase also has a requirement for a flavin cosynthetic factor, which is distinct from CSF1 used by the tetracycline dehydrogenases (Vancurova et al., 1988). Although the \textit{otc-75} mutant is blocked clearly in the conversion of ATC to DHTC, it would now appear that the mutation, which was mapped outside the 4 o'clock cluster, disrupts the biosynthesis of a cofactor required for ATC oxygenase activity and not the expression of the \textit{otcc} gene product itself. There is also no convincing mapping data available to support the published location of the \textit{otc-25} mutation (Rhodes et al., 1981) to the 10 o'clock cluster (Binnie et al., 1989). The heterologous production of OTC, however, demonstrates clearly that the \textit{otcA} locus is located within the 34 kb EcoRI fragment, flanked by the two resistance genes. Attempts are underway to locate the \textit{otcA} locus by complementation of the \textit{otc-25} mutant (Binnie, pers. comm.).

Using a series of overlapping DNA fragments from the actinorhodin and oxytetracycline clusters, it has been possible to define regions which cross hybridize (Butler et al., 1990). Substantial hybridization signals were observed between the following regions (Figure 1.11). (i) The \textit{actI} and \textit{actIII} DNA probes both hybridized to the DNA segment that complements the \textit{otcY} mutants, which is consistent with the previous proposal (Hunter, pers. comm.) that the \textit{otcY} mutants were blocked in the earliest steps in biosynthesis. (ii) An \textit{otc} DNA fragment adjoining the \textit{actII}-hybridizing region hybridizes to \textit{actVII} DNA. This suggests that there could be a gene encoding a putative cyclase in the \textit{otc} cluster, downstream of the genes for the \(\beta\)-ketoacyl synthase and ACP, as found in the \textit{act}, \textit{gra}, \textit{tcm} and \textit{whiE} PKS clusters (Section 1.12). (iii) Part of the \textit{actVA} region, which has been implicated in hydroxylation (Hopwood et al., 1985; Cole et al., 1987) hybridizes to a segment of the DNA that complements \textit{otcX} mutants. The C-4 hydroxylation of 6-methylpretetramid is one of the possible steps, which could be blocked in the \textit{otcX20} mutant. The cross-hybridization between the \textit{actVA} region and the \textit{otcX} region might indicate a segment of DNA required for hydroxylation. (iv) A region of \textit{actVB} DNA hybridizes to a segment located between the regions that hybridize to \textit{actI} and \textit{actIII}. It has been proposed that mutations in the \textit{actVB} region could be
defective in a hydroxylation step, which is separate from actVa hydroxylation (Cole et al., 1987). In the biosynthesis of OTC, the conversion of dehydratetracycline to dehydroxytetracycline requires a C-5 hydroxylation. Should the region of actVs, which hybridizes to the otc cluster, contain a structural gene involved in hydroxylation, it is possible that the region of the otc cluster that cross-hybridizes with actVs DNA could encode the otcB gene.

The otrA resistance gene product (OtrA) protects the ribosomes from translational arrest by OTC, while the product of the otrB resistance gene results in decreased accumulation (increased efflux) of the antibiotic from the mycelium (Ohnuki et al., 1985). The otrA gene has been cloned from the Pfizer production strain S. rimosus M15883 (Rhodes et al., 1984)) and sequenced (Doyle, 1987). The predicted amino acid sequence of OtrA is very similar to the deduced amino acid sequence of elongation factor G (EF-G) from E. coli (Zengel et al., 1984). It has been proposed that OtrA may function as an alternative EF-G, which substitutes for the usual EF-G that is presumed to be inhibited by tetracycline (Doyle et al., 1991).

Tandem promoters were proposed for the otrA resistance gene (Doyle, 1987). The promoter otrAp1, which is nearest the putative translation initiation codon, resembles the consensus for the major class of eubacterial promoters. The sequence of the putative otrAp2 promoter was not determined as it lay just outside the region which was cloned and sequenced. Transcripts predicted to originate from these tandem promoters were detected in RNA isolated from S. rimosus M4018, under conditions which did not support antibiotic production. As no structure resembling a rho-independent terminator was detected in the intergenic region upstream of otrA, it was suggested further that otrA may be transcribed as part of a polycistronic message along with the biosynthetic gene, otcZ, which is located immediately upstream and is transcribed from the same strand as otrA during antibiotic biosynthesis (Doyle, 1987). The presence of tandem promoters suggested that the expression of otrA could be modulated during the growth cycle by alternative forms of RNA polymerase.
1.15 Scope of Present Work

The work presented in this thesis was concerned primarily with transcriptional analysis at the "otrA" end of the oxytetracycline cluster. The specific aims were to identify and characterise all of the promoters pertaining to the expression of the otrA resistance gene, to determine the temporal activity of these promoters at different stages during the fermentation of oxytetracycline and to analyse the regions of these promoter for any putative regulatory sequences involved in otrA expression.

During the course of this project DNA segments that complemented classes of mutants blocked in OTC production (Rhodes et al., 1981) were also sequenced and analysed.
CHAPTER 2
MATERIALS AND METHODS
INTRODUCTION

This chapter contains the general procedures used in the experiments which were the basis of this thesis. The chapter is divided into three main sections for convenience; (2.1) bacterial strains and vectors, (2.2) microbiological techniques and standard media and (2.3) nucleic acid isolation and manipulation.

2.1 BACTERIAL STRAINS AND VECTORS

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

<table>
<thead>
<tr>
<th>STRAIN</th>
<th>GENOTYPE</th>
<th>REFERENCE/SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Escherichia coli strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CB51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS941</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptomyces lividans strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TK24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TK54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TK64</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Streptomyces rimosus strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M15883</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 2.1 Bacterial strains

Escherichia coli strains

<table>
<thead>
<tr>
<th>STRAIN</th>
<th>GENOTYPE</th>
<th>REFERENCE/SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB51</td>
<td>dam, ara, thi, ( \Delta(\text{lac-pro}) )</td>
<td>Chris Boyd</td>
</tr>
<tr>
<td>DS941</td>
<td>recF143, proA7, str31, thr1, leu6, tsx33, mt12, his4, argE3, lacY*, lacZ(\overline{M})15, lacI(q), galK2, ara14, supE44, xyI5.</td>
<td>Dave Sherratt</td>
</tr>
<tr>
<td>TG1</td>
<td>supE, hsd(\Delta5), thi, ( \Delta(\text{lac-proAB}) )</td>
<td>Gibson (1984)</td>
</tr>
</tbody>
</table>

Streptomyces lividans strains

<table>
<thead>
<tr>
<th>STRAIN</th>
<th>GENOTYPE</th>
<th>REFERENCE/SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK54</td>
<td>his2, leu2, spc1.</td>
<td>Hopwood et al., (1985a)</td>
</tr>
<tr>
<td>TK64</td>
<td>pro2, str6.</td>
<td>Hopwood et al., (1985a)</td>
</tr>
</tbody>
</table>

Streptomyces rimosus strains

<table>
<thead>
<tr>
<th>STRAIN</th>
<th>GENOTYPE</th>
<th>REFERENCE/SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>M4018</td>
<td>OTC(R)</td>
<td>Rhodes et al., (1981)</td>
</tr>
<tr>
<td>M15883</td>
<td>OTC(R)</td>
<td>Binnie et al., (1989)</td>
</tr>
</tbody>
</table>
### TABLE 2.2 Plasmids

<table>
<thead>
<tr>
<th>PLASMID</th>
<th>DESCRIPTION*</th>
<th>REFERENCE/SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDM130</td>
<td>573bp \textit{SmaI}-\textit{BamHI} \text{ fragment} from pGLW7 in pIJ486.</td>
<td>Doyle (1987)</td>
</tr>
<tr>
<td>pGLW7</td>
<td>2.6kb \textit{SstI2}-\textit{BamHI} \text{ fragment encoding \textit{otrA}} in pT7-1.</td>
<td>Iain Hunter</td>
</tr>
<tr>
<td>pIJ2843</td>
<td>--</td>
<td>Clayton and Bibb (1990)</td>
</tr>
<tr>
<td>pIJ486</td>
<td>--</td>
<td>Ward \textit{et al.}, (1986)</td>
</tr>
<tr>
<td>pKO500</td>
<td>--</td>
<td>Close and Rodriguez (1982)</td>
</tr>
<tr>
<td>pPFZ46</td>
<td>6.0kb \textit{SstI2}-\textit{SstI13} \text{ fragment encoding \textit{otcC}, \textit{otcZ} and \textit{otrA}} in pPFZ34.</td>
<td>Pfizer</td>
</tr>
<tr>
<td>pPFZ105</td>
<td>5.5kb \textit{PstI8}-\textit{PstI18} \text{ fragment in pPFZ34.}</td>
<td>Pfizer</td>
</tr>
<tr>
<td>pPFZ513</td>
<td>1.0kb \textit{BclI5}-\textit{PstI8} \text{ fragment in pUC18.}</td>
<td>Pfizer</td>
</tr>
<tr>
<td>pUC18/19</td>
<td>--</td>
<td>Yanisch-Perron \textit{et al.}, (1985)</td>
</tr>
<tr>
<td>pKME805</td>
<td>305bp \textit{BclI5}-\textit{KpnI7} \text{ fragment from pPFZ513 in pUC18. \textit{BclI5} site nearest the \textit{EcoRI} site in the poly linker.}</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>pKME807</td>
<td>722bp \textit{KpnI7}-\textit{PstI8} \text{ fragment from pPFZ513 in pUC18.}</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>pKME808</td>
<td>805bp \textit{BamHI6}-\textit{PstI9} \text{ fragment from pPFZ513 in pUC18.}</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>pKME917</td>
<td>1.6kb \textit{SmaI4}-\textit{SmaI9} \text{ fragment from pPFZ46 in pUC19. \textit{SmaI4} site nearest the \textit{EcoRI} site in the poly linker.}</td>
<td>Chapter 4</td>
</tr>
<tr>
<td>pKMS605</td>
<td>\textit{HindIII}/\textit{EcoRI} insert from pKME805 in pIJ486.</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>pKMS607</td>
<td>\textit{HindIII}/\textit{EcoRI} insert from pKME807 in pIJ486.</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>pKMS608</td>
<td>\textit{HindIII}/\textit{EcoRI} insert from pKME808 in pIJ486.</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>pKMSOCPl</td>
<td>259bp \textit{Sau3AI} \text{ fragment containing \textit{otcCpl}/\textit{otcXpl} promoter region in pIJ2843.}</td>
<td>Chapter 6</td>
</tr>
</tbody>
</table>

*the numbering system used to designate restriction sites in this thesis is new and is not that of Butler \textit{et al.}, (1989)
<table>
<thead>
<tr>
<th>BACTERIOPHAGE</th>
<th>DESCRIPTION</th>
<th>REFERENCE/SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>mL6A 2.1kb SmaI\4-\9 fragment in M13mp18.</td>
<td>Smal site being nearest the EcoRI site in the polylinker.</td>
<td>Pfizer</td>
</tr>
<tr>
<td>mL6C</td>
<td>As mL6A except the Smal site is nearest the EcoRI site.</td>
<td>Pfizer</td>
</tr>
<tr>
<td>mL6A 2.1kb SmaI\4-\9 fragment in M13mp18.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mL6C</td>
<td>As mL6A except the Smal site is nearest the EcoRI site.</td>
<td></td>
</tr>
<tr>
<td>mKM804 387bp HincII-BamHI\6 fragment from pGLW7 in M13mp18.</td>
<td></td>
<td>Chapter 3</td>
</tr>
<tr>
<td>mKM905 305bp BclI\5-KpnI\7 fragment from pPFZ513 in M13mp19. BclI\5 site nearest the EcoRI site in the polylinker.</td>
<td></td>
<td>Chapter 4</td>
</tr>
<tr>
<td>mKM803/903 250bp PstI\8-Smal\9 fragment from pKME917 in M13mp18/19.</td>
<td></td>
<td>Chapter 4</td>
</tr>
<tr>
<td>mKM805/905 HindIII/EcoRI fragment from pKME805 in M13mp18/19.</td>
<td></td>
<td>Chapter 4</td>
</tr>
<tr>
<td>mKM807/907 HindIII/EcoRI fragment from pKME807 in M13mp18/19.</td>
<td></td>
<td>Chapter 4</td>
</tr>
<tr>
<td>mKM817/917 HindIII/EcoRI fragment from pKME917 in M13mp18/19.</td>
<td></td>
<td>Chapter 4</td>
</tr>
<tr>
<td>mKM810/910 965bp SphI\2-SstI\3 fragment from pPFZ105 in M13mp18/19.</td>
<td></td>
<td>Chapter 6</td>
</tr>
<tr>
<td>mKM88/98 Sau3AI insert from the pIJ2843 construct in isolate 8, removed as HindIII/EcoRI fragment and inserted into M13mp18/19.</td>
<td></td>
<td>Chapter 6</td>
</tr>
<tr>
<td>mKM812/912 Sau3AI insert from the pIJ2843 construct in isolate 12, removed as HindIII/EcoRI fragment and inserted into M13mp18/19.</td>
<td></td>
<td>Chapter 6</td>
</tr>
<tr>
<td>mKM815/915 Sau3AI insert from the pIJ2843 construct in isolate 15, removed as HindIII/EcoRI fragment and inserted into M13mp18/19.</td>
<td></td>
<td>Chapter 6</td>
</tr>
</tbody>
</table>
2.2 MICROBIOLOGICAL TECHNIQUES AND STANDARD MEDIA

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, Detroit, Michigan, USA and Sigma Chemical Co. Ltd, Poole, Dorset.

2.2.1 STANDARD MEDIA FOR THE PROPAGATION OF E. COLI

L-Broth: 10g tryptone, 5g yeast extract, 5g NaCl, 1g glucose, 20mg thymine, 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 K₂HPO₄, 2g KH₂PO₄, 4g (NH₄)₂SO₄, 0.25M trisodium citrate, 0.1g MgSO₄.7H₂O, 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 Mingoli (D&M) Salts (X4): 28g K₂HPO₄, 8g KH₂PO₄, 16g (NH₄)₂SO₄ 1g trisodium citrate, 0.4g MgSO₄.7H₂O, made up to 1 litre with distilled water.

2.2.2 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, 0.5g K₂HPO₄, 0.2g MgSO₄.7H₂O, 0.01 FeSO₄.7H₂O, 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$_2$SO$_4$, 20.2g MgCl$_2$6H$_2$O, 5.9g CaCl$_2$2H$_2$O, 20g glucose, 6g proline, 0.2g casamino acids, 4ml trace elements solution (Hopwood et al., 1985a), made up to 1 litre in distilled water. R2B- 11.5g MOPS, 10g yeast extract, 203g sucrose, adjusted to pH 7.4 with NaOH, made up to 1 litre in distilled water. Equal volumes of R2A (melted and cooled to 55°C) and R2B plus 1ml of 1% (w/v) KH$_2$PO$_4$ were combined prior to use.

R9 Medium: R9A- 44g agar, 0.5g K$_2$SO$_4$, 8.2g MgCl$_2$6H$_2$O, 4.7g CaCl$_2$.2H$_2$O, 4g NaNO$_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°C) and R9B were mixed with 1ml of 1% (w/v) KH$_2$PO$_4$ 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 1 litre in distilled water.

Trusoya Medium 1 (TS1): soya flour, starch, vegetable oil, inorganic salts and MOPS (pH7.5). The exact composition of this medium cannot be revealed as it is used commercially by Pfizer for strain improvement.

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.
2.2.3 STERILISATION

All growth media were sterilised by heating to 120°C for 15 mins in an autoclave. Supplements and buffer solutions were heated to 108°C and CaCl₂ to 114°C for 10 mins. Heat-labile solutions, such as amino acids, were sterilised by filtration through Nalgene 0.22μm pore membranes.

2.2.4 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 a 5° from the horizontal) was inoculated with 150μl of a spore or mycelial fragment suspension and incubated at 30°C. 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°C. Spores could then be harvested immediately or stored at -20°C 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 aliquoted and frozen at -20°C. Aliquots were thawed and thoroughly mixed using a vortex, prior to use. The titre of the spore suspension was determined after storage at -20°C 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.

### 2.2.5 *E. coli* Growth Conditions

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 50μg.ml⁻¹). The volume of broth inoculated depended on the quantity of plasmid required. Routinely, 5ml and 200ml cultures were used for small and large scale plasmid preparations, respectively (see section 2.3.1). For the preparation of competent cells, liquid cultures of *E. coli* DS941 or CB51 were grown in L broth while *E. coli* TG1 was grown in 2YT. To maximise aeration of the culture, the volume of the Erlenmeyer flask was at least five times that of the broth. All cultures were incubated at 37°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°C.

### 2.2.6 Growth of *Streptomyces* Mycelium

Erlenmeyer 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 MgCl₂. 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 (10⁵–10⁶ 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°C at ca. 250 rpm).

### 2.2.7 Quantification of *Streptomyces* Growth

Optical density measurements gave a poor estimation of growth due to the formation of mycelial pellets and wefts in liquid cultures of
Streptomyces. Dry weights were determined, therefore, according to the method of Harvey (pers. comm.). Duplicate Whatman Glass Fibre Filters (GF/C, diameter 4.7cm) were numbered with a pencil and dried for 15 min in a microwave oven (600 W), which was set on reheat, and allowed to cool to room temperature in a desiccator for 30 min. The weights of the filters were recorded to three decimal places. Culture samples, usually 5ml, were filtered through the GF/C filters using a Millipore sintered glass vacuum filtration unit. The filters were then washed by passing 15ml of dH2O through the filtration system, dried in the microwave, allowed to cool in the desiccator and weighed as before. The dry weight (g.L\(^{-1}\)) was determined by deducting the original weight (g) of the filter from the weight recorded after filtration and then multiplying this value by the dilution factor. At least three independent dry weight determinations were used to calculate the average value for a particular time point.

2.2.8 QUANTIFICATION OF OXYTETRACYCLINE IN CULTURE SUPERNATANTS

Samples of liquid cultures (1ml) were extracted into 9ml of HCl (pH 1.7) for 30min and then filtered through Whatman No.1 papers. 1ml samples were then sent to the Pfizer Process Development Laboratories at Sandwich and quantified by isocratic, ion-paired, reverse-phase high performance liquid chromatography (hplc) using a C18 µ Bondabank column (Waters, Millipore House, Harrow, Middlesex, UK.) with a mobile phase of acetonitrile/water (3:7) containing 0.5g.L\(^{-1}\) of 1-hexanesulphonic acid, adjusted to pH 1.7 with sulphuric acid.

2.2.9 ANTIBIOTICS AND INDICATORS

The antibiotic concentrations used throughout for both liquid and plate selection were as follows:
**Chapter 2 Materials and Methods**

<table>
<thead>
<tr>
<th>DRUG</th>
<th>SELECTIVE STOCK SOLUTION</th>
<th>STORAGE TEMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>50µg.ml⁻¹ (water)</td>
<td>-20°C</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>1-900µg.ml⁻¹ (water)</td>
<td>-20°C</td>
</tr>
<tr>
<td>Thioestrepton</td>
<td>25µg.ml⁻¹ (DMSO)</td>
<td>4°C</td>
</tr>
</tbody>
</table>

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

X-gal (5-bromo-4-chloro-3-indolyl-β-galactosidase) was used in conjunction with IPTG to identify *E. coli* strains CB51, DS941 and TG1 containing pUC or M13mp vectors with inserts in their multiple cloning sites. Recombinants containing inserts are generally white while those lacking inserts are blue. X-gal was stored at a concentration of 20µg.ml⁻¹ in dimethylformamide (DMF) at -20°C while IPTG was stored at a concentration of 24µg.ml⁻¹ in dH₂O at -20°C. X-gal and IPTG were added to L-agar plates to a final concentration of 20µg.ml⁻¹ and 50µg.ml⁻¹, respectively.

### 2.2.10 PRESERVATION OF *STREPTOMYCES* AND *E. COLI* STRAINS

*Streptomycetes* spp. were preserved by freezing agar slants covered in spores at -20°C. The *Streptomycetes* spp. used in this work remain 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. *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% peptone (w/v) and frozen at -70°C. The strains were revived by scraping the surface of the frozen suspension with a toothpick and either inoculating liquid broth or streaking onto agar to isolate a single colony.

### 2.2.11 INTRODUCTION OF PLASMID DNA INTO *E. COLI*

Plasmids were introduced into *E. coli* DS941 and CB51 by genetic transformation. CaCl₂ treatment routinely yielded cells with a transformation efficiency of 10⁶-10⁷ µg⁻¹ plasmid.
Preparation of competent cells: An overnight culture of the recipient strain was diluted 1 in 100 into 30ml L-broth and incubated for 90-120 min to a density of approximately $10^8 \text{ml}^{-1}$ cells (OD$_{600}$ 0.45-0.55). The cells were harvested using a centrifuge (12000g, 5 min, 4°C) and resuspended in 10ml of ice-cold 50mM CaCl$_2$. The cells were pelleted again, resuspended in 1ml of ice-cold 50mM CaCl$_2$ and kept on ice for at least 15 min before use. At this stage glycerol could be added to the cells to 20% (v/v) and 200μl aliquots frozen at -70°C. The cells retained their competence over several months without any significant decrease in their transformation efficiency.

Transformation with plasmid DNA: 200μl aliquots of the competent cells were added to plasmid DNA (1-100ng) in a volume less than 10μl, mixed gently by inverting the microfuge tube and incubated on ice for up to 1 hour. The cells were heat-shocked (2 min, 42°C) and returned to ice for a further 15 min. An 800μl aliquot of L-broth was added and the cell suspension incubated at 37°C for 45 min to allow expression of the plasmid resistance genes. For ampicillin selection an expression stage as such was not required. The cells were spread on L-agar plates containing the appropriate antibiotic.

2.2.12 TRANSFECTION OF E. COLI TG1 WITH BACTERIOPHAGE M13

A 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.

Reagents: TFB; 10mM 2[N-morpholino]ethanesulphonic acid (MES) buffer (pH 6.3), 10mM RbCl, 45mM MnCl$_2$·4H$_2$O, 10mM CaCl$_2$·2H$_2$O, DTT/KAc; 2.25mM dithiothreitol, 40 mM potassium acetate (pH 6.0).

Preparation of competent cells: An overnight culture of E. coli TG1 grown in 2YT broth was diluted 1 in 100 into 30 ml of fresh 2YT. The culture was grown until the OD$_{600}$ was between 0.45 and 0.55. The culture was then transferred to a 50ml polypropylene centrifuge tube
(pre-rinsed with sterile distilled dH2O and pre-chilled on ice) and incubated on ice for 20 min. The cells were harvested by centrifugation (4000g, 4°C for 10 min), resuspended very gently in 2.5ml of pre-chilled TFB and incubated on ice for 15min. The cells were then treated with 100μl of dimethylsulphoxide (DMSO) and incubated for 5min on ice. This was followed by the addition of 100μl DTT/KAc and incubation on ice for a further 5min. The cells were ready to use after adding a second 100μl aliquot of DMSO and incubating on ice for 5min.

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 were added to the transfected cells, followed by 10μl of IPTG (24mg.ml⁻¹) and 50μl X-gal (20mg.ml⁻¹). The cells were then mixed and added to 2.5ml of molten water-agar (0.6% w/v, pre-cooled to 45°C), which were poured onto thoroughly dried minimal medium plates containing D & M salts, vitamin B1 (5μg.ml⁻¹) and glucose (2mg.ml⁻¹) supplements.

2.2.13 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 ZnCl₂, 200mg FeCl₃.6H₂O, 10mg CuCl₂.2H₂O, 10mg MnCl₂.4H₂O, 10mg Na₂B₄O₇.10H₂O and 10mg (NH₄)₆Mo₇O₄.4H₂O. Medium P; 5.73g N-tris[hydroxymethyl]methyl-2-aminoethanesulfonic acid (TES), 103g sucrose, 2.93g MgCl₂.7H₂O, 0.5g K₂SO₄, 3.68g CaCl₂.2H₂O, 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 K₂SO₄, 2ml trace elements (Hopwood et al., 1985a), 2.5mM MgCl₂, 2.5mM CaCl₂, KH₂PO₄ (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 microwave (600W) for 10s on the reheat setting and then mixed with 3ml of medium P.
Preparation of protoplasts: 30ml cultures were grown in the appropriate medium at 30°C. The optimal time to harvest the mycelium in order to recover the most competent protoplasts was species-dependent i.e. \textit{S. lividans} mycelium were harvested after 65hr and \textit{S. rimosus} after 48hr, respectively. The mycelium were pelleted at 12000g for 10 minutes and washed twice in 10.3% (w/v) sucrose. The pellet then resuspended in 4ml of lysozyme solution and incubated at 37°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 (12000g for 10 minutes) and washed twice in P medium. Finally, they were resuspended in 4ml of medium P, dispensed into 200μl aliquots and frozen at −70°C.

Transformation of protoplasts: The protoplasts were thawed 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°C before they were combined and 1ml of 1%(w/v)KH₂PO₄ 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 45min. They were then rotated 180°C and their relative positions reversed so that those at the front of the hood were placed at the back. After a further 45min, the plates were removed and left overnight at 30°C. 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 in this work for plasmid selection in *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°C 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*.

2.3 NUCLEIC ACID ISOLATION AND MANIPULATION

2.3.1 ISOLATION OF PLASMID DNA FROM *E. COLI* AND *STREPTOMYCES*

Protocols based on the alkaline lysis method (Birnboim and Doly, 1979) were used for the isolation of plasmid from small (5ml) or large (50-200ml) culture volumes of both *Streptomyces* spp and *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 CH₃COOK and 2M CH₃COOH. 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 isooamyl alcohol. DNAase-free RNAase; Pancreatic RNAase (RNAase A) was dissolved at a concentration of 10mg.ml^{-1} in dH₂O, heated to 100°C for 15min and allowed to cool slowly to room temperature. The RNAase was then aliquoted and stored at -20°C.

Large scale plasmid preparations: 200ml cultures of stationary phase cells were harvested using a centrifuge (12000g, 5min at 4°C). 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°C for 20min with 4mg.ml\(^{-1}\) and 1mg.ml\(^{-1}\) lysozyme solutions, respectively. *E. coli* cultures were incubated on ice for 5min. Then 8ml of Birnboim-Doly II solution were added and the solution left on ice for 5–10 min before 6ml of cold Birnboim-Doly III solution were 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°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.ethylidium bromide (CsCl.EtBr) gradient. The nucleic acid pellet was redissolved in 1 ml of dH\(2\)O and 4.5g of CsCl dissolved in 3.5ml of dH\(2\)O. 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 289,000g for 16 hours at 20°C. Two bands were visible in the gradients after centrifugation, a lower supercoiled plasmid band and an upper chromosomal and relaxed plasmid DNA band. The lower band was removed using a 1ml syringe and the EtBr removed by repeated extractions with water–saturated butanol. After dilution with 3 volumes of dH\(2\)O, 9 volumes of absolute ethanol were added. The precipitate was pelleted by centrifugation (27000g, 4°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\(2\)O. This procedure yielded very large amounts of pure plasmid DNA (up to 1mg from *E. coli* and 100μg from *Streptomyces* cultures) suitable for all *in vitro* manipulations.

Small scale plasmid preparations: This protocol was used for isolating plasmids from both *E. coli* and *Streptomyces* spp, without modification. Routinely, plasmids were isolated from 1.5ml of *E. coli* cultures and 3.0ml of *Streptomyces* cultures or 2cm\(^2\) patches of mycelial growth on agar plates. The cells were pelleted 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 thoroughly the suspension. 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 chromosomatal material was harvested by centrifugation (12000g, 4°C for 10min) in a microfuge. The supernatant was transferred to a fresh tube and extracted with half volumes of phenol.chloroform and chloroform. The nucleic acid was then precipitated by the addition of 2 volumes of ethanol and allowed to stand at room temperature for 3-5min. The precipitate was harvested by centrifugation in a microfuge (12000g, 4°C 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 dH2O containing DNAase-free RNAase (20µg.ml⁻¹). The typical yield of high-copy-number plasmids such as pUC from E. coli and pIJ486 from Streptomyces was 2-5µg of plasmid. Plasmid prepared in this way could be used for most in vitro manipulations.

2.3.2 ISOLATION OF TOTAL RNA

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 300°C. Distilled water was treated with DEPC (0.1% [v/v] of total volume) overnight and heated to 121°C for 15min in an autoclave. All solutions were prepared from DEPC-treated water and chemicals reserved for RNA work.

Reagents: 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. 10X DNAase RQ buffer; 400mM Tris·HCl (pH 7.9), 100mM NaCl, 60mM MgCl₂. Kirby mixture; 1g tri-isopropynapthalene 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 isolated always from 50ml liquid cultures 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 dH2O, which had been frozen at -20°C, rapidly lowering the temperature to ca. 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 of Kirby mix (prechilled on ice) containing 12g of 0.45mm diameter glass beads and agitated thoroughly on a vortex mixer for 1-2min. 5ml of phenol.chloroform (stored at room temperature) were then added and the mixture agitated for a further 1-2min. The homogenate was then spun in the centrifuge (9000g, 4°C for 10min) 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 1min on the vortex mixer. The phases were then separated as before. At this stage, the nucleic acid solution could be stored at -70°C without any degradation being detectable. To remove DNA from the preparation, the RNA could be pelleted by caesium chloride centrifugation and/or it could be treated with RNAase-free DNAase.

Pelleting the RNA through a dense caesium chloride cushion: Homogenate (ca. 3.5ml) from the Kirby lysis was layered carefully on top of a 5.7M CsCl-0.1 EDTA cushion (1.5ml, density 1.707 g.ml⁻¹) in a ultracentrifuge tube (5ml, Beckman, Polyallomer™). The RNA was then pelleted by centrifugation (35,000rpm, 20°C 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 dH2O 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°C for 10min). The RNA pellet was washed twice with 70% (v/v) ethanol, dried in the open microfuge tube at room temperature and dissolved in 100–200μl of dH2O. The RNA concentration was estimated spectrophotometrically, reprecipitated and stored at -70°C.

Treatment with DNAase RQ (Promega): An aliquot of the homogenate from the Kirby lysis, usually 1ml, was precipitated by the addition of 1/50 volume of 5M NaCl and 2 volumes of ethanol. The nucleic acid was harvested by centrifugation, washed twice with 70% (v/v) ethanol and dried by leaving the tube open on the bench. The nucleic acid was then redissolved in 400μl of 1X DNAase RQ buffer, DNAase added to 100U.ml⁻¹ and the reaction mixture incubated at 37°C for 20–30min. The mixture was then extracted with equal volumes of phenol.chloroform and chloroform. The remaining nucleic acid, in the aqueous phase, was precipitated by addition of 1/10 volume of 5M NaCl and an equal volume of isopropanol. After centrifugation, the pellet was washed in 70% (v/v) EtOH, dried briefly and redissolved in DEPC-treated dH2O.

2.3.3 QUANTIFICATION OF NUCLEIC ACID

The concentration and purity of the nucleic acid was determined by spectrophotometry; an A₂₆₀ of 1 is equivalent to 40μg.ml⁻¹ RNA and 50μg.ml⁻¹ double stranded DNA. Pure preparations of RNA and DNA have an A₂₆₀/A₂₈₀ of 1.8 and 2.0, respectively. Contaminating protein or phenol lowers significantly these values.

For quantitative S1 nuclease protection experiments, 25–30μg of RNA that had been purified by CsCl density centrifugation and treated with RNAase-free DNAase was resuspended in DEPC-treated dH₂O. Optical density measurements were used to quantify the amount of RNA in 5μl aliquots which had been diluted to 1ml. The measurements were repeated until three consecutive absorbance values agreed to within ±0.002 (approximately 0.1μg).
2.3.4 PRECIPITATION OF NUCLEIC ACID USING ETHANOL OR ISOPROPANOL

DNA solutions were precipitated by the addition of 1/50 volume of 5M NaCl and 2 volumes of cold ethanol or an equal volume of isopropanol. After mixing, the DNA was pelleted by centrifugation (27000g, 30 min, 4°C for volumes of 7.5-20ml or 12000g, 15 min, 4°C for small volumes in microfuge tubes). The pellet was washed in 70% (v/v) ethanol and dried briefly in a vacuum desiccator or in an open tube on the bench.

2.3.5 DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASES

Digestion of DNA was performed usually in a total volume of 20µl containing 0.25-1.0µg of DNA and 1-10 units of restriction endonuclease µg⁻¹ 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 appropriate temperature. Reactions were stopped either by the addition of gel loading buffer or by rapid heating to 70°C for 5 min, followed by cooling on ice.

2.3.6 LIGATION OF DNA FRAGMENTS

The ligation of DNA fragments was carried out usually 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 usually in 20µl of 1 X BRL ligation buffer, containing 1U of T₄ ligase µg⁻¹ of DNA. The reactions were incubated for 4 hours at room temperature or overnight at 16°C.

2.3.7 REMOVAL OF THE 5' PHOSPHATE FROM LINEARISED DNA

10X CIP Buffer: 200mM Tris.HCl (pH 8.0), 10mM MgCl₂, 10mM ZnCl₂ and 0.5mg.ml⁻¹ Bovine Serum Albumin.

Procedure: Calf Intestinal Alkaline phosphate (CIP) was used to remove the 5' phosphate from DNA. Around 5 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°C for 30min. The reaction was terminated by heating to 65°C in 1X gel loading buffer for 10min. The 5'-terminal dephosphorylated DNA was recovered from an agarose gel after electrophoresis.

2.3.8 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. Phenol:Chloroform; 50 parts phenol, 49 parts chloroform, 1 part isoamyl alcohol.

Procedure: Proteins were removed from nucleic acid solutions by extracting first with phenol:chloroform and then chloroform. 1/2 volumes of the solvents were added to the nucleic acid solution, mixed using a vortex mixer and the phases separated by centrifugation (1200g for 3min). Proteins were retained in the organic phase. Subsequently, precipitation with isopropanol or ethanol removed any solvent in the aqueous phase.

2.3.9 PREPARATION OF RADIO-LABELLED PROBES
2.3.9.1 RANDOM PRIMED DNA LABELLING METHOD

Labelling of DNA fragments with $^{32}$P followed the procedure of Feinberg and Vogelstein (1983 and 1984) using a Boehringer Mannheim Multiprime™ kit. The labelling reaction was set up in the following way:

25–50ng of denatured DNA fragment in a volume less than 10μl
(denatured by heating for 10min at 95°C with subsequent cooling on ice).
1μl of each unlabelled dNTP, from 0.5mM stocks.
2μl of reaction mixture (containing hexanucleotide mix and 10X concentrated reaction buffer).
5μl (50μCi) of [α-$^{32}$P] dCTP (3000 Ci.mmol⁻¹).
1μl (2U) of Klenow enzyme.

Made up to 20μl total volume with dH₂O.
Chapter 2 Materials and Methods

The reaction mixture was incubated at 37°C for 30 minutes, and stopped by heating to 65°C for 10 minutes. The labelled DNA fragments (and template DNA) were purified from the unincorporated dNTP's by Sephadex-G50 column chromatography.

2.3.9.2 5'-END LABELLING OF OLIGONUCLEOTIDES

Oligonucleotide probes were labelled at their 5' terminal ends using T4 polynucleotide kinase and [γ-32P]ATP.

Reagents: 10 X Kinase Buffer; 500mM Tris-HCl (pH 7.5), 100mM MgCl2, 50mM dithiothreitol, 1mM spermidine, 1mM EDTA. Stored at -20°C.

Reaction mixture:

<table>
<thead>
<tr>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X Kinase buffer</td>
</tr>
<tr>
<td>50ng of oligonucleotide</td>
</tr>
<tr>
<td>50µCi of [γ-32P] ATP (3000Ci.mmol⁻¹)</td>
</tr>
<tr>
<td>T4 Polynucleotide kinase (10U)</td>
</tr>
</tbody>
</table>

The reaction was incubated at 37°C for 30min. The unincorporated ATP was removed by Sephadex G50 column chromatography.

2.3.9.3 SEPHADEX G50 COLUMN CHROMATOGRAPHY

After the labelling reactions had been completed, 5µl of dextran blue (50mg.ml⁻¹) and 5µl of phenol red (50mg.ml⁻¹) were added to the reaction mixture. The samples were then loaded onto Sephadex-G50 (20 x 0.5cm dimension gravity column), which had been pre-equilibrated with column buffer (100mM NaCl, 10mM Tris.HCl (pH 7.5) 1mM EDTA). Fractions of approximately 500µl were collected. The dextran blue co-eluted with the DNA fragments and all other aliquots were discarded. The labelled DNA sample was boiled for 5min prior to use.

2.3.10 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 TBE buffer, however, TAE buffer was used when DNA fragments were to be isolated from the gels (see section 2.3.12).

Buffers: 10X TBE Buffer (pH 8.3); 109g Tris, 55g boric acid, 9.3g Na₂EDTA·2H₂O made up to 1 litre in distilled water, pH should be ca. 8.3. 10X TAE Buffer pH 8.2; 48.4g Tris, 16.4g Na acetate, 3.6g Na₂EDTA·2H₂O, 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.

2.3.10.1 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 TBE (TAE), boiled then cooled to 60°C. 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 TBE (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.

2.3.10.2 LARGE GELS

200ml gels were also used to ensure good separation of DNA fragments for accurate sizing and/or Southern analysis. They were made by pouring 200ml of molten agar containing 200μg EtBr, into a 16.5 X 23cm gel former with a 20 space slot former. The gels were run overnight at 20V in 1X TAE or TBE buffer in gel tanks with a capacity of 3 litres. DNA samples were mixed with 1/5 volume of 5X AGL buffer, heated to 70°C for 2min and cooled on ice before loading onto the gel.
2.3.10.3 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).

2.3.11 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 Na₂EDTA, 95% (v/v) formamide (de-ionized with a mixed-bed resin). 40% Acrylamide stock solution; 76g acrylamide (BDH. Electran Grade) and 4g bisacrylamide. Dissolved in dH₂O with stirring, brought to a final volume of 200ml, filtered through 3MM paper and stored at 4°C.

Preparation of polyacrylamide gels: 6% (w/v) denaturing polyacrylamide gels were used for sequencing and single-stranded probe isolation. The gels were prepared from the following stock solutions:

- 40% (w/v) acrylamide stock 15ml
- urea 55g
- 10X TBE 10ml
- dH₂O 35ml

ca. 100ml.

The urea was dissolved by heating the mix to 37°C and then cooled to room temperature. The sequencing gel solution could be stored at 4°C for several weeks without loss of resolution. Before pouring the gel, 300µl of freshly prepared 10% (w/v) Ammonium persulphate and 50µl of TEMED were added to 50ml of the stock solution.

Preparation of glass plates and pouring the gel: The plates (40cm X 33cm) were cleaned thoroughly with alcohol and water and assembled using three spacers (0.4mm thick) along the vertical sides and the
Chapter 2 Materials and Methods

bottom of the gel. The entire assembly was held in place by clamps. The
gel solution was poured from a beaker down one edge of the plates
while tilting the plates at an angle of about 30°. The plates were then
laid at an angle of 5° and the sharks tooth combs inserted. The gel
polymerized usually within 30min 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 containing sequencing loading buffer were heated to 95°C
for 5 mins, 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.

2.3.12 RECOVERY OF DNA FROM AGAROSE GELS

DNA fragments were recovered from TAE agarose gels using
GeneClean™ and electro-elution. The GeneClean™ kit (purchased from
Stratech Scientific Ltd) was used for recovering DNA in the size range
1-7kb. Smaller fragments were not recovered efficiently from the glass
beads and were isolated, therefore, by electro-elution. TBE gels were
avoided as the borate ion significantly reduces the amount of DNA
recovered using GeneClean™ and can influence some enzyme reactions
such as ligation.

2.3.12.1 GENECLEAN™

The TAE agarose gel was placed on a 302nm UV transilluminator
and a small block (<0.2cm³) of agarose containing the DNA fragments of
interest excised using a scalpel. The agarose block was fragmented and
2.5 volumes of NaI solution were added. The suspension was incubated
at 55°C for 5 minutes or until the agarose had dissolved completely.
5μl of "glassmilk" was added to the solution, which was mixed rapidly
on a vortex mixer and placed on ice for 5min. The glass beads with the
DNA bound were pelleted by centrifugation (5 seconds in a microfuge)
and the supernatant discarded. The glass beads were washed three
times in ice-cold NEW solution (500μl), each time agitating the suspension
using a vortex mixer and recovering the beads by centrifugation. After
the final wash, care was taken to remove all traces of NEW solution.
The DNA was recovered from the glass beads by adding 20μl of dH2O and incubating at 55°C for 5min. The glass beads were pelleted by centrifugation (12000g for 30s in a microfuge) and the supernatant retained. This final step was repeated using another 20μl of dH2O and the DNA solutions combined. The exact composition of the NaI, Glassmilk and NEW solutions were not disclosed by the manufacturers (Bio 101).

2.3.12.2 ELECTRO-ELUTION

The TAE agarose gel was placed on a 302nm UV transilluminator and a trough measuring 0.5cm wide (the length and depth was determined by the length of slot and thickness of gel respectively) was excised from the TAE agarose gel immediately in front (with respect to direction of migration) of the DNA fragments of interest. The distal edge of the trough (with respect to the DNA) was lined with dialysis tubing (pre-treated by boiling in 2% (w/v) sodium bicarbonate, 1mM EDTA) and electrophoresis continued until the desired DNA had eluted from the agarose and migrated across the trough to be held by the dialysis tubing. The polarity of the current was then reversed momentarily before being switched off and the DNA was removed from the surface of the dialysis tubing using a pipette. The recovered DNA was then extracted with phenol.chloroform and chloroform and precipitated with ethanol. To recover relatively undamaged DNA, agarose gel electrophoresis was performed in the absence of EtBr. The migration of the DNA was monitored by removing a small "reference" strip of agarose containing DNA standards and staining it in EtBr.

2.3.13 ISOLATION OF DNA FROM POLYACRYLAMIDE GELS

The following procedure, which is a modification of the technique described by Maxam and Gilbert (1971), was used to isolate radioactive probes for mapping the 5'-termini of RNA transcripts.

Reagents: Elution buffer; 500mM ammonium acetate, 10mM magnesium acetate, 1mM EDTA (pH 8.0), 10% (w/v) SDS.
Crush and Soak method: The DNA fragment to be isolated was resolved on a denaturing 6% (w/v) polyacrylamide.urea gel and located by autoradiography. A section of the gel containing the fragment of interest was removed using a scalpel blade and transferred to a microfuge tube. Using a disposable plastic pipette tip, the gel slice was crushed against the wall of the tube. The probe was eluted by adding 600µl of elution buffer and incubated at 37°C in an orbital shaker (contained within a lead pig) overnight. Fragments of acrylamide were removed using centrifugation to force the eluant through a syringe barrel packed with siliconised glass wool. The syringe was washed with 200µl of fresh elution buffer and the eluant and washings combined. The eluant was extracted twice with phenol.chloroform and twice with chloroform. The probe was then precipitated by adding 2 volumes of ethanol and harvested by centrifugation in a microfuge (12000g at 4°C for 30min). From this stage onwards RNAase free solutions and disposables were used. The pellet of single-stranded probe was rinsed once with 70% (v/v) ethanol and redissolved in 200µl DEPC-treated dH2O.

2.3.14 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°C 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°C for 2 hours.
2.3.15 SOUTHERN ANALYSIS

Southern analysis was carried out by a method adapted from Southern (1975), and described in "Blotting and hybridization protocols for Hybond-N™ membranes" published by Amersham International plc.

2.3.15.1 CAPILLARY TRANSFER

After electrophoresis and photography, the resolved DNA fragments were transferred under alkali conditions to Hybond-N™.

Reagents: Denaturing solution; 1.5M NaCl, 0.5M NaOH. Alkali transfer buffer; 0.25M NaOH, 1.5M NaCl. 20X SSC; 3m NaCl 0.3M trisodium citrate.

Procedure: The gel was rinsed in distilled water, placed in enough denaturing solution to immerse it completely and left for 30min. The gel was removed, excess liquid removed by blotting and equilibrated for 10-15min in alkaline transfer buffer. The DNA was then transferred to the nylon membrane in alkaline buffer, by capillary action (disposable nappies proved a particularly useful absorbent material for driving the transfer process). After blotting for a least 4 hours (but usually overnight), the membrane was washed briefly in 2X SSC to remove any adhering agarose.

2.3.15.2 AQUEOUS PREHYBRIDIZATION AND HYBRIDIZATION CONDITIONS

The procedures which follow were suitable for DNA immobilised on filters from colony lifts and Southern transfer.

2.3.15.2(a) HYBRIDIZATION USING RANDOM PRIMED PROBES.

Reagents: 100X Denhardts solution; 10g ficoll, 10g polyvinylpyrrolidone, 10g bovine serum albumin. Made up to 500ml with dH₂O. 20X SSC; 3M NaCl, 0.3M trisodium citrate.

Prehybridization and hybridization conditions: The prehybridization and hybridization solutions used in this work for the probing of homologous sequences contained 5X Denhardt solution, 1X SSC, 0.5% (w/v) SDS (Sigma, Molecular Biology Grade) and 100µg.ml⁻¹
sheared and denatured Salmon sperm DNA. The volume of the
hybridization solutions depended on the size of the filters; 200µl were
added for every cm² of membrane. The solutions were filtered through
0.4µm Nalgene membranes prior to the addition of Salmon sperm DNA and
random primed probe (50ng, ca. 1X 10⁹ cpm·µg⁻¹) (see section 2.3.9.1).
The prehybridization and hybridization reactions were incubated at 65°C
for 4 and 16 hours respectively.

Washing of membranes after hybridization: The nylon membrane
was washed twice in 200ml of 1X SSC, 0.5% (w/v) SDS at 65°C for 20min
each. Excess fluid was removed by blotting, the filter sealed in a plastic
bag and subjected to autoradiography. The probe could be stripped
from the filter (provided the filter was kept moist after hybridization)
by washing in 0.4M NaOH at 45°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.

2.3.15(b) CONDITIONS FOR AQUEOUS HYBRIDIZATIONS WITH
OLIGONUCLEOTIDES

This procedure was used for probing homologous DNA, which was
immobilised on Hybond-N™ membranes, with radioactively labelled
oligonucleotides of 20–30 bases.

Prehybridization and hybridization conditions: The surface area of
the filter was measured and 200µl of prehybridization and hybridization
used for every square centimetre. The prehybridization solution
contained 6X SSC, 0.05% (w/v) SDS (Sigma, Molecular Biology Grade),
0.05% (w/v) sodium pyrophosphate and 200µg.ml⁻¹ Heparin (Sigma, Grade
1). The prehybridization solution was heated to 65°C before the required
amount of heparin was dissolved. The solution was then added to the
filter and incubated at 65°C for 4 hours. The hybridization solution was
the same as the prehybridization solution, except that it contained 0.5%
(w/v) SDS and 50ng of probe (ca. 1X 10⁹ cpm·µg⁻¹). The hybridization
solution was prewarmed to 65°C before the heparin and 50ng of
oligonucleotide probe were added (see section 2.3.9.2). The filter was
incubated with the hybridization solution at 65°C for 90min.
Washing of the membrane: The membrane was washed in 150ml of 5X SSC, 0.05% (w/v) SDS for 15min at room temperature and then twice in 150 ml of the same wash solution at 60°C.

2.3.16 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 1.5ml of 2X YT broth containing 15μl of an overnight culture of E. coli TG1. This was grown at 37°C for 5-6 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 200μl 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 dH2O and extracted twice with phenol.chloroform and twice with chloroform. The DNA was then ethanol precipitated from the aqueous phase and recovered by centrifugation in a microfuge. The ssDNA was then washed with 70% (v/v) ethanol, dried in vacuo and redissolved in 20μl dH2O.

Maxipreparations: The same overall procedure was followed as for the "minipreparations", except that all the volumes were scaled up 20 fold. The cells from a 30ml culture were spun out (14000g for 2min), the supernatant recovered and respun as before. The supernatant (20ml) that remained was then precipitated with 5ml of 20% (w/v) PEG/2.5M NaCl for 10 min at room temperature and the phage harvested by centrifugation (14000g at 20°C for 15min). The phage pellet was resuspended in 1ml of dH2O, then reprecipitated and processed as for the minipreparations with the volumes scaled up accordingly.
2.3.17 DNA SEQUENCING

Dideoxy sequencing (Sanger et al., 1975) was carried out on single-stranded M13 templates using a TAQuenase™ kit supplied by United States Biochemical Corporation (USB).

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 13µl with dH₂O. After denaturing at 85°C for 2min, the primer was annealed to the template by incubating at 45°C for 10min. After annealing was complete, the tube was placed on ice and used within ca. 2 hours.

Labelling reaction: To the annealed template-primer, the following were added on ice:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-deaza-dGTP labelling mix</td>
<td>2</td>
</tr>
<tr>
<td>5µCi of [α-3²P]dATP (800Ci.mmol⁻¹)</td>
<td>0.5</td>
</tr>
<tr>
<td>Cloned Taq DNA polymerase (2 units)</td>
<td>2</td>
</tr>
</tbody>
</table>

The above were mixed and incubated at 45°C for 5min. To read sequences close to the primer annealing site a 1:5 dilution of the labelling mixture was used.

Termination reactions: Immediately after the labelling reaction was complete, 4µl aliquots were dispensed into four microfuge tubed 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 70°C for 5min. After the reactions were complete, they were allowed to cool to room temperature, 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°C for 5-10 min, cooled rapidly on ice and loaded immediately onto the gel. Approximately 2-3µl of the sequencing products were loaded per track. The compositions of the buffers and nucleotide mixes were not disclosed by the manufacturers (USB).
2.3.18 TRANSCRIPT MAPPING USING COMPLEMENTARY SINGLE-STRANDED DNA PROBES

Mapping of the 5' terminus of RNA transcripts was undertaken using single-stranded probes derived from M13 (Calzone et al., 1987), as described by Maniatis 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°C. 10X S1 nuclease digestion buffer; 2.8M NaCl, 500mM NaCH3C00 (pH 4.5), 45mM ZnSO4. Stored at -70°C. 10X Exonuclease VII digestion buffer; 500mM Tris.HCl (pH 7.8), 500mM KCl and 100mM EDTA (pH 8.0). Denatured Salmon sperm DNA; prepared by dissolving in dH2O, extracting any protein with phenol.chloroform, shearing by passage through a narrow gauge hypodermic needle and ethanol precipitating. The fragmented DNA was then redissolved in dH2O to a concentration of 10mg.ml⁻¹, boiled for 10min and stored at -20°C in small aliquots. Prior to use, the DNA was partially denatured by heating to 100°C for 5min and chilling rapidly on ice. Carrier tRNA; prepared by dissolving commercially available yeast tRNA in dH2O, extracting any protein with phenol.chloroform, precipitating and redissolving in dH2O to a final concentration of 10mg.ml⁻¹. S1 nuclease stop buffer; 500mM Tris.HCl (pH 8.0), 125mM EDTA.
Probe synthesis: 0.5 pmoles of template ssDNA was mixed with 3 pmoles of an oligonucleotide primer in a volume of 16 µl, containing 2µl 10X BRL REact 2 buffer, and then annealed by heating to 85°C for 5min and cooling slowly to 37°C. The extension reaction was then carried out at 37°C for 30 min by adding:

<table>
<thead>
<tr>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100mM dithiothreitol</td>
</tr>
<tr>
<td>50µCi of [α-32P]dATP</td>
</tr>
<tr>
<td>(sp. act. 3000 Ci.mmole⁻¹)</td>
</tr>
<tr>
<td>40µM solution of dATP</td>
</tr>
<tr>
<td>20mM solution of dCTP, dGTP and dTTP</td>
</tr>
<tr>
<td>Klenow (1 U.µl⁻¹)</td>
</tr>
</tbody>
</table>

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°C for 10min to inactivated the Klenow. 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 optimum temperature for the enzyme. The nucleic acid was then precipitated with ethanol and recovered by centrifugation in a microfuge. The DNA pellet was washed with 70% (v/v) ethanol and dried in vacuo. To purify the single-stranded probe, the pellet was resuspended in 20µl 1X sequencing loading buffer, denatured by heating to 85°C for 5min and resolved by electrophoresis on a denaturing 6% (w/v) polyacrylamide gel. The position of the probe was determined by autoradiography and eluted from the gel using the crush and soak protocol (see Section 2.3.13).

Incorporation of radio-labelled nucleotides: When the probe was used in quantitative S1 nuclease protection experiments, the percentage of radio-labelled nucleotides that were incorporated was calculated. 0.5µl samples removed from the extension mixture before the addition of Klenow and after the synthesis of the probe but before the "cold" chase were spotted onto 1 cm² square of Whatman GF/C filters in microfuge
Chapter 2  Materials and Methods  78
tubes (1.5ml). The filters were stored at room temperature until all the fluid had evaporated. Using blunt-ended forceps, the filter containing the sample removed immediately before the "cold" chase was swirled in a beaker containing 100-150ml of ice-cold 5% (w/v) trichloroacetic acid and 20mM sodium pyrophosphate for 2min. This washing step was repeated three times using fresh volumes of the acid solution. The filter was then washed in 70% (w/v) ethanol for 30s and allowed to dry at room temperature. The two filters were then placed in separate scintillation vials and the amount of radioactivity measured by Cerenkov counting. The proportion of radio-labelled precursors and specific activity of the probe were calculated using the following equations:

\[
\frac{\text{cpm in washed filter}}{\text{cpm in unwashed filter}} = \text{proportion incorporated}
\]

\[
\text{proportion incorporated} \times \text{total weight} = \text{total amount of product}
\]

\[
\frac{\text{cpm incorporated}}{\text{total amount of product}} = \text{specific activity}
\]

This procedure yielded routinely 0.1 to 0.4 pmoles of single-stranded probe with a specific activity of $10^6$-$10^7$ Cerenkov cps.pmole$^{-1}$ DNA.

Hybridization: The single-stranded probe (ca. 0.1-0.2pmoles) 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 Hybridization buffer. The solution was pipetted up and down many (20-30) times and heated to 60°C for 30min to ensure that the pellet was dissolved completely. The hybridization solution was then heated to 85°C for 10min, to denature the nucleic acids and incubated at 55-65°C overnight. After overnight incubation, the hybridization mixture was either treated with S1 nuclease or Exonuclease VII.
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 (1-2μl) were added, the reaction mixture pooled by centrifugation and incubated at 37°C for 30min. 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 \textit{in vacuo} and resuspended in 7μl 1X sequencing loading buffer. The products were run against sequencing ladders on denaturing polyacrylamide gels.

Exonuclease VII digestion: 280μl of ice cold 1X exonuclease buffer, containing 10 units of exonuclease VII, were added to the hybridization solution, mixed by inverting the tube several times, pooled by brief centrifugation and incubated at 37°C for 45min. The reaction mixture was extracted with phenol.chloroform and chloroform and processed as for the S1 digestion products above.

2.3.19 PRIMER EXTENSION MAPPING OF TRANSCRIPT 5' ENDS

The protocol described below was communicated by Lewis Wray (Boston University, USA).

Reagents: 5X Annealing buffer; 1.0M KCl, 0.1M Tris.HCl (pH 8.3). 5X Elongation buffer; 0.9M Tris.HCl (pH8.3), 0.1M MgCl₂.
Procedure: 5'-labelled oligonucleotide (ca. 0.5-1ng) and *Streptomyces* RNA (10-40µg) were co-precipitated, washed with 70% (v/v) ethanol and resuspended in 20µl 1X Annealing buffer. The sample was incubated at 80°C for 5min to denature and then incubated at 42°C for 3 hours. After the RNA-oligonucleotide annealing reaction was completed the following were added:

<table>
<thead>
<tr>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X Elongation buffer</td>
</tr>
<tr>
<td>Actinomycin D (1 mg.ml⁻¹)</td>
</tr>
<tr>
<td>2mM dATP, dCTP, dGTP, dTTP solution</td>
</tr>
<tr>
<td>RNAguard Ribonuclease inhibitor (Pharmacia, 37U.µl⁻¹)</td>
</tr>
<tr>
<td>Avian Myeloblastosis Virus (AMV) reverse transcriptase (Pharmacia)</td>
</tr>
<tr>
<td>100mM dithiothreitol</td>
</tr>
<tr>
<td>dH₂O</td>
</tr>
</tbody>
</table>

The elongation reaction was incubated at 42°C for 30min and then terminated by the addition of 10µl Sequencing Stop loading buffer (USB, TAQuenase™ kit). Immediately prior to loading, the extension products were heated to 85°C. 3-5µl of the sample was then loaded onto the sequencing gel.

2.3.20 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°C) 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.

2.3.21 DENSITOMETRY SCANNING

The intensity of autoradiographic images were determined using a Hoefer Scientific Instruments GS-300 Scanning Densitometer, integrated with an IBM GS-360 Data System.
CHAPTER 3

ANALYSIS OF PROMOTER ACTIVITY UPSTREAM OF *OTRA*
3.1 Introduction

The biosynthetic genes for oxytetracycline production are located in a single cluster, flanked by two resistance genes (Binnie et al., 1989). Both the otrA and otrB resistance genes have been cloned from S. rimosus M15883, a commercial production strain developed by Pfizer (Rhodes et al., 1984). The product of the otrA resistance gene protects the ribosomes from translational arrest by oxytetracycline, while the otrB gene product is involved in efflux of the antibiotic from the mycelium (Ohnuki et al., 1985; Doyle et al., 1988). During vegetative growth, tetracycline resistance in S. rimosus can be induced by pre-exposure to sub-lethal concentrations of tetracycline (Ohnuki et al., 1985; Lee, unpublished result).

The otrA resistance gene has been sequenced and the pattern of transcription analysed partially (Figure 3.1; from the results of Doyle, 1987). Two transcription start points were proposed for otrA, 129nt and 339-344nt upstream of the translation initiation codon, corresponding to the promoters otrAp1 and otrAp2 respectively. Transcription has been shown to terminate immediately downstream of the otrA resistance gene at a site which is predicted to form a typical rho-independent terminator structure in the messenger RNA (Doyle, 1987). The recognition sequence of otrAp1 resembles the consensus sequence for the major class of eubacterial promoters. This promoter is thought to be transcribed from the S. rimosus equivalent of the S. coelicolor RNA polymerase holoenzyme, Eo35, which is capable of transcription from the veg promoter of B. subtilis and the dagAp4 from S. coelicolor (General Introduction, 1.9) (Westpheling et al., 1985; Buttner et al., 1988). No DNA sequence was obtained for otrAp2, as it was external to the cloned DNA conferring tetracycline resistance (General Introduction, 1.14).

Owing to the commercial importance of S. rimosus M15883, in vivo manipulations can only be carried out within the confines of the Pfizer Laboratories. Consequently, previous analyses of otrA transcription used RNA isolated from the ancestral production strain S. rimosus M4018, grown on Tryptone Soya Broth (TSB), which does not support antibiotic production (Doyle, 1987). As no sequence resembling a rho-
independent terminator could be detected in the intervening region between otrA and otcZ (a biosynthetic gene located immediately upstream; Figure 3.1), it was proposed that these genes may be co-transcribed during antibiotic production, as part of a polycistronic message.

Without exception, antibiotic production genes have been found in clusters. Indeed, where all the biosynthetic genes have been identified or where heterologous expression of antibiotic production has been attained, all of the biosynthetic genes and at least one resistance gene have been mapped to a single location (Malpartida and Hopwood, 1984; Stanzak et al., 1986; Chen et al., 1986; Motamedi and Hutchinson, 1987; Binnie et al., 1989). The close proximity of resistance determinants to biosynthetic genes and the requirement for their concomitant expression during physiological differentiation has prompted speculation that the expression of these genes may be regulated coordinately (Hopwood, 1988).

3.1.1. Objectives

The transcriptional analysis of otrA promises to reveal interesting facets about the coordination of gene expression during the switch to secondary metabolism and the induction of antibiotic resistance during vegetative growth.

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

(i) To identify the promoters responsible for the transcription of the otrA resistance gene and to compare their recognition sequences with the consensus sequences for known classes of promoters.

(ii) To establish culture conditions for S. rimosus M4018 and M15883 that yielded reproducible growth and oxytetracycline production profiles.

(iii) To isolate RNA from S. rimosus cultures at different stages in the fermentation of oxytetracycline and to determine the temporal changes in the transcription of otrA.
(iv) To identify suitable RNA samples/ time points for intensive transcriptional analyses of the OTC cluster.

3.1.2 Methodology

Two strategies were employed to investigate the transcription of \( \text{otrA} \); (i) the cloning of fragments from the region immediately upstream of \( \text{otrA} \) into a promoter-probe system and (ii) high-resolution S1 nuclease protection experiments.

At the onset of this investigation (1987) only two plasmid systems were available readily for measuring \( \text{in vivo} \) promoter activity in \( \text{Streptomyces} \); pARCl, a low-copy-number vector (Horinouchi and Beppu, 1985) and pIJ486/7, two high-copy-number vectors that differ only in the orientation of their multiple cloning sites (Ward et al., 1986).

It has been known for some time that the molar ratio of transcriptional regulators to their recognition sites on DNA is of major importance in regulating gene expression (Ptashne, 1986). As the pARCl system is based on a low-copy-number plasmid, it should be useful for studying promoters that are regulated by the binding of ancillary proteins. The pARCl plasmid contains a promoterless locus specifying the synthesis of a brown pigment. Recombinants containing a promoter, which is actively directing the transcription of the pARCl reporter locus, have a phenotype which can be detected easily without selection. This system can be used, therefore, to analyse promoters with distinct temporal activities. Unfortunately, pARCl has only a single \( \text{BamHI} \) site, upstream of the reporter locus, which is suitable for inserting fragments to be tested.

The pIJ486/7 system includes several useful features; a multiple cloning site located immediately upstream of the \( \text{aphII} \) reporter gene, which encodes kanamycin resistance, an \( \text{E. coli} \) bacteriophage fd terminator upstream of the cloning sites to prevent transcriptional read-through from vector promoters and a translational stop codon, which is located 3' to the cloning site but in frame and 5' to the reporter gene,
to avoid translational fusions, which could alter the specific activity of the *aphII* gene product. Although the high copy-number of the pIJ486/7 plasmids permits the detection of weak promoters, any regulation dependent on the ratio of target DNA to ancillary protein may be disrupted.

As *S. rimosus* M4018 and M15883 produce a brown pigmented by-product during the production of oxytetracycline and are inherently resistant to high levels of kanamycin (ca. 500μg.ml⁻¹, Hunter pers. comm.), neither the pARCl nor the pIJ486/7 system can be used to measure promoter activity in these production strains. Previously, the pIJ486/7 plasmids were used to analyse the *in vivo* activity of *otrA* and due to the ease with which DNA fragments could be inserted in a specific orientation (Doyle, 1987). The assays were carried out using the heterologous host *S. lividans*, which is highly sensitive to kanamycin and can be transformed with high efficiency. In order that measurements of promoter activities could be compared directly, this approach was also adopted to analyse further the transcription of *otrA*.

The S1 nuclease protection assay (Berke and Sharp, 1977) is the most widely used technique for determining the precise locations of the 5' and 3' termini of transcripts. Several variations on the original technique have been published (for an overview see Calzone *et al.*, 1987). Single-stranded probes derived from recombinants of bacteriophage M13 were used to map the 5' termini of the *otrA* transcripts (as described by Maniatis *et al.*, 1989). This approach has several advantages; (i) the purification of single-stranded probe minimises the possibility of the complementary DNA strand (i.e. the template) competing with RNA for hybridization to the probe, (ii) the high-specific activity of the probes, which can exceed 10⁹ Cerenkov cpm.μg⁻¹ DNA, facilitates the detection of rare transcripts and (iii) the templates, which are used to produce the probe, can also be used to generate dideoxy-mediated sequencing ladders for the accurate sizing of the nuclease-resistant fragment(s).
3.2 Results

3.2.1 Construction of Recombinants for \textit{in vivo} Promoter Analysis.

DNA fragments from the region upstream of \textit{otrA} (Figure 3.1) were subcloned into the pIJ486 promoter probe vector (Ward \textit{et al.}, 1986). The copy numbers of these constructs were not significantly different, as judged by the relative yields from numerous plasmid preparations. Initially, the 305bp \textit{BclII-KpnI} (isolated as a \textit{KpnI} fragment using a site in the polylinker), the 722bp \textit{KpnI-PstI} and the 805bp \textit{BamHI-PstI} fragments from pPFZ513 (Binnie, Pfizer) were subcloned into pUC18 to produce pKME805, pKME807 and pKME808 respectively. The \textit{BclII-KpnI} insert was orientated such that the \textit{BclII} site was nearest the \textit{EcoRI} site in the polylinker. This intermediate cloning step was undertaken as \textit{E. coli} recombinants could be analysed rapidly and the subsequent subcloning step into the streptomycete vector was much simpler, that is the same preparation of \textit{HindIII/EcoRI}-digested pIJ486 could be used for all three cloning experiments.

\textit{HindIII/EcoRI} fragments from the pUC18–based constructs were then ligated into pIJ486, which had been treated with calf intestinal phosphatase, to derive pKMS605 from pKME805, pKMS607 from pKME807 and pKMS608 from pKME808 respectively. The products of each ligation were then introduced into \textit{S. lividans} TK54 by genetic transformation. For each transformation, plasmid DNA was isolated from at least three different thiostrepton-resistant transformants and digested with restriction endonucleases. Although gross rearrangements were never detected, occasionally, plasmids were found which did not contain an insert. \textit{HindIII/EcoRI} and \textit{PstI} digests of plasmid DNA from the isolates that were assayed subsequently for \textit{in vivo} promoter activity are shown in Figure 3.2. All of these constructs produced the expected band patterns when the products of endonucleolytic cleavage were separated by electrophoresis on agarose gels. The inserts were orientated such that any promoter activity capable of directing transcription of \textit{otrA} would promote expression of the \textit{aphII} reporter gene.
Figure 3.1 Promoter-probe constructs used to investigate promoter activity capable of transcribing the otrA resistance gene. Arrows indicate the orientation of the insert with respect to the promoterless aphII gene in pIJ486. Shaded boxes denote the otcZ and otrA protein coding regions. Transcripts detected by Doyle (1987) are represented by bold wavy arrows, the closed circles indicating the proposed start site. The level of kanamycin conferred by the constructs is indicated in the column at the left of the figure. Abbreviations: B, BamHI; Bc, BclII; Hc, HincIII; K, KpnI; P, PstI; Sm, SmaI.
Figure 3.2. Restriction digests to confirm the status of the inserts in independent isolates of the promoter-probe constructs.

Gel A; HindIII/EcoRI digests of the pIJ486 constructs
Gel B; PstI digests of the pIJ486 constructs
Lane 1 and 11; lambda HindIII DNA markers
Lanes 2-4; pKMS605 from isolates 1, 2 and 3
Lanes 5-7; pKMS607 from isolates 1, 2 and 3
Lanes 8-10; pKMS608 from isolates 1, 2 and 3
3.2.2 Assay of Promoter Activity from pIJ486 Constructs

Serial dilutions of spore suspensions of *S. lividans* TK54 containing each of the above constructs were spread out on Emersons agar plates to determine the spore titre. Similarly, spore suspensions of transformants containing pIJ486 and pDM130 (a pIJ486 construct containing the SmalI–BamHI fragment; Doyle, 1987) were also prepared as negative and positive controls respectively.

To assay promoter activity, approximately $10^2$ spores of each of the above transformants were streaked out on Emersons agar plates containing thiostrepton (25µgml$^{-1}$) and kanamycin at concentrations ranging from 100µgml$^{-1}$ to 900µgml$^{-1}$. Growth on the agar plates was assessed after incubation for 7 days at 30°C (Figure 3.1). The only recombinants to exhibit promoter activity were *S. lividans* TK54 containing pKMS605 and the positive control pDM130. Using RNA isolated from *S. lividans* TK64 containing pDM130, the 5' endpoints of transcripts originating from otrAþ had been mapped at 220 and 221nt to the left of the BamHI site, within the BcII site recognition sequence (Doyle, 1987). Similar levels of promoter activity were also detected using pDM110, a pIJ486-based construct containing the 221 bp BcII–BamHI fragment (Doyle, 1987). In addition to confirming the presence of transcripts originating from otrAþ, S1 nuclease mapping experiments using RNA from *S. rimosus* M4018 also detected a second signal, which was proposed to correspond to a transcript arising from a second promoter, designated otrA2. The transcriptional start site for this promoter was mapped 5 to 10nt to the left of the BamHI site within the predicted protein coding region (PPCR) of otcZ (Doyle, 1987).

At this stage in the present analysis, the simplest explanation for the increased kanamycin resistance in recombinants containing pKMS605 (>900µgml$^{-1}$) compared with those containing pDM130 (ca. 300µgml$^{-1}$) was that both otrAþ and otrA2 were directing transcription of the *aphII* reporter gene in the pKMS605 construct. The recognition sequence of otrA2 was interpreted as overlapping with the BamHI site, as the pKMS608 construct, which contains the BamHI–PstI fragment, did not confer kanamycin resistance to *S. lividans* TK54, even to concentrations
as low as 20μg/ml⁻¹. The finding that the pKMS607 construct did not contain promoter activity was not unexpected as it contains neither \( otrA\)p1 nor \( otrA\)p2.

A possible shortcoming of using the outgrowth of spore suspensions to assay for \textit{in vivo} promoter activity in the pIJ486/7 system is that constructs containing promoters, which are not transcribed during vegetative growth, may not confer kanamycin resistance to germinating spores and consequently may not be detected. In an attempt to circumvent this possibility, kanamycin-resistance assays were also carried out using mycelial inoculums from liquid cultures of \( S.\) \textit{lividans} TK54 containing pKMS607 and pKMS608 at various stages in growth, including stationary phase. Promoter activity could still not be detected and it was concluded that in the absence of any OTC-specific activator, the transcription of \( otrA\) was mediated solely by \( otrA\)p1 and \( otrA\)p2.

### 3.2.3 S1 Nuclease Mapping of the Putative \( otrA\)p2 Transcripts

High resolution S1 protection experiments were undertaken to establish the precise 5' endpoint of the putative transcripts originating from \( otrA\)p2. Single-stranded probe, complementary to the anti-sense strand of \( otrA\), was prepared from mKM905 (containing the \textit{BcII5-KpnI1} fragment with the \textit{BcII} site nearest the primer-annealing site, Chapter 4) by extension from the universal (~20) primer and digestion with HindIII (Materials and Methods; 2.3.18). This continuously-labelled probe was used in high resolution S1 mapping experiments with total RNA from three independent isolates of \( S.\) \textit{lividans} TK54 containing pKMS605, \( S.\) \textit{rimosus} M15883 (a Pfizer production strain) and \( S.\) \textit{rimosus} M15883S (a spontaneous mutant with the entire OTC cluster deleted) as a control. Transcripts originating from 5-10nt to the left of the \textit{BamHI5} site were predicted to result in protected fragments of 212-217nt using RNA from \( S.\) \textit{rimosus} M15883 and 233-238nt using RNA from \( S.\) \textit{lividans} TK54 (pKMS605). The difference in the size of the predicted protection is due to the putative transcripts from pKMS605 having additional complementarity to polylinker sequences derived from M13mp19 in the probe. These transcripts should have been resolved easily in
denaturing polyacrylamide gels. However, no RNA-protected fragments were detected, which could correspond to the protection of probe by transcripts originating from around the \textit{BamH} site (Figure 3.3, Panel A). Furthermore, repeated high-resolution S1 mapping experiments using numerous batches of RNA from \textit{S. lividans} TK54 (pKMS605) failed to detect protection of any part of the probe (result not shown). As transcripts originating from \textit{otr}Ap1 in the pKMS605 construct only have complementarity to 22-23nt of the polylinker sequence in the probe, it is possible that a RNA-protected fragment corresponding to \textit{otr}Ap1 was not detected due the poor stability of the probe-RNA complex under the conditions used for hybridization (Materials and Methods 2.3.18) and/or the low specific activity (cpm.pmole\(^{-1}\)) of the short RNA-protected fragment.

Using RNA isolated from \textit{S. rimosus} M15883 and M4018 (a former OTC production strain, Pfizer), a protected fragment of approximately 305nt was always detected (Figure 3.3, Panel B). The absence of a RNA-protected fragment using RNA from \textit{S. rimosus} M15883S indicated that this protection was specific to the OTC cluster. As the probe contains nucleotides at both the 5' and 3' end that are derived from M13mp19, it is possible to distinguish between the fragment produced by reannealing of the probe to contaminating DNA template and probe protected by transcripts originating outwith but extending through the region cloned in M13mp19. The size of the protected band corresponded exactly to that predicted for transcripts extending through the entire 305bp \textit{BcII}-\textit{KpnI} region. In light of these S1 nuclease mapping experiments, the protection interpreted by Doyle (1987) as corresponding to transcripts originating from \textit{otr}Ap2 is now considered to be full-length probe (FLP) which has had a few nucleotides removed from the 3' end by over digestion with S1 nuclease. Consideration of the combined results from the \textit{in vivo} analysis of promoter activity and the S1 nuclease protection experiments, suggests that; (i) \textit{otr}Ap1 is the only promoter capable of directing transcription of \textit{otr}A within the \textit{BcII}-\textit{Ps}I region and (ii) a promoter located at least 1.4kb upstream (beyond the \textit{Ps}I site) also directs transcription of this resistance gene.
Figure 3.3 (A). High resolution S1 mapping of transcripts within the Blog5-Kpnl7 region. Single-stranded probe (1X 10^6 Cerenkov cpm.pmoles^-1) was generated from mKM905 by extension from the universal primer and digestion with HindIII (see Materials and Methods, 2.3.18). 0.2pmoles of probe were hybridized at 55°C with 10μg of RNA from the following: (a) S. rimosus M18883, producing OTC (ca. 2mg.ml^-1) in trusoya medium 1 (TS1); (b) S. rimosus M15883S (a spontaneous mutant of M15883 deleted for the entire OTC cluster) grown on TSB; (c); (d) and (e) different isolates of S. lividans TK54 (pKMS605) grown on YEME containing thiostrepton (25μg.ml^-1). All samples were digested with S1 nuclease. An arrow indicates the position in the M13mp18 sequencing ladder coincident with the protected probe band in lane (a). Vertical lines () indicate the position at which probe protected by transcripts originating from otrAp2 were expected to migrate. Shown at the bottom of the panel is an interpretation of the S1 nuclease protection. The wavy arrow indicates the extent of probe protection by transcripts originating far upstream of otrA.
Figure 3.3 (B) High resolution S1 mapping of transcripts within the BclI-KpnI region. Single-stranded probe (1X 10^6 Cerenkov cpm pmoles^-1) was generated from mKM905 by extension from the universal primer and digestion with HindIII (see Materials and Methods, 2.3.18). 0.2 pmoles of probe derived from mKM905 was hybridized with 10µg of RNA from the following S. rimosus strains: (a) M15883S grown on TSB; (b) M4018 grown on TSB, prepared by Doyle (1987); (c) M4018 grown on TSB (this work) and (d) M15883 grown on TS1 and producing OTC (ca. 2mg.ml^-1). An arrow indicates the position in the mKM905 sequencing ladder coincident with the protected band seen in lanes (b) to (d). Shown at the bottom of the panel is the sequence of the BclI-KpnI region. As the products of the mKM905 sequencing reaction include nucleotides from the polylinker and primer, the actual 5' limit of the protection brought about by transcripts within this region (a|a|a|a) is 38nt upstream of the coincident position in the mKM905 ladder (*).
3.2.4 Transcription of otrA during OTC Production in Batch Culture

The temporal transcription of the otrA resistance gene during the production of oxytetracycline was investigated using RNA isolated from two different batch fermentations; (i) *S. rimosus* M4018 grown on Liquid Complete Medium (Figure 3.4) and (ii) *S. rimosus* M15883 grown on TS1 (a production medium developed by Pfizer; Figure 3.6). For clarity, the S1 nuclease protection analyses of RNA isolated from the different fermentations will be presented separately.

In each case, a single-stranded probe including 387nt of sequence from a *HincII* site between the *Smal*4 and the *BclI* site to the *BamHI* site (see Figure 3.1 for the location of restriction sites) was synthesized from mKM804 (Chapter 4) by extension from the universal (-20) primer and digestion with *EcoRI* (Materials and Methods, 2.3.18). Transcripts originating from otrAp1 and far upstream protected DNA fragments of 166-167nt and 387nt respectively, which were resolved easily in denaturing polyacrylamide gels (Figures 3.5 and 3.7).

To calculate the relative abundance of transcripts originating from otrAp1 and far upstream, it was assumed that radio-labelled nucleotides were evenly distributed along the entire length of the probe. Under standard conditions, the *HincII-BamHI* probes were synthesized from 0.5pmoles of mKM804 template in the presence of 16 pmoles of [α-32P]dATP and 40 pmoles of "cold" dATP. The efficiency with which radio-label was incorporated was always greater than 65%, that is at least 36 pmoles of dATP were incorporated into the probes. As 0.5pmoles of template was present in the labelling reactions, the average probe would have incorporated at least 72 dATP nucleotides. As the average G+C content of streptomycete DNA is 73% mol (Enquist and Bradley, 1971), the primers should have been extended, on average, by 530 nucleotides. Therefore, the majority of the reaction products should have contained the double-stranded *EcoRI* site used to process the probe.

The transcription of otrA at the onset of oxytetracycline (OTC) production was examined using *S. rimosus* M4018 grown on Liquid Complete Medium, which produces typically 1-2mg.ml⁻¹ of OTC after 10-
12 days (Figure 3.4). RNA isolated from *S. rimosus* M4018, before and after the first time point at which OTC was first detected in the medium, was analysed by high-resolution S1 nuclease mapping using the HincII-BamHI probe (Figure 3.5). The intensity of the bands produced by autoradiography was determined by densitometry scanning (Materials and Methods, 2.3.21). After correction for differences in the length of the radio-labelled fragments, the band corresponding to 0.01 pmoles of untreated probe in lane (a) was shown to be 2 to 10 fold more abundant than the probe bands protected by the *otrA* transcripts in samples (c) to (e). As the RNA samples were hybridized with 0.1 pmoles of probe, the molar excess of probe to *otrA* transcripts was calculated to be between 20 to 100 fold, which is within the limits recommended by Calzone and others (1987). RNA isolated from *S. rimosus* M4018 grown on TSB, which was analysed previously (Figure 3.3, Panel B), was used as a marker to indicate RNA-protected fragments corresponding to transcripts originating from *otrA* and "far upstream" (lane b).

During the period of rapid growth, namely 24 hrs after inoculation and before the onset of OTC production, transcripts originating from *otrA* accounted for greater than 95% of the total *otrA* transcription (Figure 3.5, Panel B). During the transition to OTC production, the relative abundance of the *otrA* transcripts changed dramatically. At 32 hours, when OTC was still not detected in the culture supernatant, the contribution made by *otrA* had decreased to 80%. During the early phase of OTC production, 56 hours after inoculation, transcription from the promoter(s) located far upstream (pFAR) accounted for 51% of the total transcription of *otrA*. This shift in the relative contribution of p1 and pFAR to the transcription of *otrA* was accompanied by overall decrease of approximately 80% in the abundance of *otrA* transcripts within the total RNA population (Figure 3.5, Panel B).

The transcriptional analysis described so far, including that of Doyle (1987), has utilised subclones of DNA isolated from *S. rimosus* M15883 and RNA isolated from *S. rimosus* M4018. The strain development programme, which derived *S. rimosus* M15883 from *S. rimosus* M4018
Figure 3.4. Oxytetracycline production by *S. rimosus* M4018 grown on Liquid Complete Medium. The graph is derived from 6 samples (O) taken from different cultures (50ml), inoculated at the same time with $10^7$ spores and incubated in an orbital shaker at 30°C. The • symbol indicates the concentration of OTC in cultures from which RNA was isolated and subsequently used in S1 nuclease protection experiments (Figure 3.5). The symbol † indicates the dry weight (mg.ml⁻¹) of the mycelium at different stages in the fermentation.
Figure 3.5. (A) S1 nuclease mapping of otrA transcripts in total RNA isolated from *S. rimosus* M4018 grown on Liquid Complete Medium (LCM) (see Figure 3.4). Single-stranded probe (0.2pmoles, $2.1 \times 10^7$ Cerenkov cpm.pmol$^{-1}$) synthesized from mKM804 by extension from the universal primer and digestion with EcoRI (see Materials and Methods, 2.3.18) was hybridized at 55°C with 10μg of RNA from M4018 grown on LCM for 24, 32 and 56hrs (lanes c, d and e). RNA isolated from M4018 grown on TSB was also included to provide markers for the RNA-protected fragments corresponding to transcription from otrApl and far upstream, which are indicated by p1 and pFAR respectively (lane b). The position of full-length probe (FLP) is also shown. 0.01 pmoles of untreated probe were also loaded on the gel (lane a).
Figure 3.5 (B) Histogram showing the results of quantitative analysis of the otrA transcripts originating from p1 and far upstream in total RNA isolated from S. rimosus grown on LCM. Abundance estimates were obtained by densitometry scanning of autoradiographs (Materials and Methods, 2.3.21) produced by different exposures to the denaturing polyacrylamide gel in which RNA-protected fragments had been resolved.
through several rounds of mutagenesis and selection for survivors that produced higher titres of OTC, has almost certainly lead to sequence differences within the OTC cluster of these two strains.

To obtain RNA that would be completely homologous with the DNA probes and could, thus, be used to analyse the transcription of the entire OTC cluster, a short period of time was spent at the Pfizer laboratories in Sandwich. Culture conditions were developed for *S. rimosus* M15883 that yielded reproducible oxytetracycline profiles, thus enabling RNA to be isolated at different stages during the production of OTC.

Aliquots (1ml) of a stationary phase culture of *S. rimosus* M15883 grown in Liquid Complete Medium (LCM) were used to inoculate TS1 production media (50 ml) in Erlenmeyer flasks (300 ml). The cultures were incubated on a rotary shaker with a 4cm stroke at 28°C. Samples of the supernatant were analysed frequently for OTC and RNA was isolated at different phases in the fermentation (Figure 3.6).

It was not possible to determine a growth profile for *S. rimosus* M15883, as the production medium (TS1) contained a colloidal suspension of soya flour and starch and an emulsion of vegetable oils, which interfered with the estimations of growth using optical density measurements and dry weight determinations. Furthermore, RNA could only be isolated after incubation for 18 hours, as the soya flour and starch, which were not utilised as substrates for growth during the early stages of the fermentation, formed an impervious "nougat-like" layer around the mycelium when they were harvested by either centrifugation or filtration.

High resolution S1 nuclease protection, as described in the previous section, was employed to analyse the RNA isolated from cultures of *S. rimosus* M15883 at different stages during the production phase (Figure 3.7; samples d to g). The most striking features of this analysis were (i) the absence of detectable amounts of transcripts from *otrAp1* and (ii) the transient peak in the abundance of *otrA* transcripts from
Figure 3.6. Oxytetracycline production by *S. rimosus* M13883 grown in Trusoya Medium 1. The plot is derived from 14 samples (0) taken from different (50ml) cultures, inoculated at the same time and incubated under standardised conditions. The (●) symbol indicates the concentration of OTC in cultures from which RNA was isolated for S1 nuclease protection experiments (see Figure 3.7).
Figure 3.7. (A) S1 nuclease protection by transcripts of the *otrA* resistance gene from *S. rimosus* M15883 grown on TS1. Single-stranded probe (2.9X 10⁷ Cerenkov cpm.pmol⁻¹), was generated by extension from universal primer, which was annealed to mKM804, and digestion with *EcoRI* (Materials and Methods, 2.3.18). 0.2pmoles of probe was hybridized at 52°C with 10µg of RNA from the following: (a) *S. rimosus* M15883S grown on TSB medium; (b) *S. rimosus* M4018 grown on TSB (prepared by Doyle, 1987); (c) *S. rimosus* M4018 grown on TSB medium (this work); (d); (e); (f) and (g) *S. rimosus* M15883 grown on TS1 (the OTC production medium) for 18, 46, 69, and 93 hours respectively; (h) and (i) as (b) and (c) respectively. Samples (a) to (g) were digested with S1 nuclease. Exonuclease VII was used to digest samples (h) and (i). The positions in the mKM804 sequencing ladder that are coincident with the RNA-protected fragments derived from transcripts originating from *otrA* pl and far upstream are indicated by pl and pFAR. The position of probe protected along its entire length by contaminating template DNA (FLP) is also indicated.
Figure 3.7. (B) Histogram showing the results of quantitative analysis of the otrA transcripts originating from p1 and far upstream. Abundance estimates were obtained by densitometry scanning (Materials and Methods, 2.3.21) of autoradiographs produced by different exposures to the denaturing polyacrylamide gel in which the RNA-protected fragments had been resolved (panel A).
far upstream at the onset of antibiotic production. The abundance of otrA transcripts from pFAR within the total RNA population increased by a factor of 4, between 22 and 46 hours. Over the next two time points, 69 and 93 hours, the level of these transcripts declined steadily to almost the 22 hour level. This pattern of transcription from pFAR has been reproduced in a separate S1-nuclease protection experiment (result not shown). RNA could not be isolated from cultures older than 100 hours, suggesting that the majority of protein synthesis and cellular metabolism had ceased by this point in the fermentation.

Exonuclease VII digestion of hybridization complexes formed between the HincII-BamHI probe and RNA from S. rimosus M4018 (Figure 3.7, lanes b and c) produced RNA-protected fragments of similar size to those produced by S1 nuclease digestion (Figure 3.7, lanes h and i). This indicated that the proposed transcription start sites for otrApl at 200-221nt (Doyle, 1987) were not an artifact of endonucleolytic cleavage at a site of secondary structure caused by a breakdown in complementarity between the otrA message isolated from S. rimosus M4018 and the DNA probe derived from S. rimosus M15883. Exonuclease VII can only digest unprotected probe from the 3' and 5' ends to within 5-7nt of the DNA-RNA hybrid (Rose and Botstein, 1983). Consequently, the Exonuclease VII-resistant fragments (lanes h and i) are 10-14nt longer than the corresponding S1 nuclease-resistant fragments (lanes b and c).

3.3 Discussion

The nuclease protection experiments and in vivo promoter analysis described in this chapter, have shown clearly that the otrA resistance gene is transcribed from otrApl and a promoter(s) located far upstream, the former resembling the prokaryotic consensus sequence (Doyle, 1987). It is now believed that overdigestion of a DNA-probe, which was protected by transcripts originating from far upstream of otrA, with S1 nuclease led previously to the erroneous identification of a promoter that overlapped the BamHIs site (Doyle, 1987).
Interestingly, the maximum level of kanamycin resistance conferred by pKMS605 (containing the 305bp BcIIS-KpnI region) is at least three fold greater than that conferred by pDM130 (containing the SmaI4-BamHIss region). The pIJ486-construct pDM110, that contains the BcIIS-BamHIss region, conferred the same level of resistance as pDM130 (Doyle, 1987). The increased kanamycin levels specified by pKMS605 appeared to be coincident with the inclusion of the 91bp BamHIss-KpnI region in the cloned DNA.

Tetracycline resistance in S. rimosus ATCC 10970 and S. rimosus M4018 can be induced by pre-exposure to sub-lethal concentrations of tetracycline (Ohnuki et al., 1985; Lee, unpublished results). The otrA homologue from S. rimosus ATCC 10970, designated tetA, has been cloned in Streptomyces griseus using a plasmid vector. The demonstration that expression of tetA was also inducible in this heterologous host suggested that; (i) any DNA segment(s) required for the regulation of tetA expression was also cloned in the plasmid, (ii) S. griseus contained the necessary factor(s) required for the induction of tetA expression and (iii) that the induction mechanism(s) was not unique to S. rimosus or oxytetracycline biosynthesis. The induction of high levels of tetracycline resistance conferred by tetA has not been investigated further.

In E. coli and Salmonella typhimurium the transposon Tn10 carries two genes, tetA and tetR, which confer high-level, inducible resistance to tetracycline (Kaneko et al., 1985). The amino acid sequence of the deduced tetA protein has high identity to the predicted amino acid sequence of the tetB (otrB) gene product from S. rimosus (Reynes et al., 1988). The protein products of both tetA of Tn10 and tetB of S. rimosus have been shown to mediate the efflux of tetracycline from the cell (Kaneko et al., 1985; Ohnuki et al., 1985). The tetR gene encodes a repressor protein which regulates negatively transcription of both the tetA and tetR genes, which are transcribed from divergent, overlapping promoters. Tetracycline binds to the tetR repressor reducing its affinity for the tet operator and thus, induces expression of both tetA and tetR (Hillen et al., 1984). Sequencing of the act cluster of S. coelicolor
within a region that confers resistance to actinorhodin has identified PPCRs, which are predicted to encode proteins with high amino acid sequence identity to the Tn10 tetA and tetR products (Hopwood, pers. comm.). It is possible that actinorhodin acts as an inducer of resistance in *S. coelicolor*.

Sequencing of the otc cluster of *S. rimosus* around the resistance determinants has not identified any PPCR which is predicted to encode a regulatory protein that is homologous to the tetR protein from Tn10. It is unlikely, therefore, that the DNA fragments from *S. rimosus*, which confer tetracycline resistance in *S. griseus* (Ohnuki *et al.*, 1985), encode a regulatory gene similar to tetR of Tn10. The mechanism by which tetracycline resistance in *S. griseus* can be induced remains to be elucidated. *S. griseus* may encode a regulatory protein, which is similar to tetR and binds to a regulatory sequence within the promoter region of tetA (otrA) from *S. rimosus* and thus blocks transcription. However, this proposal raises a number of questions. What are the host-encoded gene(s) that the putative regulator protein would regulate normally? How do sub-lethal concentrations of tetracycline induce tetracycline resistance? Does tetracycline bind the putative regulatory protein, decreasing its affinity for an operator? *S. griseus* is not known to produce any secondary metabolites, which are related structurally to tetracycline. What is the structure of the inducer? Can sub-lethal concentrations of other antibiotics induce tetracycline resistance in transformants of *S. griseus* containing resistance determinants from *S. rimosus*? Although there are precedents for repressor/induction mechanism in regulating tetracycline resistance in Tn10 and possibly actinorhodin resistance in *S. coelicolor*, it is possible that the inducible tetracycline resistance conferred by *S. rimosus* determinants in *S. griseus* is mediated by an alternative, unknown mechanism. Analysis of the otrA promoter region has not identified any sequence, which is similar to the tet operator in Tn10 (Hillen *et al.*, 1986).
Chapter 3  Analysis of Promoter Activity Upstream of otrA

As the minimum length of cloned DNA that could confer inducible tetracycline resistance was 2.5kb (Ohnuki et al., 1985) and the protein coding sequence of otrA extends for 2kb (Doyle, 1987), any regulatory sequences required for inducible expression are probably contained within 500bp upstream of translation start codon of tetA. Assuming that the transcriptional organisation and regulation of OTC biosynthesis are similar in S. rimosus M15883 and ATCC 10970, it is possible that the pKMS605 construct, which contains sequence extending from otrAp1 to 435bp upstream of the protein coding region sequence of otrA, contains any regulatory sequences required for the inducible expression of otrA. This could be tested easily by introducing pKMS605 into S. griseus and determining if kanamycin resistance, which is conferred by the aphII reporter gene, can be induced by pre-exposure to sub-lethal concentration of tetracycline (and/or oxytetracycline). It would also be interesting to determine if expression of the aphII gene can also be induced by tetracycline in the S. lividans transformants containing pKMS605.

Although the induction of expression of otrA or tetA has not been investigated further, the significantly lower kanamycin resistance levels conferred by pDM130 and pDM110 compared with pKMS605 could result from the disruption of the mechanism(s) which regulates expression of otrA. As there are 220bp between the transcription start sites of otrAp1 and the BamHIa site, sequence(s) distant from this promoter could be implicated in the regulation of otrA expression. Although, the binding of regulatory protein(s) to recognition sequences on DNA far from the point of transcription initiation was discovered initially in eukaryotic cells, similar regulatory interactions have been found in prokaryotes (for an overview see Magasanik and Neidhardt, 1987; Magasanik, 1989) and may be important in regulating the transcription from otrAp1. It is envisaged generally that DNA looping (Ptashne, 1986) facilitates proteins bound at a distance from the promoter sequence to interact either with other regulatory proteins bound near to the promoter or with the RNA polymerase directly. For example, the repression of transcription of the lacZYA operon in E. coli requires the cooperative binding of the lacI repressor to two operator sites (for review see
Gralla, 1989). The primary operator $O_1$ overlaps with the promoter sequence (Gilbert et al., 1976), while the secondary operator $O_2$ is located within the lacZ gene, 401bp downstream of the transcription start point (Reznikoff et al., 1974). The lac repressor has been shown to bind simultaneously to $O_1$ and $O_2$, causing the intervening DNA to be looped out (Kramer et al., 1988). The cooperative binding of the repressor to both of these operator sites is thought to repress transcription of lacZ in two ways; (i) by tightening the interaction of the repressor bound at $O_1$, thereby, preventing RNA polymerase from binding to the promoter and (ii) by preventing the transcription of lacZ proceeding beyond the secondary operator ($O_2$), within the $\beta$-galactosidase gene (Flashner and Gralla, 1988). An analogous arrangement of operator sites has also been found in the ara operon (Huo et al., 1988; Hamilton and Lee, 1988) and the gal operon (Irani et al., 1983; Haber and Adhya, 1988). The mechanisms which have been proposed to regulate transcription of the ara and gal operons also involve the cooperative binding of repressor protein and the looping out of DNA. Should transcription from otrA$\text{p}1$ be subject to repression (although there is no experimental evidence to distinguish between an activation or repression mechanism), it is difficult to envisage how the addition of an upstream region would lead to "derepression".

The disparity in the level of resistance to kanamycin conferred by pDM130 and pKMS605 could be explained should maximum transcription from otrA$\text{p}1$ require an interaction with a protein bound to DNA far upstream (at least 200bp) of the transcription start point. In prokaryotes, only a single example of a pathway-specific activator protein, which binds at a distance from promoter sequences, has been described. The promoter of the $\text{gln}A\text{LG}$ operon is controlled by the nitrogen regulator NR$_1$ (also called NtrC or GlnG), which is activated by phosphorylation in response to nitrogen starvation (for review see Magasanik, 1988). The looping of DNA is thought to facilitate the activation of RNA polymerase holoenzyme Eo$^{54}$, which is prebound to the promoter, by facilitating an interaction with phosphorylated NR$_1$ (P-NR$_1$) bound about 100bp upstream of the transcription start point (Sasse-Dwight and Gralla, 1988). The binding site for P-NR$_1$ can be moved more
than a kilobase away from the promoter and still facilitate transcriptional activation (Reitzer and Magasanik, 1986). A similar activation mechanism has also been identified in other systems regulated by nitrogen, for example, the transcriptional activation of the formate dehydrogenase gene in *E. coli* (Gralla, 1990). It has been proposed that Eoσ54 is a specific receptor for the NR1 regulator (Gralla, 1989).

The analysis of transcriptional activation by the OmpR protein has revealed that activation can be mediated from locations several turns of the DNA helix away so long as the binding site remains on the original side of the DNA (Maeda *et al.*, 1988). However, the ability to activate transcription diminished significantly with distance. It was suggested that activator proteins such as CRP and OmpR, which activate Eoσ70, may be unable to act over long distances in *E. coli* (Gralla *et al.*, 1989). To date, the transcriptional activation of the stable RNA (ribosomal and transfer RNA) operons of *E. coli* by the binding of FIS to DNA determinants up to 130bp upstream of the transcription start sites (Nilsson *et al.*, 1990) represents the furthest distance that a protein bound to DNA has been shown to activate transcription by Eoσ70.

In *Bacillus subtilis*, hyperprotease production (*hpr*) mutants have been isolated, which result in increased expression of the neutral protease (*nprE*) and alkaline protease (*aprE*) genes (Henner *et al.*, 1988). Northern analysis has shown that the *hpr* mutations affect the expression of the *aprE* gene at the level of transcription (Ferrari *et al.*, 1988). Deletion analysis of *aprE*'-*lacZ* fusions in *hpr* mutants has revealed that stimulation by *hpr* is lost somewhere between 200 and 400bp upstream of the transcription initiation site. Although the mechanism by which *aprE* expression is activated remains to be elucidated, this example nevertheless demonstrates the importance of DNA determinants located "far upstream" in regulating gene expression. Therefore, it is a formal possibility that maximum expression of *otrA* requires regulatory sequences located within the *BamHIE-KpnH* region, at least 220bp upstream of the transcription start point of *otrApl*. 
The differences in kanamycin levels between pDM110 and pKMS605 could also reflect differences in the topology of the DNA in the immediate vicinity of otrAp1. It is well established that promoter function in vivo and in vitro can be influenced strongly by the level of DNA supercoiling (for review, see Drlica, 1987). For example, the specific induction of transcription of the proU loci of Escherichia coli and Salmonella typhimurium is mediated by changes in the DNA superhelicity in vivo in response to osmotic stress (Higgins et al., 1988). Analysis of the leu-500 promoter of S. typhimurium (Richardson et al., 1988) has shown that local rather than global changes in DNA topology can be important in activating transcription. Whilst changes in the topology of DNA will clearly be influenced by changes in the activities of topoisomerase I and gyrase, the local topology may also be dependent on the binding of proteins such as HU and H1 (the general histone-like DNA-binding proteins) or more specifically by factors similar to FIS (Factor for Inversion Stimulation: described by Kahman et al., 1985; Johnson and Simon, 1985) or IHF (Integration Host Factor: described by Craig and Nash, 1984).

Although a number of inverted repeats and direct repeats, which could be the binding sites for regulatory proteins, are present upstream of otrAp1 in the region inserted in the pKMS605 construct (result not shown), no sequence similarity has so far been detected to any consensus binding sites for DNA-binding proteins. It would not be constructive to discuss the repeats located upstream of otrA in any detail, as streptomycetes have an inherent capacity to form such structures at high frequency, due to the high G+C composition of the DNA and the restricted use of synonymous codons.

The above is not intended to be an exhaustive discussion of the mechanisms which could account for the disparity in the levels of kanamycin resistance conferred by pKMS605 and pDM130. It is intended merely to demonstrate the scope of possible regulatory interactions. As the first step in analysing the tentative regulation of transcription, a series of promoter-probe constructs containing otrAp1, with the upstream region truncated to different extents, is currently being made to delineate the sequence(s) required for the maximal activity of otrAp1 in vivo (McDowall, unpublished work).
Using high-resolution S1 nuclease protection experiments to analyse RNA samples isolated from cultures at different stages in the fermentation of oxytetracycline, it is possible to quantitate the relative changes in the abundance of transcripts from otrAp1 and far upstream. However, only general inferences can be made about the absolute levels of otrA transcripts within the mycelium, as the rates of accumulation and turnover of rRNA, tRNA and mRNA in streptomycetes have not been studied in any detail. Physiological studies using Escherichia coli as a model have indicated that the rate of synthesis of rRNA and tRNA is regulated and corresponds to the cellular need for ribosomes (for an overview see Jinks-Robertson and Nomura, 1987). As the growth rate of E. coli slows, the demand for protein synthesis decreases and there is a corresponding decrease in the number of ribosomes per cell. As rRNA and tRNA account for the majority of RNA in E. coli (greater than 95% in cells growing exponentially) a slowing of the growth rate results in an overall reduction in the total RNA per cell. The growth rate dependence of ribosome accumulation is almost certainly a ubiquitous feature of bacterial growth, due to the central role of ribosomes in the synthesis of other cellular components and the large amounts of cellular energy and metabolic intermediates required to produce ribosomes.

A striking reduction in protein, DNA and RNA synthesis has been detected in the transitory phase between vegetative growth and morphological development in Streptomyces coelicolor A3(2), when grown on solid medium (Granozzi et al., 1990). This reduction in macromolecular synthesis is similar to the change seen in E. coli cultures entering stationary phase. In S. coelicolor A3(2), however, the synthesis of macromolecules resumes at the onset of morphological development and is restored nearly to the values measured during vegetative growth (Granozzi et al., 1990). The production of secondary metabolites coincides typically with morphological development. Indeed, substantial genetic evidence suggests that both processes share elements of a complex regulatory network and at least in principle could be triggered by a common signal (General Introduction, 1.9). The accumulation of RNA during the fermentation of oxytetracycline may also occur in two distinct phases.
Analysis of the transcription of otrA, using RNA isolated from *S. rimosus* M4018 grown on Liquid Complete Medium, has shown that this resistance gene is transcribed almost exclusively from otrAp1 during intense vegetative growth. The relative contribution of otrAp1 diminishes, however, as the growth rate slows and the onset of oxytetracycline production approaches. Early in the production phase, transcripts originating from otrAp1 and far upstream are almost equal in abundance.

The changes in the relative abundance of transcripts of otrA were seen against an overall decrease in the abundance of these transcripts within the total RNA population. Any change in the overall RNA content of the *S. rimosus* M4018 mycelium, as the culture approaches stationary phase and the onset of antibiotic production, is likely to be a reduction in the total amount of rRNA, tRNA and mRNA (associated with vegetative growth). The decrease in the abundance of otrA transcripts, within the total RNA population at this early stage in the fermentation, therefore, probably represents a real decrease in the actual number of transcripts within the mycelium. However, before credible predictions can be made about the absolute levels of transcripts, it will be necessary to study the rates of accumulation and turnover of mRNA, rRNA and tRNA at different stages in the life cycle of *Streptomyces*.

The analysis of transcription during the production phase of a fermentation failed to detect transcription from otrAp1 in RNA isolated from *S. rimosus* M15883 grown on TS1. Transcripts originating from far upstream increased in abundance during the early stages of production (relative to total RNA), peaking at about 20 hours after oxytetracycline was first detected in the culture broth, before decreasing in abundance to levels seen at the onset of production. Should a second round of RNA synthesis occur at the onset of oxytetracycline production in liquid cultures of *S. rimosus*, as described for the morphological development of *S. coelicolor* A3(2) on solid medium (Granozzi *et al.*, 1990), then the transient increase in the abundance of transcripts from far upstream, within the total RNA population, may represents an actual increase in
their numbers within the mycelium. It is proposed that the transient peak in the abundance of otrA transcripts from far upstream within the total RNA population probably reflects transient activity at the cognate promoter(s), located far upstream.

A similar observation has been reported recently by Guthrie and Chater (1990). The transcripts of an undescribed gene in the undecylprodigiosin cluster of S. coelicolor increased briefly in abundance after the most rapid growth phase had passed (Guthrie and Chater, 1990). The level of transcription in the undecylprodigiosin cluster was estimated by the insertion of a promoterless xyIE gene (described in Chapter 5) into a previously undescribed region implicated in the early steps of biosynthesis, using a bacteriophage-based integrative vector.

The results presented in this chapter indicate that the expression of the otrA resistance gene is under complex transcriptional control. The two transcripts of otrA which have been identified are transcribed from promoters with distinct temporal activities. In Bacillus subtilis a cascade of alternative sigma factors enables RNA polymerase to recognise unique classes of promoters and thereby activate selective sets of sporulation genes, in the correct order and at the appropriate time (Losick and Pero, 1981). The finding that the switch from continued extension of aerial hyphae to their development into chains of spores, is controlled by an alternative sigma factor (σ^whiC) in S. coelicolor, suggests that a similar regulatory cascade could be important in controlling the differentiation of Streptomyces (Chater et al., 1989).

The putative -35 region (CTGACC) and -10 region (TACCGT) of otrAp1 resembles the -35 region (TTGACA) and -10 region (TATAAT) of the consensus for the major class of eubacterial promoters (Hawley and McClure, 1983). Correspondingly, this promoter has a distinct temporal activity associated with vegetative growth. Using the pKO500 promoter-probe vector (Close and Rodriguez, 1982), the BamH15-BclI5 fragment containing the otrAp1 promoter was shown to be capable of directing transcription in E. coli (McDowall, unpublished result). Therefore, it
would appear that the otrAp1 promoter is recognised by a form of RNA polymerase in Streptomyces which is responsible for the transcription of the general housekeeping genes.

Comparative analysis of promoters responsible for the transcription of genes involved in morphological and physiological differentiation could define common elements involved in the control of Streptomyces development. Therefore, the promoter responsible for transcription of otrA from far upstream was isolated and characterised as the next step in dissecting the coordination of oxytetracycline resistance and production (see Chapter 5).
CHAPTER 4
THE OTCZ GENE: NUCLEOTIDE SEQUENCE, DEDUCED FUNCTION AND TRANSCRIPTION
4.1 Introduction

The *otc-151* mutant is representative of the *otcZ* class of mutants that are blocked in "early" oxytetracycline (OTC) biosynthesis, before the formation of anhydrotetracycline (ATC). The specific enzymatic step which is blocked has not been assigned, as the *otcZ151* mutant failed to produce OTC in cosynthesis tests when it was combined with representatives from other groups of "early" mutants on agar strips and was also unable to convert purified "early" intermediates to OTC when these were supplied to liquid cultures (Rhodes *et al.*, 1981).

Antibiotic production was restored in the *otcZ151* mutant by the introduction of a plasmid containing the 4.4kb *SphI*-*SpHII* region (Figure 4.1) (Butler *et al.*, 1989). The mutation on the chromosome was either complemented *in trans* by the expression of a functional gene from the plasmid or it was repaired by a recombination event(s) between homologous sequences on the chromosome and DNA cloned on the plasmid. Using a series of plasmids that contained truncated derivatives of an insert, which in full-length form could restore antibiotic production in *otcZ151*, the mutation was mapped close to the "left" of the *PstI* site (Butler *et al.*, 1989).

The extreme "left" of the 4.4kb *SphI* region (*SphI*-*BclI*) contains the 5' end of the *otRA* resistance gene, which has been sequenced (Doyle, 1987). The *otcc* gene, encoding the ATC oxygenase, has also been localised within the *SphI*-*SpHII* region using "reverse genetics" (Binnie *et al.*, 1989). Purification of the ATC oxygenase allowed "reverse translation" of the amino acid sequence from the amino-terminus. An oligonucleotide containing the predicted nucleotide sequence was synthesized chemically and used as a probe to locate the 5' end of the *otcc* gene to the "left" of the *SpHII* site. Limited sequencing suggested that the potential protein coding region (PPCR) of *otcc* read from "right" to "left". From the molecular weight of the purified ATC oxygenase (52 kD), the 3' end of the *otcc* gene was predicted to be between the restriction sites *SmaI* and *PstI*. 
Chapter 4  The otcZ gene 114

Figure 4.1. Restriction map of the "left" end of the oxytetracycline cluster. The location of otrA and the probable locations of otcC and otcZ are indicated. For clarity only the restriction sites that are relevant to the delineation of the position of otcZ are shown. B, BamHI; Bc, BclI; H, HindIII; K, KpnI; S, SstI; Sm, Smal; Sp, SphI; P, PstI.
Chapter 4  The otcZ gene  115

The above results delineated the possible location of the DNA segment, which "complements" the otcZ151 mutation, to the 1.3kb BciII-Smal region. The determination of the DNA sequence in this region was considered a priority for the following reasons.

(i). The DNA sequence could be used to predict the precise genetic organisation of this region and the primary structure of the otcZ gene product.

(ii). The inferred protein sequence could then be compared with entries in computer-held databases, such as GenBank, EMBL and NBRF, for similarity to proteins of known biological function. Homology with other proteins might provide insight into the enzymatic step blocked in the otcZ151 mutant.

(iii). A knowledge of the genetic organisation, in conjunction with an analysis of transcription, would be necessary to identify regulatory sequences involved in controlling the expression of the oxytetracycline genes.

4.2 Results
4.2.1 Sequencing Method

Modified DNA polymerase from *Thermus aquaticus* (Taq polymerase) was used to sequence the otcZ gene. This thermo-stable enzyme has a temperature optimum of 80°C, which is considerably higher than that of other commercially-available DNA polymerases for dideoxy sequencing, such as modified T7 DNA polymerase (supplied by United States Biochemical Corporation as SEQuenase™) and modified DNA polymerase I from *Escherichia coli* (supplied by BRL as Klenow), which are used at recommended incubation temperatures of 42°C and 30°C respectively. The high incubation temperature of Taq polymerase was found to be a considerable advantage when sequencing DNA from streptomycetes, which typically has a base composition of 73% mol G+C (Enquist and Bradley, 1971). Inverted repeat sequences occur frequently in *Streptomyces* DNA and these can hybridize to form highly stable G+C-rich secondary structures.
Chapter 4  The otcZ gene  116

The effects of secondary structure can be manifested at two different stages during sequencing. Firstly, stable hairpin loops in the template can cause the DNA polymerase to "stall" during the sequencing reaction and as a result artifactual bands are produced in all four tracks of the sequencing ladder. Secondly, the products of the sequencing reaction may form regions of G+C-rich secondary structure. This can cause bands to be compressed, and result in part of the sequencing ladder being unreadable over a localised area.

Dideoxy sequencing of DNA from Streptomyces has been undertaken previously at Glasgow using Klenow DNA polymerase (Doyle, 1987). Secondary structure posed a significant problem and it was necessary to use elevated incubation temperatures, formamide gels and the chemical degradation method of Maxam and Gilbert (1977) to complete the sequence of the otrA resistance gene. A number of alternative polymerases for DNA sequencing have become readily available since the sequencing of otrA. These include SEQuenase™ II and TAQuenase™ (modified Taq polymerase), both supplied by USB.

Before a strategy for the sequencing of otcZ was devised, the ability of these polymerases to produce readable sequence from a single-stranded template (mKM905) known to cause the Klenow enzyme to "stall" during sequencing was investigated using the standard protocols supplied with the USB kits (Figure 4.2). The SEQuenase™ enzyme apparently had a considerable problem reading through regions of streptomycete DNA at the recommended incubation temperature of 42°C (Figure 4.2, Panel A). At the higher incubation temperature used with Taq polymerase (70°C) the G+C-rich secondary structures (Figure 4.2(B)) were presumably less stable and consequently the progression of the sequencing reaction was not impeded. The TAQuenase™ sequencing kit was used, therefore, for the sequencing of otcZ. Termination mixes containing the nucleotide analogue 7-deaza dGTP were used routinely to minimise the possibility of band compressions in the sequencing ladders. This analogue pairs weakly with conventional bases and thus, reduces the possibility of secondary structure formation in the products of the sequencing reaction (Mizusawa et al., 1986).
Figure 4.2. (A) Comparison of the sequencing reactions using SEQuenase™ and TAQuenase™. (B) Secondary structure predicted by STEMLOOP in the region of the premature terminations. Small arrows (►) indicate the precise positions at which the SEQuenase enzyme stalls on the template. For the SEQuenase™ reactions, the radio-label was [35S] α-dATP. [32P] α-dATP was used for the TAQuenase™ reactions.
It has been found subsequently that other groups have obtained good quality sequence data from streptomycete DNA using SEQuenase™, when the reactions are carried out at elevated incubation temperatures (ca. 55°C) and used in conjunction with single-stranded DNA-binding protein from bacteriophage T4 (Leadlay, pers. comm.).

4.2.2 Sequencing Strategy

A summary of the sequencing strategy is shown in Figure 4.3. Three internal fragments from the 1.3kb BcII-SmaI region were ligated to M13mp18 and M13mp19 respectively, which had been digested with the appropriate restriction endonucleases, and introduced into E. coli TG1 by genetic transformation (Section 2.3.12). The single-stranded templates were isolated from phages containing recombinant DNA molecules (Section 2.3.16). The 250bp PstI-SmaI fragment from pKME917 (pUC19 containing the SmaI+9 fragment) was inserted into M13mp18 and M13mp19 to generate mKM803 and mKM903 respectively. The 722bp KpnI-PstI and the 305bp BcII-KpnI fragments were obtained from the pUC18-based constructs, which were intermediates in the introduction of the same fragments into the pIJ486 promoter-probe vector (Section 3.2.1). HindIII/EcoRI fragments from these constructs were then inserted into M13mp18 and M13mp19 to derive mKM807 and mKM907 from pKME807 and mKM805 and mKM905 from pKME805 respectively.

The 722bp KpnI-PstI inserts in the mKM807 and mKM907 templates could not be sequenced completely using the universal primer (UP), as only around 300 nucleotides could be read from the standard sequencing reactions. Oligonucleotides 807a and 907a, which incorporated sequence obtained from extension of the universal primer were used, therefore, as primers to extend the sequencing of the mKM807 and mKM907 templates respectively. In turn, sequence data generated from the 807a and 907a primers was used to design oligonucleotides 807b and 907b respectively, which allowed the mKM807 and mKM907 templates to be sequenced completely.
Figure 4.3. M13 templates used to sequence the 1.3kb BcII-Smal region. The template designations are given above the lines (|—|) used to illustrate the extent of the cloned DNA. The arrows indicate the direction and extent of primed synthesis during nucleotide sequence analysis. The oligonucleotides used to prime the extension reactions are given above the arrows. UP, represents the universal primer. For clarity, only restriction sites relevant to the subcloning of the BcII-Smal region are shown.
Figure 4.4. Comparison of sequence ladders produced from mKM907 and the complementary template mKM807. A line (|) indicates the extent of the sequencing artifact seen in the mKM907 ladder but not evident at the corresponding position in the mKM807 ladder.
Only a single stretch of 5 to 6 nucleotides from the mKM903 template could not be read unambiguously due to the comigration of bands from different tracks in the sequencing ladder (Figure 4.4). Interestingly, this region (1110–1130nt) is not predicted to form significant secondary structure. The sequence data from the complementary strand (mKM803) was excellent and no attempts were made to overcome this artifact. Upon completion of this work, non-overlapping regions of sequence remained at the KpnI and PstI boundaries. Two other M13mp18-derived constructs were used to sequence across these restriction sites. The mKM817 construct contained the 1.6kb SmaI–SmaI fragment, which was subcloned from pKME917 as a HindIII/EcoRI fragment, with the SmaI site nearest the EcoRI site in the polylinker. The mKM971 construct contained the same insert but in the opposite orientation. The mKM817 template, in conjunction with the oligo 807b, was used to sequence across the KpnI boundary. The PstI boundary was sequenced using mKM917 and oligo 907b, thereby completing the sequencing of the 1.3kb BclI–SmaI region.

The sequence is comprised of 1,265bp and has a base composition of 72.57 mol% G+C. The annotated sequence, including the amino acid sequences of potential protein coding regions, a putative ribosome binding site (RBS), the oligonucleotides used for sequencing and the locations of restriction sites referred to in the text are shown in Figure 4.5.

The sequencing reactions and related gels were performed and run as detailed in Materials and Methods (Section 2.3.17). All of the templates were sequenced an average of three times. The autoradiographs were read by eye and entered into computer files using the sequence assembler/editor program SEQED (Devereux et al., 1984). Each autoradiograph was interpreted at least twice, each reading being entered into independent files. Duplicate files were then compared using the GAP program to eliminate "input" errors.
Chapter 4  The otcZ gene 124

Figure 4.5. Annotated sequence of the 1,265bp BcII-SmaI region. The predicted amino acid sequences of otcZ and the C-terminus of otcC are shown below the corresponding nucleotide sequences. The shaded boxes (-----) indicate the nucleotides incorporated into the oligonucleotides which were used as primers for sequencing. A possible ribosome binding site (RBS) for otcZ is also indicated. For clarity, only restriction sites referred to in the text are indicated.
4.2.3 Computer-Assisted Sequence Analysis

Computers have an integral role in the analysis of DNA sequence, due to the enormous amount of data that has been generated and the complex algorithms employed in sequence analysis and comparison. The 1.3kb BclI-Smal sequence was analysed using the "Sequence Analysis Software Package, Version 6.2" from the Genetics Computer Group at the University of Wisconsin (UWGCG) (Devereux et al., 1984). This package has been compiled specifically for molecular geneticists wishing to analyse and compare sequence data rapidly and accurately. Access to the three main databases; EMBL, NBRF and GenBank was available through this system.

The UWGCG package was used successfully to predict the potential protein coding region of otcZ and the amino acid sequence of the gene product, to identify similarity with other proteins and consequently to suggest a possible enzymatic function for the predicted protein. The individual UWGCG programs which were used are listed below.

WORDSEARCH: searches for any sequence similarity between the query sequence and entries in the databases using a Wilbur and Lipman algorithm (1983).

SEGMENTS: aligns and displays segments of similarity found by WORDSEARCH.

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 Pearson and Lipman algorithm (1985).

COMPARE: compares two protein sequences for points of similarity.

DOTPLOT: displays graphically the points of similarity found by COMPARE.


STEMLOOP: finds inverted repeats within a sequence which could form secondary structure.
Before the analysis of the otcZ sequence is described, it is appropriate that a few of the terms used commonly in the proceeding sections are explained. An "algorithm" is defined as the description of how to solve a problem presented in terms that could be solved using a computer. DNA sequence, in common with its biological role, is analysed as a linear array of information. A "window" is a fixed length of sequence within which a statistical calculation is performed. This calculation is specified by the algorithm. A "scoring matrix" is a two dimensional array of scores which represent the degree of "relatedness" between entities being compared. For example, the Dayhoff matrix, which is used by COMPARE, contains values which represent the degree of "evolutionary relatedness" for every possible pairwise comparison of amino acids. Matrices are extremely useful for identifying imperfect homologies between amino acid sequences.

4.2.3.1 Potential Protein Coding Regions

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 ORFs 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%). Not surprisingly perhaps, the GUG codon is noticeably more frequent in Streptomyces, being present in approximately 18% of genes compared to the 3% in E. coli. Translation initiation at other codons has not been identified in Streptomyces genes. In E. coli a very small number of proteins are initiated, however, 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) in the 1.3kb BcII-Smal fragment were identified using the FRAME algorithm (Bibb et al., 1984). FRAME analysis exploits the non-random distribution of bases within codons to identify PPCRs. Streptomyces DNA has typically a base composition of 73% mol G+C (Enquist and Bradley, 1971), which is reflected in the biased usage of synonymous codons. The average mol G+C composition in the first, second and third positions of streptomycete codons are 70%, 45% and 90%, respectively (calculated from a file of 63
genes compiled by Bibb, unpublished result). Intervening regions between coding sequences have a random distribution of bases within consecutive triplets. The FRAME program scans 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. 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 repeated. This procedure is continued until the entire sequence has been analysed.

FRAME revealed two PPCRs within the 1.3kb BclI-SmaI region (Figure 4.6). The heavy line is the plot of the mol% G+C composition at every third base, in a window of 50 triplets, starting at base 3 of the BclI-SmaI sequence. Similarly, the beaded (---) line is the plot starting at base 2 and the thin line is the plot starting at base 1. The analysis of the graphic output will be considered from the BclI site (base 1). The FRAME plots are immediately diverse and start to converge around position 920. The heavy line plot commencing from base 3, averages around 45% mol G+C and thus, corresponds to the second position of the codons in the PPCR. Similarly, the beaded line originating from base 2, with its average G+C composition of 95%, corresponds to the third codon position and the thin line from base 1, which fluctuates around 75%, corresponds to codon position 1. Therefore, the open reading frame for PPCR1 is translated from right to left in the BclI-SmaI sequence in nucleotide frame 2 (N2<).

Two potential start codons, a AUG at 1066nt and a GUG at 1042nt and a UGA stop codon at 34nt respectively, were identified for PPCR1 (graphically represented above the FRAME plot). The AUG codon is preceded by a reasonable ribosome binding site (AAGGAG) with a high degree of complementarity to a region close to the 3' end of 16S rRNA from S. lividans (sequenced by Bibb and Cohen, 1982) (Figure 4.5). This putative RBS is separated from the AUG triplet by 4nt and is well within the limits of 5-9nt for Shine-Dalgarno interactions (Gold et al, 1987). No significant complementarity to the 3' terminus of S. lividans 16S rRNA was found within the sequence immediately upstream of the GUG.
Figure 4.6. FRAME plot (Bibb et al., 1984) of the 1,265bp otcZ region using a "window" of 50 triplets. The thin line plot (—) represents the first triplet position and its associated reading frames (N1> and N2<); the beaded line (•—•) represents the second triplet position and its related reading frames (N2> and N3<) and the heavy line (—) indicates the third triplet position and its associates reading frames (N3> and N1<). Symbols > and < indicate AUG codons in the forward and reverse frames respectively, Symbols > and < indicate GUG codons and the symbol | indicates stop codons.
The AUG codon at 1066nt was the probable translation initiation codon for PPCR1. Translation initiation is considered in more detail within Chapter 7. The FRAME analysis starting at the BcII site was immediately diverse, as the non-coding region between the BcII site and the 3' end of otcZ was very short (31bp) and consequently the "windows" (150nt) that covered this region were primarily influenced by the coding sequences of PPCR1.

Although the plots started to converge at base position 920, an equilibrium position was never reached around 73 mol% G+C composition. Instead the plots diverged once more. This divergence suggested the presence of a second short PPCR at the extreme right end of this region. Preliminary sequence data from the Pfizer Laboratories at Sandwich (Butler, unpublished result), has identified PPCR2 as the 3' end of the otcC gene, which encodes the ATC oxygenase (see Section 4.1). The translation stop codon for otcC is at base 1129, which is in the same frame as PPCR1. The intervening region of 60bp between PPCR1 and PPCR2 is much shorter than the window (150nt) used for FRAME analysis. Therefore, the FRAME values were strongly influenced by sequences in PPCR1 and PPCR2. This explains why an equilibrium position around 73 mol% G+C, which is normally typical of non-coding sequence, was never reached. The same PPCRs in the BcII-SmaI region were also identified using the TESTCODE program from the UWGCG package (Devereux, 1984).

The otcZ15I mutation, which was mapped close to the "left" of the PstI site (Butler et al., 1989), probably lies within PPCR1. Therefore, PPCR1 will be referred to as the otcZ structural gene. The otcZ gene was predicted to encode a protein of 344 amino acids with a molecular weight of 37.36kD. As expected for a streptomycete gene, otcZ has an average mol G+C distribution in the first, second and third codon positions of 75.6%, 46.4% and 94.5% respectively. Comparison of the codons used in otcZ with a codon preference table compiled from 63 Streptomyces genes (Bibb, unpublished data), revealed that the biased usage of synonymous codons in otcZ was in accordance with the codon preferences observed for other streptomycete genes (Figure 4.7).
Figure 4.7. (A) Codon usage in the otcZ structural gene (B) Codon usage in a file of 63 streptomyces genes (compiled by Bibb, unpublished data). The numbers represent the frequency with which the codons are used. The frequency is expressed as a percentage (%) of the total number of codons in the file.
4.2.3.2 Amino Acid Sequence Comparisons

Using WORDSEARCH and SEGMENTS, comparison of the predicted amino acid sequence of otcZ to protein sequences held in the NBRF-Protein database (Release 24, March 1990), revealed that a hydroxyindole O-methyltransferase (HIOMT) from bovine pineal glands (Ishida et al., 1987) had the highest sequence similarity. Comparison of the amino acid sequence of otcZ with entries in the EMBL (Release 19, July 1990) and GenBank (Release 60, June 1989) databases using TFASTA did not identify any other entries with sequence similarity to the otcZ gene product (OtcZ). The bovine hydroxyindole O-methyltransferase is a protein of 350 amino acids with a molecular weight of 37.78kD, similar to that of OtcZ. HIOMT is involved in the biosynthesis of melatonin, which has been studied intensively by biochemists and physiologists. The amino acid sequence of HIOMT was predicted from the DNA sequence of a cDNA clone isolated by screening a library, which had been constructed in the lambda gt expression vector, with monoclonal antibodies to purified HIOMT (Ishida et al., 1987).

The tcmLa ORF4, from the tetracenomycin C-producer, S. glaucescens, was reported by Hutchinson at the 19th UCLA Symposium in Colorado (1990) to encode a protein which also has sequence similarity to the bovine HIOMT. The predicted product of tcmLa ORF4 has a molecular weight of 55.93kD and contains 494 amino acids (Hutchinson, pers. comm.). The sequence similarity is limited to the carboxy-terminal two-thirds of HIOMT. However, the amino-terminal region, extending to one-third of the predicted product of tcmLa ORF4, has high sequence similarity to the amino-terminal half of the predicted products of act ORF4 and gra ORF4 and the full-length product of whiE ORF6 (Hopwood and Sherman, 1990). Both act ORF4 and gra ORF4 are postulated to encode bifunctional cyclase/dehydrases (Hopwood and Sherman, 1990), which are required for the formation of the correct isochromanequinone nucleus, as they complement actVII mutants that accumulate an aberrantly-cyclised shunt product, mutactin (Zhang et al., 1990). The amino-terminus of the predicted protein product of tcmLa ORF4 and the product of whiE ORF6 may also contain cyclase activity. The reader is directed to the Introduction (Section 1.12) for a more detailed discussion of the proposed role of the above gene products in the synthesis of polyketides.
Production of tetracenomycin C was restored when \textit{tcmla} ORF4 was introduced on a plasmid into \textit{tcmII} and \textit{tcmv} mutants, which accumulate C-3 demethyl intermediates (Hutchinson, pers. comm.) The gene product of \textit{tcmla} ORF4 appeared, therefore, to have C-3 O-methyltransferase activity, in addition to its putative role in cyclization. Recently, Hutchinson and coworkers were surprised to find that deletion of the proposed ribosome binding site and translation initiation codon for \textit{tcmla} ORF4, did not interfere with the complementation of \textit{tcmII} and \textit{tcmv} mutants (Hutchinson, pers. comm.). An AUG codon, which is located approximately one-third of the way into \textit{tcmla} ORF4 (amino acid 137), is preceded by a reasonable ribosome binding site and may act as an initiation codon for the translation of the C-3 O-methyltransferase. The boundary between the putative "cyclase" and "methyltransferase" coding regions is currently being re-sequenced and expression studies carried out to establish if \textit{tcmla} ORF4 is, in fact, two juxtaposed PPCRs. For the purposes of comparison with HIOMT and OtcZ, only the predicted product of \textit{tcmla} ORF4 from methionine 137 (OMT) was considered. Optimum alignment of the amino acid sequences of OtcZ, HIOMT and \textit{tcmla} OMT by rationalising all the possible pairwise comparisons using WORDSEARCH and SEGMENTS (Figure 4.8), revealed that OtcZ has 20.3% (71/350) and 23.5% (84/358) overall amino acid sequence identity to HIOMT and \textit{tcmla} OMT respectively.

COMPARE and DOTPLOT were used to identify the most conserved regions in the above protein sequences. COMPARE analyses two protein sequences, creating a file of points of similarity between them which can be displayed using DOTPLOT. The COMPARE program finds points of similarity using a "window"/"stringency" match criteria. The sequences are compared in all possible registers and where the quality of the matches within a "window" of amino acids is greater than or equal to a value specified by the "stringency", the point is written to a file. Matches are determined using the Dayhoff matrix which contains a numerical value representing the evolutionary distance for every possible amino acid comparison (Devereux \textit{et al.}, 1984).
Figure 4.8. Alignment of the predicted amino acid sequences of tcmla OMT, OtcZ and HIOMT. Lines (———) indicate the two regions identified by COMPARE and DOTPLOT as being the most evolutionarily conserved (Figure 4.9). Asterisks (*) and the symbol (●) indicate the positions of the conserved glycines and aspartate, respectively, which are thought to be involved in the binding of S-adenosylmethionine. Amino acids were grouped as follows; (GASTP), (ILVM), (FYW), (QNED), (HKR) and (C).
Figure 4.9. COMPARE/DOTPLOT analyses of the deduced protein product of otcZ versus (A) HIOMT (Ishida et al., 1987) and (B) the deduced protein product of tcm1a ORF4 (amino acids 137-494) (Hutchinson pers. comm.). The stringency was 20 for a window of 30 amino acids (for definitions of window and stringency, see Devereux et al., 1984).
Using these match criteria it is possible to determine imperfect homologies and show evolutionary relationships. COMPARE puts only one point in the file at the middle position of the window whenever the stringency is met.

DOTPLOTS of the comparisons between OtcZ, HIOMT and tcmIa OMT highlighted two regions, shared by all three proteins, which appeared to be particularly well conserved. Within region A, OtcZ (amino acids 6 to 53) shows 22.0% (11/50) and 36.7% (18/49) amino acid sequence identity to HIOMT (2-51) and tcmIa OMT (159-207) respectively. Within Region B, OtcZ (amino acids 171 to 246) has 34.7% (26/75) and 37.0% (27/73) amino acid sequence identity to HIOMT (175-247) and tcmIa OMT (324-398) respectively (Figure 4.8).

S-Adenosylmethionine is the mononucleotide coenzyme that donates the methyl group in the hydroxyindole O-methyltransferase reaction (Axelrod and Weissbach, 1961; Nakane et al., 1983)). A "fingerprint" sequence has been identified that is characteristic of the nucleotide-binding domains of many proteins (Wierenga and Hol, 1983). Although the nucleotide-binding sequence was identified initially in proteins that bind NAD+ (Wierenga and Hol, 1983), this motif has also been found in proteins which bind nucleotides such as FAD, GTP and ATP (Ogawa et al., 1987).

X-ray crystallographic studies, have shown that amino acids within the "fingerprint" sequence interact with the adenosine 5'-diphosphate moiety in the nucleotides (Wierenga and Hol, 1983). The dinucleotide-binding domain consists of a region of β-sheet followed by a region of α-helix and then another region of β-sheet, making up a βαβ unit with the following features: (i) the sequence G-X-G-X-X-G in the region joining the first β-strand and the amino terminus of α-helix, (ii) the presence of a hydrophilic residue at the beginning of the first β-sheet region and (iii) the presence of an acidic residue at the carboxyl terminus of the second β-sheet. In addition, several hydrophobic residues are located at specific positions in the hydrophobic core of the βαβ unit. The first and third invariant glycines
of the nucleotide-binding βαβ units, allow the polypeptide chain to make a sharp turn between the first β-strand and the α-helix. The presence of a side chain at the first glycine position would interfere with the binding of the ribose moiety. The second glycine allows the pyrophosphate moiety to approach the amino terminus of the α-helix, which it binds through hydrogen bonding. For an extensive comparison of nucleotide-binding domains from proteins with diverse functions, the reader is directed to Ogawa et al. (1987).

Examination of the alignment of the amino acid sequences of OtcZ, tcmLa OMT and HIOMT, has identified a "consensus" sequence within Region B that resembles the nucleotide-binding "fingerprint" and may form a S-Adenosylmethionine (SAM)-binding domain (Figure 4.8). Owing to the speculative nature of the UWGCG programs available for predicting secondary structure, it was not possible to establish if this "consensus" sequence forms a βαβ domain. The putative SAM-binding sequence contains three invariant glycines (G-G-X-X-G), which are present in the "fingerprint" sequence for NAD⁺-binding proteins (Wierenga and Hol, 1983). The absence of an amino acid, separating the first and second glycines of the putative SAM-binding domain, may reflect the absence of a pyrophosphate moiety in S-Adenosylmethionine. A conserved aspartate residue is located 19 residues towards the carboxyl-terminus from glycine 3 (Figure 4.8). In the corresponding position of the "fingerprint" sequence, an acidic residue, such as aspartate, is proposed to form strong hydrogen bonds with the 2'-hydroxy group of the adenine-ribose. It should be noted that the assignment of a SAM-binding sequence to region B is highly speculative, as SAM-binding domains remain to be characterised by X-ray crystallography and molecular modelling. To date, the analysis of amino acid sequence in region A has not identified any similarity to consensus motifs for groups of functionally-related proteins.

4.2.4 Analysis of the Transcription of otcZ

To establish if transcription originated from a promoter located close to the predicted translation initiation codon of otcZ, S1 nuclease protection analysis was undertaken (Figure 4.10) using a single-stranded
Figure 4.10. High resolution S1 mapping of transcripts within the PstI-SmaI site. Single-stranded probe (1.2X 10^7 Cerenkov cpm.pmole^-1) was generated from mKM803 by extension from the universal primer (-20) and digested with EcoRI (see Materials and Methods 2.3.19). The probe (0.1pmoles) was hybridized at 52°C with ca. 10μg of total RNA from the following (a) S. rimosus M15883S, (b) S. rimosus M4018 grown on Tryptone Soya Broth (16hrs), (c) S. rimosus M4018 producing OTC (ca. 60μgml^-1) on Liquid Complete Medium (56hrs) (see Figure 3.4) and (d) S. rimosus M15883 producing OTC (ca. 0.7mgml^-1) on Trusoya Medium 1 (46hrs) (see Figure 3.6). The sequencing ladder was produced from mKM803 (see Section 4.2.2) by extension from the -40 primer. The position in the mKM803 sequencing ladder that is coincident with the RNA-protected fragment derived from transcripts originating far upstream of the SmaI site is indicated by pFAR. The position of probe protected along its entire length by contaminating template DNA is also indicated (FLP).
probe that included 250bp of sequence from the \textit{PstI}s to the \textit{SmaI}s site. The predicted start codon for \textit{otcZ} is 51bp upstream of the \textit{PstI}s site. The probe was generated from mKM803 template, which was used previously for sequencing (Figure 4.3), by extension from the universal primer (-20) and digestion with \textit{EcoRI} (Materials and Methods 2.3.18). Nucleotides derived from M13mp18 sequence accounted for 32nt at the 5’ end and 14nt at the 3’ end of the 297nt probe respectively.

The probe was hybridized with a selection of RNA samples isolated from \textit{S. rimosus} in which the relative abundances of transcripts from \textit{otrAp1} and pFAR had been shown to be very different (Section 3.2.4). Transcripts originating from \textit{otrAp1} were by far the most abundant (>95%) in total RNA isolated from M4018 grown on Tryptone Soya Broth, while transcripts from \textit{otrAp1} and pFAR had similar abundances in RNA isolated from M4018 grown in Liquid Complete Medium for 56hrs. In total RNA isolated from M15883 grown on Trusoya Medium 1 for 46hrs, only transcripts originating from pFAR could be detected. The probe was also hybridized with total RNA isolated from \textit{S. rimosus} M15883S, as a negative control.

A single fragment was protected by transcripts in all of the above RNA samples (except the negative control), which co-migrated with position 200 (numbering from the \textit{PstI} site) in the sequencing ladder. As mKM803 was sequenced using the -40 primer, the extension products, which formed the sequencing ladder, contained 52 nucleotides of M13mp18 sequence at the 5’ end. Therefore, the RNA-protected fragments were 252nt long. This length of protection corresponds almost exactly to that predicted for transcripts extending through the entire 250bp \textit{PstI}s-\textit{SmaI}s region. The absence of a RNA-protected fragment in the M13883S sample indicated that the transcript(s) providing protection were specific to the oxytetracycline cluster.

In all four samples, including the control, a fragment of 297 nucleotides was detected that co-migrated with position 245 in the sequencing ladder. This length of protection corresponds to full-length probe. Protection of the probe over its entire length probably results from contamination of the probe with a small amount of complementary template DNA.
4.3 Discussion

Methylation is believed to occur at two steps in oxytetracycline biosynthesis; (i) the C-methylation of the virgin tetracyclic nucleus to form 6-methylpretetramid and (ii) the N-methylation of 4-amino anhydrotetracycline (4-amino ATC) to form anhydrotetracycline (ATC) (Figure 4.11) (Rhodes et al., 1981). The latter methylation is blocked in the otcD4 mutant, which was able to convert purified ATC, but not 4-amino ATC, to oxytetracycline (OTC). The otcZ151 mutant could convert 4-amino ATC to OTC and thus, could not be blocked in N-methylation (Rhodes et al., 1981).

It is predicted that the otcZ mutant would accumulate 6-demethyl derivatives of oxytetracycline intermediates. Interestingly, purified 6-demethylchlorotetracycline (6-demethyl CTC) and 6-demethyltetracycline (6-demethyl TC) from a mutant of S. aureofaciens were shown to have similar levels of in vivo antimicrobial activity as chlorotetracycline and tetracycline respectively (McCormick et al., 1957). This suggests that the absence of a methyl group on C-6 is not sufficient, on its own, to explain the antibiotic-negative phenotype of the otcZ151 mutant. However, the detailed chemical analysis of mutants of S. glaucescens blocked in tetracenomycin C production has shown that mutations which block the addition or modification of a particular side group can have "knock on" effects at later stages in the biosynthesis of this antibiotic (Motamedi et al., 1986; Yue et al., 1986). For example, tcmII mutants, which are thought to be deficient in C-3 O-methyltransferase activity (Section 4.2.3.2), accumulate intermediates and shunt products that are not oxidised at C-2, C-4a and C-12a, as well as not being methylated at C-3. In the article by McCormick (1975), a full description of the mutant of S. aureofaciens was not provided. It is possible that a range of 6-demethyl intermediates were produced. The C-6 methyl group may be important for substrate recognition by enzymes which act later in the pathway and modify or add groups required for antimicrobial activity. It is also possible that the absence of a 6-methyl group could affect the reactivity of some of the later steps (Stark, pers. comm.). Analysis of the ATC oxygenase from S. rimosus has revealed that the enzyme cannot hydroxylate 7-chloro-ATC, which is the natural substrate
Figure 4.11. Biosynthesis of oxytetracycline (OTC). The steps at which various OTC-negative mutants are blocked are indicated. From Rhodes et al., 1981.
of the enzyme from *S. aureofaciens* (Novotna, pers. comm.). It is possible that other subtle differences exist between the enzymology of tetracycline/chlorotetracycline production in *S. aureofaciens* and oxytetracycline biosynthesis in *S. rimosus*. Should the disruption of C-6 methylation in *S. aureofaciens* not hinder significantly the flux of intermediates through subsequent steps in CTC and TC biosynthesis, it is possible that the same situation might not be true of oxytetracycline production in *S. rimosus*. To date, there are no reports of mutants of *S. rimosus*, which produce 6-demethyloxytetracycline. Therefore, the proposal that the *otcZ* gene encodes a methyltransferase, which adds the C-6 methyl group to the tetracycline nucleus, is not inconsistent necessarily with the antibiotic-negative phenotype of the *otcZ151* mutant. Chemical analysis of intermediates and shunt products accumulated in antibiotic-negative mutants of *S. rimosus* would complement the detailed genetic analysis of the oxytetracycline cluster.

In the previous chapter, transcriptional analysis indicated that a promoter beyond the *PstI* site was responsible for the transcription of the *otrA* resistance gene during the production of oxytetracycline. The limited transcriptional analysis described in this chapter suggests that there is no promoter within the *PstI-SmaI* region, between the end of *otcC* and the start of *otcZ*. It would appear, therefore, that *otrA*, *otcZ* and possibly *otcC* are transcribed as part of a polycistronic message during antibiotic production. The cotranscription of *otrA* with biosynthetic genes probably ensures that resistance to oxytetracycline increases in parallel with antibiotic biosynthesis. Furthermore, these results suggest that the delayed onset of oxytetracycline production can be attributed, at least in part, to the control of gene expression at the level of transcription. It remains a formal possibility, however, that the catalytic activity of some of the enzymes involved in the production of OTC may also be regulated. If the production of antibiotic is not synchronised, inducible resistance (discussed in Section 3.3) could be a means of protecting vegetative mycelium not devoted to antibiotic production from exogenous tetracycline produced by mycelium in other regions of the same colony or other colonies. The transcription of the "left" end of the oxytetracycline cluster is consider further in the following chapter.
CHAPTER 5

CHARACTERISATION OF

THE OTCC/OTCX PROMOTER REGION
5.1 Introduction

The study of endospore formation in *B. subtilis* in response to nutrient deprivation has been a paradigm of developmental regulation in bacteria. Alternative sigma (σ) factors and thus, alternative forms of RNA polymerase holoenzymes, which recognise different class of promoters, have an integral role in "re-programming" transcription during morphological development in this prokaryote. It has been implied often that alternative σ factors may have a similar role in the morphological and physiological development of *Streptomyces*. Recently, this hypothesis gained credence with the finding that the switch from the extension of the aerial hyphae to the development of chains of spores is controlled by a sigma factor (σ^whis^) (Chater et al., 1989).

The expression of several antibiotic pathways would appear to be regulated by the controlled expression of ancillary factors, which either act along with the RNA polymerase to facilitate transcription at the end of the growth phase or antagonise RNA polymerase-promoter interactions during primary metabolism. Pathway-specific regulatory genes have been identified within the clusters of genes for the biosynthesis of actinorhodin (Rudd and Hopwood, 1979), methylenomycin (Chater and Bruton, cited in Hopwood et al., 1986), bialaphos (Anzai et al., 1987), streptomycin (Distler et al., 1988), milbemycin (Malpartida et al., 1987) and tylosin (Hershberger, 1989). These specific examples are discussed in more detail within the General Introduction, Section 1.6. At present, it is open to debate whether the "housekeeping" RNA polymerase or another form, associated with a minor σ factor, recognises the promoters of genes involved in antibiotic production.

It has been established for some time that the ratio of regulatory proteins to their cognate binding sites on DNA is of key importance in regulating gene expression (Ptashne, 1986). The molecular analysis of physiological and morphological differentiation in *S. coelicolor* A3(2) has provided several examples which illustrate this point. The introduction of extra copies of the actII region, which contains a gene encoding an activator of transcription, lead to the significant over-production of actinorhodin (Hopwood et al., 1986). A fragment of DNA from *B. subtilis*
carrying a promoter recognised by \( \sigma^{\text{whic}} \) caused reduced sporulation when introduced into \( S. coelicolor \) A3(2) on a high-copy-number plasmid (Chater et al., 1989). The latter example is reminiscent of the reduction in antibiotic production, the "switch-off" phenomenon, which was observed when certain regions of the oxytetracycline cluster were introduced into \( S. rimosus \) M15883 on high-copy-number vectors (Butler et al., 1989). The same regions did not "switch-off" oxytetracycline production when they were introduced into the same production strain on a low-copy-number vector. From these results, it could be interpreted that the regions which cause antibiotic production to be "switched off" contain a recognition site for a positive activator that is essential for transcription of the \( otc \) genes. The positive activator would be titrated out by the introduction of the cognate regulatory sequence on high-copy-number vectors. To date, however, there is no genetic evidence to support this proposal.

This chapter describes the isolation and characterisation of a promoter region which directs transcription of a putative \( otcC-otcZ-otrA \) polycistronic message that extends "leftwards" (as the cluster is drawn) to the terminator at the 3'-end of \( otrA \) during oxytetracycline production.

A promoter can be defined as the region of DNA involved directly in the binding of RNA polymerase to initiate transcription. Although RNA polymerase interacts with DNA from approximately 50bp upstream (\(-50\)) to about 20bp downstream (+20) of the transcription start point, sequence-specific contacts are not found ordinarily downstream of the transcription start site (Simpson, 1982). In general, the promoters recognised by different forms of RNA polymerase holoenzyme (\( E_\sigma \)) have characteristic consensus sequences (Grossman and Losick, 1986). Although the conserved sequences for each promoter class differ from each other, two hexameric regions of nucleotide conservation, which precede the transcription start site, are found typically in each promoter class. The conserved hexameric regions of the consensus sequence for the major class of eubacterial promoters, which are recognised by \( E_\sigma^{70} \) in \( E. coli \) and \( E_\sigma^{43} \) in \( B. subtilis \), are centred approximately 35bp and 10bp preceding the transcription start site and
are separated from each other by 16-18bp (Hawley and McClure, 1983; Moran et al., 1982). These hexamers are referred to as the "-35" (TTGACA) and "-10" (TATAAT) regions respectively. Mutational analysis and chemical-protection experiments have established that promoter recognition is largely (although not exclusive) determined by base-specific contacts within these two conserved hexameric regions of the promoter sequence (Simpson et al., 1982; Grossman and Losick, 1986).

A consensus sequence has been derived for ten promoters from *Streptomyces*, which resemble the major class of eubacterial promoters (Hopwood et al., 1986). In the *Streptomyces* consensus, the sequence of the -35 region and the spacing of the conserved hexameric regions are identical to the consensus for the major class of eubacterial promoters. However, the -10 region of the *Streptomyces* consensus (TAGGAT) differs slightly from that of *E. coli* and *B. subtilis*. This disparity could be a consequence of one or more of the following.

(i) In *E. coli* and *B. subtilis* a single abundant σ species directs transcription of the general "housekeeping" enzymes. In contrast, *S. coelicolor* contains a family of four distinct, but closely related, σ polypeptides that have high amino acid sequence identity to the major σ species found in *E. coli* (σ70) and *B. subtilis* (σ43) in the regions which are thought to make specific contacts with promoters (Takahashi et al., 1988; Buttner et al., 1990). It has been proposed that these σ factors recognise different classes of promoters with slight differences in the sequence or spacing of the -35 and -10 recognition sites (Westpheling and Brawner, 1989). It is possible, therefore, that the *Streptomyces* consensus (Hopwood et al., 1986) was derived from four classes of promoters. Only one of these classes may be recognised by a σ factor that is the equivalent of the major sigma factor found in *E. coli* or *B. subtilis*.

(ii) The actual consensus sequence of promoters which direct the transcription of general "housekeeping" genes in *Streptomyces* may be slightly different due to the significant diversion in the G+C composition of *Streptomyces* compared with *E. coli* or *B. subtilis*.

(iii) The sample of promoters used to construct the *Streptomyces* consensus may not have been representative.
5.1.1 Overall Strategy

The results presented in Chapters 3 and 4 suggested that otrA, otcZ and possibly otcC were transcribed as part of a polycistronic message during antibiotic production. Transcriptional and functional analysis of the gene clusters for actinorhodin (Malpartida and Hopwood, 1986), methylenomycin (Chater and Bruton, 1985) and tylosin (Hershberger, 1989) has indicated that many of the genes within these respective clusters are organised in large transcription units extending up to 9.5kb (for further discussion, see Section 1.5). The otcC, otcZ and otrA genes are contained within a relatively short region of DNA of approximately 5.5kb. As the transcription and genetic organisation of the otc cluster upstream of otcC has not been studied intensively, it was considered possible that other genes could be co-transcribed with otcC, otcZ and otrA during antibiotic production. It was calculated, however, that there was a reasonable probability that the promoter(s), which directed transcription of otrA, otcZ and possibly otcC would be located within the 5.5 kb region that extends from the PstI site, which is internal to otcC, to the PstI site that is located approximately 4.5 kb upstream of the 5'-end of otcC.

As transcripts had been shown to terminate at the 3' end of otrA (Doyle, 1987), perhaps a straightforward approach to delineate the location of the promoter(s) directing transcription of otrA during antibiotic production would have been to blot total cellular RNA to nitrocellulose (Alwine et al., 1977; Fourney et al., 1988) and hybridize with radio-labelled probe derived from otrA. However, Northern analysis has never been used successfully within the laboratory at Glasgow (Alves, 1989; Doyle, pers. comm.). This strategy was not used, therefore, to locate the 5'-end of the polycistronic message. The overall approach that was taken during this work, was to screen for promoter activity in a library of Sau3AI fragments from the 5.5kb PstI-PstI region, which had been constructed in the promoter-probe vector pIJ2843 (Clayton and Bibb, 1990), and then to locate the promoters in the otc cluster by Southern analysis (see section 2.3.15).
The pIJ2843 vector system can be used to measure in vivo promoter activity in *Streptomyces* spp, including the *S. rimosus* production strains. Neither the pIJ486/7 nor the pARC1 systems, which were described previously, can be used in *S. rimosus* for the reasons discussed in Section 3.1.2. The pIJ2843 vector is based on the low-copy-number SCP2* replicon and utilises the *xylE* gene from *Pseudomonas putida* as the reporter of transcription (Clayton and Bibb, 1990). The product of the *xylE* gene is a catechol dioxygenase, which converts the colourless substrate catechol to an intensely yellow hydromuconic semi-aldehyde (Zukowski et al., 1983). Catechol dioxygenase activity can be measured using a simple colourimetric assay. Using a similar vector, the activity of *galp1-xylE* fusions in crude extracts of *S. coelicolor* A3(2) was shown to reflect faithfully the intracellular levels of RNA, as determined by quantitative dot blots (Ingram et al., 1989). Several useful features have been incorporated into the pIJ2843 vector. A multiple cloning site (MCS) upstream of the promoterless *xylE* gene assists the insertion and excision of fragments of DNA to be tested. The incorporation of a terminator from the coliphage fd upstream of the MCS prevents transcripts from vector promoters extending to the reporter gene. Translational stop codons have been created between the MCS and the *xylE* reporter gene in all three forward frames to prevent translational fusions which could affect either the expression of the *xylE* gene or the stability and/or specific activity of the *xylE* gene product. The terminator to from the coli phage lambda has also been inserted downstream of the reporter gene to prevent transcriptional read-through from the "test" promoter affecting the expression of genes in vector sequences and thus, possibly vector stability.

5.2 Results and Discussion

5.2.1 Construction of a Promoter Library In pIJ2843

The pIJ2843 vector was digested with *BamHI* and treated with CIAP to remove the 5'-phosphate from the cohesive ends generated by linearisation of the plasmid (Section 2.3.7). To verify the efficiency of dephosphorylation and to ensure that the vector had retained its ability to ligate with "insert" DNA, the following ligation reactions were set up;
(i) pIJ2843 (100ng) treated with BamHI and CIAP plus the 221bp BcII-BamHI fragment (3.5ng) containing otrA, (ii) pIJ2843 (100ng) treated with BamHI and CIAP and (iii) as (ii) but not incubated with T4 ligase. The molar ratio of insert to vector in the test ligation (pIJ2843 plus "insert" DNA) was approximately 2:1.

Samples (1/10th) of the above ligation mixtures were introduced to S. lividans TK24 by genetic transformation. Serial dilutions of the transformed protoplasts were then incubated on regeneration plates to determine the efficiencies of transformation (Table 5.1).

Table 5.1 Assay of the Efficiency of Phosphatase Treatment.

<table>
<thead>
<tr>
<th>Transformation efficiency (transformants,ug⁻¹ vector)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supercoiled pIJ2843</td>
</tr>
<tr>
<td>Dephosphorylated vector</td>
</tr>
<tr>
<td>Self-ligated dephosphorylated vector</td>
</tr>
<tr>
<td>Dephosphorylated vector ligated to test</td>
</tr>
<tr>
<td>&quot;insert&quot; (221bp BcII-BamHI fragment)</td>
</tr>
</tbody>
</table>

The low number of transformants recovered after self-ligation of the linearised vector, which had been treated with CIAP, suggested that the 5'-phosphate groups had been removed with reasonable efficiency. As the transformation efficiency of the dephosphorylated vector, which had been ligated with the test "insert", was at least an order of magnitude greater than the control ligation, the linearised vector appeared to be able to ligate with "insert" DNA. The similar transformation efficiencies of 5'-dephosphorylated vector, which had been incubated with and without ligase, indicated that the background transformants were due primarily to incomplete digestion of the vector with BamHI (only 0.5ng of undigested plasmid would have been required to produce 10⁴ transformants). Comparison of the transformation efficiencies suggested that greater than 95% of the transformants from the test ligation should contain recombinant molecules.
Using sequence data for over 8kb of the oxytetracycline cluster, it was determined that Sau3AI cleaves streptomycete DNA on average every 160bp. As restriction sites for Sau3AI occur frequently, it was possible that these sites might occur in sequences required for transcription. To increase the probability of isolating promoters and any cognate regulatory sequences on the same fragment, the 5.5kb PstI fragment from pPFZ105 (a Pfizer construct) was only digested partially with Sau3AI.

Conditions were established for the partial cleavage of the 5.5kb PstI fragment with Sau3AI by adding a fixed amount of enzyme and monitoring the progress of the reaction. In a pilot experiment, 500ng of the 5.5kb PstI fragment was resuspended in 29μl of 1X Sau3AI restriction buffer (as supplied by BRL) and equilibrated to 37°C for 3-5min, before the addition of 0.5 units of Sau3AI. After incubation for 5, 15, 30, 45 and 60mins, 6μl aliquots were removed, added immediately to 4μl of 5X AGL buffer, incubated at 70°C for 15min to heat inactivate the restriction endonuclease and placed on ice until all the samples had been processed. The reaction products were then separated by electrophoresis on a 6% (w/v) polyacrylamide gel and compared with 100ng of the PstI fragment which had been digested completely with Sau3AI. Using this system, products that were digested completely (15-800bp) were resolved easily (results not shown).

Digestion of the 5.5kb PstI fragment was almost complete after incubation for 60min. After incubation for 45min, the fluorescence (as a result of ethidium bromide staining) of bands corresponding to products which were completely digested was approximately half that of the corresponding fragments in the sample which had been digested with an excess of Sau3AI. Although products which were completely digested could be seen in the sample digested for 30min, the majority of the fluorescence came from fragments greater than 900bp (result not shown). As the regulation of prokaryotic gene expression involves typically the binding of ancillary proteins to regulatory sites close to the promoter (usually less than 150bp from the transcription start site), it was anticipated that promoters and cognate regulatory sequences
could be isolated together on fragments smaller than 900bp. The 5.5kb 
PstI fragment was eventually digested for 37min. All the products of 
complete digestion were represented and the majority of the digestion 
products were less than 900bp (result not shown).

For shotgun cloning, the recommended concentration of vector in 
the ligation reactions is 0.4-0.8 fmoles.μl⁻¹ ("M13 Cloning/Dideoxy 
Sequencing" published by BRL). Correspondingly, the concentration of 
piJ2843 used in the ligation reactions was 6ng.μl⁻¹ (0.8 fmoles.μl⁻¹). 
Assuming that partial digestion produced fragments with an average 
length greater than 320bp, fewer than 16 Sau3AI fragments should 
have been produced, on average, from a single PstI fragment. To 
suppress the formation of concatemers, which would subsequently 
hinder the positioning of promoters using Southern analysis, the molar 
ratio of insert-to-vector in the ligations was <1:1. It was calculated 
that 3ng.μl⁻¹ of the products from the partially digestion of the 5.5kb 
PstI fragment would correspond to a concentration of no greater 
than 0.8 fmoles.μl⁻¹. The ligation reaction was set up, therefore, using 
120ng of piJ2843 (treated with BamHI and CIAP) and 3ng of Sau3AI 
fragments. As a control, 120ng of dephosphorylated piJ2843 was ligated 
to itself.

Aliquots (1/10th) of the ligation reactions were introduced to S. 
lividans TK54 protoplasts, which could be transformed by supercoiled 
piJ2843 with an efficiency of 3.97X 10⁷ transformants.μg⁻¹ of piJ2843. 
From the test ligation (piJ2843 and the 221bp BclII-BamHI fragment), it 
was predicted that 12ng of dephosphorylated vector ligated to the 
partially digested Sau3AI fragments would produce approximately 4X 
10³ transformants. Aliquots (10X) containing 1/120 dilutions of the 
transformed protoplasts were incubated on regeneration plates. As 
expected, approximately 50 transformants grew on each plate (5X 10⁵ 
transformants.μg⁻¹ DNA). The transformation efficiency for self-ligated 
dephosphorylated vector was 1.56X 10⁴.μg⁻¹. Consideration of these 
transformation efficiencies suggested that greater than 95% of the 
transformants would contain recombinant vector molecules.
5.2.2 Assay For in Vivo Promoter Activity

The promoter responsible for the transcription of otrA, otcZ and possibly otcC has a distinct temporal activity associated with the slowing of growth rate and the onset of oxytetracycline (OTC) production (see Section 3.2.4). On solid media (Regeneration agar and Soya Mannitol agar), the production of OTC by S. rimosus is coincident with morphological differentiation. The demonstration that OTC can be produced by S. lividans when the entire cluster is cloned on a low-copy-number plasmid suggests that this heterologous host recognises the promoters responsible for the transcription of the OTC cluster (Binnie et al., 1989). It is not known, however, if the production of OTC by S. lividans follows the same temporal pattern established for S. rimosus. Assays for in vivo promoter activity were carried out, therefore, after incubation of the transformants on regeneration plates (R2 medium) for 3, 4 and 5 days at 30°C. The regeneration plates were sprayed with an aerosol of 0.5M pyrocatechol (supplied in crystalline form by Sigma) and incubated at room temperature for 30min. For each time point, six well-isolated xylE+ transformants with a range of pigmentation were selected for further analysis. Approximately 10% of the transformants were xylE+

Secondary screening of the 18 isolates for catechol dioxygenase activity after growth on Regeneration agar and Hopwood’s Minimum agar (both containing thiostrepton, 25μg.ml⁻¹) for 3, 4 and 5 days failed to detect any catechol dioxygenase activity in isolate 5. All the other isolates were highly pigmented at each of the time points. S. lividans TK54 transformed with pIJa2843 did not exhibit any catechol dioxygenase activity.

5.2.3 Characterisation of the xylE+ Transformants

Digestion of plasmid DNA from each of the 18 S. lividans transformants with HindIII and EcoRI released the "insert" fragment along with the xylE reporter gene from the rest of the vector. The fragments were resolved on a 0.8% (w/v) agarose gel (Figure 5.1) and then transferred to a Amersham Hybond-N™ membrane by capillary blotting (Southern, 1975) under alkaline conditions (see Section 2.3.15).
Figure 5.1. *Hind*III/*Eco*RI digests of independent isolates of the pIJ2843 constructs. The number above the lanes designates the isolate which was the source of the plasmid. Lambda *Hind*III markers were used to size the *Sau*3AI inserts.
Cleavage of the recombinant vector molecules with SphI and PstI indicated that gross rearrangements of the vector had not occurred (result not shown). The recombinant plasmids contained inserts ranging in size from less than 300bp to approximately 9.0 kb in isolate 18. The most likely explanation for the large insert present in this isolate is that a concatemer of large Sau3AI fragments has been cloned in this recombinant.

To localise the inserts containing promoter activity, digests of the 5.5kb PstI fragment (SphI, SstI, Sphl/KpnI, KpnI and Sau3AI) were separated on a 0.8% (w/v) agarose gel and transferred to nylon membrane as described above. HindIII/EcoRI fragments from recombinant vectors with small Sau3AI inserts were selected preferentially for Southern analysis. The HindIII/EcoRI fragment from isolate 12, which contained the reporter gene and a Sau3AI insert of less than 300bp, was used to synthesize a radio-labelled probe by the "Random Primer" method (see Section 2.3.9.1). The probe hybridized to the 920bp Sphl2-Sstl3 region (Figure 5.2, panel B). The Sphl-Sstl fragment was then isolated and used to prepared a radio-labelled probe. This probe hybridized with the plasmid DNA of only eight isolates (Table 5.2), indicating that more than one promoter element must exist in the 5.5kb region. To determine the minimum number of DNA elements within this region, which could direct the transcription of xylE, further rounds of hybridization were undertaken. Using the approach described above, the HindIII-EcoRI fragment from isolate 14, which did not hybridize with the Sphl-Sstl probe and contained a small Sau3AI insert (<300bp), was used to probe digests of the PstI region. This probe hybridized to DNA from the 200bp KpnI-SstI region (Figure 5.2, panel C). In turn, the KpnI-SstI fragment was used to probe plasmid DNA from the 18 xylE+ isolates. This probe hybridized to the plasmid DNA from seven xylE+ transformants, including two which had been identified using the Sphl-Sstl probe (Table 5.2). As several of the transformants contained plasmids with large Sau3AI inserts (>1kb), it was not unexpected that certain plasmids would hybridize with more than one probe from the 5.5kb PstI region. After this round of analysis, plasmid DNA from only five xylE+ transformants had not hybridized with
Figure 5.2. Location of the Sau3AI fragments, which were cloned in the xyIE-Positive transformants within the 5.5kb *PstI* region. (A) Idealized representation of the fragments that were resolved by agarose gel electrophoresis and transferred to a Hybond-N™ membrane. The numbers indicate the restriction sites used to derive the fragments. (B) Autoradiograph of the Hybond-N™ membrane after hybridization with a radio-labelled probe derived from the insert in the pIJ2843-based construct of isolate 12. (C) and (D) as (B) except hybridized with probe derived from isolates 14 and 18 respectively. The location of the Sau3AI insert within the 5.5kb *PstI* region is indicated below each autoradiograph. Abbreviations: *K*, KpnI; *P*, PstI; *S*, Saci; *Sm*, SmaI; *Sp*, SphI.
<table>
<thead>
<tr>
<th>Isolate</th>
<th>Approx. Insert Size (kb)</th>
<th>Probe Hybridizations</th>
<th>otcZ</th>
<th>otcC</th>
<th>1kb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SphI/ SstI</td>
<td>KpnI/ SstI</td>
<td>PstI</td>
<td>P</td>
<td>SmK</td>
</tr>
<tr>
<td>1</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>&lt;0.3</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>5*</td>
<td>&lt;0.3</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>&lt;0.3</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>&lt;0.3</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>&lt;0.3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4.0</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>&lt;0.3</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>&lt;0.3</td>
<td>+&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>&lt;0.3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>&lt;0.3</td>
<td>-</td>
<td>+&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>&lt;0.3</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3.5</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4.0</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>&gt;9.0</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1. Summary of the hybridizations between the plasmid DNA in the xyle-positive transformants and probes derived from fragments of the 5.5kb PstI region. The line (----) indicates the probable location of the Sau3AI fragments, which were cloned in the transformants, within the PstI region.

Abbreviations: K, KpnI; P, PstI; S, SacI; Sm, SmaI; Sp, SphI. *Secondary screening indicated that isolate 5 was xyle-negative. <sup>c</sup>HindIII/EcoRI fragments, which contained the Sau3AI insert, from the plasmid DNA of these isolates were used originally to probe the 5.5kb PstI region.
either of the probes. A radio-labelled probe was synthesized, therefore, from isolate 11, one of the remaining isolates that contained a small Sau3AI insert. This probe hybridized to DNA from the 400bp SstI-PstI region (Figure 5.2, panel D). In turn, the SstI-PstI fragment was used to probe plasmid DNA from the 18 xylE+ isolates. The SstI-PstI probe hybridized to all of the remaining xylE+ isolates and several which had been identified previously (Table 5.2).

From the hybridization data, it was possible to deduce the most probable locations of the Sau3AI fragments, which were cloned in the xylE+ transformants (including isolate 5) within the 5.5kb PstI region (Table 5.2). It was necessary, however, to postulate that several of the recombinant vectors contained large concatamers of Sau3AI fragments in order to reconcile the hybridization data with the size estimates of the Sau3AI inserts.

The hybridization of three Sau3AI inserts from different transformants to three discrete areas within the 5.5kb PstI region indicated that a minimum of three DNA elements within the PstI region could promote expression of xylE in the pIJ2843 constructs. As different promoters could be directing transcription in recombinant plasmids that hybridize with the same probe, more than three DNA elements from the PstI8-i7 region could be directing the transcription of xylE in the pIJ2843 constructs.

5.2.4 Location and Characterisation of the Promoter Region Responsible for the Transcription of otrA During Antibiotic Production

The main objective of shotgun cloning Sau3AI fragments into pIJ2843 was to delineate the location of the promoter(s) responsible for the transcription of otcC, otcZ and otrA during antibiotic production. The location of promoter activity to the SphI2-SstI3 fragment suggested that such a promoter(s) would be located between the 5'-end of otcC and the SstI3 site. Therefore, further analysis of the xylE+ transformants was not undertaken at this stage.
The following is the rationale for concluding that the SphI-SstI fragment contained a promoter capable of directing transcription of otcC, otcZ and otrA. It was considered improbable that the SphI-SstI region could contain only promoter activity capable of directing transcription in the opposite direction to otcC. The mRNAs of divergent non-overlapping genes within other antibiotic clusters do not generally overlap to any great extent, for example; the erythromycin resistance gene (ermE) and an unidentified ORF of Saccharopolyspora erythraea (Bibb et al., 1986; Bibb and Janssen, 1987) and the tyiH,D region and tyiF region of Streptomyces fradiae (Hershberger, 1989). The organisation of transcription units within antibiotic clusters is discussed in detail within Section 1.5. Should the SphI-SstI region contain bidirectional non-overlapping open reading frames, it would be highly likely that the promoters responsible for their transcription would be located between the 5'-end of otcC and the SstI site.

The strong codon bias of Streptomyces genes (Bibb et al., 1984) would appear to place severe restrictions on the coding capacity of the five reading frames which overlap protein coding regions. Only a couple of publications have presented tentative evidence for the presence of bidirectional, overlapping PPCRs in Streptomyces; ORF85 and ORF79 in the plasmid pIJ101 (Kendall and Cohen, 1988) and ORFs 1215 and 330 in the Insertion Element IS110 (Bruton and Chater, 1987). Interestingly, both of these examples are derived from extrachromosomal elements, which might be the subject of strong selective pressure to minimise the amount of DNA required to encode essential functions. To date, bidirectional PPCRs which overlap extensively have not been found within the gene clusters for antibiotics. It was considered unlikely, therefore, that a divergent mRNA originating within the SphI-SstI region would overlap with a large polycistronic message that originated upstream of the SstI13 site and extended to otrA. Following this line of argument, it was concluded that a promoter(s) located between the 5'-end of otcC and the SstI13 site should direct transcription towards otcC. It was also considered possible that another convergent gene(s) within the SphI-SstI region could precede otcC in the putative transcription unit that extends to the 3'-end of otrA.
Chapter 5  The otcC/otcX Promoter Region  157

To corroborate, the above deductions and to facilitate the orientation and location of the Sau3AI inserts in the pIJ2843 constructs, the SphiI2–SstI13 region was sequenced. Detailed analysis of the sequence is described in the following chapter. However, a schematic representation of the genetic organisation in this region is shown in Figure 5.3. The SphiI–SstI region contains two potential protein coding regions (PPCRs) which are encoded on the opposite strand to otcC, otcZ and otrA. The predicted AUG start codon for otcX PPCR1 is 120bp to the right of the Sphi site. As the translation initiation codon for otcC is 35bp to the left of the SphiI2 site, the non-coding region between otcC and otcX PPCR1 is only 155bp. The recognition sequences for promoters are not found generally within protein coding regions. Therefore, it was considered probable that the promoter for the putative otcC–otcZ–otrA polycistronic message would be contained within the intergenic region between otcC and otcX PPCR1.

5.2.4.1 Location of the 5' Termini of the Divergent Transcripts

Primer extensions were used to map and quantitate the 5'-termini of the putative otcC–otcZ–otrA polycistronic message and the divergent transcript from the opposite strand. The technique involves hybridizing the RNA of interest to an excess of a synthetic oligonucleotide, which is radio-labelled at the 5'-end. Reverse transcriptase is then used to extend the primer to produce DNA complementary to the RNA (cDNA). The same oligonucleotide can also be used to generate dideoxy-mediated sequencing ladders (Sanger et al., 1977), which can be used to size the cDNA. Using this approach, it is essential that the oligonucleotide primers bind to a site within 50-150 nucleotides of the 5'-terminus of the mRNA. Primers that hybridize to more-distal sites generate frequently heterologous extension products due to the tendency of reverse transcriptase to pause in regions of high secondary structure within the RNA. This is a particularly important consideration when analysing streptomycete RNA, which contains a preponderance of inverted repeats due to the high G+C content (Enquist and Bradley, 1971).
Figure 5.3. Schematic representation of the genetic organisation within the SphI12-SstI13 region. The closed boxes (■) indicate the positions of the PPCRs. Abbreviations: A, AvaI; S, SacI; Sp, SphI; S3, Sau3AI. Modified from Figure 6.3.
Reverse transcriptase is available from two sources, Molony Murine Leukemia Virus (M-MuLV) and Avian Myeloblastosis Virus (AMV), which have optimum temperatures of 37°C and 42°C respectively. Initial attempts at primer extension using M-MuLV reverse transcriptase were unsuccessful. Although a major extension product could be identified, a heterogeneous population of extension products was generated (result not shown). Under the conditions recommended by Maniatis et al. (1989), MuLV reverse transcriptase appeared to have considerable problems extending through regions of streptomycese RNA.

Homogeneous extension products were obtained, however, using a protocol communicated by L.V. Wray, Jr (Boston, University), which used AMV reverse transcriptase to extend the primer. The decreased stability of secondary structure in RNA at the higher incubation temperature used with AMV reverse transcriptase may explain the observed decrease in heterogeneous extension products.

The 5' termini of transcripts for several genes associated with secondary metabolism are also the first nucleotide of the proposed translation initiation codons, for example, the 23S rRNA methylase gene from *S. erythraeus* (Bibb et al., 1985), the streptothricin acetyl transferase (*sta*) gene from *S. lavendulae* (Horinouchi et al., 1987), the aminoglycoside phosphotransferase (*aph*) gene from *S. fradiae* (Janssen et al., 1989) and the gene (*afsA*) responsible for A-factor biosynthesis in *S. griseus* (Horinouchi et al., 1989). Oligonucleotide primers, C68 and X84, were synthesized, therefore, that hybridized to sites in the RNA 68 and 84 nucleotides downstream from the translation initiation codons of *otcC* and *otcX* PPCR1 respectively (Figure 5.6). Two M13mp18-based templates mL6A and mL6C (Butler, unpublished), which differed only in the orientation of their inserts, were used to generate sequencing ladders for sizing the cDNA which were synthesized from the divergent transcripts. Both templates contained the 2.05 kb *SmaI*-14 fragment; the *SmaI* site being nearest the binding site for the universal primer (UP) in mL6A (for restriction sites, see Figure 5.10). The oligonucleotides, C64 and X84, were also used to generate sequencing ladders from mL6A and mL6C respectively.
Primer extension analysis identified single 5'-termini for both the putative \textit{otcC-otcZ-otrA} message and the divergent transcript in total RNA isolated from \textit{S. rimosus} M15883 grown on TS1 for 46hrs (Figure 5.4, Panels A and B). The leader sequences of the divergent mRNA for \textit{otcC} and \textit{otcX} were 30 and 25 nucleotides respectively. The transcripts were separated by 100bp: they were non-overlapping (Figure 5.6).

Analysis of hypothetical transcripts in the intergenic region between \textit{otcC} and \textit{otcX} PCR1 using the FOLD program (Devereux \textit{et al.}, 1984; see Section 3.2.4) did not identify any significant secondary structure which could possibly block the extension of the complementary DNA on hypothetical transcripts originating outwith this region. Exonuclease VII protection analysis of the \textit{otcC/otcX} promoter region was carried out, however, to corroborate the primer extension data.

A single-stranded probe (220nt), complementary to the sense strand of \textit{otcC}, was generated from mL6A by extension from oligonucleotide C64 (which was labelled at the 5' end) and digestion of the product with \textit{AvaI} (see Section 2.3.18). Similarly, single-stranded probe (199nt), complementary to the sense strand of \textit{otcX}, was produced by extension from 5'-labelled oligonucleotide X84, which was hybridized to mL6C, and digestion of the product with \textit{SphI}. As these particular probes do not contain sequences derived from coliphage M13mp18 at their 5'-termini, the position in the sequence ladder with which they co-migrate corresponds to the actual transcription start site. Exonuclease VII analysis (Figure 5.5) revealed the same transcription start points observed by primer extension, indicating that the extension were not interrupted by secondary structure in the mRNA. As Exonuclease VII cannot digest DNA to a flush end (Rose and Botstein, 1983), the major Exonuclease VII-resistant fragments migrated with an apparent size 5-7nt larger than the cDNA products of primer extension.
Figure 5.4. Primer extension analysis of the 5' ends of the putative otcC-otcZ-otrA message and the divergent transcript. (A) The extension product from oligonucleotide C68 hybridized with 10μg of total RNA isolated from S. rimosus M15883 grown on TS1 for 46hrs (see Figures 3.6 and 3.7). (B) as (A) except extensions were from oligo. X84. The arrows (> ) indicate the transcription start points of otcCpl and otcXpl in the sequencing ladders derived from oligo. C68 hybridized to mL6A and oligo. X84 hybridized to mL6C respectively. No extension products were generated from total RNA isolated from S. rimosus M15883S.
Figure 5.5. Exonuclease VII analysis of the divergent transcripts. (A) Single-stranded probe (1.3X $10^7$ Cerenkov cpm.pmole$^{-1}$) complementary to the anti-sense strand of otcC was generated by extension from 5'-labelled oligo. C64, which was annealed to mL6A, using unlabelled nucleotides and digestion with Aval (see Section 2.3.18). 0.3pmoles of probe was hybridized at 37°C with 10μg of total RNA isolated from S. rimosus M15883S (lane a) and S. rimosus M15883 grown on TS1 for 46hrs (lane b). (B) Single-stranded probe (0.9X $10^7$ Cerenkov cpm.pmole$^{-1}$) was synthesized from mL6C by extension from 5'-labelled oligo. X84 and digestion with SphI. 0.2pmoles of probe was hybridized to the same selection of RNA as in (A).
5.2.4.2 Quantification of the 5' Termini of the Divergent Transcripts.

Primer extensions were used to quantitate the 5'-termini of the putative otcC-otcZ-otrA polycistronic message and the divergent transcript in RNA isolated from S. rimosus M15883 (see Figure 3.6) at various points during the fermentation of oxytetracycline. Apart from taking rigorous steps to ensure that the same amount of RNA (10μg) in each sample was hybridized to the oligonucleotide primer, the extension reactions were undertaken as described previously (Section 5.2.4.1).

The amount of cDNA synthesized from total RNA, which was isolated from cultures grown on TS1 for 46hrs, was approximately 4-4.5 fold greater than the amount produced from RNA isolated after 93hrs (Figure 5.7, Panel A). The 46hr sample was isolated early in the production phase when the concentration of OTC in the media was low, approximately 0.6mg.ml⁻¹. The 93hr sample represented the mid-point in the fermentation when the concentration of OTC was 3.2mg.ml⁻¹. The amount of cDNA synthesized from RNA isolated after 69hrs was slightly more than that produced from the 93hr sample. Previously, a single-stranded probe was used to measure quantitatively the amount of transcription between the 3' end of otcZ and the 5' end of otrA in the same RNA samples (Section 3.2.4). The relative amounts of single-stranded probe protected by transcripts within the different RNA samples (see Figure 3.7B) were similar to the amount of cDNA synthesized on the otcC transcript at the 5' end (Figure 5.7). These results were consistent with otcC being the first gene of a transcription unit that also includes otcZ and otrA.

Under similar fermentation conditions, the specific activity of anhydrotetracycline oxygenase, which is encoded by otcC, has been shown to rise rapidly during the transition from the period of intense growth to the stationary phase, reach a maximum level approximately 20 hours after OTC is first detected in the culture broth and then decline, within 20 hours, to a level seen at the onset of production (Butler, unpublished result). The changes in the specific activity of ATC oxygenase, which reflect approximately the changes in the abundance of
Figure 5.6. Annotated sequence of the otcCp1/otcXp1 promoter region showing the position of the putative translation start codon, ribosome binding sites, transcription start sites and promoter sequences for otcC and otcX.
Figure 5.7. Quantitative measurement of cDNA synthesized from the divergent transcripts at the 5' end. (A) Oligonucleotide C64 (1ng), which was labelled at the 5' end, was annealed to 10μg of total RNA isolated from *S. rimosus* M4018 grown TSB, *S. rimosus* M15883 grown on TS1 for 46hrs, 69hrs and 93hrs (see Figure 3.6). (B) as (A) except oligo. X84 was annealed to the RNA samples. The primers were extended by AMV reverse transcriptase (see Section 2.3.19). The intensity of the images produced by autoradiography of the 6% (w/v)polyacrylamide gel used to resolve the cDNAs was determined by densitrometry scanning (see Section 2.3.21).
the 5' end of the putative otcC-otcZ-otrA message, provided support for the proposal (Section 4.3) that the delay in the onset of oxytetracycline production can be attributed, at least in part, to the temporal regulation of transcription. It remains a formal possibility, however, that the catalytic activity of anhydrotetracycline oxygenase and possibly other enzymes involved in the production of OTC may also be regulated.

Changes in the abundance of the 5'-terminus of the divergent transcript followed a similar pattern to those determined for the putative otcC-otcZ-otrA polycistronic message (Figure 5.7, Panel B). Assuming that the primers, which were used to analyse the transcripts, were labelled with similar efficiency (no significant difference in the amount of radioactivity in the oligonucleotides was detected after removal of the unincorporated τ-32P ATP by column chromatography), the similarity in the intensity of the cDNA products produced by autoradiography suggests that in corresponding RNA samples the putative otcC-otcZ-otrA polycistronic message and divergent transcripts were present in similar amounts. The promoters for these transcripts would appear, therefore, to have similar temporal regulation.

Primer extension analysis using RNA isolated from cultures of S. rimosus M4018 grown on TSB, in which transcripts originating from otrApl predominate (see Figure 3.7A) failed to detect any transcripts other than those originating from otcCpl and otcXp1 (results not shown). It would appear, therefore, that single promoters direct transcription of the otcC-otcZ-otrA polycistronic message and the divergent transcript.

5.2.5 Comparison of the otcCpl/otcXp1 Promoter Sequences

Alignment of the DNA sequence from -50 to +1 of otcCpl and otcXp1 revealed greater than 38% identity, without the need to insert gaps (Figure 5.8). Several other interesting features were also identified, which could be involved in the initiation and regulation of transcription.
**Figure 5.8.** Sequence comparisons of the otcCp1 and otcXp1 promoters. Transcription start sites and putative -10 and -35 regions are underlined. Asterisks (*) indicate conserved nucleotides. Bold arrows (——>) indicate the location of the tandem repeats. The *E. coli* and *Streptomyces* consensus sequences shown are those of Hawley and McClure (1983) and Hopwood *et al.* (1986).

**Figure 5.9.** Alignment of possible regulatory sequences in phosphate-regulated promoters of *Streptomyces*. The consensus sequence for *pho* boxes in *E. coli* is that of Makino *et al.* (1990). The tandem repeat in the *E. coli* *pho* box is indicated by arrows (——>). The nucleotides in bold type are conserved with respect to the consensus *pho* box in *E. coli*. The alignment of the promoters for the *pabS*, ORFX and *aphD* genes was taken from Liras *et al.* (1990).
Centred around -10 there are hexamers in otcCp1 (CAGCAT) and otcXp1 (TACGCT) which match reasonably well the -10 consensus sequence (TATAAT) for the major class of eubacterial promoters. The tentative -10 regions of these promoters show even higher similarity to the -10 region (TAGGAT) of the Streptomyces consensus (see Section 5.1) (Hopwood et al., 1986).

An inspection of the promoter sequences 16-18bp upstream of the "-10" hexamers, revealed no similarity to the "-35" region of the consensus for the major class of eubacterial promoters (TTGACA). Centred around -36, however, hexamers in otcCp1 (TGGACA) and otcXp1 (TCGAAG), which are spaced 21bp upstream of the tentative -10 region, did show reasonable similarity to the consensus -35 region. Whilst, such similarities would be dismissed in E. coli due to the uncharacteristic distance between the putative recognition sequences (Mandecki and Reznikoff, 1982; Mulligan et al., 1985; Stefano and Gralla, 1982), there is a precedent for this spacing in certain promoters from B. subtilis, which are regulated temporally. The spoIE promoter (Guzman et al., 1988) and spoIIG promoter (Kenney and Moran, 1987) both contain hexamers which show high similarity to the -35 and -10 regions of the consensus for the major class of eubacterial promoters. However, these regions have the atypical spacings of 21bp for the spoIE promoter and 22bp for the spoIIG promoter respectively. Similarly, the spacing of the putative -35 and -10 regions of the gap1 promoter from the Streptomyces galactose operon is unusually long (21 or 24bp) (Westpheling and Brawner, 1989).

For gap1, it was proposed that the major form of RNA polymerase could be made to recognise promoters with "-10" and "-35" regions that were separated by an atypical distance by regulatory proteins that interacted either with the promoter or with the RNA polymerase (Westpheling and Brawner, 1989). Another possibility is that one or more of the Streptomyces α-like polypeptides encoded by the homologues of rpoD (the hrd genes- Takahashi et al., 1988; Buttn er et al., 1990) could recognise promoters with sequences similar to the major eubacterial class, but separated by atypical distances (for discussion of
transcription initiation, see Section 1.9). It is possible, therefore, that the hexameric regions identified in otcCp1 and otcXp1 could be recognised by RNA polymerase. Transcription from otcCp1 and otcXp1 occurs predominantly during the early phase of antibiotic production. The unusual spacing of the putative -35 and -10 regions in these promoters may reflect this distinct temporal activity.

Between -41 and -23 of otcXp1 a perfect tandem repeat of an octamer (GCTCGAA) was identified. The repeat sequences were separated by 3bp. Similarly, a hyphenated tandem repeat sequence was found in otcCp1 (Figure 5.8). Direct repeats with a similar signature to those found in otcCp1 and otcXp1 are the recognition sites for transcription activators such as OmpR, ToxR and PhoB in E. coli (Mizuno and Mizushima, 1986; Miller et al., 1987; Makino et al., 1986).

To date, the best characterised of these regulatory proteins is PhoB, which positively regulates the phosphate regulon in E. coli, during phosphate limitation (see Section 1.7) (Makino et al., 1986; Makino et al., 1988). PhoB recognises and binds to a conserved regulatory sequence, designated the pho box, activating transcription (Makino et al., 1988). The consensus sequence for the pho box consists of a tandem repeat of a heptamer (CTGTCAT) separated by 4bp. It has been suggested that each pho box may be bound by two PhoB subunits. Each PhoB subunit probably makes contact with the same side of the DNA molecule, as one turn of the helix of B-form DNA is 10.5bp long, similar to the distance between the centres of the repeated sequences (10bp). The pho boxes of E. coli promoters regulated by phosphate are located at the same position, 10bp upstream of the -10 regions. It has been proposed that PhoB protein bound to these pho boxes, substitutes for a typical -35 region in the phosphate-controlled promoters of E. coli (Makino et al., 1986, 1988).

Similar to the pho boxes, the tandem repeats in the promoter sequences of otcCp1 and otcXp1 were found 10bp upstream of the tentative "-10" regions. The centres of these repeats were also separated by 10bp. It is conceivable, therefore, that the tandem repeats
between -41 and -23 of otcCp1 and otcXp1 may be recognition sites for a common regulatory protein that binds at the onset of oxytetracycline production activating transcription.

As the otcCp1 promoter directs transcription of the ATC oxygenase, which has been shown to be regulated by phosphate (Behal 1979; Butler pers. comm.), the tandem repeats of otcCp1 and otcXp1 were aligned with the consensus for the E. coli pho boxes (Makino et al., 1990) and pho-like sequences from the regulatory regions of Streptomyces promoters (Figure 5.9) (Liras et al., 1990), which are also subject to phosphate control (see Section 1.7). The tandem repeats of otcCp1 and otcXp1 have 44% and 33% sequence identity to the E. coli consensus sequence respectively. Although the pho-like sequences of the pabS, sphD and ORFX promoters from Streptomyces have higher similarity to the E. coli consensus (>65%), the otcCp1 and otcXp1 matches are considered to be significant. The tandem repeats of both the otc promoters share the most highly conserved nucleotides of the E. coli consensus (CT—A—). It is proposed that the tandem repeats of otcCp1 and otcXp1 be referred to as opl (oxytetracycline pho-like) boxes.

The S. rimosus production strain M15883 was derived from a soil isolate, S. rimosus G7, through an extensive strain improvement programme, involving many rounds of mutagenesis and selection for survivors that produced higher titres of OTC. Transcription of the otc cluster may have been a target of strain improvement. It is possible, therefore, that any pho boxes in the oxytetracycline cluster may have diverged significantly from that of the original S. rimosus strain G7. This may explain why the tandem repeats of otcCp1 and otcXp1 show less homology to the E. coli consensus than the other pho-like sequences so far identified in Streptomyces. This possibility is currently being investigated by sequencing the otcCp1/otcXp1 promoter region from S. rimosus G7, which has been amplified using the polymerase-chain reaction (Hunter, unpublished work).
Chapter 5  The otcC/otcX Promoter Region  171

The similarity of the opl boxes, which overlap with the putative -35 regions of the otcCp1 and otcXp1, to the pho boxes in E. coli and the suppression of oxytetracycline production by high levels of phosphate, suggests that the onset of antibiotic production in S. rimosus may be controlled by a protein which under phosphate limitation activates transcription by binding to promoters within the otc cluster. This proposal could provide an explanation for the reduction in oxytetracycline production, which is observed when certain regions of the otc cluster are introduced into the production strains on high-copy-number plasmids, but not on low-copy-number plasmids (Butler et al., 1989). Analysis of the DNA segments at the otrA end of the oxytetracycline cluster, revealed that all of the fragments containing otcCp1 and otcXp1 caused "switch-off" when they were cloned on the high-copy-number plasmids (Figure 5.10). It is possible that plasmid-borne copies of the tandem repeats in these promoters sequestered the putative regulatory protein, which is required for the transcription of the otc cluster. The activator protein may be the Streptomyces-equivalent of the PhoB protein found in E. coli.

An overview of the Pho regulon in E. coli is presented in the General Introduction (Section 1.7) and therefore, will not be described at length in the following discussion. The phosphate regulon in E. coli was characterised originally by analysing the expression of the alkaline phosphatase gene, phoA (for a lucid review and references, see Stock et al., 1989). The phoA gene product hydrolyses phosphoesters to Pi in the periplasm and can be assayed easily by a simple plate test (Torriani and Rothman, 1961). In wild-type cells, the expression of phoA is suppressed by high levels of phosphate. Many of the loci involved in the uptake of phosphate were identified by characterising mutants, which expressed constitutively alkaline phosphatase. It would be interesting, therefore, to study the expression of alkaline phosphatase in Streptomyces. Should the otc genes be part of a phosphate regulon in Streptomyces, it would be predicted that mutations which cause the constitutive synthesis of the alkaline phosphatase at a high-level activity may also result in the constitutive production of OTC. Such mutants would demonstrate convincingly the integration of phosphate assimilation and OTC production.
Figure 5.10. The "Switch-Off" phenomena at the "otrA" end of otc cluster. Lines (——) indicate the extent of the DNA segments which cause a significant reduction in the biosynthesis of oxytetracycline when introduced into the production strain, S. rimosus, on high-copy-number plasmids. A bold line (■■■) indicates the location of the otcCp1/otcXp1 promoter region. Abbreviations: B, BamHI; Bc, BclII; K, KpnI; P, PstI; S, SacI; Sm, SmaI; Sp, SphI.
The opl boxes may provide an alternative startpoint for investigating the mechanism(s) which controls the onset of oxytetracycline production in *S. rimosus*. By combination of sequencing and transcription analysis it should be possible to establish if opl boxes are a common feature of promoters in the oxytetracycline cluster. Should this be the case, then gel retardation assay using opl boxes as substrates may facilitate the purification of the putative response regulator OpLB, that presumably activates transcription. The "reverse translation" of amino acid sequence from the amino-terminus of the purified protein, would allow an oligonucleotide to be generated, which should facilitate the isolation of the gene for putative regulator to be isolate from a DNA library of *S. rimosus* M15883.

Finally, it is interesting to speculate why the production of an antibiotic may be part of a phosphate regulon. On solid medium, the production of antibiotic often coincides with the switch to morphological development, when the aerial mycelium is envisaged to gain sustenance from the lysing substrate mycelium. It has been argued that the general role of antibiotic production is to prevent the invasion of the lysing substrate mycelium by microorganisms such as bacilli (Chater, 1984). The production of antibiotics does not always accompany the morphological development of *Streptomyces*, for example, *S. rimosus* produces OTC on L-agar plates, but does not sporulate, while sporulation but not antibiotic production occurs on plates containing tryptone soya agar (TSB, containing 2% (w/v) agar) (Hunter, pers. comm.). In soil, antibiotic production may not be necessarily a prelude to the morphological development of *Streptomyces*. When growth is limited by low levels of phosphate, the role of oxytetracycline may be to kill other microorganisms, resulting in the release of catabolites, including phosphate, which could be used to sustain the growth of the producing organism.
CHAPTER 6
THE \textit{SPHI}_{12}-\textit{STI}_{12} REGION: NUCLEOTIDE SEQUENCE AND LOCATION OF DNA ELEMENTS DIRECTING EXPRESSION OF \textit{XYLE} IN pIJ2843 CONSTRUCTS
6.1 Introduction

The 965bp SphI12-SstI13 fragment lies within the 5.5kb PstI18-PstI18 segment of DNA, which complements the otcX20 mutation (Figure 6.1) (Butler et al., 1989). The otcX class of mutants along with two other classes, otcY and otcZ, are blocked in the biosynthesis of OTC before the formation of 4-amino-anhydrotetracycline (Rhodes et al., 1981).

The otcY mutants are blocked probably in the earliest steps of OTC biosynthesis, as the DNA which complements the otcY90 mutant has homology with regions of the actinorhodin cluster that encode functions involved in synthesis of the polyketide backbone and its cyclisation to form the isochromanequinone nucleus (see Section 1.14; Butler et al., 1990). Consistent with this proposal, mutants in the class represented by otcY90 were pale and non-pigmented, while many of the intermediates and shunt products of the pathway are highly coloured (Hunter, pers. comm.).

The position of the otcZ151 mutation has been mapped close to the "left" of the PstI18 site (Butler et al., 1989). Sequencing of the BclI18-Smal18 region (Chapter 4), which includes the PstI18 site, identified a potential protein coding region (PPCR). The predicted product of the otcZ gene has high sequence identity to a methyltransferase from bovine pineal glands (Ishida et al., 1987) and a predicted gene product from the region of the tetracenomycin cluster of S. glaucescens, which complements tcmII and tcmV mutants that accumulate C-3 demethyl intermediates (Hutchinson pers. comm.). It is possible, therefore, that the otcZ151 mutation blocks the methylation of C-6 in the tetracyclic nucleus to form 6-methylpretetramid (6-MPT) (see Section 4.2.3.2).

As the otcY and otcZ mutants have been assigned provisionally to stages before the formation of 6-MPT, the otcX20 mutant would appear to be blocked in one of the following: (i) C-4 hydroxylation to form 4-hydroxy-6MPT, (ii) C-6 oxidation to form 4-keto-anhydrotetracycline (4-amino-ATC) or (iii) C-4 transamination to form 4-keto-ATC (see Section 1.14).
It has not yet proven possible to complement the \textit{otcX20} mutation with subclones of this \textit{PstI} fragment (Butler \textit{et al.}, 1990). The 5' end of the \textit{otcZ} structural gene lies at the extreme "left" of the \textit{PstIa-PstIb} region (Figure 6.1). The \textit{otcC} gene, which encodes the anhydrotetracycline (ATC) oxygenase, is located between the \textit{SphI12} and \textit{SmaI9} restriction sites (Binnie \textit{et al.}, 1989; Section 4.1). As the \textit{otcX20} mutant can produce OTC in cosynthesis tests with the \textit{otcZ151} mutant (Rhodes \textit{et al.}, 1981) and mutations in the \textit{otcC} gene would be blocked in the conversion of anhydrotetracycline to dehydroxytetracycline, the location of the \textit{otcX20} mutation can be delimited to the region between the \textit{SphI12} and \textit{PstIb} sites (Figure 6.1).

Interestingly, the \textit{SphI12-SstI13} fragment within the 5.5kb \textit{PstI} region has been shown to hybridize at high stringency with a 1.15kb \textit{PstI-BglIII} fragment from the \textit{actVa} region of the actinorhodin cluster of \textit{S. coelicolor A3(2)} (Butler \textit{et al.}, 1990; Butler pers. comm.). The \textit{actVa} region (Figure 6.2) is thought to encode functions involved in the hydroxylation of the isochromanequinone nucleus. The introduction of the \textit{actVa} region into the medermycin producer, \textit{Streptomyces AM-7161}, on low-copy-number vectors (Figure 6.2) led to the production of a hydroxylated form of medermycin, called mederrhodin A (Hopwood \textit{et al.}, 1985b). A mutant in the \textit{actVa} class has been shown also to accumulate kalafungin, which lacks the corresponding hydroxyl group that is present normally on actinorhodin (Cole \textit{et al.}, 1987). Taking all this evidence into account, it could be deduced that the \textit{otcX20} mutant may be blocked in the C-4 hydroxylation of 6-MPT. It is possible, therefore, that the hybridization between the 965bp \textit{SphI12-SstI13} fragment from the \textit{otcX} region and the 1.15kb \textit{PstI-BglIII} fragment from the \textit{actVa} region is due to a segment of DNA, which encodes a common function (possibly hydroxylation), which is necessary for the biosynthesis of both oxytetracycline and actinorhodin. The 1.15kb \textit{PstI-BglIII} fragment from the \textit{actVa} region has not yet been sequenced completely (Caballero, pers. comm.).
Figure 6.1. The otcX region of the oxytetracycline cluster. The hatched boxes indicate the PPCRs of otcC (Butler, unpublished work) and otcZ (Chapter 4). Abbreviations: K, KpnI; P, PstI; S, SacI; Sm, SmaI; Sp, SphI.

Figure 6.2. Physical and genetic organisation of the act cluster in S. coelicolor A3(2). The closed box indicates the location of the fragment in the actVa region, which hybridizes with the 965bp SphI-SstI fragment in the otcX region of S. rimosus (Butler, unpublished work). (i) Restriction map of the act region (from Malpartida and Hopwood, 1984). (ii) The positions of the coding regions of the act genes and the expected limit of the act cluster (bold line) were determined by the complementation of mutations using segments of the cluster cloned on plasmid vectors (Malpartida and Hopwood, unpublished result). (iii) The lines show the extent of DNA fragments that resulted in the hydroxylation of medermycin when introduced into Streptomyces AM-7161 on low-copy-number vectors (Hopwood et al., 1983). Abbreviations: B, BamHI; Bg, BgIII; R, EcoRV; P, PstI. Modified from Hopwood et al., (1986).
The $Sph_{12}^{-}-Sst_{13}$ fragment was sequenced principally to allow the orientation of the Sau3AI inserts, which were cloned in the xyIE-positive transformants and described in Chapter 5. It was hoped, however, that the comparison of predicted gene products of PPCR(s) in the $Sph^{-}-Sst$ fragment to entries in the NBRF-protein and GenBank databases would identify similarities to polypeptides with known functions, perhaps confirming the role of this region in hydroxylation.

6.2 Results

6.2.1 DNA Sequencing of the $Sph^{-}-Sst$ Region

The sequencing strategy for the $Sph_{12}^{-}-Sst_{13}$ region (Figure 6.3) was similar to that used for the otcZ gene (Section 4.2.2). The 965bp $Sph_{12}^{-}-Sst_{13}$ fragment from pPFZ105 was ligated to similarly-digested M13mp18 and M13mp19 (Yanisch-Perron et al., 1985) and introduced into E. coli TG1 by genetic transformation (Section 2.2.12). Single-stranded DNA was then isolated from M13 phages, which contained recombinant DNA molecules (Section 2.3.16). Taq polymerase, as supplied in the TAQuenase™ kit by USB (Section 2.3.17), was used to sequence the single-stranded templates mKM810 and mKM910, which were derived from the insertion of the $Sph^{-}-Sst$ fragment into M13mp18 and M13mp19 respectively. Nucleotide mixes containing 7-deaza-dGTP (Mizusawa et al., 1986) were used to minimise the formation of secondary structure in the extension products, which would lead to band compressions in the sequencing ladders. As the standard sequencing ladders could only be read for around 350nt, oligonucleotides were synthesized and used as primers to extend the sequence.

The $Sph_{12}$ boundary was sequenced from both strands using the templates, mL6A and mL6C, and oligonucleotides, C64 and X84, which were used to map the 5' termini of the divergent transcripts (Section 5.2.4.1). Sequencing of the Sau3AI insert from the pIJ2843 construct in isolate 15 (templates mKM8X15 and mKM9X15) spanned the $Sst_{13}$ site.

In total, 1166bp was sequenced between the annealing site of oligo. C64 (114bp to the left of the $Sph_{12}$ site) and the Sau3AI$_{14}$ site (87bp to the right of the $Sst_{13}$ site). Over 85% of the DNA was sequenced on both strands. The sequence was determined unambiguously in the regions, which were sequenced from only one strand.
Figure 6.3. Strategy used to sequence the 965bp SphI~SstI Region. The lines indicate the region of DNA cloned in the M13 constructs. The arrows show the direction and extent of the nucleotide sequence analysis. The oligonucleotides used to prime the extension reactions are given above the arrows. Abbreviations: UP, universal (~20) primer; oligo., oligonucleotide.
Chapter 6 The SphI-SstI Region

The average mol G+C composition of this region is 72.0%. The annotated sequence, including the predicted amino acid sequences of potential protein coding regions, potential ribosome binding sites, the oligonucleotides used to prime the sequencing reactions and the restriction sites referred to in the text are shown in Figure 6.4.

6.2.2 Features of the Nucleotide Sequence

The 1166bp region containing the SphI-SstI fragment was analysed primarily using the "Sequence Analysis Software Package" (Version 6.2) from the Genetics Computer Group at the University of Wisconsin (UWGCG). The individual UWGCG programs have been described already (Section 4.2.3) and thus, will not be discussed in detail within this section.

The sequence was analysed for potential protein coding regions (PPCRs) using the program FRAME (Bibb et al., 1984). A non-random distribution of G+C nucleotides was detected in three different segments of the 1166nt sequence (Figure 6.5). Analysis of the sequence for unequal usage of synonymous codons using the UWGCG program TESTCODE also indicated that the same regions encoded proteins (result not shown). Analysing the sequence from the annealing site of oligo-C64, the FRAME plots were diverse immediately. By nucleotide position 100, the plots had converged to around 68% mol G+C, suggesting that a PPCR was present at the extreme left of the sequence. This PPCR corresponds to the 5' end of the otcC gene (Section 4.1) (Binnie et al., 1989). A potential ribosome binding site (GAGGAG) with a high degree of complementarity to the 3' end of the 16S rRNA from S. lividans (sequenced by Bibb and Cohen, 1982) was located 5nt upstream of the AUG start codon (69-67nt) for otcC, which is in nucleotide frame N2< (Figure 6.5). The predicted amino acid sequence from the 5' end of otcC is in complete agreement with the sequence obtained from the amino terminus of the gene product, ATC oxygenase (Binnie et al., 1989).

From base position 100, the FRAME plots diverge once more. The thin line plot, which commences from base 1 fluctuates around 70% and thus, corresponds to the first position of the codons in this PPCR.
**Figure 6.4.** Annotated sequence of the 965bp SphiI-SstI13 region. The predicted amino acid sequence of otcX PPCR1 and PPCR2 are shown above the corresponding nucleotide sequences. Also indicated are the nucleotides incorporated into the primers used to further the sequence, possible ribosome binding sites (RBS) for otcC, otcX PPCR1 and PPCR2 and the transcription start points for otcCpl and otcXpl. For clarity, only restriction sites referred to in the text are indicated. Continued on pages 181 and 182.
The beaded line (---) originating from base 2 reaches an average G+C composition of 50% by base position 200 and thus, corresponds to the second codon position. Similarly, the heavy line from base 3, which fluctuates around 95%, corresponds to codon position 3. Thus, the open reading frame for otx PPCR1 is translated from left to right in nucleotide frame 1 (N1'), which is divergent to otcC.

The AUG codon at base position 235 is probably the translation start codon for otx PPCR1, as it is within the region of plot divergence and it is preceded by a potential ribosome binding site (GAGGA), which is located 6nt upstream. PPCR1 ends at base position 663 with a UGA codon. As expected for a gene from Streptomyces, otx PPCR1 has an average mol G+C distribution in the first, second and third codon positions of 69.9%, 49.0% and 92.3% respectively. In accordance with the assignment of PPCR1 to the region between base positions 235 and 663, the FRAME plots start to converge around base position 500. The FRAME plots do not fluctuate, however, around 70% mol G+C composition, instead they diverge once more. The thin line plot, which commences from base 1, reaches a position were it fluctuates around 50% mol G+C composition, suggesting that it corresponds to the second codon position of another PPCR. The first codon position was assigned to the heavy line plot originating from base 3, as it fluctuates around 75%. Similarly the beaded line (---) from base 2 fluctuates around 90% mol G+C composition and was assigned to the third codon position. The open reading frame for otx PPCR2 is translated, therefore, from left to right in nucleotide frame N>3.

Within the region of plot divergence (base position 600-700) an AUG codon (660-662) and a GUG codon (690-692) were located within N>3. The translation start codon of otx PPCR2 was assigned to the AUG codon as it was preceded (6nt upstream) by a reasonable RBS (GGAGG). The GUG codon was not preceded by a sequence with significant complementarity to the 3' end of 16S rRNA from S. lividans (Bibb and Cohen, 1982). PPCR2 does not contain an inframe stop codon. This incomplete PPCR has a typical average mol G+C composition in the first, second and third codon positions of 77%, 54.2% and 90.5% respectively.
Figure 6.5. FRAME plot (Bibb et al., 1984) of the 965bp SphI\textsubscript{12}-SstI\textsubscript{13} region using a "window" of 50 triplets. The thin line plot (——) represents the first triplet position and its associated reading frames (N1\textgreater{} and N<2); the beaded line (---) represents the second triplet position and its associated reading frames (N2\textgreater{} and N3<) and the heavy line (——) indicates the third triplet position and its associated reading frames (N>3 and N1<). Symbols > and < indicate AUG codons in the forward and reverse frames respectively. Symbols > and < indicate GUG codons and the symbol | indicates stop codons.
The putative AUG start codon (660-662) for PPCR2 overlaps with the UGA stop codon (661-663) for PPCR1. It is possible, therefore, that PPCR1 and PPCR2 are translationally coupled similar to those found in a variety of prokaryotic operons (Zalkin and Ebbole, 1988; Harms et al., 1988). Such an arrangement would facilitate the synthesis of proteins encoded by PPCR1 and PPCR2 in stoichiometric amounts. The products of these PPCRs could be components of a multi-peptide complex. The translation initiation of otcX PPCR1 and PPCR2 is considered further in Chapter 7.

6.2.3 Analysis of the Deduced Amino Acid Sequence of PPCR1 and PPCR2

The predicted gene product of PPCR1 is composed of 142 amino acids and has a calculated Mr of 15,851. The incomplete PPCR2 encodes 169 amino acids. Comparison of the codon usage in PPCR1 and PPCR2 with a codon preference table compiled from 63 streptomyces genes (Bibb, unpublished data) revealed that the biased usage of synonymous codons within the PPCRs of otcX was in accordance with the codon preferences observed for other Streptomyces genes (Figure 6.5).

In an attempt to determine the possible functions of the polypeptides encoded in the 1166bp region containing the SphI12-SstI13 fragment, the predicted amino acid sequences were compared with the entries in the NBRF-protein database (Release 25, June 1990) and the GenBank database (Release 60, June 1989) using the UWCGG programs WORDSEARCH/SEGMENTS and TFASTA. No striking similarities were found.

6.2.4 Location of the DNA elements from the SphI-SstI region which direct transcription of xylE in pIJ2843 Constructs

In the previous chapter, the PstI5-18 region was digested partially with Sau3AI, ligated with the promoter-probe vector pIJ2843, which had been digested with BamHI, and then introduced into S. lividans TK54 by genetic transformation. Several transformants contained recombinant vectors with inserts from the SphI-SstI region, which were capable of directing transcription of the xylE reporter gene.
Figure 6.6. (A) Codon usage in a file of 63 streptomycete genes (compiled by Bibb, unpublished data). (B) Codon usage in *otcX* PPCR1. (C) Codon usage in the incomplete PPCR2. The numbers represent the frequency with which the codons are used. The frequency is expressed as a percentage (%) of the total number of codons in the file. Amino acids are indicated by the standard single letter code and stop codons by asterisks.
To delineate the locations of the promoter elements within the SphI-SstI region, the Sau3AI fragments from recombinant vectors in the xylE-positive transformants 8, 12 and 15 (see Table 5.1) were sequenced. The Sau3AI inserts were removed on HindIII-EcoRI fragments and ligated with similarly-digested M13mp18 and M13mp19. The HindIII-EcoRI fragments also contain the xylE reporter gene, which is located between the former BamHI site (used to insert the Sau3AI fragments) and the EcoRI site. The ligation mixes were then introduced into E. coli TG1 by genetic transformation (Section 2.2.12). Single-stranded templates were isolated from M13 phage containing recombinant DNA molecules (Section 2.3.16). The HindIII-EcoRI fragment from the pIJ2843 construct in isolate 8 was used to derive mKMX88 and mKMX98 from M13mp18 and M13mp19 respectively. Similarly mKMX812/912 and mKMX815/915 were derived from the HindIII-EcoRI fragments of isolates 12 and 15 respectively. The Sau3AI inserts in the M13mp19 constructs could not be sequenced from the universal (-20) primer, as the xylE reporter gene was located between the primer annealing site and the insert. These constructs were sequenced, therefore, from an oligonucleotide, which anneals to a complementary sequence on the antisense strand of xylE, immediately downstream of the multiple cloning site in pIJ2843 (Figure 6.7).

Sequencing revealed that the recombinant vectors in isolates 8 and 12 both contained the Sau3AIc-e fragment with the Sau3Aic site nearest the xylE reporter gene. Should the DNA element, which is directing the transcription of xylE in the pIJ2843 constructs in isolates 8 and 12, be active in the chromosome of S. rimosus it would direct transcription towards the SphI12 site. In other words, it would produce transcripts from the antisense strand of otcX PPCR1. Similarly, the pIJ2843 construct in isolate 15 contains the Sau3AIr-g fragment, with the Sau3Aic site nearest the xylE reporter gene. Should the promoter element in this insert be active in the chromosome it would direct the synthesis of transcripts from the antisense strand of otcX PPCR2.
Chapter 6 The SphII–SstI Region

5′...AAGCTTGATGCCTGCAGGACTACTAGAGTCACTTTCGCCACGTTG

| H | Sp | P | Ac | X | B |

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pIJ2843 Reverse Primer Site, 24-mer</td>
</tr>
</tbody>
</table>

RBS

xyLE

MetAsnLys...

GCGGAACAAACCTGACAAACATGAACTATGAAGGTGACGTCATGAACAAA...GAATTC...3′

Figure 6.7. Multiple Cloning Site in pIJ2843. The DNA strand, which is shown, is generated when the HindIII–EcoRI fragment is inserted into M13mp19 and isolated as single-stranded DNA. The nucleotides in bold type are the three translation stops codons. Abbreviations: B, BamHI; Ac, AccI; E, EcoRI; H, HindIII; P, PstI; S, SacI; Sp, SphI; X, XbaI.
6.2.5 Analysis of transcription within the Sau3Alc-ε and Sau3AIf-g regions

Single-stranded probe, which was complementary to the sense strand of otcX PPCR1 within the Sau3Alc-ε region, was generated by annealing the pIJ2843 Reverse Primer (Figure 6.7) to mKMX912 (Figure 6.3), extending from the oligonucleotide in the presence of radio-labelled nucleotides and digesting with HindIII (Materials and Methods, 2.3.18). Similarly, a probe complementary to sense strand of otcX PPCR2 within the Sau3AIf-g region was generated from mKMX915. These continuously-labelled probes were hybridized initially at 37°C with total RNA from S. rimosus M4018 grown on TSB, S. rimosus M15883 grown on LCM, which produces low titres of OTC and S. rimosus M15883 grown on the high-titre production medium, TS1. This selection of RNA was the same as that described in Section 3.2.4. After digestion with S1 nuclease, no RNA-protected fragments were detected in any of the samples. This procedure was repeated, carrying out the hybridizations at 22°C. Although transcripts overlapping with the probes by as little as 15-20nt should have formed stable hybrids, no RNA-protected fragments were detected after digestion with S1 nuclease.

6.3 Discussion

It has been proposed that overlapping bidirectional protein coding regions are present in Streptomyces. For example, ORF85 and ORF79 in the plasmid pIJ101 (Kendall and Cohen, 1988) and ORFs 1215 and 330 in the Insertion Element IS110 (Bruton and Chater, 1987). In both of these cases, however, there is a compromise in the codon usage of the protein coding regions in the area of overlap. The average mol G+C composition in the first, second and third codon positions are approximately 70%, 45%, and 70% respectively. The first codon position of the forward PPCR is the third codon position of the reverse PPCR and vice versa. The average distribution of G+C residues in the codons of otcX PPCR1 and PPCR2 is typical of Streptomyces genes. As the typical codon usage in PPCR1 and PPCR2 would necessitate the use of rare codons in the antisense RNA it was considered unlikely that transcripts originating from the element(s) cloned in isolates 8, 12 and 15 would be translated efficiently, unless the antisense transcripts encoded very
short polypeptides (ca. 50 amino acids). The FRAME analysis would not detect necessarily a compromise in the distribution of G+C residues in the different codon positions within a limited segment of overlapping PPCRs. An analysis of the 1166bp region did not identify, however, any reverse ORFs (reading right to left) containing inframe translation start codons at the 5' end, which were preceded by consecutive nucleotides with complementarity to the 3' end of 16S rRNA from S. lividans (Bibb and Cohen, 1982). As no evidence has been obtained for the presence of protein coding regions in the reverse frames of the SphII-SstI region, it seemed unlikely that antisense transcripts arising from the promoter elements cloned in isolates 8, 12 and 15 would be messenger RNAs.

Increasingly, small RNA molecules are being implicated in regulating translation, as well as participating in the control of transcription, transposition, phage development, DNA replication, plasmid copy number and plasmid transfer (for review and examples, see Inouye and Delihas, 1988). Often an antisense RNA transcript, which is complementary to the translation start codon and Shine-Delgarno (SD) sequence, binds the mRNA preventing the initiation of translation, for example: pOUT RNA binds to the SD sequence and AUG start codon of transposase mRNA and thereby, regulates Tn 10 transposition (Simons and Kleckner, 1983); conjugal transfer of plasmid R100 is regulated by two finP transcripts which are envisaged to block translation of the traJ message by binding to the SD region (Dempsey, 1987) and the translation of the antiterminator (ant) mRNA in bacteriophage P22 is inhibited by the binding of a small RNA, which shields the AUG codon and SD sequence (Liao et al., 1987). Not all antisense RNA transcripts regulate transcription by preventing ribosome binding, for example, in bacteriophage lambda OOP RNA prevents cII expression by hybridizing to the 3' end of the message and creating an RNAase III-sensitive site (Krinke and Wulff, 1987). As the putative antisense transcripts in the otcX region are unlikely to encode polypeptides, they could conceivably regulate the translation of a otcX message, which originates from otcXp1 (Figure 6.4), by an unknown mechanism.
Chapter 6 The SphI12-SstI13 Region 191

The identification of conserved pho-like boxes in the promoters otcCp1 and otcXp1 (Section 5.2.7), however, has already lead to the proposal of a plausible model for regulating the expression of genes in the otc cluster. Transcription from otcCp1 and otcXp1 and possibly other promoters within the otc cluster, which direct transcription of genes whose products are involved directly in the biosynthesis of OTC, may be activated by the binding of a regulatory protein within the promoter region. Although similar models involving trans-acting regulatory proteins have been proposed to regulate the biosynthesis of other antibiotics (for review, see Section 1.6), to date, no articles have been published which suggest a role for antisense RNA in regulating the expression of enzymes of secondary metabolism at the level of translation in Streptomyces.

In Bacillus subtilis, the genes encoding enzymes which have roles in the early steps of arginine biosynthesis are located in a single cluster, with the gene order argC-argA-argE-argB-argD-cpa-argF (Mountain, et al., 1986). Transcriptional analysis in vitro revealed a promoter upstream of argC, which directed transcription rightwards. The insertion of small DNA fragments into the coding region of argC had a polar effect, which disrupted the expression of argF (Smith, unpublished result). As this arg cluster is not thought to contain a regulatory locus, it was deduced that argC-argA-argE-argB-argD-cpa and argF were part of a single transcriptional unit (Baumberg and Smith, pers. comm.) However, the shotgun cloning of random fragments from a segment of DNA, which contained the proposed argC-argA-argE-argB-argD-cpa-argF operon, into a promoter–probe vector identified four internal fragments that were capable of directing transcription of the reporter gene (Baumberg, unpublished result). As there are no precedents for the antisense RNA regulation of gene expression in B. subtilis, it was proposed that the cloned elements may be "phantom" promoters, which have no role in the transcription or regulation of the arg operon in the chromosome (Baumberg, pers. comm.). By analogy with the arg operon, it is a formal possibility that the plasmid-active promoter (PAP) elements cloned in isolates 8, 12 and 15 are not active in the chromosome of S. rimosus. The inability to detect the 5' ends of transcripts originating from the Sau3Alc and Sau3Alr regions in total RNA isolated from S. rimosus at different stages in the fermentation
of OTC is consistent with the PAP elements in these regions being "phantom" promoters. As will be discussed later, there are other possible reasons why the 5' ends of the putative antisense transcripts were not detected. At this stage in the analysis it is not clear what roles, if any, the promoter elements, cloned in isolates 8, 12 and 15 have in the biosynthesis of OTC.

The 259bp Sau3AIb-c fragment containing the otcCp1/otcXp1 promoter region (see Figure 6.4) has been inserted into the BamHI site of pIJ2843 to derive pKMSOCP1, which has the Sau3AIb site nearest the xylE reporter gene (McDowall and Kok, unpublished work). Transformants of S. lividans TK54, which contain pKMSOCP1, only express xylE after incubation for approximately six days on Emersons agar (containing thiostrepton, 25μg.ml⁻¹) at 30°C (McDowall and McGregor, unpublished result). A similar delay in transcription from the otcCp1 promoter is also observed when the same transformants were incubated on R2 medium (McDowall and Kok, unpublished result). In the previous chapter, the assays for transformants containing recombinant pIJ2843 vectors with promoter elements capable of directing transcription of the xylE reporter gene were carried out after incubation for 3, 4 and 5 days on regeneration plates at 30°C (Section 5.2.2). At each time point, six xylE⁺ isolates were selected for further analysis. This screening strategy would be biased, therefore, against promoters which were only active after incubation for 6 days, given that all the isolates which were selected (with the exception of isolate 5) were highly pigmented after spraying with catechol at each of the time points (Section 5.2.3). As the otcXp1 and otcCp1 promoters have been shown to have similar temporal activities in S. rimosus (Section 5.2.6), it is perhaps not surprising that isolates containing pIJ2843 constructs with the Sau3AIb-c fragment in either orientation were not selected.

The S1 mapping experiments described in Section 6.2.5 failed to detect transcripts originating from either the Sau3AIc-e or Sau3AIr-g regions in total RNA isolated from S. rimosus at different stages in the fermentation of oxytetracycline. The possibility that the elements cloned in isolates 8, 12 and 15 were "phantom" promoters has already been discussed. It is also possible that the 5' ends of the putative antisense transcripts were not detected, as they did not overlap to any
great extent with the single-stranded probes (Section 6.2.5). Interestingly, the Sau3AIc fragment cloned in isolates 8 and 12 was the product of partial digestion. It is possible that the internal Sau3AIc site, which is located only 31bp to the right of the Sau3AIc site, lies within sequences required to direct transcription of xyIE in the pIJ2843 constructs. Should this be the case, then the transcription start point of the promoter element may well be located close to the Sau3AIc site and thus transcripts originating from this element might not have hybridized efficiently to the probe generated from mKMX912 (Figure 6.3).

Primer extension from oligonucleotide C64 hybridized to total RNA isolated from S. rimosus (Section 5.2.5) should have detected the 5' ends of transcripts originating within the Sau3AIc region, provided transcription did not terminate before the primer annealing site. Similar to the rho factor-independent transcription termination sites in the enterobacteria (for review, see Galloway and Platt, 1986), the putative transcription termination sites of many streptomycete genes contain G+C-rich dyad symmetry (for review and references, see Seno and Baltz, 1988). Streptomycete terminators do not contain, however, the string of uridine residues, which follows the region of dyad symmetry at the 3' end of transcripts in the enterobacteria. No sequences with G+C-rich dyad symmetry typical of streptomycete terminators were detected in the intervening region between the Sau3AIc site and the annealing site for oligo. C64 (Figure 6.4). In the enterobacteria, another class of termination sites, which do not possess dyad symmetry and a string of uridine residues, can also cause RNA polymerase to terminate transcription. Although these termination sites are generally A+T-rich and require additional protein factors, there is no unifying sequence or structure for this second class of terminators (for reviews, see Platt and Bear, 1983). As yet, no factor-dependent termination sites have been characterised in Streptomyces. It is remains a formal possibility, therefore, that transcripts originating from the element(s) cloned in isolates 8 and 12 terminate at an unidentified factor-dependent terminator before the annealing site for oligo. C64. Another explanation for why the 5' ends of the putative antisense transcripts were not detected using primer extension analysis, is that the cloned promoter elements might only be active transiently. This might be expected if the antisense transcripts have a role in regulating the expression of the
Chapter 6 The SphI_{12}-SstI_{13} Region

_otcX_ locus and thus, the biosynthesis of OTC. No attempt, as yet, has been made to correlate the presence of catechol dioxygenase activity in plate assays (Section 5.2.3) with the intracellular levels of _xyIE_ transcripts, originating from cloned promoters in pIJ2843 constructs. The detection of catechol dioxygenase activity in isolates 8, 12 and 15 after incubation for 3, 4 and 5 days (Chapter 5, 5.2.3) is not inconsistent necessarily, therefore, with the promoter elements being active transiently.

On balance, the analyses described above suggest that the elements cloned in transformants 8, 12 and 15 are unlikely to direct the synthesis of mRNAs, which encode products involved directly in the biosynthesis of OTC. Although the above discussion does not lead to any firm conclusion about the role of these promoter elements, it highlights the range of possibilities which must be considered in undertaking further transcriptional analysis of the _otcX_ locus. The lines of investigation which are being considered to characterize further the promoter elements cloned in isolates 8, 12 and 15 are described in the concluding chapter.

It was hoped that the comparison of the amino acid sequences of the predicted products from _otcX_ PPCR1 and PPCR2 with entries in the NBRF-protein and EMBL databases would identify similarity to polypeptides with known functions and thus suggest a possible role for the _otcX_ locus in the biosynthesis of OTC. Unfortunately, no similarities to any entries in the databases were detected.

The _actVa_ region in the actinorhodin cluster of _S. coelicolor_ is currently being sequenced (Caballero, pers comm.). Comparison of sequence in the 0.96kb _SphI_{12}-SstI_{13}_ fragment from the _otcX_ region (Figure 6.1) and the 1.15kb _PstI_{5}-BglII_{5}_ fragment from the _actVa_ region (Figure 6.2) should reveal the DNA determinants which permit hybridization between these regions at high stringency.
CHAPTER 7

TRANSLATION INITIATION
Chapter 7  Translation Initiation  196

7.1 Introduction

_E. coli_ has been used as a model for studying translation in prokaryotic systems. Translation initiation is specified by sequences in messenger RNA other than start codons. The initiation codons of many genes from _E. coli_ are preceded by a short sequence of nucleotides that are directly complementary to some part of the -ACCUCCUUA-OH nucleotide sequence, lying at the extreme 3' end of the 16S ribosomal RNA of _E. coli_ (Shine and Dalgarno; 1974, 1975).

The 16S rRNA is an integral component of the 30S ribosomal subunit, which interacts with mRNA in conjunction with ribosomally associated proteins and initiation factors. Shine and Dalgarno postulated that a preliminary interaction between sequences at the 3' end of 16S rRNA and complementary sequences on the mRNA could target a 30S ribosomal subunit towards an AUG or GUG translation initiation codon. This interaction would allow the formation of a 30S subunit-mRNA-<sub>3</sub>m<sub>7</sub>g<sub>5</sub>RNA complex, which was positioned correctly and could give rise to an active in-frame ribosome on addition of the 50S ribosomal subunit. The translation of mRNA can be affected by the complementarity of the Shine-Dalgarno (SD) sequence and the spacing of this sequence from initiation codons (for review, see Gold and Stormo, 1987).

For the Gram-positive genera, _Bacillus_ and _Staphylococcus_, it has been proposed that a high degree of complementarity between the sequence at the 3' end of 16S rRNA and sequence immediately upstream of the translation initiation codon on the mRNA (equivalent to the SD sequence in _E. coli_) is required for efficient translation (McLaughlin et al., 1981). Although the sequence at the 3' end of the 16S rRNA of these Gram-positive genera is identical to that of _E. coli_ (McLaughlin et al., 1981), genes from _E. coli_ are poorly expressed in _B. subtilis_. It has been proposed that the principal obstruction to the expression of _E. coli_ genes in _B. subtilis_ occurs at the level of translation, as the SD sequences of _E. coli_ have significantly lower complementarity to the 3' end of 16S rRNA than do those of _Bacillus_ and _Staphylococcus_ (Ehrlich and Sgaramella, 1978).
Chapter 7 Translation Initiation 197

The inability of B. subtilis to translate E. coli gene starts has been correlated with the absence of the largest 30S ribosomal protein found in E. coli, protein S1 (Higo et al., 1982). Recently, the removal of the S1 protein from E. coli ribosomes was shown to result in the preferential translation of mRNA from Bacillus (Roberts and Rabinowitz, 1989). It was concluded that the S1 protein permits ribosomes to initiate translation at codons which are preceded by Shine-Dalgarno sequences with low complementarity to the 3' end of 16S rRNA.

Comparisons of the sequences preceding the translation start codons of genes from Streptomyces has indicated that the binding of ribosomes to Streptomyces mRNA does not require the degree of complementarity required apparently in Bacillus and Staphylococcus (Hopwood et al., 1986; Seno and Baltz, 1988; this work). Unlike Bacillus and Staphylococcus, the 30S subunits of Streptomyces ribosomes contain a S1 protein (Roberts and Rabinowitz, 1988). In these respects, the initiation of translation in Streptomyces is analogous to the process in E. coli.

Although, the SD sequence can promote the initiation of translation, it is not essential in all cases. A number of E. coli genes lack SD sequences and yet are translated efficiently. Transcripts of the lambda cI gene which originate from the Pm promoter, for example, lack a SD sequence and have the A of the AUG initiation codon at their 5' terminus. An examination of the sequences following the translation initiation codons of E. coli genes has shown that nearly all contain, at least, three to four conserved nucleotides that are complementary to some part of the 5'-AAAUGAAGAGUUUGA- sequence located at the 5' terminus of 16S rRNA (Petersen et al., 1988).

It has been proposed that sequences at the 5' end of 16S rRNA may be involved in mRNA recognition through a mechanism analogous to the well established "Shine and Dalgarno" interaction. In support, cI transcripts that lack SD sequences contain nucleotides downstream from the start codon which are complementary to a discrete region at the 5' end of the E. coli 16S rRNA (Petersen et al., 1988).
At present, many laboratories interested in *Streptomyces* are sequencing genes involved either in morphological differentiation or in secondary metabolism. Using the FRAME program (Bibb et al., 1984), it is usually relatively straightforward to identify the approximate boundaries of potential protein coding regions (PPCR). FRAME analyses is unable, however, to determine the precise bases at which a PPCR starts, particularly when two or more in-frame start codons are closely spaced (for further description, see Section 4.2.3.1). It has not always been possible to assign unambiguously translation initiation codons on the basis of complementarity to the 3’ sequence of 16S rRNA.

For *E. coli* genes, the statistical analysis of sequences "downstream" from the proposed translation initiation codons (Petersen et al., 1988) provided additional criteria to distinguish *bona fide* translation initiation codons from "false" starts in uncharacterised sequence data. Additional criteria would also aid the identification of translation initiation codons in *Streptomyces*. For simplicity, the term "gene start" (used by Petersen et al., 1988) will be used to describe the region of mRNA, which is thought to be involved in the initiation of translation, at the start of a coding sequence.

The objective of this preliminary study was to establish if sequences downstream from the putative initiation codons of a selection of *Streptomyces* genes could form complementary base pairs with sequence at the 5’ end of 16S rRNA and could be involved, in ribosome binding. The majority of the tentative gene starts which were examined were taken from the oxytetracycline cluster of *S. rimosus*.

7.2 Results and Discussion

7.2.1 Comparison of 16S rRNA Sequences from *Streptomyces* and *E. coli*

16S rRNAs from *S. coelicolor* and *S. lividans* have been sequenced (Baylis and Bibb, 1987; Suzuki and Yamata, 1987) and show greater than 99.5% identity. The 5’ and 3’ sequences of the *Streptomyces* 16S rRNA, corresponding to the *E. coli* sequences implicated in the binding of
the ribosome to mRNA, are shown in Figure 7.1, Panel A. The 3’
sequences of the *Streptomyces* 16S rRNAs are almost identical to the
equivalent *E. coli* sequence involved in Shine-Dalgarno base pairing.
This is consistent with the finding that most of the *Streptomyces* genes
which have been sequenced possess a reasonable Shine-Dalgarno
sequence between 1 and 18 nucleotides before the start codon (see Table
7.1 for examples).

Alignment of the 16S rRNA sequences from *Streptomyces* revealed
a slight difference in the 5’ end points of *rrnD* from *S. coelicolor* and
*rrnB* from *S. lividans*. The 5’ termini of these 16S rRNAs may actually
be different, alternatively, experimental error could be the cause of the
slight discrepancy in their start points. For clarity, the 5’ end of 16S
rRNA from *S. lividans*, which is two nucleotides longer than the 5’ end
of 16S rRNA from *S. coelicolor* and one nucleotide longer than the *E. coli*
16S rRNA, was used as the reference sequence in the following
comparisons. It was assumed that sequences at the extreme 5’ end of
16S rRNA would be highly conserved in all *Streptomyces* spp.

The sequence of the 5’ end of *Streptomyces* 16S rRNA only shows
identity with nucleotides 8 to 16 of the *E. coli* sequence. The first 7
nucleotides of the *E. coli* 16S rRNA, which are implicated in
"downstream" recognition, have no identity to the corresponding
sequence in *Streptomyces*. The difference in sequence at the extreme 5’
end of 16S rRNA may reflect the disparity in the preferred codon usage
of *Streptomyces* and *E. coli*, as any interaction involving 16S rRNA
within the coding sequence would be expected to be compatible with the
coding requirements of the mRNA.

7.2.2 Complementarity Between Downstream Sequences and the 5’ End of
16S rRNA

From the work of Petersen and co-workers (1988), which inspired
this particular investigation, only the first 16nt of the 16S rRNA from *E.
coli* were implicated in complementary base pairing with "downstream"
sequences. Correspondingly, only the first 17nt of 16S rRNA from
*Streptomyces* were analysed. In *E. coli*, nucleotides capable of forming
Figure 7.1. (A) Alignment of the 5' and 3' sequence of 16S rRNA from *S. coelicolor* (Baylis and Bibb, 1987), *S. lividans* (Suzuki and Yamata, 1987) and *E. coli* (Brosius et al., 1981). A cross (+) indicates a nucleotide conserved between the three 16S rRNA sequences and the asterisks (*) shows a nucleotide conserved between only the *Streptomyces* sequences. (B) Alignment of the 5' end of 16S rRNA from *Streptomyces* with complementary sequences in the tentative "downstream" regions of *otrA* (Doyle, 1987), *otcC* (Butler, unpublished work), *otcX ORF1* and ORF2 (this work), *otcZ* (this work), *actIII* (Hallam et al., 1988), φC31 c (Sinclair and Bibb, 1988), *dagA* (Buttnер et al., 1987), *dhq* (White, unpublished work), *gra ORF1* and ORF2 (Sherman et al., 1989), *pgm* (White, unpublished result), *tcmI ORF1* and ORF2 (Bibb et al., 1989) and *whiG* (Chater et al., 1989). The line (I—I) indicates the extent of the proposed complementary base pairing, using the criteria described in Table 7.1.
Table 7.1 Selection of gene starts from *Streptomyces* spp that show some degree of complementarity to a discrete region at the 5' end of 16S rRNA.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>ggauuccGAAccccucc..aug.aacaagcUGAAAUcugggcauccuggcc..</td>
</tr>
<tr>
<td>otrA</td>
<td>uuggaucaAGGAGcguu..aug.cgguaagCGUGGgauauggcgggcg</td>
</tr>
<tr>
<td>otcX1</td>
<td>ggaccggAGGACgucu..aug.CCGUuaccccaagaaggaaauccauac</td>
</tr>
<tr>
<td>otcX2</td>
<td>caggagcGGAAGacgcc..aug.agggaacucgcccgaauuCCCGUGggau</td>
</tr>
<tr>
<td>otcZ</td>
<td>cccagaccGAGGAucc..aug.acciaccaccgucgccCCCGUGgcc</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>actIII</td>
<td>aggcagGGAGGGAacac..aug.ccacacggacUCGGaagucgcacug</td>
</tr>
<tr>
<td>c (sC31)</td>
<td>caacGAaGGGcggacc..aug.aagcgggucACUCUGcgccgacgc</td>
</tr>
<tr>
<td>dagA</td>
<td>cgaagAAGGAacgacau..gug.guccaccaCGUGAucuacuaagug</td>
</tr>
<tr>
<td>dhqX</td>
<td>ucacgcaccGGGAgacc..gug.ccCCGacccuggcacaagccggcauc</td>
</tr>
<tr>
<td>gra ORF1</td>
<td>caguacAGGAGaccacg..gug.acccgacgguaUGAucaccggguc</td>
</tr>
<tr>
<td>gra ORF2</td>
<td>ccgccggAGGAGGccgc..gug.aggacCCUGAaccgccgggacgc</td>
</tr>
<tr>
<td>ORF1215*1</td>
<td>ccacuccGAGGccgac..gug.acciaccucUGAaagggccgcauguc</td>
</tr>
<tr>
<td>ORF1215*2</td>
<td>ccacuccGAGGccgac..gug.acciaccucUGAaagggccgcauguc</td>
</tr>
<tr>
<td>pgm</td>
<td>uggauuuaccgcucGGAAGc..aug.gccagcagccCGUacaagcuauccu</td>
</tr>
<tr>
<td>tcmI ORF1</td>
<td>cgaaggggUGAGGcugag..aug.accCGGacgcggagaacggguggug</td>
</tr>
<tr>
<td>tcmI ORF2</td>
<td>ccggaauuuccggUGGAu..gug.guacgCGUGUacaaagcgcgcgc</td>
</tr>
<tr>
<td>whiG*1</td>
<td>accgcacBGAGcAcga..aug.cccagacacUCGCggttcgacgc</td>
</tr>
<tr>
<td>whiG*2</td>
<td>gcccgcgacGGUGGcagc..gug.cgacccgCCGccccctgcagcccgac</td>
</tr>
</tbody>
</table>

**C**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>afsA*1</td>
<td>aug.gacgcggaggccggagggugugcatccgggggaucgag ..............CCGorUGG..or..CCCGUGg........</td>
</tr>
<tr>
<td>afsA*2</td>
<td>auggacgcGGAGGccgag..gug.guacatcCGCGUGgggaucgagauggu</td>
</tr>
<tr>
<td>sta*</td>
<td>aug.acciaccacccauggcagcaguacgaauucgcagccg ...............UGG..or..CGU..or..yUCCG........</td>
</tr>
<tr>
<td>aph</td>
<td>aug.gacgcagcagcuuugcgccggaaguaccgcccagc ...............CGU..or..CCG..or...CCG........</td>
</tr>
</tbody>
</table>

The Shine–Dalgarno (SD) and "downstream" sequences are shown in upper case. Nucleotides that extend the SD and downstream sequence if G:U base pairing is allowed are underlined. One mismatch was permitted in each sequence. The symbol (+) indicates translation start points confirmed by Nh-sequencing of the gene product. The symbol (*1) indicates the authors first choice for the start codon. Likewise the authors choice (sC31) indicates an alternative translation start point. In examples, were only three consecutive were identified, only the first triplet occurring downstream of the start codon is indicated. For references to sequences see text and Figure 7.1.
favourable interactions with 5' sequences from 16S rRNA were distributed between 4 to 21 nucleotides downstream of the start codon.

In this study, nucleotides as far as 39nt downstream of the proposed translation initiation codons of *Streptomyces* genes (+39) were analysed initially for complementarity. Similar to the situation in *E. coli*, it would appear that the coding regions of *Streptomyces* genes also contain a sequence of three or more contiguous nucleotides downstream of the translation initiation codon which are complementary to the 5' end of *Streptomyces* 16S rRNA (Table 7.1). Of the oxytetracycline genes, which have been sequenced completely (Table 7.1, Panel A), the putative translation initiation codons were designated initially after consideration of codon usage (Bibb *et al.*, 1984) and potential SD sites (Shine and Dalgarno, 1974). In support, all of these potential gene starts contain at least four consecutive nucleotides downstream of the start codons, which are complementary to the 5' end of 16S rRNA.

Sequences in *otcX* PPCR1 (UCCGUGgg) and *otcZ* (CCGUGg) have been identified, between +21 and +28 nucleotides, which could form highly favourable interactions with nucleotides at the 5' end of *Streptomyces* 16S rRNA. On the basis of this finding, it is suggested that putative "downstream" interactions should be considered as far downstream as +28 in *Streptomyces* genes. In this study, all of the gene starts which were examined contained at least three consecutive nucleotides downstream of the proposed start codon that were complementary to the 5' end of 16S rRNA.

Although, the most probable translation initiation codons for the oxytetracycline genes were assigned relatively easily, without consideration of "downstream" sequences, it has not always been possible to do so solely on the basis of suitably spaced Shine–Dalgarno sequences. The translation start codon for *whiG*, for example, could not be assigned unambiguously, as two potential start codons were preceded by suitably spaced sequences that had complementarity to the 3' end of 16S rRNA (*whiG* and *whiG*, Panel B) (Chater *et al.*, 1989). Analysis of these potential gene starts identified five consecutive nucleotides in
whiG\textsuperscript{1} and three consecutive nucleotides in whiG\textsuperscript{2}, respectively, which were complementary to the 5' end of 16S rRNA. Although, the whiG\textsuperscript{1} gene start is predicted to form the most favourable interaction with the 30S ribosomal subunit, it remains a formal possibility that the whiG\textsuperscript{2} start could also be a site for the initiation of translation.

The proposed gene start for ORF1215 of the element IS110 (Bruton and Chater, 1987) contains 4 consecutive nucleotides (6, if G:U base pairing is considered) downstream of the tentative AUG start with complementarity to the 5' end of 16S rRNA (Table 7.1. Panel B). Although, the tentative start codon is preceded by a region with complementarity to the 3' end of 
Streptomyces
16S rRNA, the precise region of complementarity is unconventional. It was noticed, however, that an in-frame GUG codon, 24 nucleotides upstream of the proposed AUG codon, is preceded by a reasonable SD-like sequences and contains four consecutive nucleotides downstream of this codon that are capable of forming complementary base pairing. On the basis of the overall complementarity to the 5' and 3' ends of 16S rRNA, it is suggested that the GUG codon upstream of the AUG codon should be considered as a potential start codon. This proposal is consistent with the FRAME plots (Bruton and Chater, 1987) which start to diverge at this point in the sequence.

As in 
E. coli,
 it is not uncommon in 
Streptomyces
for the first nucleotide of the translation start codon to be also the 5' terminus of the mRNA, for example; the streptothricin acetyl transferase (sta) gene from 
S. lavendulae
(Horinouchi et al., 1987), the aminoglycoside phosphotransferase (aph) gene from 
S. fradiae
(Janssen et al., 1989) and the gene responsible for A-factor biosynthesis (afsA) of 
S. griseus
(Horinouchi et al., 1989). Analysis of the regions downstream of the proposed translation start codons for these genes has identified in each case, at least, three consecutive nucleotides which could form complementary base pairs with the 5' end of 16S rRNA (Table 7.1, Panel C). Interestingly, in the afsA message there is a GUG, located six codons downstream from the AUG codon at the 5' end. This GUG codon is preceded by five consecutive nucleotides that could form
conventional base pairs with the 3' end of 16S rRNA (afsA<sup>2</sup>) (Horinouchi et al., 1989). Downstream of the GUG codon are 5 consecutive nucleotides (if G:U base pairing is considered) which could base pair with the 5' end of 16S rRNA. On the basis of complementarity of surrounding sequences to the 16S rRNA, the GUG codon might be expected to be the primary codon at which translation of the afsA message is initiated (afsA<sup>2</sup>). Site-directed mutagenesis has been used to create a frame shift between the AUG codon (that is the 5' codon of the afsA message) and the GUG codon (Horinouchi et al., 1989). It was inferred from the ten-fold decrease in the expression of afsA in the mutant, that the AUG codon was the primary start codon, but, that translation at a low level could be initiated at the downstream GUG codon (Horinouchi et al., 1989). An alternative hypothesis is that the GUG gene start, which has extensive complementarity to both the 5' and 3' end of 16S rRNA, was the primary site for translation initiation and mutagenesis upstream of the GUG codon severely affected the mRNA-ribosome interaction in an unknown manner. Amino-terminal sequencing of the afsA gene product would resolve these two possibilities.

It has been proposed that the lack of SD sequences in the mRNA of the above examples may be related to an unknown mechanism of differential gene expression (Horinouchi et al., 1989). The "downstream" sequences of the above gene starts do not appear to contain significantly longer runs of consecutive nucleotides with complementarity to the 5' end of 16S rRNA, compared to those with SD sequences. In these examples, it is possible that other unknown factors are required to target 30S ribosomal subunits towards the translation codon at the extreme 5' end of the message. Although this preliminary analysis has focussed solely on the sequence content of gene starts, it should be noted that the secondary structure of mRNA around the start codon is also believed to have an important role in controlling translation. The reader is directed to papers by Iserentant and Fiers (1980) and de Smit and van Duin (1990) for discussion and references on the role of secondary structure in translation initiation.
7.2.3 Nucleotides of 16S rRNA implicated in binding to "downstream" regions

Alignment of the tentative "downstream" regions with the complementary sequences in the 5' end of 16S rRNA, revealed that only 7nt (AGGCACU) of the 17nt at the 5' end of Streptomyces 16S rRNA were repeatedly implicated in downstream-30S ribosomal subunit interactions (Table 7.1). Stated more precisely, the sequences of mRNA complementary to the 16S rRNA contain, in part, several conserved nucleotides from the string T.CaT.CaT.G.T.G.AaG within the first 25 nucleotides of the coding sequence immediately following the initiation codon. The alternative nucleotides within the string arise from possible G:U base pairing.

The nucleotides AGGCACU at the 5' end of 16S rRNA from Streptomyces, which are tentatively involved in mRNA recognition are not present at the equivalent position in 16S rRNA from E. coli. This is not surprising, perhaps, as the mol% G+C composition of these two bacteria are very different. It is conceivable that sequences at the 5' end of 16S rRNA, which are involved in "downstream" recognition, have evolved to accommodate the particular codon usage of individual bacterial genera. It would be interesting, therefore, to compare sequences at the 5' end of 16S rRNA and gene starts from a variety of genera with markedly different G+C compositions, to establish if there is any correlation with the biased usage of synonymous codons.

Although many Streptomyces promoters are not transcribed in E. coli, the binding of Streptomyces mRNA to E. coli ribosomes appears to proceed normally when transcription is facilitated by an adjacent E. coli promoter (Hunter and Baumberg, 1989). For example, the otrAp1 promoter and the 5' end of the otrA resistance gene (nucleotides +1 to +1146) can direct expression of a OtrA:LacZ fusion protein in E. coli (McDowall, unpublished result). The expression of Streptomyces genes in E. coli (including examples where an E. coli promoter is used to direct transcription) suggests that "downstream" sequences in Streptomyces gene starts, which only rarely show any complementarity to the 5' end
of *E. coli* 16S rRNA, are not the principal determinants that direct the 30S ribosomal subunit to the correct translation initiation codon on mRNA. In general, an interaction with the Shine-Dalgarno sequence almost certainly directs the 30S ribosomal subunit to *bona fide* translation initiation codons (Gold and Stormo, 1987). Nevertheless, the statistical analysis of *E. coli* genes has revealed that contiguous nucleotides in the "downstream" region with complementarity to the 5' end of 16S rRNA are a conserved feature of authentic translation initiation codons. Downstream interactions may have an important role, therefore, in determining the efficiency of translation, perhaps by stabilising mRNA-30S subunit interactions during the formation of the initiation complex.

7.3 Scope of Future Studies

This limited study has shown that consecutive nucleotides downstream of the translation start codons in *Streptomyces* gene starts have the potential to bind the 5' end of 16S rRNA. This proposed interaction is envisaged to be important in determining the efficiency of translation.

In *E. coli*, the probabilities of contiguous nucleotides located downstream of translation initiation codons occurring at random with complementarity to sequences at the 5' end of 16S rRNA were calculated using a simple equation (Petersen et al., 1988). It was assumed that the four nucleotides A, C, G and T were distributed randomly at the beginning of the coding regions. This would clearly not be a valid assumption for the analysis of "downstream" interactions in *Streptomyces*. The average mol G+C composition in the first, second and third position of *Streptomyces* codons is 70%, 45% and 90% respectively: this is clearly non-random.

The second approach, taken by Petersen and colleagues to analyse "downstream interactions" statistically, could be adopted to analyse *Streptomyces* gene starts. They created a file of "false" gene starts by searching the EMBL database for SD-like sequences followed
by suitably spaced AUG codons, which were not thought to be authentic sequences for the initiation of translation. The distribution of sequences complementary to the 5' end of 16S rRNA in files of approximately 150 "true" and "false" gene starts were then compared. In this way, the occurrence of complementary sequences was shown to be statistically significant in the downstream regions of real start codons. This analysis implied that the codon usage at the start of E. coli genes was biased to accommodate favourable base pairing with the 16S rRNA. The usage of synonymous codons in Streptomyces is already strongly biased towards those containing G and C residues. The nucleotides GGCACU at the 5' end of 16S rRNA, which are the principal nucleotides implicated in the base pairing with the downstream sequences (involved in greater than 50% of the proposed interactions, Figure 7.1) are representative of the average mol% G+C composition of the Streptomyces genome. It is possible, therefore, that the general biased usage of synonymous codons in Streptomyces, contributes to the high frequency with which consecutive nucleotides, complementary to the 5' end of 16S rRNA, are found downstream of translation start codons. With an increasing number of genes from Streptomyces being sequenced, currently around 50 have been published, it should soon be possible to analyse Streptomyces gene starts statistically, in a similar manner to E. coli, and determine if the observed complementarity in downstream regions of Streptomyces genes could be used as an additional criterion to elucidate authentic translation start points.
CHAPTER 8
CONCLUDING REMARKS
CONCLUDING REMARKS

The primary objective of the work described in this thesis was to characterise the transcription of the otrA resistance gene during the fermentation of oxytetracycline (OTC). The otrA gene was transcribed from two promoters, otr4p1 and otcCp1, which were regulated differentially. During the period of vegetative growth that preceded antibiotic production, otrA was transcribed principally from the otr4p1 promoter, which is located in the intergenic region between the 5' end of otrA and the 3' end of the biosynthetic gene otcZ. The relative amount of transcripts originating from otr4p1 decreased within the total RNA population as the growth rate slowed. At the onset of antibiotic production, the otrA resistance gene was shown to be transcribed from beyond the 3' end of otcZ. A promoter was located upstream of otcC, which is the biosynthetic gene situated upstream of otcZ. The otcCp1 promoter showed maximal activity during the early phase of OTC production. As S1 nuclease mapping experiments did not identify transcripts originating from immediately upstream of otcZ, it was concluded that during antibiotic production otrA was transcribed probably as part of a otcC-otcZ-otrA polycistronic message. Such an arrangement would ensure that the level of resistance to OTC increases in accordance with the level of antibiotic production. It remains to be shown directly, however, that a 5.3kb transcript extends from the promoter otcCp1 to the rho-independent terminator, which is located immediately downstream of otrA.

Northern analysis (Alwine et al., 1977; Fourney et al., 1988) may appear to be a straightforward technique for confirming the presence of the 5.3kb otcC-otcZ-otrA polycistronic message. For unknown reasons, however, a reliable protocol has not been developed for analysing RNA from Streptomyces by Northern analysis. Recently, an alternative approach has been developed to map transcripts from Streptomyces at low resolution (Smith, unpublished work). Samples of total RNA are hybridized (as described for high-resolution S1 mapping) to a series of single-stranded M13 constructs, which contain overlapping inserts from the region of interest. The RNA samples are then digested with S1
nuclease. The RNA/DNA duplexes, which are not digested by S1 nuclease, are then resolved on agarose gels by electrophoresis. The RNA-protected fragments are then transferred to nylon membranes by capillary blotting under alkali conditions (as for Southern analysis). The immobilised DNA can then be analysed using radio-labelled oligonucleotides or probes produced by the random primer method. It is possible to map precisely the position of transcription units, by analysing the RNA-protected fragments with a number of different probes. The most time consuming part of this technique is the insertion of the DNA fragments from the region of interest into M13mp18 and M13mp19. Already much of the oxytetracycline cluster has been subcloned into these vectors to enable sequencing (Butler, unpublished work) and this technique may be useful for determining the organisation of transcription units within the otc cluster.

Sequencing of the Smal\textsuperscript{-}BcII\textsubscript{s} fragment identified a single PPCR between the 3' end of otcC and the 5' end of otrA. The predicted amino acid sequence of the derived gene product was very similar to the amino acid sequence of a methyltransferase from bovine pineal glands (Ishida et al., 1987) and a polypeptide predicted from a PPCR in a segment of the tetracenomycin cluster from S. glaucescens, which complements tcm\textsubscript{II} and tcm\textsubscript{V} mutants that accumulate C-3 demethyl intermediates (Hutchinson, pers. comm.). S-adenosylmethionine (SAM) has been shown to be the methyl group donor in the methylation reaction catalysed by bovine pineal gland. A possible SAM binding site was identified in a region which was highly conserved in all three polypeptides. It was concluded that the otcZ gene encoded a polypeptide that catalysed the C-6 methylation of the tetracyclic nucleus to from 6-methylpretetramid. It is predicted that otcZ mutants will accumulate C-6 demethyl intermediates. The otcC gene, which encodes the ATC oxygenase, was not described in any detail within this thesis, as it was sequenced in collaboration with M.J. Butler, formerly of Pfizer.

Two PPCRs, which were encoded on the opposite strand to otcC, otcZ and otrA, were identified within the SphI\textsubscript{II}2-SstI\textsubscript{II}3 fragment of the otc\textsubscript{Y} region. Comparison of the amino acid sequences of the predicted
products from these PPCRs with entries in the NBRF-protein and EMBL databases did not identify any similarity to polypeptides with known functions. The Sphi-SstI fragment has been shown to hybridize to a region within the actIa region of the actinorhodin cluster of *S. coelicolor* (Butler *et al.*, 1990), which has been implicated in hydroxylation (Hopwood *et al.*, 1985; Cole *et al.*, 1987). The PPCRs in the Sphi-SstI fragment may encode polypeptides with functions required for the C-4 hydroxylation of 6-MPT. Should this be the case, mutants altered in the functions of these polypeptides would be predicted to accumulate C-4 dehydroxyl intermediates.

Sequencing of DNA regions in the otc cluster, which hybridize with the actI and actIII probes (described by Malpartida *et al.*, 1987), has identified a number of PPCRs that are believed to encode polypeptides required for the biosynthesis of the polyketide backbone (Sherman, pers. comm.; Thamchaipenet, pers. comm.) Using the sequence data available for the otc cluster, it is now possible to introduce defined mutations into a number of PPCRs. Site-directed mutagenesis could be used to create a series of in-frame mutations, which would hopefully inactivate the gene products without disrupting transcription and translation. The mutated PPCR and flanking sequences would then be introduced into *S. rimosus* on derivatives of the temperature-sensitive vector pSG5 (Muth *et al.*, 1989), which contain resistance markers that can be select in *S. rimosus* (Hunter, unpublished work). Replacement of the wild-type gene with the mutated gene would be achieved in two steps. It should be possible to isolate recombinants produced by a single recombination event between sequences cloned on the plasmid and homologous sequences on the chromosome. This would be achieved by incubating transformants at temperatures which restrict the replication of the plasmid, while still maintaining selection for plasmid markers. After culture of these recombinants at the restrictive temperature, isolates would be screened for the loss of the antibiotic resistance conferred by the plasmid and the reciprocal recombination event that removes the wild-type gene. A collaboration has been established with Dr R.A. Hill from the Chemistry Department at the University of Glasgow to analyse the polyketide products and shunt
metabolites produced by defined mutants. By combining expertise in analysing polyketide chemistry and the molecular genetics of *S. rimosus*, it should be possible to establish or confirm the function of individual genes in OTC biosynthesis.

Comparison of the promoters *otcCp1* and *otcXp1*, both of which are most active at the onset of antibiotic production, identified a conserved sequence that resembled *pho* boxes in *E. coli* (Makino *et al.*, 1986; 1988). This sequence has been designated the oxytetracycline *pho*-like box (*opl* box). Similar motifs have been identified in other promoters from *Streptomyces*, which are regulated by phosphate (Liras *et al.*, 1990). This conserved sequence is believed to be the binding site for a regulatory protein which activates transcription from promoters under conditions of low phosphate.

The isolation and characterisation of the regulatory protein, which is believed to bind *opl* boxes, represents the next major challenge in understanding the molecular mechanism that regulates OTC production. The isolation of total RNA from cultures of *S. rimosus* M15883 at different stages during the fermentation of OTC has shown that the *otcC-otcZ-otrA* polycistronic message and possibly other transcripts of the oxytetracycline cluster are most abundant at the onset of OTC production. The regulatory protein which is proposed to bind promoters in the *otc* cluster and activate transcription is predicted to be most abundant and/or active early in the production of OTC. One of the advantages of studying antibiotic production in *S. rimosus* M4018 and M15883, which have been studied intensively due to the commercial importance of oxytetracycline, is that reproducible fermentations have been established which yield high levels of OTC. The ability to predict when OTC will be produced in batch cultures should ease the isolation of reasonable amounts of active regulatory protein, which in turn will aid the characterisation of *opl*-protein interactions.

Possibly, the most perplexing result of this work was the finding that the DNA elements which directed transcription of *xyIE* in pIJ2843
constructs and hybridized to the *SphI*-*SstI* region did not correspond to either *otc* or *otcX*. Should the plasmid-active-promoter (PAP) elements, which were cloned in *S. lividans* transformants 8(12) and 15, be active in the chromosome, they would direct the synthesis of RNA that was complementary to the mRNA for *otcX* PPCR1 and PPCR2. It was concluded that these antisense RNA transcripts were unlikely to be mRNAs, as PPCRs were not identified on the opposite strand to *otcX* PPCR1 and PPCR2.

Using total RNA isolated from *S. rimosus* at different stages in the fermentation of OTC, S1 nuclease mapping experiments failed to detect transcripts arising from the regions cloned in isolates 8 and 15. Primer extension analysis could be used to establish the transcription start points of the PAP elements, by annealing the reverse primer for pIJ2843 to total RNA isolated from cultures of isolates 8 and 15 (at the earliest possible time point when *xylE* was being expressed) and extending the primer using reverse transcriptase. One possible explanation for why primer extension analysis of the transcripts originating from *otc* did not detect transcripts originating from the PAP elements, is that they terminated before the annealing site of the primer. Should the primer extension analysis of total RNA from *S. lividans* transformants 8 and 15 identify transcripts with a reasonable distance (>40 nucleotides) between the 5' end and sequences derived from pIJ2843, it would be possible to use oligonucleotides that were complementary to sequences close to the 5' end of the transcripts to identify short transcripts originating from the PAP elements in total RNA isolated from *S. rimosus* by primer extension. It is also possible that the antisense transcripts in the *otcX* region were not detected, as the PAP elements are only active transiently. Not knowing the role of these transcripts, it is impossible to predict when they might be synthesized. An alternative approach to detecting transcription from the PAP elements may be to insert a promoterless copy of the *xylE* reporter gene into the chromosome downstream of the PAP elements. The *xylE* reporter gene would be integrated as proposed previously for replacing wild-type genes in the *otc* cluster with genes mutated by site-directed mutagenesis.
Further analysis of oxytetracycline production in *S. rimosus* using molecular genetics promises to reveal many interesting facets about the regulation of antibiotic biosynthesis and its integration into the complex life cycle of *Streptomyces*. Genetic manipulation will also be an important tool for accessing the contribution of genes within the *otc* cluster to the biosynthesis of oxytetracycline.
BIBLIOGRAPHY


Bibliography


Bibliography


