MOLECULAR ANALYSIS OF AN OXYTETRACYCYLINE

RESISTANCE GENE FROM <u>Streptomyces rimosus</u>

A thesis submitted for the degree of

Doctor of Philosophy at the

University of Glasgow.

Deirdre Doyle

Institute of Genetics

University of Glasgow

Church Street

Glasgow

Sept 1987

(i)

ProQuest Number: 10948163

ProQuest 10948163

Dedicated to mum

and dad.

CONTENTS

Abbreviations

Chemicals:

APS - ammonium persulphate

ATP - adenosine triphosphate

CIP - calf intestinal phosphatase

CsCl - caesium chloride

DMF - dimethylformamide

DMS - dimethylsulphate

DNA - deoxyribonucleic acid

dATP - deoxyadenosine triphosphate

dCTP - deoxycytidine triphosphate

dGTP - deoxyguanosine triphosphate

dTTP - deoxythymidine triphophate

ddNTP - dideoxyNTP

DTT - dithiotreitol

EDTA - ethylenediaminetetra-acetic acid (sodium salt)

EtBr - ethidium bromide

EtOH - ethanol

HZ - hydrazine

IPTG - isopropylthio-B-D-galactoside

RNA - ribonucleic acid

mRNA - messenger RNA

rRNA - ribosomal RNA

SC - sodium cacodylate

SDS - sodium dodecylsulphate

Sod. Cit. - sodium citrate

TEMED - tetramethylethylene-diamine

Tris - tris (hydroxymethyl) amino ethane

XGAL - 5-bromo-4-chloro-3-indolyl-B-galactoside

Antibiotics:

Ap - ampicillin

km - kanamycin

tsr - thiostrepton

OTC - oxytetracycline

Measurements:

bp - base pairs

kb - kilo bases

V - volts

W - watts

Ci - curies

mCi - millicuries

uCi - microcuries

U - unit

g - centrifugal force equivalent to gravitational acceleration

mg - miligrammes

ug - microgrammes

ng - nanogrammes

l - litres

ml - millilitres

ul - microlitres

nM - nano metres

OD - optical density

M - molar

mM - millimolar

nM - nanomolar

pM - picomolar

^{o}C - degrees centigrade

RT - room temperature

mins - minutes

hrs - hours

Miscellaneous

ss - single stranded

ds - double stranded

E - core RNA polymerase

Acknowledgements

SUMMARY

An oxytetracycline (OTC) resistance gene, otcI, has been cloned previously from Streptomyces rimosus 15883, a commercial producer of the antibiotic. This gene confers protection to the hosts ribosomes from arrest of translation by OTC. The otcI gene has been sequenced and its transcriptional regulation pattern has been analysed.

The otcI structural gene encodes a polypeptide of 71.329KD. The predicted amino acid (a.a.) sequence is highly homologous with the tetM polypeptide, from a Streptococcal transposon Tn1545. The tetM polypeptide also protects the protein synthesing machinery from the action of tetracycline. The first 130 amino acids of each polypeptide are 67% homologous and smaller regions of extensive homology are dispersed throughout the sequences. The amino terminal end of elongation factor Tu (EF-Tu) from E.coli is also homologous. It exhibits 36% homology and if chemically equivalent amino acids are considered, 48% homology to the otcI polypeptide. This homology includes the characterised GTP binding site of EF-Tu. The conservation of a.a. between otcI, tetM and EF-Tu suggests that the amino terminal end of otcI has a similar function to the amino terminal end of the EF-Tu. Possibly the otcI polypeptide acts as an alternative EF-Tu in S.rimosus which protects protein synthesis from the inhibitory effects of oxytetracycline.

Two transcriptional start sites have been identified for the otcI structural gene. Transcription initiates 129bp (otcIp1) and 339-344bp (otcIp2) upstream of the translational initiation codon. otcIp1 resembles an E.coli type consensus promoter and otcIp2 has not been sequenced it is external to the cloned DNA. Both transcripts were identified by S1 mapping with RNA from S.rimosus.

Six closely clustered divergent transcriptional start sites (the otcpA1-6 promoters) have been identified with RNA from S.lividans TK24 containing a clone of the region 5' to the otcI structural gene and

11% of the coding region. These promoters overlap the otcIp1 start site and the transcript overlaps the otcZ coding region (one of the OTC production genes) which is transcribed in the same direction as otcI. Promoter activity has been recorded from promoter probe constructs of this region but the divergent transcriptional activity is repressed in constructs which contain 45% of the otcI coding sequence. The transcriptional start sites of otcpA1-6 have not been identified with RNA from S.rimosus which is consistent with the promoter probe data. In promoter probe constructs otcIp1 is partially repressed by promoter activity from otcpA1-6. otcpA1 is very similar to a group of streptomycete promoters which are probably transcribed by core RNA polymerase in association with sigma factor 49. The otcpA1-5 promoters show no homology to any recognized promoter sequence.

The transcriptional complexity of this region indicates several levels of regulation · i.e. the number and variety of promoter signals, repression of the divergent transcript by sequences within the coding region of otcI (possibly the otcI polypeptide itself) and partial repression of otcIp1 by the transcriptional activity of overlapping promoters otcpA1-6. The function of the divergent transcript is unknown but it is possible that it regulates the expression of otcZ. This regulation might be mediated by an anti-sense RNA or if the transcript is translated, by the polypeptide. Such control of the expression of an OTC biosynthetic gene would ensure that a cell was expressing otcI and therefore resistant to OTC before the onset of OTC production.

CHAPTER 1

INTRODUCTION

1.1 GENERAL INTRODUCTION

Streptomycetes form a genus of gram-positive soil bacteria which contains many dfferent species. Historically S.coelicolor has been studied most but the vast range of secondary metabolites produced by all streptomycetes has stimulated interest in a large number of species. S.coelicolor has an estimated genome size of approximately 10^4 kilobases (kb) (Antonov et al 1978) and exhibits a complex growth cycle involving spore germination, growth to substrate mycelium, production of aerial mycelium and spore formation (Chater and Merrick 1979). The range of secondary metabolites produced by streptomycetes includes more than 60% of the known natural antibiotics, anti-cancer agents, anti-heliminthics, growth promotants and herbicides. Antibiotic production has been correlated with the development of aerial mycelium and spore formation and both processes indicate the need for the differential regulation of gene expression. The industrial potential of Streptomyces has stimulated interest in understanding the processes of cellular differentiation and secondary metabolite production, though many aspects of Streptomyces biology are of general interest. In comparison to other prokaryotic organisms, streptomycetes have a large genome size and they exhibit gross DNA macrolesions. Streptomycete DNA has an average G+C content of 72% (Enquist and Bradley 1971), which differs substantially from the DNA of more intensely studied prokaryotes e.g. E.coli-51% (Singer and Ames 1970) and B.subtilis-40% (Singer and Ames 1970).

Streptomyces have unusually plastic genomes and extensive genome alterations have been observed in many strains. Amplifications of 100-500 copies per chromosome of DNA segments ranging from 5-20kb have been recorded (Robinson et al 1981) and some of the amplifiable units are sandwiched between directly repeated sequences (Fishman and Hershberger 1983). The stability of amplified sequences varies from stably maintained amplifications to those which progressively deamplify during growth (Hershberger and Fishman 1985). Deletions of various sizes have also been observed, ranging from 0.3kb (Fisher et al 1987) to greater than 100kb (Hasegawa et al 1985). Deletions are often found in association with amplifications (Schrempf 1983) but

they also occur independently.   Some deletions result in the loss of
phenotypic markers.   For example the argG phenotype in S.lividans is
due to a deletion, with an associated amplification of neighbouring
DNA (Piendl et al 1987).   However, two independent deletions of
greater than 100kb and greater than 70kb analysed in S.glaucsens did
not affect the prototrophy of the strain (Hasegawa et al 1985). Young
and Cullum (1987) have proposed a hypothetical mechanism for the large
scale deletions which occur.   Their model requires that the
amplifiable unit is bordered by directly repeated sequences and
predicts that deletion events could occur at either end of the
amplified unit.   The model suggests that the units are amplified by a
rolling circle replication mechanism after a recombination event
between the directly repeated sequences.   The replication structure
may be resolved by homologous recombination or illegitimate
recombination which would lead to a deletion occuring on one side of
the amplified sequence.   The stimul i responsible for these major
genome alterations are unknown but the majority of genome alterations
have been observed after stressing cells by protoplasting, protoplast
fusion or treatment with DNA intercalating dyes.   A   direct
relationship between these events and the DNA rearrangements has yet
to be demonstrated.   A wide variety of gross genome alterations have
been analysed, many of which have no adverse effect on the cell but
the function of such a fluid genome has yet to be determined.

Circular chromosome maps have been obtained for several Streptomyces
strains.   The most complete linkage map is of S.coelicolor (Chater
1984) but the less extensive linkage maps of other strains show many
similarities with the S.coelicolor linkage map (Friend and Hopwood
1971).   The vast majority of markers map in two large clusters,
diametrically opposed, which flank two silent regions.   Some
functionally related genes map at diametrically opposed positions and
it has been suggested that a chromosomal duplication occurred in an
evolutionary ancestor of the streptomycetes.   Streptomyces genomes
have considerably greater coding capacity than E.coli and B.subtilis,
but it is unclear what the functions of the extra 6000 kilobase pairs
are.

Streptomyces DNA has a high G+C content (72%) in common with other

bacteria such as, Micromonospora (72%), Myxococcus (68%), Actinoplanes (74%), Nocardia (70%) (Singer and Ames 1970). G+C content varies widely among bacterial species, from 25% to 75%, and it is likely that these differences have been caused by mutation pressure. Singer and Ames (1970) propose that bacterial species exposed to sunlight evolve high G+C contents to avoid thymine specific damage from the ultraviolet (UV) radiation from sunlight. They cite many cases of bacteria which reproduce aerially, aquatic bacteria and bacteria protected by carotenoids, that have a high G+C content and there is a strong positive correlation of G+C content with exposure to sunlight. However other protection mechanisms to UV damage exist. UV light is absorbed by cellular constituents such as tryptophan and RNA and cells with greater cellular mass will screen out more UV light. DNA repair mechanisms also play a major role in repairing damaged DNA. It is difficult to assess the efficency of these mechanisms particularly as the repair systems of many bacteria are poorly understood. Singer and Ames (1970) have proposed that another factor, the evolution of the base ratio of the DNA, such that fewer targets are exposed to the mutagen.

The codon usage of any species reflects the relative abundance of G+C base pairs to A+T base pairs (Muto and Asawa 1987). Species with high A+T contents preferentially select synonomous codons with A/T in the third degenerate position of a codon. Mycoplasma capricolum (29% G+C) has a 10% G+C content at the third position of codons and Micrococcus luteus (75% G+C) a 90% G+C content. Streptomyces coding regions have a similar skewed occurrence of G or C base pairs, which makes a coding region instantly recognizable (Bibb et al 1984).

1.2 Developments in the molecular biology of Streptomyces

Hopwood's group at the John Innes Institute have pioneered both the genetical and molecular biological research in Streptomyces. The development of protoplasting and regeneration techniques enabled transformation of Streptomyces species (Bibb et al 1978). A gene conferring resistance to methylenomycin (Mm) from S.coelicolor was the first gene cloned from a streptomycete (Bibb et al 1980). Total DNA from S.coelicolor A3(2) was ligated to the low copy number vectors

Spores of each transformed strain were tested for resistance to Mm and the ability to elicit lethal zygosis (a small area of growth inhibition due to the presence of a plasmid containing cell on a plasmid free lawn). More sophisticated vectors are now available which have antibiotic resistance genes to select transformants and positive cloning vectors which select for transformants containing inserts. Many of the low copy number vectors now used in cloning, such as pIJ61 and pIJ922 are derived from the early cloning vectors, SLP1.2 and SCP2 respectively. pIJ101 is a high copy-number vector isolated from S.violaceroruber (Kieser et al 1982) which has been used to produce a large number of wide host range cloning vectors. Many phages have been isolated from different Streptomyces species and one of the most developed is the broad host range actinophage ϕC31 (Chater et al 1981). ϕC31 has the ability to form lysogens or undergo a lytic life cycle after entering a cell. Lysogeny follows a similar pattern to phage lambda in E.coli. There is a specific integration site in the chromosome (attB) and a specific site (attP) in the phage genome which interacts with the chromosomal attachment site. The establishment and maintenance of lysogeny requires the c-gene product which also represses lytic phage functions. ϕC31 has been developed extensively as a cloning vector. Recombinant phages deleted for attP can integrate into an ϕC31 lysogen by recombination to form double lysogens or by homologous recombination between the chromosome and DNA cloned in the ϕC31 vector. Vectors deleted for the c-gene can accommodate larger inserts, of up to 8 kb, and may be used in the double lysogen system. ϕC31 vectors are useful for the functional analysis of DNA by the insertional inactivation of transcription units (Seno et al 1984) and in mutational cloning (Chater and Bruton 1983). Mutational cloning is a useful technique as it circumvents the need to isolate mutants in order to screen for a cloned gene. The mutations arise when the insert DNA in ϕC31 is completely internal to the transcription unit and recombination between the insert DNA and chromosomal DNA disrupts the transcription unit.

Great advances have been made since the early gene cloning reports and large numbers of antibiotic resistance genes, some nutritional requirement genes, developmental genes, a few genes coding for

Actinorhodin

Medermycin

Mederrhodin A

Dihydrogranticin

Dihydrogranatirhodin
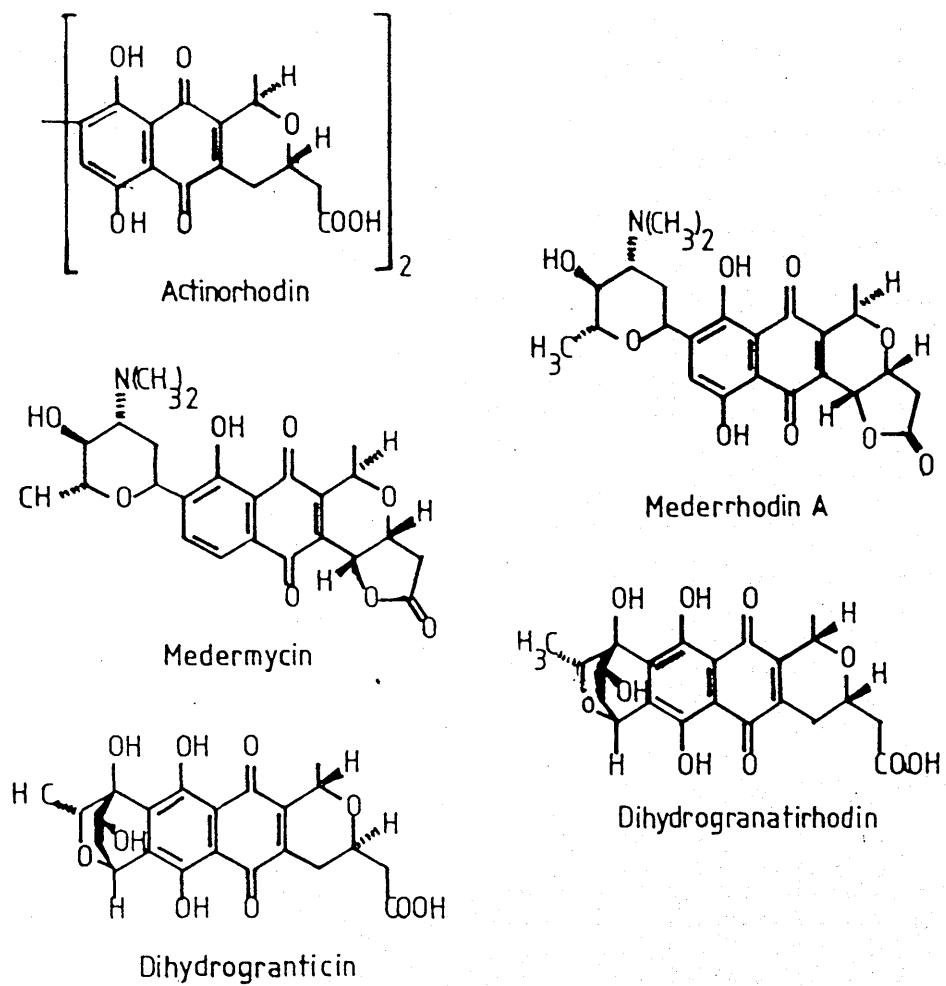
Fig 1.1 Structures of the antibiotics: actinorhodin, medermycin, and dihydrograticin and the hybrid antibiotics mederrhodin and dihydrogranaticin.

for antibiotic production have been cloned. Both bovine and human genes have been expressed in S.lividans. Gray et al (1984) expressed bovine growth hormone using the promoter of the S.fradiae aph gene and Pulido et al (1986) expressed a biologically active form of human interferon ∝2 also controlled by the aph promoter but using the ribosome binding site of E.coli membrane lipoprotein. The expression of heterologous genes in S.lividans indicates the potential of Streptomyces as hosts for the production of commercially useful proteins.

1.3 Characterization of antibiotic pathways and production of novel compounds.

The demand for antibiotics has stimulated antibiotic research in two directions: the characterisation of cloned antibiotic pathways and the production of new antibiotics. Novel structures produced to date have depended on the alteration of existing compounds, for example penicillin G and penicillin V are manufactured by feeding analogs of normal precursors to the production process. More modern methods of producing hybrid structures involve introducing cloned antibiotic pathways or individual genes from antibiotic pathways into strains producing related antibiotics. Medermycin and actinohordin (Act) are structurally related but different compounds and are both classified as isochromane quinone antibiotics. The Streptomyces strain AM-7161, which produces medermycin, was transformed with subclones of the act gene cluster. Some of the transformants analysed continued to produce medermycin and in addition a new compound named, mederrhodin A (Fig 1.1) (Hopwood et al 1985). Another novel product dihydrogranatirhodin was produced in a similar experiment (Hopwood et al 1985). S.violaceroruber, the producer of dihydrogranaticin (an isochromane quinone), was transformed with the complete act biosynthetic pathway. Transformants produced actinohordin and the novel compound dihydrogranatirhodin (Fig 1.1). The novel products mederrhodin A and dihydrogranatirhodin have minor alterations to their structures. Large scale screening programmes would be necessary to identify hybrid compounds with commercially useful structures. The screening factor limits the usefulness of this strategy but a more

act

| VI | Va | II | III | I,VII,IV,Vb |

22 kb

late   early

tcm

?   ?

II IV V   Ia Id VI (Ib) (Ic III)   VII tcm^r

14 kb

middle   early   late
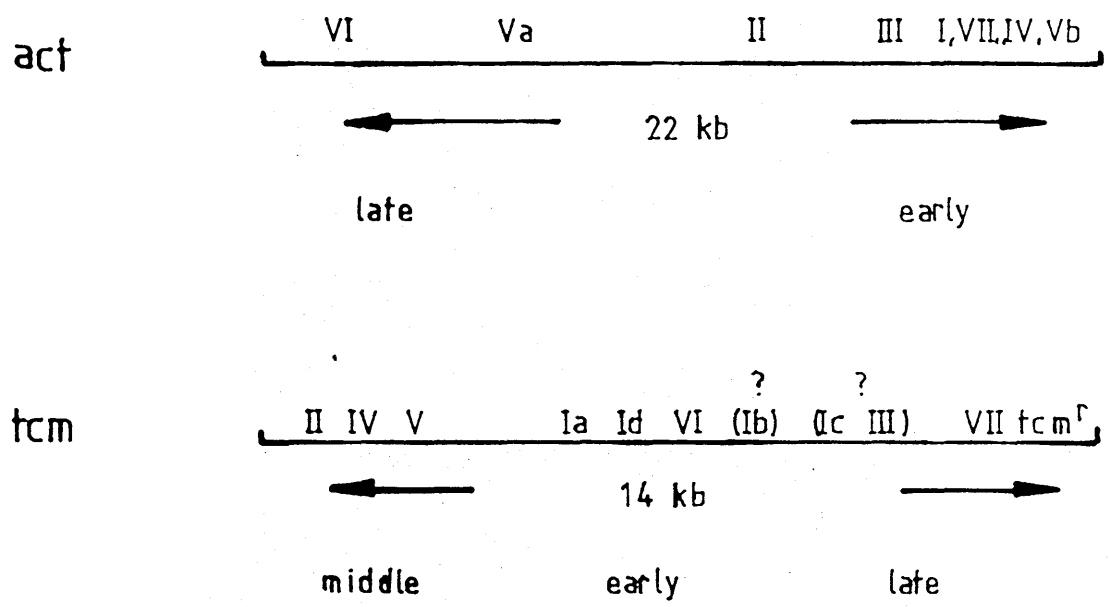
Fig 1.2 Diagram of the actinorhodin and tetracenomycin biosynthetic
gene clusters, displaying the similarities in gene organization.

complete biochemical and genetical understanding of antibiotic biosythesis will identify the most suitable approaches for the directed generation of novel antibiotics and overproduction of existing antibiotics. The general genetic engineering approach discussed above may be refined by altering steps in the biosynthetic pathway e.g. by directed mutagenesis of cloned production genes and manipulation of regulatory mechanisms. This new approach to developing novel antibiotics still depends on existing structures but the versatility of the system should enable it to produce radically different structures in the future.

To date eight complete or partial antibiotic pathways have been cloned i.e. pathways synthesizing actinohordin (Malpartida and Hopwood 1984), methylenomycin (Chater and Bruton 1985), streptomycin (Ohnuki et al, 1985), undecylprodigiosin (Feitelson et al, 1985), biolaphos (Murakami et al, 1986), erythromycin (Stanzak et al, 1986), oxytetracycline (Butler et al, 1986) and tetracenomycin (Motamedi and Hutchinson 1987). The cloning of these pathways has been facilitated by the clustering of the genes for antibiotic biosynthesis. The organization of the gene clusters for actinorhodin (Malpartida and Hopwood 1986) and tetracenomycin biosynthesis (Motamedi and Hutchinson 1987) has been analysed in detail and share some features. The early steps of actinohordin biosynthesis actI, actIII, actVII, actIV and actVb are located on the right side of the cluster and the late genes actVa, actVI on the left side separated by the regulatory gene actII. The early genes of the tetracenomycin biosynthetic pathway tcmIa, tcmId and tcmVi are centrally located and flanked on the left by tcmII, tcmIV and tcmV which mediate reactions in the middle of the pathway and on the right by the late genes tcmVII and tcmIII (Fig 1.2). Preliminary evidence suggests there is a regulatory gene between the early and late genes. Motamedi and Hutchinson (1987) propose that there is a functional advantage to this type of architecture. They suggest that as antibiotic synthesis occurs through a multistep biochemical pathway, sequential expression of the enzymes in the pathway should help efficient synthesis of an antibiotic. Their hypothesis is supported by Malpartida and Hopwood (1986), who demonstrated that the early and late act genes are transcribed by two polycistronic messages (Fig 1.2). The architecture of the OTC

biosynthetic pathway is also very similiar to both the Act and Tcm pathways (Butler et al 1986), supporting the observation that the organization of genes in these pathways has been conserved. These three antibiotics are part of the polyketide group and the proposed polyketide synthase genes of the Tcm and OTC pathways share homology as judged by DNA hybridisation to the actI probe whose gene product encodes the synthase enzyme (Malpartida et al 1987). DNA from many other polyketide producing species hybridise to the actI probe as do some non-producer strains which may contain a cryptic copy of a polyketide antibiotic pathway. These cross hybridization studies will facilitate the cloning of most polyketide antibiotics whether expressed or cryptic and it seems plausible to extend this cloning technology to other antibiotic groups such as B-lactams and aminoglycosides.

1.4 Gene Expression in Streptomyces

Gene expression is controlled by transcription and translation signals, usually 5' to the coding sequence. These control elements have been studied extensively in E.coli and a consensus sequence established which core RNA polymerase (RNAP) recognizes in association with sigma factor-70. Sigma factor-70 is the major factor used to direct transcription from E.coli promoters but it is not the only species. Two other minor species have been identfied. Sigma 32, recognizes a small set of promoters involved in the heat shock response of E.coli (Grossman 1984) and the ntrA gene product is a sigma factor which directs RNAP to bind to the glnA promoter (Hirschman et al 1985). Gene expression may also be positively or negatively controlled by activator or repressor proteins which act on specific or small subsets of genes (Raibaud and Schartz 1984). Five bacterial and two phage encoded sigma factors have been identified in B.subtilis (Losick and Pero 1981). The B.subtilis sigma factor 43 in association with core RNAP recognizes promoter sequences with the same conserved elements as the E.coli consensus sequence recognized by sigma 70. The remaining sigma factors in assosiation with core RNAP all direct transcription initiation from different sequences but very little sequence data is available to establish any real consensus. Many of the sigma factors are expressed during vegetative cell growth

out some e.g. sigma-29 appear during the first hours of sporulation (Grossman and Losick 1986). However the specifc role of these sigma species in vegetative and sporulative gene expression have not yet been investigated.

The first studies on gene expression in Streptomyces demonstrated that the E.coli recA, E.coli lacUV5, S.marcesens trp and the B.subtilis penP promoters were expressed in S.lividans (Bibb and Cohen 1982). The reciprocal experiment i.e. cloning of fragments of Streptomyces DNA in E.coli demonstrated that Streptomyces transcriptional control signals were rarely recognized in E.coli. Jaurin and Cohen (1985) have carried out a more comprehensive analysis of Streptomyces promoters which function in both E.coli and Streptomyces. They analysed the sequence and transcriptional activity of five such promoters, named Streptomyces-E.coli-type promoters (SEP). These promoters were selected for their activity in E.coli and compared well with the wild type E.coli ampC average strength promoter. However transcription activity measurements of SEP in Streptomyces indicate they are poorly expressed. The wild type E.coli ampC promoter has a 20-fold greater expression level in Streptomyces than the SEP. As the SEP -10 and -35 sequences match the E.coli consensus promoter sequence well, the evidence suggests that the SEP sequences have regulatory elements, absent in ampC, which are recognized in S.lividans. The transcriptional start points of three SEP were determined in E.coli and multiple direct repeats and a conserved hexameric sequence were found in the sequence upstream from the start point, which may play a regulatory role. Statistical analysis of the pool of clones which yielded the five SEP analysed, estimates there are approximately 200 SEP in the S.lividans genome. This suggests they form a minority of promoter signals in S.lividans.

Other Streptomyces promoters which are expressed in E.coli include three promoters from the plasmid pIJ101 (Buttner and Brown 1987, Deng et al 1986). Promoters pIJ101A and pIJ101C initiate transcription at the same start point in E.coli and Streptomyces, indicating that the same transcriptional signals are recognized in both genera. Jaurin and Cohen (1984) demonstrated that Streptomyces RNAP responds to up-promoter mutations of the E.coli ampC promoter. However, the degree

9

of increased activity varied from the E.coli results, indicating that the RNA polymerases have different specificties for the same promoter sequence. The evidence suggests that the basic prokaryotic consensus sequence recognized by E.coli sigma factor 70 and B.subtilis sigma factor 43 exists in Streptomyces. However, these sequences are not ubiquitious as in E.coli and the Streptomyces RNAP holoenzyme does not have exactly the same recognition specificity.

Other Streptomyces promoters which are not functional in E.coli show little or no homology to the E.coli consensus sequence. A group of these promoters exhibit homology to the sequence recognized by sigma factor 37 from B.subtilis. Two sigma factors have been identified in S.coelicolor (Westpheling et al 1985). One of these factors specified transcription from the ctc promoter of B.subtilis, usually transcribed by E-sigma 37 and also specified transcription from the endoH promoter of S.plicatus which exhibits homology to the ctc promoter. The second sigma factor specified transcription of the B.subtilis veg promoter, which has the general E.coli consensus sequence. No other sigma factors have been isolated but it appears probable that Streptomyces has an analogous system, of multiple sigma factors, to B.subtilis.

One of the more unusual properties of Streptomyces promoters is the high frequency of tandem promoters and the preponderance of multiple transcription start sites. Many of the genes analysed to date, have two transcription start points separated by up to 300bp. Other genes have multiple transcription start points. The dag promoter exhibits four distinct transcription starts (Buttner et al 1986) and an open reading frame associated with erythromycin resistance gene has three transcription start sites (Bibb and Janssen 1987). This variety of control signals may help to regulate the expression of these genes which code for antibiotic resistance, an extracellular agarase and glycerol and galactose utilization genes. These genes are all good candidates for differential expression.

Translational initiation has also revealed some unusual facets in Streptomyces. Translational initiation depends on an ATG or GTG start codon, and usually the ribosomal complex recognizes a region 5' to the start codon called the ribosome binding site (RBS) which displays

(A)
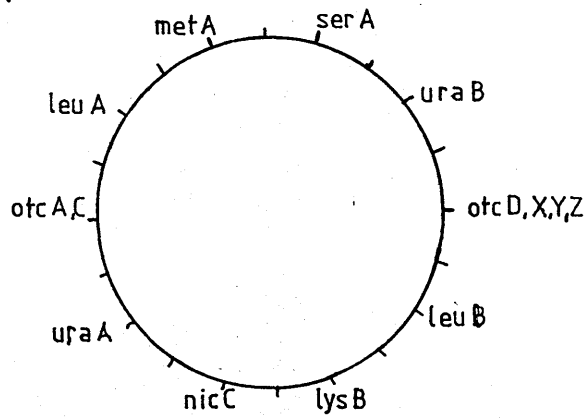
(B)
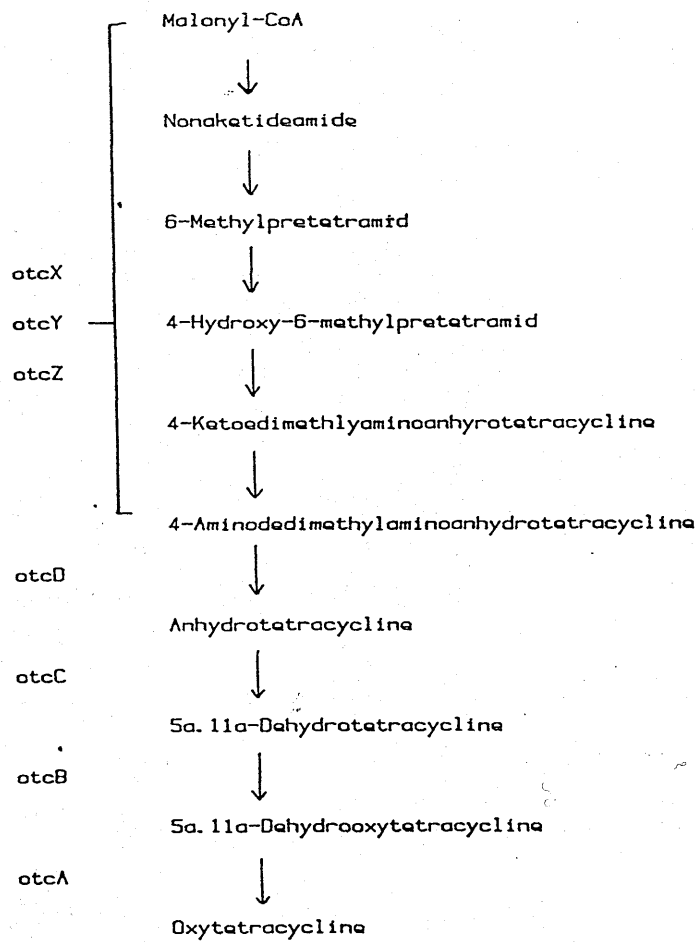
Fig 1.3 (A) Diagram of the biosynthesis of oxytetracycline. The positions of the blocked mutants are indicated.

(B) The location of OTC negative mutations on the genetic map of S.rimosus. The loci are arbitrarily spaced at equal intervals

complementarity to the 3' end of the 16S rRNA. In some species e.g. B.subtilis and Staphylococcus a high degree of homology is required. However, E.coli and Streptomyces are not so stringent in their requirements. The aph and sta genes have identical transcriptional and translational start sites and the ermEP1 initiates trancription one base before the translational start site. The requirement for a ribosome binding site, in these examples, appears to be non-existent but the role of sequences 3' to the translational start site may be important (Scherer et al 1980, Gold et al 1981). This appears to be a rare event amongst other bacterial species and it is difficult to assess its significance in Streptomyces until a larger spectrum of cloned Streptomyces genes have been studied.

1.5 Cloning and characterisation of the oxytetracycline (OTC) biosynthetic pathway from S.rimosus

The pathway has been partially characterised both genetically and biochemically by Rhodes et al (1981). They produced blocked mutants in the pathway by NTG and UV mutagensis and classified these mutants by cosynthesis tests and feeding intermediates of the OTC pathway to blocked mutant strains. These techniques defined the organization of 6 mutants in OTC biosynthesis. Two of the blocked mutants were assigned to specific steps in the conversion of anhydrotetracycline (ATC) to OTC and the other four mutants were blocked before ATC. The order of otcX, otcY and otcZ has not been determined but otcD is blocked in the conversion of 4-amino-ATC to ATC (Fig 1.3). The mutants were mapped to diametrically opposite clusters on the S.rimosus linkage map (Fig 1.3). Two OTC resistance genes map with the early gene cluster.

OTC resistance clones were isolated by screening a gene library made in an OTC sensitive host, S.rimosus M15883 (Rhodes et al 1984). Two unrelated clones were isolated, as judged by restriction analysis and DNA/DNA hybridisation. This suggested that there were two resistance genes and possibly two mechanisms of resistance. Several antibiotic producers have two resistant genes: S.fradiae, a neomycin producer, has two enzymes to inactivate the antibiotic (Davies et al 1979) and S.vinaceus, a viomycin producer, has an inactivating enzyme and

11
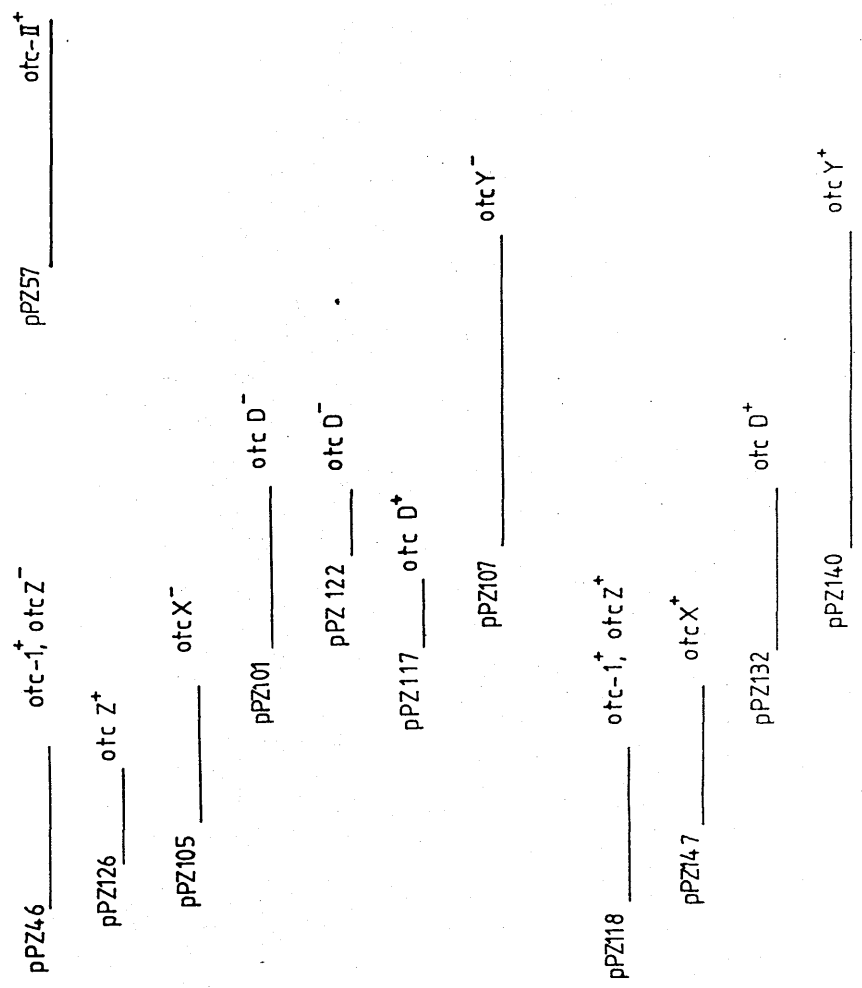
Fig 1.4 Subclones of the 4 o'clock gene cluster and their ability to complement blocked mutants. These high and low copy-number clones indicate the effect of copy-number on complementation of blocked mutants in OTC production.

pPZ15

pPZ24

pPZ19

otc-1   otcZ

otc-1   otcZ

otc-1   otcZ

vector   B - BamH1   Bgl - BglII   C - ClaI   S - SstI

insert   H - HindIII   P - PstI
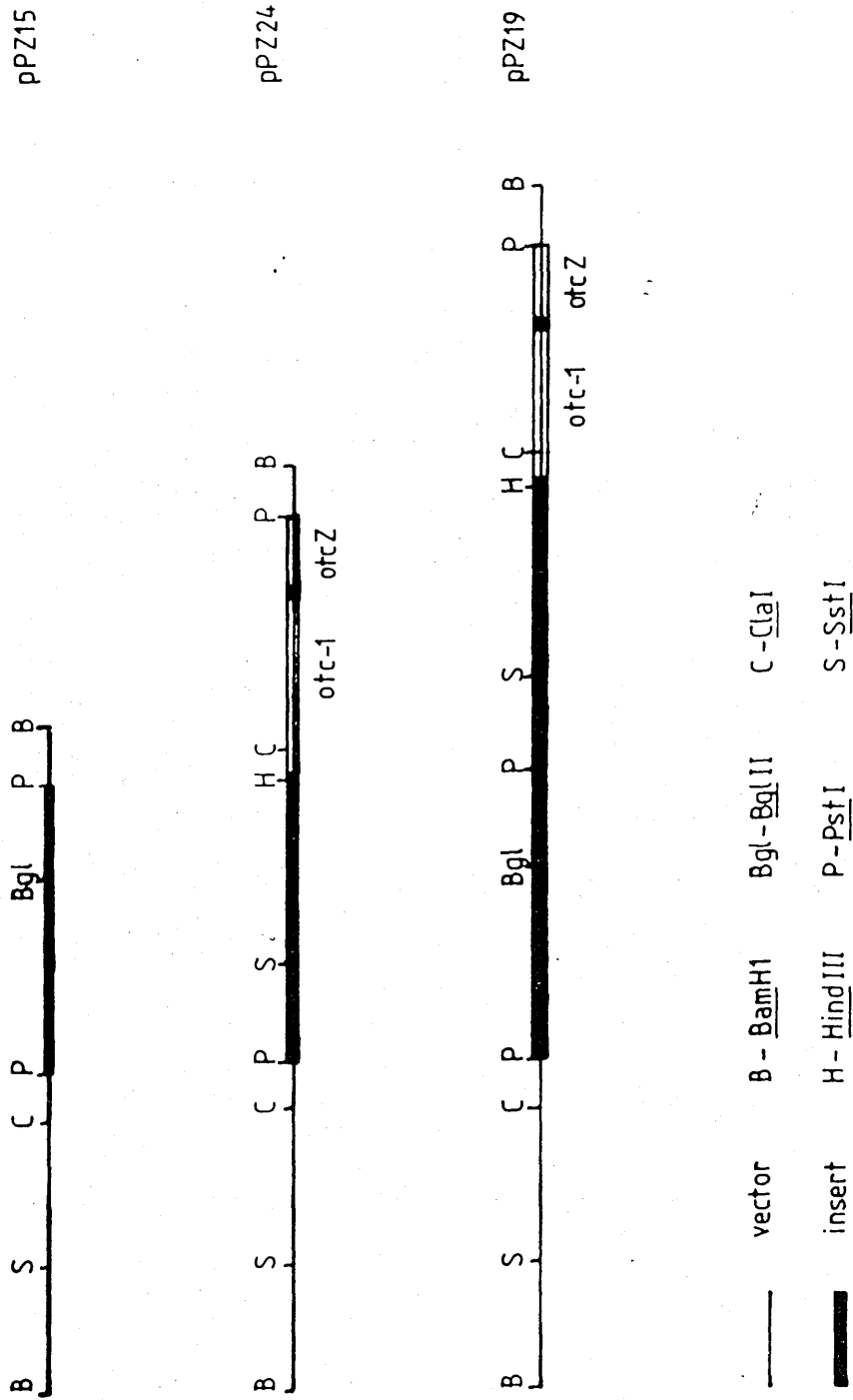
Fig 1.5 Diagram of pPZ15, pPZ24 and pPZ19 which contains the inserts of both pPZ15 and pPZ24. The coding regions of otcZ and otcI are indicated by open boxes with a line through the centre.

S.kanamyceticus, a kanamycin producer (Nakano et al 1984). The resistance genes otcI and otcII are closely linked and also map with the early gene cluster making them useful as probes to clone this cluster. A cosmid library of S.rimosus was constructed in a bifunctional Streptomyces/E.coli cosmid vector pPZ74 (Chambers and Hunter 1984) in E.coli LE392 and probed with otcI (pPZ24) and otcII (pPZ57). Two cosmid clones pPZ499 and pPZ500 were isolated. Analysis of these clones, which are overlapping, mapped the otcI and otcII clones to the ends of the gene cluster (Fig 1.4). Subclones of the cosmid inserts were made in high copy-number vectors and transfomed into a restriction deficient S.rimosus strain (Hunter and Friend 1984). Some of these subclones caused repression of OTC production possibly indicating the presence of regulatory elements which when present on high copy number vectors repress the pathway. When inserts were cloned on low copy number vectors they did not affect OTC production in the restriction deficient S.rimosus and were therefore used to complement the four blocked mutants otc-D4,otc-X20,otc-Y90 and otc-Z151 (Fig 1.4) (Butler et al 1986). Regulatory regions have been defined by further subcloning and are localized between otcZ and otcX and between otcD and otcY (Fig. 1.4). A third regulatory element was isolated with the initial otcI clone. pPZ24 encodes for otcI and otcZ but does not complement otc-Z151 unless the insert is cloned on a low copy number vector. pPZ19 contains the pPZ24 insert and an additional 4.7KB fragment (also cloned independently in pPZ15) which is not contiguous on the chromosome (Fig 1.5). pPZ19 complements the mutant otc-151 and therefore it has been suggested that this 4.7KB insert moderates the effect of the regulatory element found between otcI and otcZ. The 4.7KB fragment also segregates with a brown pigmented phenotype and has not yet been assigned a chromosomal map position.

1.6 Mechanism of tetracycline resistance and tetracyclines mode of action

Two OTC resistance genes have been cloned independently by Ohnuki et al (1985), from a S.rimosus ATCC 10970 strain and by Rhodes et al (1984) from S.rimosus 15883. Drocourt (pers. comm. 1987) has also cloned an OTC resistance gene from S.rimosus and S.aureofaciens.

production of OTC as have the strains used by Drocourt whereas S.rimosus ATCC 10970 is a soil isolate. A comparison of the OTC resistance genes from all strains may identify some effects of mutagenesis. Ohnuki et al (1985) have designated the genes tetA and tetB, which are homologous to otcI and otcII respectively, as judged by restriction analysis. Drocourt (pers. comm. 1987) has named the resistance clone tet347 and it is homologous to tetB and otcII, judged by restriction analysis. Ohnuki et al (1985) have further characterised tetA and tetB according to their level of tetracycline resistance and possible resistance mechanisms in the heterologous host S.griseus. tetA permitted growth at a tetracycline concentration of 250ug/ml, after preexposure to 50ug/ml tetracycline and tetB permitted growth at 300ug/ml also after preexposure to 50ug/ml. Further increased pre-exposure levels did not increase the maximal inhibitory concentration of tetracycline. The parent strain permitted growth at 800ug/ml after pre-exposure to 50ug/ml of tetracycline and this concentration could be raised to 1200ug/ml after pre-exposure to 100ug/ml of tetracycline. Growth was recorded as an increase in the OD_{640} of the culture. The expression of both resistance genes appeared to be induced 2-fold in the presence of tetracycline but it is difficult to relate the concentration of tetracycline adminstered to the culture to the normal levels of intracellular tetracycline present in a cell, which are unknown. Therefore it remains difficult to assess what effect tetracycline has on the expression of the resistance genes in their natural environment.

Analysis of the mechanisms of resistance has indicated that tetA protects protein synthesis from tetracycline and tetB mediates reduced accumulation of tetracycline. Benveniste and Davies (1973) proposed an evolutionary relationship between the aminoglycoside inactivating enzymes in bacteria containing R-factors and similar enzymes found in Streptomyces, suggesting that genetic information for antibiotic resistance has been transferred from streptomycetes to non-antibiotic producing bacteria accounting for their antibiotic resistance phenotypes. According to the Benveniste and Davies (1973) hypothesis the tetB resistant mechanism could be the common ancestor of the tetracycline resistance mechanism employed by many plasmids and

mechanism of these plasmids and transposons is active efflux of the
antibiotic across the membrane. The mechanism is under the control of
a repressor protein which prevents transcription of the tetracycline
resistance gene, from both its promoters. In the presence of
tetracycline, the antibiotic binds to the repressor, outcompeting the
operator sites and releases the resistance gene from repression
(Hillen et al 1984). The mechanism of tetB control has yet to be
established but it is involved in reducing the accumulation of
tetracycline in the cell.

tetA resistance was localized in vitro to the crude ribosome fraction
with some slight activity in the S150 (cytoplasmic) fraction (Ohnuki
et al 1985). However crude ribosomes washed with 1M ammonium chloride
were no longer resistant to tetracycline indicating the resistance was
not mediated by modified ribosomal proteins or rRNA. Antibiotic
resistant protein synthesis has been associated with tight ribosomes
in vitro (Cundliffe and Thompson 1979, Skinner and Cundliffe 1982),
however the data suggests that tetA mediates its resistance through an
unidentified association with the ribosome. Tetracycline resistance
determinants from Streptococcus, tetM and tetN also mediate resistance
at the level of protein synthesis (Burdett 1986). Mikulik et al
(1971) have examined the interaction of tetracycline with the protein
synthesing system of S.aureofaciens, also a producer of tetracycline.
The protein synthesising machinery of S.aureofaciens is more resistant
to tetracycline in vitro than that of E.coli, exhibiting 30%
inhibition of translation at 10ug/ml while E.coli was inhibited 80% at
this tetracycline level. They also observed tetracycline-ribosome
aggregates and the highest level of binding was equivalent to 320
molecules of tetracycline per ribosome. Tetracycline binds all over
the ribosome though much of this binding is reversible. One strong
binding site has been located on the 50S subunit and three on the 30S
subunit (Fey et al 1973). The mode of action of tetracycline is to
inhibit the binding of aminoacyl(AA)-tRNA.EF-Tu.GTP ternary complex to
the 30S.mRNA complex and this implies that the 30S subunit may play a
role in the action of the drug. Williams and Smith (1979) have
isolated chromosomal mutants in B.subtilis causing resistance to
tetracycline. One of these mutations maps to the ribosomal protein

14

altered electrophoretic mobility in the mutant strain. Ribosomes from the mutant strain show an increased level of resistance to tetracycline during in vitro translation. These studies have been followed up by Mikulik et al (1983) who demonstrated that tetracyclne producing strains of S.aureofaciens have altered electrophoretic mobility of the S10, S2 and L19 ribosomal proteins. Reconstitution studies with E.coli 30S ribosomal subunits lacking S10, have shown that these particles bind tRNA less efficently than complete particles (Nomura et al 1969, Lake 1985). S10 and tetracycline both play a role in the binding of AA-tRNA.EF-Tu.GTP to the ribosome, however S10 is required for the complexes to bind and tetracycline prevents binding. The interaction of the 16SrRNA molecule with tetracycline has also been examined (Moazed and Noller 1987). The anticodon of A site bound AA-tRNA is spatially proximal to residues in the 1400 region of 16SrRNA. Experimental results demonstrate that cleavage at residue 1493 is blocked by bound AA.tRNA, as expected and it is also blocked in the presence of tetracycline. This result indicates that tetracycline interacts with the ribosome at some of the sites usually occupied by AA-tRNA. Residue 892 also exhibited strong protection against DMS modification in the presence of tetracycline, the authors suggested that this interaction reflects the folding of the 16S rRNA molecule.

The aims of this project were to characterise the transcriptional control mechanism of the otcI gene and to sequence the gene with a view to gaining a better understanding of its mechanism of resistance.

Promoter probe vectors have been used to identify promoter activity 5' to the otcI structural gene and the transcriptional start sites have been identified by high resolution S1 mapping. The results of these experiments were compared to other sequenced streptomycete promoters and the promoter structure of other streptomycete genes, specifically antibiotic resistance genes.

The otcI gene has been sequenced and compared with the DNA sequence of tetM from Streptococcus, which also confers resistance to ribosomes. The DNA and predicted amino acid sequence have been analysed for any

sites and homology to other genes in an attempt to predict the
mechanism of ribosomal protection.

CHAPTER 2

MATERIALS AND METHODS

2.1 Bacterial Strains

The <u>E.coli</u> strains used were all derivatives of <u>E.coli</u> K12.   The <u>Streptomyces</u> strains used were from Pfizer or John Innes Institute, Norwich.   All strains are listed below.

NAME	GENOTYPE	REFERENCE/SOURCE
JM101	<u>supE</u>,<u>thi</u>, (<u>lac</u> <u>pro</u>A,B)/F`, <u>tra</u>D36, <u>pro</u>A,B, <u>lac</u>IΔZ M15	Yanisch-Perron (1985)
DS903	<u>rec</u>F143, <u>pro</u>A7, <u>str</u>31, <u>thr</u>1, <u>leu</u>6, <u>tsx</u>33, <u>mtl</u>2, <u>his</u>4, <u>arg</u>E3, <u>lac</u>y1, <u>gal</u>K2, <u>ara</u>14, <u>supE</u>44, <u>xyl</u>5	DaveSherratt
DS941	DS903 but <u>lac</u>I , <u>lac</u>ZΔM15, <u>lac</u>Y⁺	Dave Sherratt
CB51	<u>dam</u>⁻, <u>ara</u>, <u>thi</u>, (<u>lac</u>,<u>pro</u>)	Chris Boyd
<u>S.rimosus</u> 4018	derived from a w.t. <u>S.rimosus</u> G7 by strain improvement for OTC production.	Pfizer
<u>S.rimosus</u> M15883S	OTC sensitive, <u>in vivo</u> deletion of M15883 which is derived from <u>S.rimosus</u> G7 by strain improvement for OTC production.	Pfizer
<u>S.lividans</u>	<u>str</u>-6	John Innes Institute

2.2 Plasmids and bacteriophages

The plasmids and bacteriophages used in this study including those whoseconstructionis described in the thesis,are listed inTable 2.1.

2.3 Chemicals

CHEMICALS	SOURCE
General chemicals	B.D.H.,HopkinsandWilliams,

PLASMID/ BACTERIOPHAGE	DESCRIPTION	SOURCE/REFERENCE
pGLW7	2.6kb BamHI/SstI fragment encoding otcI in pT7-2	Iain Hunter
pPZ46	6kb SstI fragment encoding encoding otcI in pPZ34	Iain Hunter
pUC18/19	-	Yanisch-Perron (1985)
M13mp18/19	-	Yanisch-Perron (1985)
pIJ486/487	-	Ward et al (1986)
pDM101	216bp BamHI/BclI fragment from pGLW7 cloned into pUC18	Chapter 3/4
pDM102	357bp BclI/SmaI fragment from pGLW7 cloned into pUC18	Chapter 3
pDM103	573bp BamHI/SmaI fragment from pGLW7 cloned into pUC18	Chapter 4
pDM104	1229bp BclI/BclI fragment from pGLW7 cloned into pUC18	Chapter 4
pDM105	277bp SmaI/SstI fragment from pGLW7 cloned into pUC18	Chapter 3
pDM106	1234bp BamHI/SalI fragment from pGLW7 cloned into pUC18	Chapter 4
pDM112/3	the pDM101 insert cloned into M13mp18/19	Chapter 3
pDM122/3	the pDM102 insert cloned into M13mp18/19	Chapter 3
pDM152/3	the pDM105 insert cloned into M13mp18	Chapter 3
pDM172/3	746bp SmaI fragment from pGLW7 cloned in both orientations in M13mp18	Chapter 3
pDM182/3	1056bp SmaI fragment from pGLW7 cloned in both orentations in M13mp18	Chapter 3
pDM192/3	1496bp BamHI/SphI fragment from pGLW7 cloned into M13mp18/19	Chapter3
pDM1721-3	ExoIII deletion derivatives of pDM172	Chapter 3

PLASMID/ BACTERIOPHAGE	DESCRIPTION	SOURCE/REFERENCE
pDM1731-3	ExoIII deletion derivatives of pDM173	Chapter 3
pDM1821-4	ExoIII deletion derivatives of pDM182	Chapter 3
pDM1831-4	ExoIII deletion derivatives of pDM183	Chapter 3
pDM110/1	the pDM101 insert cloned into pIJ486/7	Chapter 4
pDM130/1	the pDM103 insert cloned into pIJ486/7	Chapter 4
pDM141	the pDM104 insert cloned into pIJ487	Chapter 4
pDM160/1	the pDM106 insert cloned into pIJ486/7	Chapter 4

	Kochlight Laboratories, May and Baker.
Media	Difco, Oxoid
General Biochemicals	Sigma, Pharamacia, B.R.L.
Hydrazine	Kodak
Dimethyl sulphate	Aldridge
Piperidine	Fluka
Agarose	B.R.L.
Acrylamide	Electran
Radiochemicals	N.E.N.
Antibiotics	Sigma, Squibb and Sons

2.4 Culture Media

L-Broth: 10g tryptone, 5g yeast extract, 5g NaCl, 1g glucose, 20mg thymine, made up to 1 litre in distilled water and adjusted to pH7.0 with NaOH

L-agar: As L-broth without glucose and the addition of 15g/l agar

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

Minimal Agar: 7g K_2HPO_4, 2g KH_2PO_4, 4g NH_4SO_4, 0.25 trisodium citrate, 0.1g $MgSO_4.7H_2O$, 17.5g agar made up to 1 litre in distilled water.

Davis and Mingoli (D&M) salts (X4): 28g K_2HPO_4, 8g KH_2PO_4, 4g $(NH_4)_2SO_4$, 1g sodium citrate, 0.4g $MgSO_4.7H_2O$, made up to 1 litre with distilled water.

Supplements: When required supplements were added to the above minimal medium at the following concentrations:

glucose 2mg/ml   thiamine vitamin B1 20ug/ml
amino acids 40ug/ml

Tryptone Soya Broth (T.S.B.): 30g of Oxoid tryptone soya broth powder made up to 1 litre in distilled water.

Yeast extact-Malt extract (Y.E.M.E.): 3g Difco yeast extract, 5g Difco bacto peptone, 3g Oxoid malt extract, 10g glucose, 340g sucrose made up to 1 litre in distilled water. After autoclaving add 1/20 volume of 100mM $MgCl_2$. For preparing protoplasts also add 1/20 volume of 10% glycine.

R2 Medium: R2A- 44g agar, 0.5g K_2SO_4, 20.2g $MgCl_2.6H_2O$, 5.9g $CaCl_2.2H_2O$, 20g glucose, 6g proline, 0.2g casamino acids, 4ml trace elements solution, 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.
Combine equal volumes of R2A (melted and cooled to 55°C) and R2B plus 1ml of 1%KH_2PO_4 prior to use.

R9 Medium: R9A- 20g glucose, 0.5g K_2SO_4, 8.2g $MgCl_2$, 4.7g of $CaCl_2$, 4g sodium nitrate, 1g potassium chloride, 0.4g magnesium sulphate, 0.8g casamino acids, 2ml of 1% $FeSO_4$, 4ml trace element solution, 44g agar, made up to 1 litre in distilled water.
R9B- 11.5g Mops, 10g yeast extract, 410g sucrose, adjusted to pH 7.4 with NaOH and made up to 1 litre in distilled water.
Combine equal volumes of R9A (melted and cooled to 55°C) and R9B prior to use.

Trace Element Solution: 40mg $ZnCl_2$, 200mg $FeCl_36H_2O$, 10mg $CuCl_22H_2O$, 10mg $MnCl_24H_2O$, 10mg $Na_2B_4O_710H_2O$, 10mg $(NH_4)_6MO_7O_244H_2O$, made up to 1 litre in distilled water.

MM Medium: 10g agar, 0.5g l-asparagine, 0.5g K_2PO_4, 0.2g $MgSO_4.7H_2O$, 0.01g $FeSO_4$ made up to 1 litre with distilled water and adjusted to pH 7.0-7.2 with NaOH.

MMT Medium: Supplement 200ml MM medium with 4ml of 30% Difco casamino acids and 1.5ml of the following solution - 1000mg l-arginine, 750mg

l-cystine, 750mg l-histidine, 750mg l-phenylalanine, 750mg l-proline, 150mg adenine, 150mg uracil, 10mg nicotinamide in 100ml of distilled water.

2.5 Sterilization: All growth media were sterilized by autoclaving at $120^{\circ}C$ for 15mins; supplements and buffer solutions at $108^{\circ}C$ and $CaCl_2$ at $114^{\circ}C$ for 10mins. Amino acid and other heat sensitive solutions were filter sterilized.

2.6 Buffers and Solutions:

Electrophoresis

10X TBE Buffer pH 8.3: 109g tris, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$ made up to 1 litre in distilled water, pH is 8.3.

Agarose gel loading buffer: 0.025% bromophenol blue, 0.025% xylene cyanol, 25% ficoll, 0.5% S.D.S., 50mM EDTA.

Sequencing gel loading buffer: 0.1% bromophenol blue, 0.1% xylene cyanol, 10mM Na_2EDTA, 95% formamide (de-ionized with a mixed-bed resin).

DNA Manipulations

10X Core Buffer: 500mM Tris-HCl (pH 8.0), 100mM $MgCl_2$, 500mM NaCl. Stored at $4^{\circ}C$.

10X Sph1 Buffer: 500mM Tris-HCl (pH 7.5), 60mM $MgCl_2$, 500mM NaCl, 60mM 2-mercaptoethanol. Stored at $4^{\circ}C$.

10X Sma1 Buffer: 150mM Tris-HCl (pH 8.0), 60mM $MgCl_2$, 150mM KCl. Stored at $4^{\circ}C$.

10X Ligation Buffer: 660mM Tris-HCl (pH 7.6), 66mM $MgCl_2$, 100mM DTT. Stored at $-20^{\circ}C$.

ATP Stock Solution 10mM: Dissolve 6mg of ATP in 0.8ml of dH_2O. Adjusted to pH 7.0 with NaOH and adjusted the volume to 1ml with dH_2O. Dispensed the solution into 10ul aliquots and stored at -70^oC.

TE Buffer: 10mM Tris-HCl (pH 7.0-8.0), 1mM EDTA (pH 8.0).

TES Buffer: 20mM Tris-HCl (pH 7.5), 10mM NaCl, 0.1mM EDTA

10X Sequencing Buffer: 70mM Tris-HCl (pH 7.5), 70mM $MgCl_2$. Stored at -20^oC.

10X Annealing Buffer: 100mM Tris-HCl (pH 7.5), 100mM $MgCl_2$.

10X Kinase Buffer: 500mM Tris-HCl (pH 7.6), 100mM $MgCl_2$, 50mM DTT, 1mM spermidine, 1mM EDTA. Store at -20^oC.

10X Klenow Buffer: 100mM Tris-HCl (pH 7.5), 100mM $MgCl_2$, 500mM NaCl. Stored at 4^oC.

1X Exonuclease III Digestion Buffer: 66mM Tris-HCl (pH 8.0), 0.66mM $MgCl_2$. Stored at 4^oC.

1X Exonuclease III Stop Buffer: 0.2M NaCl, 5mM EDTA (pH 8.0). Stored at 4^oC.

Low Salt Buffer: 0.2M NaCl, 20mM Tris-HCl (pH 7.4), 1mM EDTA.

High Salt Buffer: 1M NaCl, 20mM Tris-HCl (pH 7.4), 1mM EDTA.

Maxam and Gilbert Sequencing Solutions

DMS Buffer: 50mM Sodium cacodylate (pH 8.0), 1mM EDTA.

DMS Stop Solution: 1.5M NaOAc (pH 7.0), 1M mercaptoethanol, 100ug/ml tRNA.

Hydrazine Stop Solution: 0.3M NaOAc, 0.1M EDTA, 25ug/ml tRNA.

S1 Nuclease Mapping Solutions

Hybridization Solution: 40mM PIPES adjusted to pH 7.4 with NaOH, 400mM NaCl, 1mM EDTA, 80% formamide (de-ionized with mixed-bed resin). Stored at -70°C.

5X Digestion Buffer: 280mM NaCl, 30mM NaOAc (pH 4.4), 4.5mM $Zn(OAc)_2$. Stored at -70°C.

Termination Solution: 2.5M NH_4OAc, 50mM EDTA. Stored at 4°C.

Phenol: All phenol used in the purification of DNA or RNA contained 0.1% 8-hydroxyquinoline and was buffered against 0.5M Tris-HCl (pH 8.0)

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

Birnboim Doly I (BDI): 50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA. Add lysozyme to 1mg/ml immediatley before use, if necessary.

Birnboim Doly II (BDII): 0.2M NaOH, 1% S.D.S., stored in a plastic container.

Birnboim Doly III (BDIII): 5M KOAc (pH 4.8); mix equal volumes of 3M CH_3COOK and 2M CH_3COOH, pH should be 4.8.

STET Buffer: 8% Sucrose, 5% Triton X-100, 50mM EDTA, 50mM Tris-HCl (pH 8.0).

PEG Solution: 20% PEG 8000, 2.5M NaCl.

Kirby Mix: 1g Tri-iso propylnapthalenesulphonate, 6g 4-amino salycilate (Na salt), 50mM Tris-HCl (pH 8.3), 6mls phenol, made up to 100mls in distilled water.

<u>Streptomyces</u> Transformation Solutions:

Transformation Mix (T-Mix): 2.5% sucrose, 100mM $CaCl_2$, 2.5mM K_2SO_4, 1ml trace elements, 50mM maleic acic, adjust to pH 8.0 with 1M Tris. Add 25% (w/v) PEG 1000 before use.

Medium P: 5.73g TES, 103g sucrose, 2.93g $MgCl_2.7H_2O$, 0.5g K_2SO_4, 3.68g $CaCl_2.2H_2O$, 2ml trace element solution. Adjust to pH 7.4 with NaOH and make up to 1 litre in distilled water.

Lysozyme Solution: 10% sucrose, 25mM TES buffer (pH 7.2), 2.5mM K_2SO_4, 2ml trace elements, 2.5mM $MgCl_2$, 2.5mM $CaCl_2$. Add 0.3mg/ml lysozyme and 0.5ml of KH_2PO_4/100ml solution immediatley before use.

2.7 Antibiotics and Indicators: The antibiotic concentrations used throughtout for both liquid and plate selection were as follows:

NAME	SELECTIVE CONCENTRATION	STOCK SOLUTION	STORAGE TEMPERATURE
Ampicillin	50ug/ml	20mg/ml (water)	-20°C
Kanamycin	5-250ug/ml	20mg/ml (water)	-20°C
Oxy-tetracycline	50ug/ml	5mg/ml (10mM HCl)	fresh
Thiostrepton	25ug/ml or	10mg/ml (DMSO)	4°C

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

X-gal (5-bromo-4-chloro-3-indolyl-B-galactosidase) was used in conjunction with the host strains JM101 and DS941 and the pUC and M13mp vectors providing a screen for plasmids with inserts in the polylinker region. Clones containing inserts are generally white and

24

clones lacking inserts are blue. X-gal was stored at a concentration of 20mg/ml in DMF at -20° and added to L-agar plates to a final concentration of 20ug/ml.

2.8 E.coli Growth Conditions:

Liquid culture for transformation and DNA preparations were routinely grown in L-broth at 37° with vigourous shaking. JM101 liquid cultures for ssDNA preparations were grown in 2YT at 37° in a rollerdrum for 5-6hrs.

Growth on plates was on L-agar or minimal media plus supplemnts; antibiotics were added as required. Plates contained 25mls of agar solution and were incubated overnight. All dilutions were carried out in L-broth. Strains were stored in 20% glycerol and 1% peptone at -70°C. Innocula from these stocks were streaked out on L-agar plates, incubated and single colonies were re-streaked on selective plates.

2.9 Streptomyces Growth Conditions:

Liquid culture for protoplasting, DNA and RNA preparations were grown in Y.E.M.E. for S.lividans strains and T.S.B. for S.rimosus strains. Cultures were incubated at 30°C with vigourous shaking for approximatley 2 days. The cultures were innoculated from slopes stored at -20°C or spore suspensions also stored at -20°C.

Strains were stored on agar slopes (15mls) which were germinated for 4-5 days and then frozen at -20°C. Spore suspensions in 80% dH^{2}O and 20% glycerol were also stored at -20°C.

Growth on protoplast regeneration plates was at 30°C and antibiotics were overlayed on plates after 16-24hrs of incubation. Plates were further incubated for 3-4 days. S.lividans protoplasts were regenerated on R2 agar plates and S.rimosus on R9 agar plates.

2.10 E.coli in vivo Techniques:

Ethanol precipitation of DNA: The DNA solution was made 0.3M in NaOAc and 2 volumes of cold ethanol added. After mixing, the DNA was precipitated by cooling on ice for up to 1 hour and pelleted by centrifugation (27000g, 15min, 4°C for large volumes or 12000g, 15min, 4°C for small volumes in eppendorf centrifuge tubes). The pellet was usually washed in 70% ethanol and dried briefly in a vacuum drier.

Restriction of DNA: Restrictions were usually performed in a total volume of 20ul containing between 0.25ug and 1ug of DNA, 2ul of 10X restriction buffer and 1 unit/ug DNA of enzyme, the volume being made up with distilled water. For larger scale restrictions 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 rapid cooling on ice.

Ligation of DNA fragments: The restriction fragments to be ligated were mixed such that the insert was in 3 times excess over the vector (10 times excess for blunt end ligations) and made up to 20ul by the addition of 2ul 10 X ligation buffer, 2ul 4mM ATP and distilled water. T4 DNA ligase was added (0.01 units/ug DNA for 'sticky' end ligation and 1 unit/ug DNA for 'Blunt' end ligation) and the solution ligated for 1 hour at room temperature (overnight at 16°C for 'blunt' end ligations). Aliquots of the ligation mix were used to transform competent cells.

Calf Intestinal Phosphatase (CIP) treatment: To increase cloning efficiency, CIP was used to remove the 5'-terminal phosphate groups from the linearised vector to prevent recircularisation of the vector. This technique is particularly useful if there is no direct selection for the insert. CIP works in core buffer and was used at a concentration of 1-2 units/ug DNA. It was added directly to the restriction digest for the final 15min of the incubation and was heat killed in the manner described for restriction enzymes. If the DNA was been prepared for kinase end labelling, it was also phenol extracted.

Transformation with plasmid DNA: Plasmids were introduced to different strains by genetic transformation. An overnight culture of the recipient was diluted 1 in 100 into 100ml L-broth and was grown to a density of approximately 10^8 cells/ml (about 90min - 2 hours). The cells were harvested (12000g, 5min, 4^oC) and resuspended in 50ml of cold 50mM $CaCl_2$. The cells were pelleted again, resuspended in 5ml of cold 50mM $CaCl_2$ and kept on ice for at least 15min before use. 200ul aliquots of the competent cells were added to the plasmid DNA, mixed gently and left on ice for up to 1 hour. The cells were heat shocked (2min, 42^oC) and returned to the ice for a further 15min. The cells were plated out on the appropriate selections. For transformation to ampicillin resistance, no expression time was necessary. Unused cells were frozen at -70^oC in 20% glycerol.

2.11 Streptomyces in vivo Techniques:

Preparation of Protoplasts: 25ml cultures were grown in the appropriate medium for 30-36hrs. Mycelia were pelleted at 12100g for 10 mins and washed twice in 10.3% sucrose. The mycelia were resuspended in 4ml of lysozyme solution and incubated at 37^oC for 15-30 mins. The mycelia were examined microscopically to monitor protoplast formation. Add 5mls of P medium, triturate a couple of times and filter the protoplasts through cotton wool. Pellet the protoplasts at 12100g for 10 mins and wash twice in P medium. Resuspend protoplasts in 4mls, dispense into 200ul aliquots and freeze at -70^oC.

Transformation of Protoplasts: Add DNA to protoplasts and within 30secs add 0.5mls T-mix. Dilute protoplasts in medium P to 2mls after 10sec. Plate out on regeneration plates which have been dried overnight.

2.12 In Vitro Techniques:

Plasmid DNA preparation: Two methods were used to obtain DNA from

cells.

Birmboin and Doly (1979) DNA preparation: 200 ml cultures of stationary phase cells were harvested by centrifugation (12000g, 5min at 4°C). The pellet was resuspended in 4ml of Birmboin-Doly I solution and incubated on ice for 5min for _E.coli_ cultures and incubated at 37°C for 15min with 100ug/ml lysozyme for _Streptomyces_ cultures. 8ml of Birmboin-Doly II solution were added and the solution left on ice for 5min before 6ml of cold Birmboin-Doly III solution was added, gently mixed and left on ice for a further 5min. The cell debris and most chromosomal DNA was removed by centrifugation (32000g, 5min at 4°C) and the plasmid DNA precipitated by addition of an equal volume of isopropanol followed by centrifugation at 39200g for 15min. This DNA was further purified by banding on a CsCl/EtBr gradient. The DNA was resuspended in 1ml of TE and 4.5g of CsCl dissolved in 3.5ml of TE. The DNA and CsCl solutions were added together with 250ul of EtBr (10mg/ml). The gradients were centrifuged in a Beckman VTi65 vertical rotor at 289,000g for 4hrs at 20°C. Two band were visible, 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 butanol extractions (using water saturated butanol). The salts were removed by dialysis in 2 X 500ml 1 X TE. The DNA was then ready for use.

Mini DNA preparation using the method of Holmes and Quigley (1981): 1.5ml of an overnight culture containing the plasmid of interest was harvested by centrifugation in a 1.5ml eppendorf tube and resuspended in 350ul of STET buffer. 25ul of STET buffer containing lysozyme at a concentration of 10mg/ml was added and the tube vortexed briefly. This solution was boiled for 40sec and centrifuged in an eppendorf microfuge for 15min at 4°C. The pellet was discarded using a toothpick and 40ul of 3M NaAc and 400ul of cold isopropanol was added, followed by microcentrifugation for 7min which precipitated the nucleic acid. The pellet was washed twice in 70% ethanol and dried briefly in a vacuum drier before being resuspended in 20ul - 50ul 1 X TE. This DNA was suitable for digestion and other _in vitro_

manipulations.

Total RNA Preparation (Hopwood et al 1985a): RNase is a very persistent enzyme and precautions were taken against contamination of equipment and solutions. All glassware was incubated overnight at 300°C. Distilled water was treated with DEPC (0.1% of total volume) overnight and autoclaved. All solutions were made up from DEPC treated water and using previously unused chemicals.

100ml cultures of <u>Streptomyces</u> mycelia were harvested by filtration. The mycelia were resupended in 12g of 0.45mm glass beads, 5mls of Kirby mix and vortexed for 2 mins and centrifuged at 6000g for 5 mins. The supernatant was poured off into an equal volume of phenol/chloroform, vortexed and centrifuged at 6000g for 5 mins. The aqueous phase was removed and again added to an equal volume of phenol/chloroform, vortexed and centrifuged at 6000g for 5 mins. The aqueous phase was removed and precipitated in an equal volume of isopropanol and 1/10 volume of 3M NaOAc. The isopropanol solution was centrifuged at 12100g for 5 mins and the pellet resuspended in 1ml of DEPC treated dH_2O.

The RNA preparation at this stage consists of total nucleic acid. The DNA and small ribosomal RNA species were largely removed from the preparation by sucessive salt precipitations. 3mls of 4M NaOAc were added to 1ml of the RNA preparation, incubated at -20°C for 1hr and centrifuged at 11400g for 5mins. The pellet was washed in 3M NaOAc, centrifuged at 11400g for 5mins, washed in 100% EtOH and re-centrifuged at 11400g for 5mins. The pellet was dried and resuspended in 1ml of DEPC treated water and the salt precipitation repeated. The final resuspended pellet was added to 1ml of isopropanol and stored at -70°C.

The concentration and purity of the RNA preparation were checked by taking O.D. readings at 260nm and 280nm.

1 O.D.(260nm)= 40ug/ml

260/280 = 2.1   Indicates a preparation almost completely free of protein.

Preparation of Labelled DNA Probes:

Filling recessed 3' Ends of Double Stranded DNA with Klenow Fragment of E.coli DNA Polymerase I:

The DNA was restricted with the appropriate enzyme, precipitated and resuspended in dH_2O. The reaction contained DNA at a concentration of 1-500ug/ml, unlabelled 1mM nucleotides, 25-50uCi of labelled nucleotides, 1/10 volume of 10X klenow buffer and 1U/ug DNA of Klenow. The reaction was incubated at r.t. for 15mins and heat killed at 70°C for 10mins.

Labelling the 5' Ends of DNA with T4 Polynucleotide Kinase:

The DNA was restricted at the appropriate restriction site, phosphatased, phenol extracted, precipitated and resuspended in dH_2O. The reaction contained DNA (1-50pM of 5' ends), 50uCi of $<^{32}P>ATP$, 20U T4 polynucleotide kinase and made up to a final volume of 50ul with dH_2O. The reaction was incubated at 37°C for 30mins, stopped by the addition of 200ul of 2.5M NH_4OAc, 1ul tRNA (1mg/ml), 750ul 95% EtOH and precipitated. The pellet was washed in 70% EtOH, dried and resuspended in dH_2O.

2.13 Gel Electrophoresis

Both DNA and RNA were visualized on horizontal neutral agarose gels. 0.8% (w/v) gels were most commonly used and 1-2% (w/v) gels were used for fragments of <1.5KB.

Mini gels were used for fast analysis of DNA digested with restriction enzymes or precipitated. A BRL model H6 gel kit was used. 0.16g agarose was added to 20ml of TBE, boiled and cooled to 60°C. 200ng/ml

EtBr was added and the molten agarose poured into a 7.6cm X 5.1cm gel caster and a 8 well slot former (4.1 X 0.8 mm wells) was placed in position. After the gel had set the slot former was removed and the gel paced in the tank with 500ml of TBE. The gel was electrophoresed for 30-45mins at 100V. The separated DNA molecules were visualized on a 254nM UV transilluminator.

Full size gels were used to isolate and size DNA molecules. A Yale gel tank (23cm X 12.5cm) was used. 150-200mls of agarose made up in TBE as before, was poured into the gel apparatus with the required slot former in place. When the gel was set, the slot former was removed and TBE overlayed on the gel to prevent it drying out during electrophoresis. The gel was loaded with samples and electrophoresed overnight at 70V or at 200V for 3-4hrs. The seperated DNA molecules to be isolated from the gel were visualized with a 302nM hand held UV lamp or on the 254nM UV transilluminator.

Alkaline agarose gels: 2.4g of agarose was added to 200mls of 50mM NaCl, and 1mM EDTA. The solution was boiled and cooled to 50°C before pouring the gel. After the gel had set it was soaked in alkaline gel electrophoresis buffer (30mM NaOH, 1mM EDTA) for 30mins. Samples to be electrophoresedwere resuspended in alkaline loading buffer (50mM NaOH, 1mM EDTA, 2.5% Ficoll, 0.025% bromocresol green). The gel was electrophoresed at 70V overnight.

Photographing Gels:

EtBr gels were viewed on a 254nM UV transilluminator and photographed using a Polaroid MP.4 Land camera and Polaroid Type 667 film or Polaroid pos-neg film. The camera was fitted with a Kodak Wratten filter no.9.

Sizing of Restriction Fragments:

The size of linear restriction fragments was estimated from graphs of the $\log_{10}$ molecular size plotted against the distance travelled of the

31

standards on the gel, according to the formula

$$logM = C \times 1/D \qquad \text{(Helling et al 1974)}$$

M = Molecular size in base pairs

C = Constant

D = Distance Migrated

Molecular weight standards were obtained by restriction of lambda cI857Sam7 (Philippsen et al 1978; Haggerty and Sheif 1976), the pUC plasmids (Yanisch-Perron et al, 1985) and OX174 RF DNA (Sanger et al, 1977).

Sequencing Gels: 6% (w/v) denaturing polyacrylamide gels were most commonly used and made up from the following stocks- 40% acrylamide stock (acrylamide: bisacrylamide ; 19:1), urea, 10X TBE.

<u>6% Acrylamide Gel</u>   <u>6% Acrylamide, 30% Formamide Gel</u>

6% Acrylamide Gel	6% Acrylamide, 30% Formamide Gel
15ml 40% acrylamide	15ml 40% acrylamide
54g urea	54g urea
10ml 10X TBE	10ml 10X TBE
	30ml formamide

Made up to a final volume of 100mls in dH_2O. The urea was dissolved by heating at $37^{o}C$ and the sequencing gel solution stored at $4^{o}C$. Before pouring the gel 300ul of 10% APS and 50ul of TEMED were added to 50ml of the stock solution.

Prerparation of Glass Plates and Pouring the Gel: The plates (40cm X 20cm) were cleaned thoroughly with alcohol and water and assembled using two (40cm X 20cm X 0.4mm) spacers and taped together. The gel solution was poured from a beaker down one edge of the plates while tilting the plates at an angle of about $30^{o}C$. The gel was laid at an angle of $5^{o}C$ and the comb inserted to the top of the well teeth. The gel polymerized at R.T. usually within 30mins but formamide gels took

at least 1hr.

Electrophoresis of Sequencing Gels: The gel was pre-electrophoresed for 30min at a constant power of 40W. Prior to loading the samples, they were heated to 95°C for 5mins, placed on ice and loaded on to the gel. 6% gels were electrophoresed for 1.75-2hrs to read the first 100 nucleotides and for 4.5-5hrs to read up to 400 nucleotides. 6% (w/v) polyacryamide 25% (v/v) formamide gels were electrophoresed for 2.5-3hrs to read the first 100 nucleotides on a gel.

Isolation of DNA from Agarose Gels:

The DNA fragments isolated were visualized with a hand held 302nM UV lamp and a small well the width of the band and 1-0.5cm long was cut out of the agarose just below the DNA band. A piece of dialysis tubing (washed in TBE) the width of the band and 2.5cm long was used to line the well and secured in place by tucking it underneath the gel. The well was filled with TBE and the gel electrophoresed at 200V for 10-20mins. The power was shut down and the TBE in the well and the dialysis tubing removed into an eppendorf. The dialysis tubing was washed down with the TBE to remove DNA adhering to it and removed from the eppendorf.

Removing Impurities from DNA after Isolation from Agarose: 2mls of high salt buffer was passed through an elutip, followed by 5mls of low salt buffer. The DNA in TBE was passed through the elutip , followed by 3mls of low salt buffer and then 0.2mls of high salt buffer. 0.2mls of dH_2O was added to the eluate from the high salt wash and precipitated with two volumes of EtOH. The sample was resuspended in a small volume of dH_2O, 10-50ul of dH_2O. The yield was checked by looking at the sample on a mini-gel.

2.14 DNA Sequencing:

M13 dideoxy sequencing (Sanger et al 1977a) and Maxam and Gilbert (1977) chemical sequencing methods were both used.

2.14.1 M13 Dideoxy Sequencing:

Preparation of s.s. Template DNA: An overnight JM101 culture diluted 1:10 in 2TY was innoculated with a white plaque and grown at $37^{O}C$ for 5-6hrs.   The culture was transferred to a 1.5ml eppendorf and centrifuged in a microfuge for 5mins.   The supernatant was transferred to another 1.5ml eppendorf and re-centrifuged.   The supernatant was transferred again to a 1.5ml eppendorf containing 300ul of PEG solution, vortexed and let stand at R.T. for 15mins.   The eppendorf was centrifuged for 10mins and the supernatant discarded.   The eppendorf was respun to remove all traces of PEG and the viral pellet resuspended in 100ul of TES buffer.   50ul of phenol was added to the sample which was vortexed for 15-20secs and centrifuged for 3mins. 90ul of the aqueous phase was transferred to another eppendorf, extracted with 50ul of chloroform, centrifuged for 3mins and 70ul of the aqueous phase removed to a fresh eppendorf to be precipitated. The s.s.DNA template was resuspended in 20ul of dH_2O.

Annealing Primer to Template DNA:

The annealing reaction is performed in a 500ul eppendorf, placed in a water filled test tube which was in a small water bath at $95^{O}C$.   The reaction was incubated for 5mins, the water bath turned off and the temperature allowed to equilibrate to R.T..   Oligonucleotides specific to insert sequences were annealed at $60^{O}C$ for 2hrs.

Annealing reaction mix:

5ul	s.s.DNA template
1ul	M13 primer (2ng)
1.5ul	10X klenow buffer
2.5ul	dH_2O

Nucleotide Stocks for Sequencing Reactions: dNTP and ddNTP solutions were made up from 10mM stock solutions. The dNTP mixes (N^{O}) were made up as follows:

	A^O	C^O	G^O	T^O
0.5mM dCTP	20ul	1ul	20ul	20ul
0.5mM dGTP	20ul	20ul	1ul	20ul
0.5mM dTTP	20ul	20ul	20ul	1ul
10X polymerase	20ul	20ul	20ul	20ul
reaction buffer				

The ddNTPs were made up at the following concentrations; 0.1mM ddATP, 0.4mM ddTTP, 0.1mM ddCTP, 0.1mM ddGTP. Equal volumes of N^O mixes and ddNTP were combined e.g. 10ul A^O mix and 10ul ddATP were added together, stored at -20^OC and termed the A mix.

Sequencing Reaction: 2ul of A mix was added to the A reaction tube and so forth for the T, C and G reaction tubes. 2ul $-^{35}$S dATP, 1ul 0.1mM DTT and 1ul of Klenow fragment (1.5U/ul) were added to the annealed primer template sample, mixed gently and 3ul aliquots dispensed to each of the reaction tubes A, T, C and G which contained the appropriate N mix. The reactions were incubated at 37^OC for 20mins, 1ul of 0.5mM dATP was added to each tube and the incubation continued for a further 15mins. The reaction was stopped by the addition of 4ul of formamide dye, if the sample was to be electrophoresed immediatley, otherwise the sample was stored at -20^OC.

2.14.2 Maxam and Gilbert Chemical Sequencing:

^{32}P end labelled DNA fragments (about 10^4 cpm) were used in the reactions. The four reactions employed cleaved the DNA specifically at 1) G residues 2) G+A residues 3) C+T residues 4) C residues. The modification and cleavage steps are tabulated below.

A+G	G	C+T	T
1ul carrier DNA	200ul S.C. 50mM	10ul dH$_2$O	15ul 5M NaCl
(1mg/ml)	5ul ^{32}P-DNA	10ul ^{32}P-DNA	5ul ^{32}P-DNA
8ul Sod. Cit. 50mM			

| 10ul ^{32}P-DNA | Add 1ul DMS | Add 30ul HZ | Add 30ul HZ |
| 4ul dH$_2$O | | | |

| 80°C for 8min | R.T. for 1.5min | R.T. for 2.5min |
| | Add 50ul DMS stop | Add 200ul HZ stop |

250ul 0.3M NaOAc	750ul 95% EtOH (0°C)
2ul tRNA (1mg/ml)	
750ul EtOH	
precipitate	precipitate
1ml EtOH wash	250ul 0.3M NaOAc, 750ul 95% EtOH
	precipitate
	1ml EtOH wash
vaccum dry	vaccum dry

Add to all tubes 100ul of 1M piperidine, resuspend pellet and heat at 90°C for 30mins. Add 500ul EtOH, 20ul 3M NaOAc, 5ul carrier DNA, 100ul dH$_2$O and precipitate. The pellet was washed in 70% EtOH, vaccum dried and resuspended in formamide dye.

2.15 S1 nuclease mapping (Hopwood et al 1985a):

Hybridisation: 1-100ng of DNA (in a small volume) was added to 40ug of RNA, stored as an isopropanol suspension (50% isopropanol, 0.3M NaOAc). The sample was centrifuged for 10min and the pellet dried in a vacuum desicator. The pellet was resuspended throughly in 10ul of hybridisation solution and incubated for 10min at 85°C in a dri-block. The temperature was reduced to the appropriate hybribisation temperature (5°C above the T$_m$[DNA/DNA]) and the samples incubated for 3 hours.

S1 nuclease digestion: 300ul of chilled (0°C) S1 nuclease digestion solution containing 50-100 units of S1 nuclease was added to each sample. The samples were vortexed and placed on ice imediately. After processing, all the samples were incubated at 37°C for 30min.

36

75ul of S1 termination solution and 7.5ul of tRNA (1mg/ml) was added to each sample and mixed by vortexing. 400ul of isopropanol was added and the nucleic acid precipitated by centrifugation for 10mins after being chilled on ice for 5min. The pellets were vacuum dried and resuspended in the appropriate gel loading buffer prior to electrophoresis on alkaline agarose gels or denaturing polyacrylamide gels.

2.16 Unidirectional digestion with exonuclease III to create targeted deletion derivatives of M13 clones

10ug of double stranded RF DNA was digested with PstI and BamHI. The sample was precipitated and resuspended in 50ul of ExoIII digestion buffer. 67 units of ExoIII were added and digestion proceded at room temperature. 2.5ul aliquots were removed at 3min intervals and mixed with 7.5ul of ExoIII stop buffer and incubated at 70°C for 10min. The samples were precipitated, dried and resuspended in 50ul S1 digestion buffer containing 15 units of S1 nuclease and incubated at 37°C for 30min. 6ul of S1 termination buffer was added to each sample and the aliquots extracted with 25ul Phenol followed by 25ul Chloroform. The samples were precipitated and resuspended in 10ul of a 10mM Tris.HCl (pH8.0), 7mM $MgCl_2$ solution. 0.1 unit of Klenow was added and the samples incubated for 10min at 37°C. 1ul of a 0.1mM mix of the four deoxyribonucleotideswas added and the incubation continued for a further 2min. 40ul of 1X ligation buffer with 1 unit of T4 DNA ligase was added and incubated at room temperature for 3-6 hours. 20ul of each time point sample was transformed into DS941 competent cells . Plaques resulting from the transformation were screened for deleted derivatives. The ExoIII digestion was found to proceed at about 50bp/min.

CHAPTER 3

Sequence of the BamHI/SstI fragment which encodes otcI

Two OTC resistance genes have been cloned from S.rimosus, otcI and otcII (Rhodes et al 1984). Alleles of these genes have been cloned independently by Ohnuki et al (1985) and named tetA and tetB, equivalent to otcI and otcII respectively. A third group have also cloned an OTC resistance gene from S.rimosus, named tet347, which is equivalent to otcII and tetB (Drocourt pers. comm. 1987). tetA and otcI were shown to mediate tetracycline resistance at the level of protein synthesis (Ohnuki et al 1985, Doyle et al 1986) and tetB functions by reducing the accumulation of tetracycline in the cell (Ohnuki et al 1985). Studies of tetracycline resistance in gram negative bacteria have revealed five classes of tetracycline resistance determinants as judged by DNA/DNA hybridisation and levels of tetracycline and minocycline resistance. However all classes mediate tetracycline resistance by active efflux of the drug out of the cell (Mendrez et al 1980). Three other resistance determinants designated tetL, tetM and tetN have been identified in Streptococcus. They can also be distinguished by DNA/DNA hybridisation and levels of sensitivity to tetracycline antibiotics (Burdett et al 1982). The tetL determinant probably mediates resistance to tetracycline by active efflux as cells containing this determinant do not accumulate tetracycline (Burdett 1986). Both tetM and tetN determinants have no effect on the accumulation of tetracycline in cells but their protein synthesizing machinery is resistant to tetracycline (Burdett 1986). The tetM determinant has been sequenced and encodes a protein with a predicted molecular weight of 72.574KD (Martin et al 1986). The calculated hydropathy value of the tetM gene product would predict that it is a soluble protein. In vitro transcription/translation of DNA expressing the otc-1 gene product predicted that it encodes a 70KD protein (Doyle et al 1986).

The DNA sequence of the otc-1 gene would enable a direct comparison of these two proteins to be made. Both proteins mediate tetracycline resistance by interacting with the ribosome. However it is not known if the resistance mechanism is the same. An analysis of DNA and amino acid homologies should indicate if there is any functional homology between the two gene products. The sequence 5' to the coding region,

identify regulatory sequences controlling the expression of otc-1. Therefore the 2657bp BamHI/SstI DNA fragment which encodes otc-1 has been sequenced.

Sequencing strategies have been widely developed since the original random or shotgun approachs were used. Unidirectional digestion with exonuclease III (ExoIII) of large clones produces a set of nested deletions suitable for sequencing (Henikoff 1984). M13 dideoxy sequencing of deletions produced in this way is a very efficent method of sequencing large fragments of DNA. This method was used to sequence segments of DNA which did not contain any convenient restriction sites for subcloning and the remainder of the sequence was derived from subclones.

DNA with a high %GC (72% for streptomycetes) content is intrinsically more difficult to sequence. Inverted repeat sequences occur frequently and can hybridise to form secondary structures such as hairpin loops. G-C base pairing involves three hydrogen bonds which forms a stronger bond than A-T base pairing, involving two hydrogen bonds. The large fragment of DNA polymerase I (Klenow enzyme) commonly used for M13 dideoxy sequencing can exhibit some difficulty in reading through regions of secondary structure and may produce artifactual bands in all four sequencing tracks. Secondly the product of the sequencing reaction may form regions of secondary structure which produce compressions of bands on the sequencing gel and result in part of the sequence being unreadable. Strategies to overcome these problems, including formamide gels and Maxam and Gilbert (1977) sequencing, were employed and are discussed in the results and discussion sections.

3.2 Sequencing Strategy

Ten subclones of the BamHI/SstI DNA fragment containing the otcI gene (Fig 3.1a: pDM112, pDM113, pDM122, pDM123, pDM172, pDM173, pDM182, pDM183, pDM152 and pDM153) served as templates in dideoxy sequencing reactions. Approximatley 60% of the BamHI/SstI fragment was sequenced in this way. Deletion derivatives of pDM172, pDM173, pDM182 and

40

(A)

pOM193

pOM192

pOM112   pOM122   pOM172   pOM182   pOM152

pOM113   pOM123   pOM173   pOM183   pOM153

pOM1721   pOM1821

pOM1722   pOM1822

pOM1723   pOM1823

pOM1731   pOM1824

pOM1732   pOM1831

pOM1733   pOM1832

pOM1833

pOM1834

(B)

BamHI   BclI   SmaI   1Kb   SmaI   SphI   2Kb   SmaI   SstI

(C)

250bp
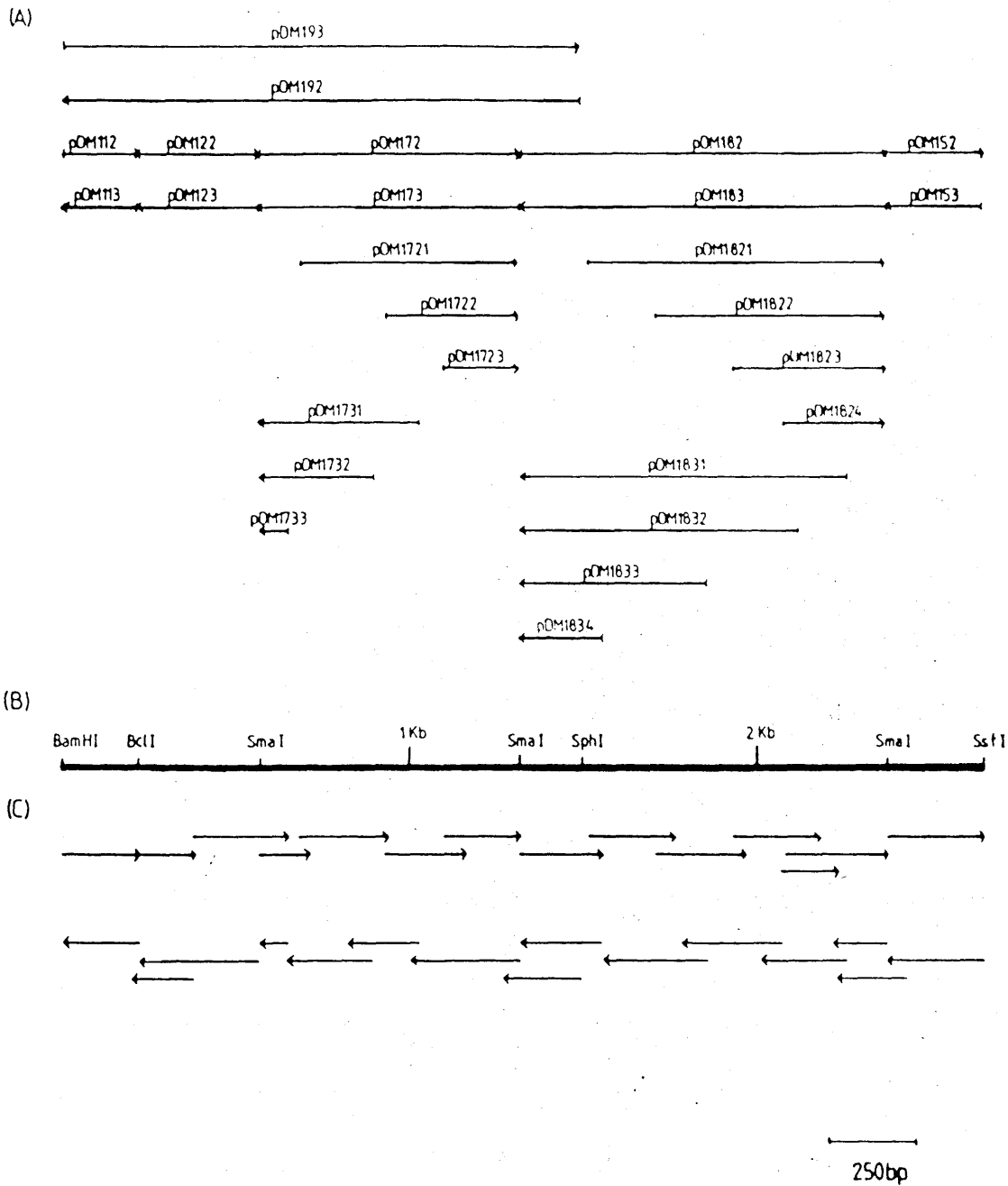
Fig 3.1 M13 Clones and Sequencing Strategy

(A) M13 clones constructed to sequence the BamH1/Sst1 fragment. The
arrowheads indicate the orientation with respect to the sequencing
primer.

(B) Restriction map of the relevant restriction sites for subcloning
the 2657bp BamH1/Sst1 fragment.

(C) Sequencing strategy.

pDM183 (Fig 3.1a: pDM1721, pDM1722, pDM1723, pDM1731, pDM1732, pDM1733, pDM1821, pDM1822, pDM1823, pDM1824, pDM1831, pDM1832, pDM1833 and pDM1834) were constructed by exonuclease III digestion and used to determine the remainder of the sequence. Upon completion of this work, four non-overlapping regions of the sequence remained at restriction site boundaries. These four boundary regions occured at a BclI site at 216bp, a SmaI site at 573bp, a SmaI site at 1319bp and a SmaI site at 2375bp. The BclI (216bp) and SmaI (573bp) junctions were sequenced using oligonucleotides which primed the sequencing reactions at 383bp and 371bp respectively and pDM192 and pDM193 served as templates. The junction at the SmaI site at 1319bp was sequenced across using the SphI/BamHI clone pDM192. The junction at the SmaI site at 2375bp was sequenced by chemical sequencing of a 362bp DdeI fragment from 2062bp-2427bp.

3.2.1 Construction of subclones from the BamHI/SstI fragment in M13

All clones are diagrammed in Fig 3.1. pDM112 and pDM113 were derived from pDM101. The BamHI/BclI insert from pDM101 (Section 4.2.1) was subcloned as a HindIII/EcoRI fragment in M13mp18 and M13mp19 vectors, giving pDM112 and pDM113 respectively. pDM122 and pDM123 were derived from pDM102. pDM102 was constructed by subcloning the 357bp BclI/SmaI fragment into pUC18. The BclI/SmaI insert was subcloned on a HindIII/EcoRI fragment in M13mp18 and M13mp19 vectors, giving pDM122 and pDM123 respectively. pDM172 and pDM173 were constructed by subcloning the 746bp SmaI fragment from pGLW7 into M13mp18 in both orientations. pDM182 and pDM183 were constructed by subcloning the 1056bp SmaI fragment from pGLW7 into M13mp18 in both orientations. pDM152 and pDM153 were derived from pDM105. pDM105 was constructed by subcloning the 277bp SmaI/SstI fragment in pUC18. The SmaI/SstI insert from pDM105 was subcloned on a HindIII/EcoRI fragment in M13mp18 and M13mp19 vectors, giving pDM152 and pDM153 respectively. pDM192 and pDM193 were constructed by subcloning the BamHI/SphI fragment from pPZ24 into M13mp18 and M13mp19 respectively.

3.2.2 Construction of ExoIII deletions from pDM172, pDM173, pDM182 and pDM183

(A)

EcoRI

SphI

HinDIII

BglII

pDM182

8.3Kb
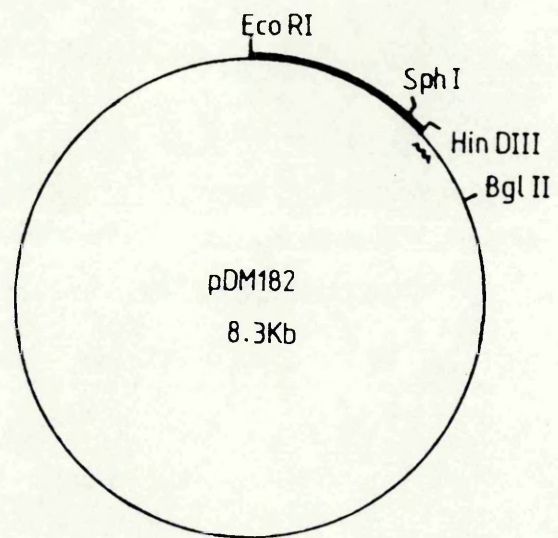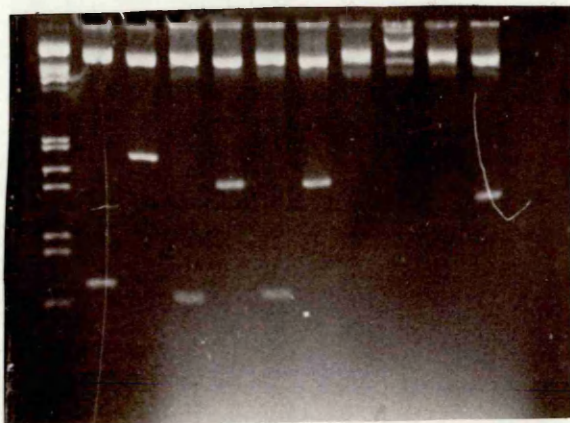
(B)   1 2 3 4 5 6 7 8 9 10 11

(C)   1 2 3 4 5 6 7 8 9 10 11
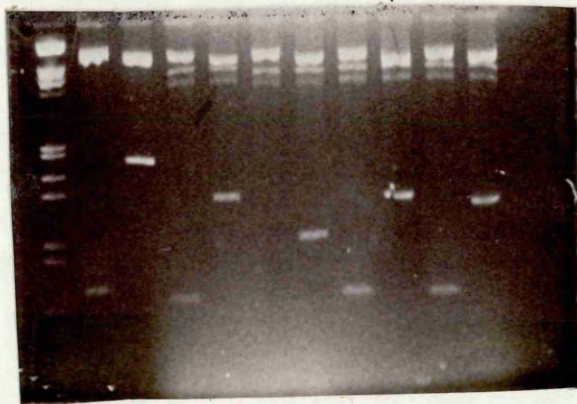
Fig 3.2 Analysis of ExoIII Deletion Clones

(A) Plasmid map of pDM182. Restriction sites used to analyse the size of the deletion, EcoRI and BglII, and to confirm the U.P. site is intact, HindIII and BglII, are indicated.

(B) **Lane 1:** HindIII/EcoRI lambda DNA markers.

   Lane 2: HindIII/BglII restriction digest of pDM182

   Lane 3: EcoRI/BglII restriction digest of pDM182

   Lane 4,6,8,10: HindIII/BglII restriction digest of pDM182 derivatives a,b,c and d respectively, digested for 4min with exoIII at R.T.

   Lane 5,7,9,11: EcoRI/BglII restriction digest of pDM182 derivatives a,b,c and d respectively.

(C) All lanes identical except that the pDM182 derivatives were treated with exoIII for 8mins at R.T.

The exoIII digestion procedure (2.16) was followed and several plaques were screened from each time point. The DNA was restricted and examined by agarose gel electrophoresis to determine the size of the deletion and confirm the presence of the universal priming site. Fig 3.2 shows a typical example of the deletions recovered. The majority of the deletions were in the expected size range assuming a digestion rate of approximately 50 bases per minute. Two random size classes were also observed:

(a) deletions which had proceeded in both directions and therefore appeared to have progressed beyond the expected size for a particular time point. Incomplete digestion with the enzyme used to produce the 3' overhang probably caused these deletions and they were discarded because they had lost the universal priming site.

(b) a second class of deletions was observed which had an intact universal priming site but had progressed further than expected for a particular time point. Henikoff (1984) suggests that these deletions result from random degradation prior to ligation possibly caused by preexisting nicks in the RF DNA, by restriction enzyme impurities or by breaks occurring during S1 nuclease digestion. These clones were still useful as alternative deletions not included in the experimental design. The selected ExoIII deletion clones pDM1721-3, pDM1731-3, pDM1821-4 and pDM1831-4 are shown in Fig 3.1.

3.3 Sequencing Methods

The M13 dideoxy sequencing method was employed almost exclusively and was carried out as described in section 2.14.1. Minor alterations of the method described in the B.R.L. sequencing handbook were required due to the high G+C content of the DNA.

3.3.1 Dideoxy Concentrations

The B.R.L. M13 sequencing handbook recommended concentrations of 0.1mM ddATP, 1.0mM ddTTP, 0.3mM ddCTP and 0.5mM ddGTP were used. However this protocol produced premature termination in the T, C and G tracks using streptomyete DNA templates. G tracks were particularly
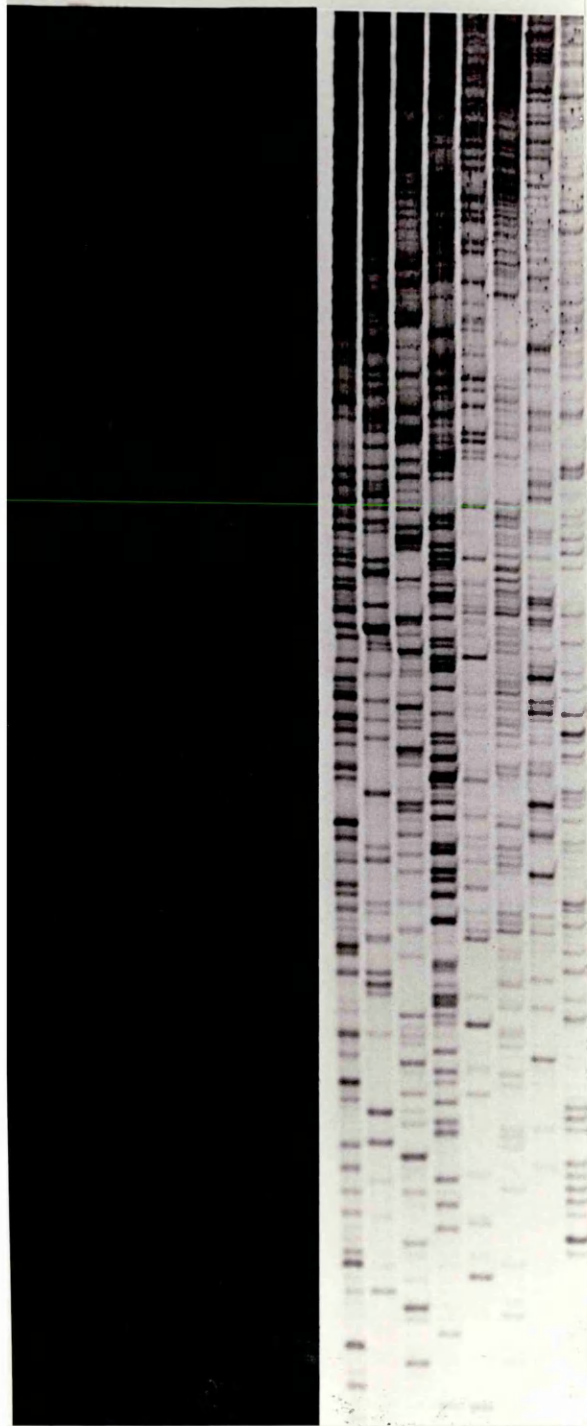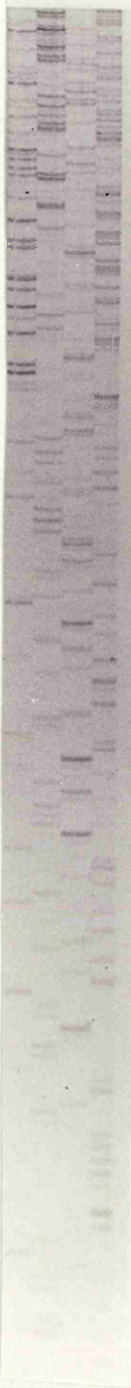
Fig 3.3 Sequence primed from oligo-86 and oligo-89 on template DNA from pDM193 and pDM192 respectively.

problematic failing to produce sequence beyond 150-200 bases. The concentrations of ddTTP, ddCTP and ddGTP were reduced to 0.4mM, 0.1mM and 0.1mM respectively. At these concentrations, C and G tracks ran out before the A and T tracks but sequences of up to 300 bases were readable and the conditions were not optimized further.

3.3.2 Annealing of oligonucleotides

The universal 17bp primer was sucessfully annealed by incubation at 90°C for 5min followed by gradual cooling to room temperature. The oligonucleotides 86 and 89 were annealed to the sequencing template by the same method.

oligo-86   TGGCTTGCCGGCGTCA
                400   384

. oligo-89   AAGCTGAATCTGGGCAT
                355   371

The sequence derived from the oligonucleotide primers was unreadable (Fig 3.3(a)), producing many spurious bands. The banding pattern indicated that the oligonucleotides might be hybridizing to more than one site on the template DNA. The annealing temperature was raised from R.T. to 60°C to increase the stringency of the annealing reaction and the sequencing reaction was carried out at 37 °C as before. The sequence primed from oligo-89 was readable but many spurious bands remained in the sequence primed from oligo-86 (Fig 3.3(b)). Oligo-86 was again annealed to the template DNA at 60°C and the sequencing reactions carried out at 60°C for 1min followed by a 3min chase, to eliminate any secondary annealing during the sequencing reaction. This procedure produced an unambiguous sequence pattern (Fig 3.3(c)).

3.4 Trouble Shooting

Some regions of the sequence were difficult to read, due to areas of compression on gels or simply small regions of sequence which were ambiguous after sequencing both strands. These problems were overcome with formamide gels, which help to denature secondary structures which

43
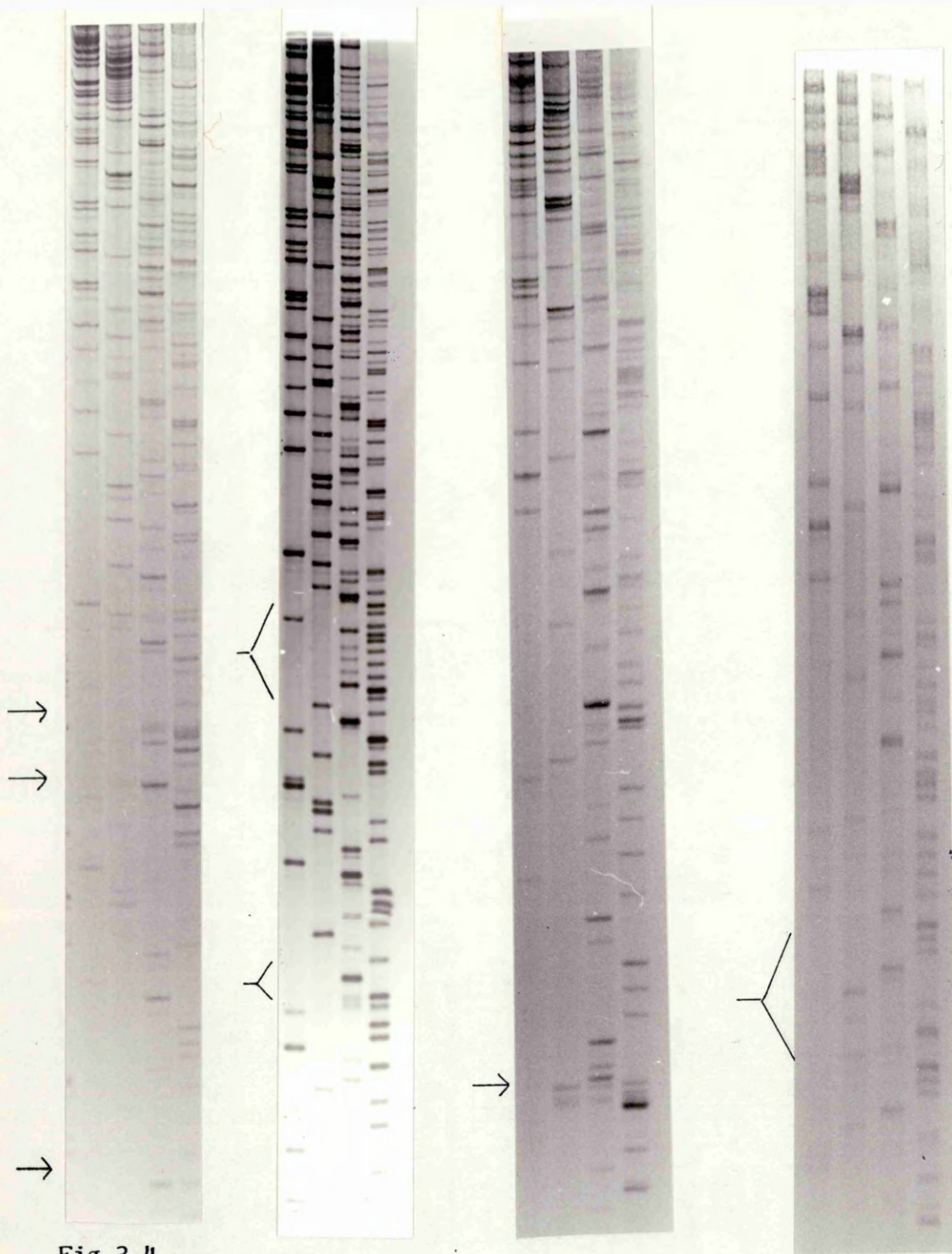
Fig 3.4

(A) Sequence of pDM1733 electrophoresed on a 6% polyacrylamide
    (p.a.) gel
(B) Sequence of pDM1733 electrophoresed on a 25% formamide, 6% p.a.
    gel
(C) Sequence of pDM1722 electrophoresed on 6% p.a. gel
(D) Sequence of pDM1722 electrophoresed on a 40% formamide, 6% p.a.
    gel

→ - indicates regions of compression

< - indicates opened out compressed regions

used as an additional tool to clarify any ambiguous sequence. Reverse transcriptase was used in place of Klenow enzyme and 7 de aza-G in place of dGTP but neither of these alterations had any significant effect on resolving ambiguous sequence.

3.4.1 Formamide Gels

Formamide increases the denaturing capacity of a gel and is useful for analysing sequence compressions caused by secondary structure. Various concentrations of formamide were tested for the clarity of the sequence obtained. Formamide increases the electrophoresis time of a gel and can decrease the resolution. 40%(v/v) formamide sequencing gels were electrophoresed for twice the usual time and produced rather fuzzy bands. 25%(v/v) formamide sequencing gels gave good resolution though the electrophoresis time was increased by approximately one third. Fig 3.4 shows three areas of sequence where compressions were observed. In pDM1733 two areas of compression were observed on a 6% (w/v) sequencing gel at 1116-1124 and at 1153-1165 (Fig 3.4(a)). The sequence was run on a 25%(v/v) formamide sequencing gel which completely resolved the compression at bases 1153-1165 and partially resolved the compression at bases 1116-1124 (the sequence was confirmed on the other strand). In pDM1722 a compression of 16 bases from 867-882 was resolved on a 40% formamide sequencing gel (Fig 3.4(b)).

3.4.2 Maxam and Gilbert chemical sequencing

Two regions of ambiguous sequence were not completely resolved by the incorporation of formamide in the sequencing gels. The C at 845bp was unclear on both strands and the sequence from 2100-2124bp was difficult to interpret. These regions were subsequently sequenced by the Maxam and Gilbert chemical sequencing method. The C at 845bp was confirmed when a 475bp XhoII/BstEII (776-1251bp) fragment, 5' end-labelled at the XhoII site, was sequenced. The second region from 2100-2124bp is within a 365bp DdeI fragment at 2026-2427bp. This fragment was isolated from an agarose gel, 5' end-labelled and the strands separated on a 5% (w/v) non-denaturing polyacrylamide gel.

2100-2124 resolved the ambiguities encountered in the dideoxy sequence. The sequence from the second strand provided sequence across the SmaI junction at 2375bp and confirmed the sequence already established.

3.5 Confirmation of sequence by analysis of predicted restriction sites

The following restriction sites predicted from the DNA sequence were confirmed by analysis of restriction digests on agarose and acrylamide gels.

Bst EII - 1251bp, 1375bp, 1887bp

Dde I - 327bp,2062bp, 2427bp, 2620bp

Hae III - 50bp, 199bp, 528bp

Hinf I - 296bp, 478bp

Mlu I - 923bp, 1103bp

Sst I - 58bp, 225bp

Taq I - 84bp, 566bp

These sites were identified by restriction of the BamHI/SstI fragment while preparing probes for S1 analysis described in chapter 5 and preparing DNA for chemical sequencing and confirm the sequence at these sites. A complete list of restriction sites is included in the appendix.

3.6 Discusssion

3.6.1 Sequence Methodology for streptomycete DNA

Both dideoxy and chemical sequencing methods have been used. The M13 dideoxy system is quicker than the chemical sequencing procedure and does not use any of the hazardous chemicals required for chemical sequencing. The dideoxy system is also more versatile. It does not depend on restriction sites to create clones, demonstrated by the production of deletion clones with ExoIII and it can employ oligonucleotides homologous to sequences, but internal to the cloned

UNIVERSITY OF GLASGOW

THESIS ACCESS DECLARATION

Candidate's Name (BLOCK LETTERS) ...Deirdre Doyle..................

Thesis title .Molecular Analysis of an oxytetracycline resistance

.gene from Streptomyces rimosus...............................

     I understand that in the interests of good scholarship, theses of the University are normally made freely available for consultation in the University Library, or within another library, immediately after deposit but that a candidate may stipulate a period of either one or three years after deposit during which his or her written consent must be sought before such access is given. (A candidate is usually advised by the supervisor if commercial or patent reasons make this restriction desirable.)

     I therefore agree to grant access and to permit copies to be made for other libraries or individuals without my specific authorisation:

    *A.   Immediately on deposit   OR

X *B.   One year after deposit   OR

    *C.   Three years after deposit.

     I further accept that candidates who stipulate written consent but from whom no reply has been received within three months of a request from the University Library posted to their last known address will be assumed to have ceded this power to the Library Committee, to be exercised in consultation with the Higher Degrees Committee of the Faculty.

Signed ...Deirdre Doyle......................

Date ....11/6/87........................

* PLEASE CIRCLE   A   OR (B) OR   C

7742
25·1·88

without any subcloning. The dideoxy concentrations in the reaction have been manipulated to sequence streptomycete DNA. The high G+C content of the DNA causes premature termination when the normal dideoxy concentrations are used because although the incorporation of ddG and ddC occurs at the same frequency there is an increased concentration of C and G residues. A decrease in the concentration of dideoxy nucleotides, decreases the rate of incorporation and therefore extends the length of the sequence. However if the concentrations of dideoxy nucleotides is too low sequence information may be lost from the shorter fragments. Intermediate dideoxy concentrations were established for sequencing streptomycete DNA which resulted on average in 250-300 bases of sequence per clone. If longer sequences were required from a clone two sets of dideoxy concentrations might be used to extend the range of sequence available.

Secondary structure causes two types of problems when sequencing by the dideoxy method. The first problem is independent of the sequencing method and occurs when the labelled DNA forms secondary structures which are not denatured in the gel system. This problem can be overcome by including formamide in the gel, which increases its denaturing capacity. The results in 3.4.1 demonstrate the efficency of the system. However, not all ambiguities were completely resolved on formamide gels (Fig 3.4(b)).

The second sequencing problem is a consequence of the dideoxy sequencing system. Some DNA sequences, for example consecutive G residues, prevent the Klenow enzyme from reading through the sequence and usually intense bands appear in all four tracks followed by faint sequence if any at all. Runs of G residues (up to 5) were sucessfully sequenced but it was noticeable that bands sometimes occurred in all four lanes at specific bases and at low intensity which indicates that the Klenow enzyme was having some difficulty reading through these regions. The sequence from bases 2100-2124 was not completely resolved on formamide gels but compressed regions were apparent from the spacing of the bands. Further ambiguities arose from extra bands which were unevenly spaced even in relation to the compressed areas. It appeared that the Klenow enzyme found it

46

```
         2109        2119                        5115      5125
M13mp18     TGACCCCGTTAAAACTT        M13mp18     CTGGTGAATCTGCCAAT
            ****  ***   **  **                   * ********   **
oligo-86 ← TGACGCCGGCAAGACCA        oligo-89    AAGCTGAATCTGGGCAT →
         3'                5                    5'                3'


         6670        6680                        493       503
M13mp18     TGGCTACAGGAAGGCCA        otcI        GAGCGGCAGCGCGGCAT
            ** *  * * *** ***                    *** * * *  *****
oligo-86 ← TGACGCCGGCAAGACCA        oligo-89    AAGCTGAATCTGGGCAT →
         3'                5'                   5'                3'


         456         466                         694       704
otcI        GGACGCCGGCACCACGA        otcI        CTGCGCAGGCTGGGCAT
            **********  ** *                     **  * ********
oligo-86 ← TGACGCCGGCAAGACCA        oligo-89    AAGCTGAATCTGGGCAT →
         3'                5'                   5'                3'


         172        182                          5087      5097
otcI        AGTCGCCGTCAAGCCCT        M13mp18     AATGTCCCTTTTATTAC
            * *****  **** **                     * ** ** *   * ****
oligo-86 ←TGACGCCGGCAAGACCA        U. P.      ← ACTGGCCGTCGTTTAC
         3'                5'                   3'                5


         825         835                         810       820
otcI        GGACGCCGGCACGCCGC        otcI        ACTGTCCGCCGTGGCGG
            **********  *  *                     ****  ***  ***
oligo-86 ← TGACGCCGGCAAGACCA        U. P.      ← ACTGGCCGTCGTTTAC
         3'                5'                   3'                5
```

the complement of
Fig 3.5   Diagram displaying the mismatch homology of ∧oligo-86,
oligo-89   and the universal 17bp sequencing primer with the   M13mp18
DNA sequence and the DNA sequence of the pDM192 insert.

This homology indicates that the oligonucleotides are
complementary to the relevant strand.

→ = direction of elongation of sequence reaction.

difficult to read through the sequence (there is a run of 5 G residues at bases 2124-2128) and that the formation of secondary structures was causing compressions. The chemical sequencing method is less affected by the secondary structure of the DNA to be sequenced, because it modifies and cleaves each base randomly. The chemical sequence of this region was much clearer with no spurious extra bands, though as expected the compressions were still visible. Chemical sequencing is a useful alternative to resolve sequence which Klenow enzyme is unable to elucidate. Other strategies for resolving secondary structure compressions, which have not been investigated, include the use of dITP as a substitute for dGTP in the dideoxy mix and the elimination of cytosine rings after the chemical reactions are complete. The inosine base forms weaker secondary structures than guanine (Amersham Sequencing Manual) and removal of the cytosine ring in chemical reactions prevents any G:C secondary structures forming (Ambartsumyan and Mazo 1980).

Sequencing with custom-built oligonucleotides as primers enables the extension of a sequence from any part of a clone without resorting to further subcloning. The results in section 3.3.2 indicate that the oligonucleotides must be designed with care to avoid annealing to alternative primer sites. The banding patterns in Fig 3.3(a) suggested that the oligonucleotides were partly homologous to more than one sequence in the clone. Analysis of the homologies between the oligonucleotides and the sequence of pDM192 revealed sequences which were up to 76% (13/17 bases) homologous (Fig 3.5). Although oligonucleotides may anneal to regions with which they have less than 100% homology, in order to prime the sequencing reaction efficently they should be completely homlogous at the 3' end of the oligonucleotide. Oligo-86 has 13/17 bp matched with the insert sequence from 456-472bp and has ten consecutive matched base pairs to within one base of the 3' end. It also has a 12/17 bp match with two other regions in the insert and a 11/17 bp match to two regions of the M13 vetor sequence which includes the base at the 3' end (Fig 3.5). Oligo-89 has a 11/17 bp match with three regions of the insert DNA and has eight consecutive base pairs of homology at one of these regions which includes the 3' end of the oligonucleotide. It also has a 11/17 bp match with two regions of the M13 vector sequence, again including

the 3' end of the oligonucleotide (Fig 3.5). The universal primer (U.P.) has a 10/17 bp match with the insert DNA at four places but only in one case does the homology extend to the 3' end. The U.P. has a 11/17 bp match with a region of M13 vector sequence. The homology in this case is dispersed along the oligonucleotide and includes the 3' end (Fig 3.5).

The U.P. does not anneal erroneously to the M13 sequence even though it has a 11/17 bp match to one region, but oligo-89 which also has 11/17 bp matchs to pDM192, produced a shadow sequence superimposed on the correctly primed sequence. However one of the homologous regions, at 694-710bp, has eight base pairs of consecutive homology starting at the 3' end. It is possible that the stability created by the consecutive homologous base pairs, which is 62% GC rich, enabled a sequencing reaction to be initiated from this hybrid. It was not possible to ascertain which secondary priming site oligo-89 annealed to at room temperature because the signals were too weak and probably masked by the more intense correctly primed sequence. Therefore it was not possible to confirm from which site the shadow sequence was initiated but when oligo-89 was annealed at $60^{\circ}C$, a clean sequencing pattern was produced (Fig 3.3b). Presumably, at the elevated temperature the hybrids formed between oligo-89 and the alternative priming sites were destabilized. Oligo-86 has a 13/17 bp match to a secondary priming site and has a long run of consecutive homologous base pairs. The ten consecutive base pair homology stretches from the middle of the oligonucleotide to within one base pair of the 3' end. One would not expect a sequencing reaction to be primed from this hybrid but the mismatched bases are G and T residues and therefore it is possible that they form a bond. The sequence obtained after a $60^{\circ}C$ annealing was still contaminated with spurious bands but annealing at higher temperatures such as $65^{\circ}C$, produced no sequence, probably because no stable hybrids were formed. The possibility that oligo-86 was annealing to secondary sites during the sequencing reaction was investigated by conducting the sequence reaction at $60^{\circ}C$ also. Fig 3.3(c) indicates that this strategy was sucessful, suggesting that oligo-86 annealed to the alternative sites while the reaction was prepared and/or while the reaction was incubating.

```
        D  P  A  T  T  T  V  H  D  V  L  Q  R  L  N  H  G  L  I  R
> GGATCCGGCCACCACGACGGTGCACGACGTACTGCAACGGCTCAACCACGGCCTCATCCG
          10        20        30        40        50        60

      G  D  T  S  S  S  S  V  E  E  Y  R  A  E  I  E  E  A  G  F
> CGGCGACACGTCGTCCAGCAGCGTCGAGGAGTACCGCGCGGAGATCGAGGAGGCCGGCTT
          70        80        90       100       110       120

      R  V  R  Q  A  E  R  I  D  N  L  L  G  D  W  L  I  V  A  V
> CCGGGTGCGGCAGGCGGAGCGCATCGACAACCTCCTCGGTGACTGGCTGATAGTCGCCGT
         130       140       150       160       170       180

      K  P  *
> CAAGCCCTGACCCTGCGTGGCCGCCCTTGCTACCGTGATCATGCCCGCGGGTGTGTTCCC
         190       200       210       220       230       240

> GCCGGGCGGATCACAGCGACGTCAGAGAGGGCGGAACGATGCGGCAGGCGACCACGAATC
         250       260       270       280       290       300

                                              M  N  K  L
> CGGCAGGAAAGGTCCGCATCACGACGCTAAGGAATCCGAACCCCCTCCATGAACAAGCTG
         310       320       330       340       350       360

      N  L  G  I  L  A  H  V  D  A  G  K  T  S  L  T  E  R  L  L
> AATCTGGGCATCCTGGCCCACGTTGACGCCGGCAAGACCAGCCTCACCGAGCGCCTGCTG
         370       380       390       400       410       420

      H  R  T  G  V  I  D  E  V  G  S  V  D  A  G  T  T  T  T  D
> CACCGCACCGGTGTGATCGACGAGGTCGGCAGCGTGGACGCCGGCACCACGACGACCGAC
         430       440       450       460       470       480

      S  M  E  L  E  R  Q  R  G  I  T  I  R  S  A  V  A  T  F  V
> TCGATGGAGCTGGAGCGGCAGCGCGGCATCACCATCCGGTCCGCCGTGGCCACGTTCGTC
         490       500       510       520       530       540

      L  D  D  L  K  V  N  L  I  D  T  P  G  H  S  D  F  I  S  E
> CTGGACGATCTCAAGGTCAACCTCATCGACACCCCGGGCCACTCCGACTTCATCTCCCAG
         550       560       570       580       590       600

      V  E  R  A  L  G  V  L  D  G  A  V  L  V  V  S  A  V  E  G
> GTCGAGCGGGCGCTCGGGGTGCTCGACGGCGCGGTCCTGGTGGTCTCGGCCGTCGAGGGC
         610       620       630       640       650       660

      V  Q  P  Q  T  R  I  L  M  R  T  L  R  R  L  G  I  P  T  L
> GTCCAGCCGCAGACCCGCATCCTGATGCGGACCCTGCGCAGGCTGGGCATTCCCACGCTG
         670       680       690       700       710       720

      V  F  V  N  K  I  D  R  G  G  A  R  P  D  G  V  L  R  E  I
> GTCTTCGTCAACAAGATCGACCGGGGCGGCGCGCGTCCCGACGGTGTGCTGCGGGAGATC
         730       740       750       760       770       780

      R  D  R  L  T  P  A  A  V  A  L  S  A  V  A  D  A  G  T  P
> CGCGACCGGCTCACCCCCGCCGCGGTGGCACTGTCCGCCGTGGCGGACGCCGGCACGCCG
         790       800       810       820       830       840

      R  A  R  A  I  A  L  G  P  D  T  D  P  D  F  A  V  R  V  G
> CGGGCCCGCGCGATCGCGCTCGGCCCGGACACCGACCCGGACTTCGCCGTCCGGGTCGGT
         850       860       870       880       890       900

      E  L  L  A  D  H  D  D  A  F  L  T  A  Y  L  D  E  E  H  V
> GAGCTGCTGGCCGACCACGACGACGCGTTCCTCACCGCCTACCTGGACGAGGAACACGTA
         910       920       930       940       950       960
```

```
        L  T  E  K  E  Y  A  E  E  L  A  A  Q  T  A  R  G  L  V  H
>  CTGACCGAGAAGGAGTACGCGGAGGAACTGGCCGCGCAGACCGCGCGCGGTCTGGTGCAC
          970         980         990        1000        1010        1020

        P  V  Y  F  G  S  A  L  T  G  E  G  L  D  H  L  V  H  G  I
>  CCGGTGTACTTCGGGTCCGCGCTGACCGGCGAGGGCCTGGACCATCTGGTGCACGGCATC
         1030        1040        1050        1060        1070        1080

        R  E  L  L  P  S  V  H  A  S  Q  D  A  P  L  R  A  T  V  F
>  CGGGAGTTGCTGCCGTCCGTGCACGCGTCGCAGGACGCGCCGCTGCGGGCCACCGTGTTC
         1090        1100        1110        1120        1130        1140

        K  V  D  R  G  A  R  G  E  A  V  A  Y  L  R  L  V  S  G  T
>  AAGGTGGACCGTGGCGCGCGCGGCGAGGCCGTCGCGTACCTGCGGCTGGTCTCCGGCACG
         1150        1160        1170        1180        1190        1200

        L  G  T  R  D  S  V  T  L  H  R  V  D  H  T  G  R  V  T  E
>  CTGGGCACCCGCGATTCGGTGACGCTGCACCGCGTCGACCACACCGGCCGGGTCACCGAG
         1210        1220        1230        1240        1250        1260

        H  A  G  R  I  T  A  L  R  V  F  E  H  G  S  A  T  S  E  T
>  CACGCCGGACGCATCACCGCGCTGCGGGTCTTCGAGCACGGGTCGGCCACCAGCGAGACC
         1270        1280        1290        1300        1310        1320

        R  A  T  A  G  D  I  A  Q  A  W  G  L  E  D  V  R  V  G  D
>  CGGGCGACCGCCGGGGACATCGCGCAGGCGTGGGGCCTGAAGGACGTACGGGTCGGTGAC
         1330        1340        1350        1360        1370        1380

        R  A  G  H  L  D  G  P  P  P  R  N  F  F  A  P  P  S  L  E
>  CGGGCCGGGCACCTCGACGGTCCCCCGCCGCGCAACTTCTTCGCGCCGCCCAGCCTGGAG
         1390        1400        1410        1420        1430        1440

        T  V  I  R  P  E  R  P  E  E  A  G  R  L  H  A  A  L  R  M
>  ACCGTGATCAGGCCGGAGCGCCCGGAGGAAGCGGGACGGCTGCACGCCGCGCTGCGCATG
         1450        1460        1470        1480        1490        1500

        L  D  E  Q  D  P  S  I  D  L  R  Q  D  E  E  N  A  A  G  A
>  CTGGACGAGCAGGACCCCTCGATCGACCTGCGGCAGGACGAGGAGAACGCGGCCGGCGCG
         1510        1520        1530        1540        1550        1560

        V  V  R  L  Y  G  E  V  Q  K  E  I  L  G  S  T  L  A  E  S
>  GTGGTCCGCCTCTACGGGGAGGTGCAGAAGGAGATCCTCGGCAGCACGCTCGCGGAGTCC
         1570        1580        1590        1600        1610        1620

        F  G  V  R  V  R  F  D  P  T  R  T  V  C  I  E  K  P  V  G
>  TTCGGCGTACGGGTGCGCTTCGACCCGACCCGTACGGTCTGCATCGAAAAGCCCGTGGGG
         1630        1640        1650        1660        1670        1680

        T  G  E  A  L  I  E  L  D  T  R  T  H  N  Y  F  W  G  A  P
>  ACCGGCGAGGCGCTGATCGAGCTGGACACGCGGACGCACAACTACTTCTGGGGCGCACCG
         1690        1700        1710        1720        1730        1740

        W  V  C  A  S  D  R  P  S  P  A  R  A  I  T  F  R  L  A  V
>  TGGGTCTGCGCGTCGGACCGGCCGAGCCCCGGCGCGGGCGATCACGTTCCGTTTGGCGGTG
          1750        1760        1770        1780        1790        1800

        E  L  G  S  L  P  L  A  F  H  K  A  I  E  E  T  V  H  T  T
>  GAACTGGGCTCGCTCCCCCTGGCCTTCCACAAGGCCATCGAGGAGACGGTGCACACCACC
          1810        1820        1830        1840        1850        1860
```

```
         L  R  H  G  L  Y  G  W  Q  V  T  D  C  A  V  T  L  T  R  T
  > CTGCGGCACGGTCTGTACGGCTGGCAGGTCACCGACTGCGCCGTCACCCTGACCCGTACC
          1870      1880      1890      1900      1910      1920

         G  V  R  S  P  V  S  A  A  D  D  F  R  K  A  N  A  R  L  V
  > GGCGTTCGCAGTCCGGTCAGCGCGGCCGACGACTTCCGCAAGGCCAACGCGCGCTTGGTC
          1930      1940      1950      1960      1970      1980

         L  M  D  A  L  G  R  A  G  T  E  V  H  E  P  V  S  S  F  E
  > CTGATGGACGCGCTCGGCAGGGCCGGTACGGAGGTGCACGAGCCGGTCAGCTCCTTCGAA
          1990      2000      2010      2020      2030      2040

         L  E  V  P  A  A  R  L  S  P  V  L  A  K  L  A  E  L  G  A
  > CTGGAGGTGCCCGCCGCCCGGCTCAGCCCGGTACTTGCGAAACTCGCGGAACTGGGCGCG
          2050      2060      2070      2080      2090      2100

         T  P  G  V  P  T  A  E  G  D  V  F  R  L  E  G  T  M  P  T
  > ACGCCCGGTGTGCCCACGGCCGAGGGGGACGTCTTCCGCCTGGAGGGCACGATGCCGACC
          2110      2120      2130      2140      2150      2160

         S  L  V  H  D  F  N  Q  R  V  P  G  L  T  Q  G  E  G  V  F
  > AGCCTCGTGCACGACTTCAACCAGCGGGTTCCCGGACTGACCCAGGGCGAGGGCGTGTTC
          2170      2180      2190      2200      2210      2220

         L  A  E  H  R  G  Y  R  P  A  V  G  Q  P  P  V  R  P  R  P
  > CTGGCCGAGCACCGGGGGCTACCGGCCCGCCGTCGGACAGCCGCCCGTGCGGCCGCGGCCC
          2230      2240      2250      2260      2270      2280

         E  G  P  N  P  L  N  R  D  E  Y  I  L  H  V  L  K  R  V  ::
  > GAGGGGCCCAACCCGCTCAACCGCGACGAGTACATCCTGCACGTGCTCAAGCGCGTGTGA
          2290      2300      2310      2320      2330      2340

  > ACCGGCACCGGACCGCGCCCCGCGCGGATTTCCGCCCGGGGCGCGGTCCGGTGGTTTCAG
          2350      2360      2370      2380      2390      2400

  > GTGCCGGCCTGTACCTGGAGGCGGACCTCAGGCCGTGGATACGGGCTGTCCTGCCGTGCG
          2410      2420      2430      2440      2450      2460

  > TACGGGCCGGTGTTCCGGGAGATTCCGAGGATCAGCATGGATTTCGCCTTGTTCACGGCA
          2470      2480      2490      2500      2510      2520

  > TCGATCCGGGAAACCGCGCCCAGCTTTTCGAAAACATTGCTCAAATGGCGTTTCACGGTC
          2530      2540      2550      2560      2570      2580

  > GCTTCGGTGATGGTGAGCCGGGCGGCGATCTGGCTGTTGCTCAGCCCGCGCGCCACGTTC
          2590      2600      2610      2620      2630      2640

  > TCCATTATCTGGAGCTCGGATCCGGCCACCACGACGGTGCACGACGTACTGCAACGGCTC
          2650
```

Fig 3.6 DNA Sequence and predicted amino acid sequence of the BamHI/SstI fragment.

Oligonucleotides synthesized as sequencing primers should be designed with these problems in mind especially for streptomycete DNA which contains many of repeated sequences. If the whole sequence is available it is possible to search for mismatch homology to a proposed oligonucleotide primer. However in most cases this approach would not be possible. To maximize the uniqueness of the primer a long oligonucleotide with a low GC content (50-55%) should be chosen for sequencing streptomycete DNA. The examples of oligo-89 (47% G+C) and oligo-86 (64% G+C) demonstrate that an oligo with a high G+C content forms more stable hybrids.

3.6.2 Analysis of Sequence

The sequence has been analysed by two methods to identify open reading frames (orf's) which encode proteins. Orf's were identified using the university computer by translating the DNA sequence and searching the 3 possible reading frames for translational stop codons, TAA, TAG and TGA. Two orf's were identified. Orf 1 occurs at the 5' end of the sequence and terminates at a TGA codon at 190bp (Fig 3.6). Orf 2 starts at an ATG at 349bp and terminates at a TGA codon at 2340bp, with the potential to encode a protein of 71.392KD (Fig 3.6). The potential of the orf's to code for proteins was assessed by two computer analyses, Fickett (Fickett 1982) and Frame (Bibb et al 1984). Fickett measures the asymmetry in the distribution of each base among the three coding positions and the overall A, C, G and T content. The analysis then weights each of these 8 parameters on their ability to predict coding sequence in isolation and uses all this information to predict if a sequence is coding or non-coding. The second computer analysis called Frame, simply measures the frequency of G and C bases in each of the positions within a codon (Bibb et al 1984). Frame is particlarly relevant to DNA sequences with extremely high or low %GC contents because it is also possible to determine the coding strand of such sequences.

The Fickett analysis indicated that both orf's were coding and that the intervening sequence between the open reading frames from 191-348bp was non-coding. The analysis also indicated a third potential coding sequence at the extreme 3' end of the sequence (Fig 3.7). A

TESTCODE, (-) STRAND

SEQ.FOTC
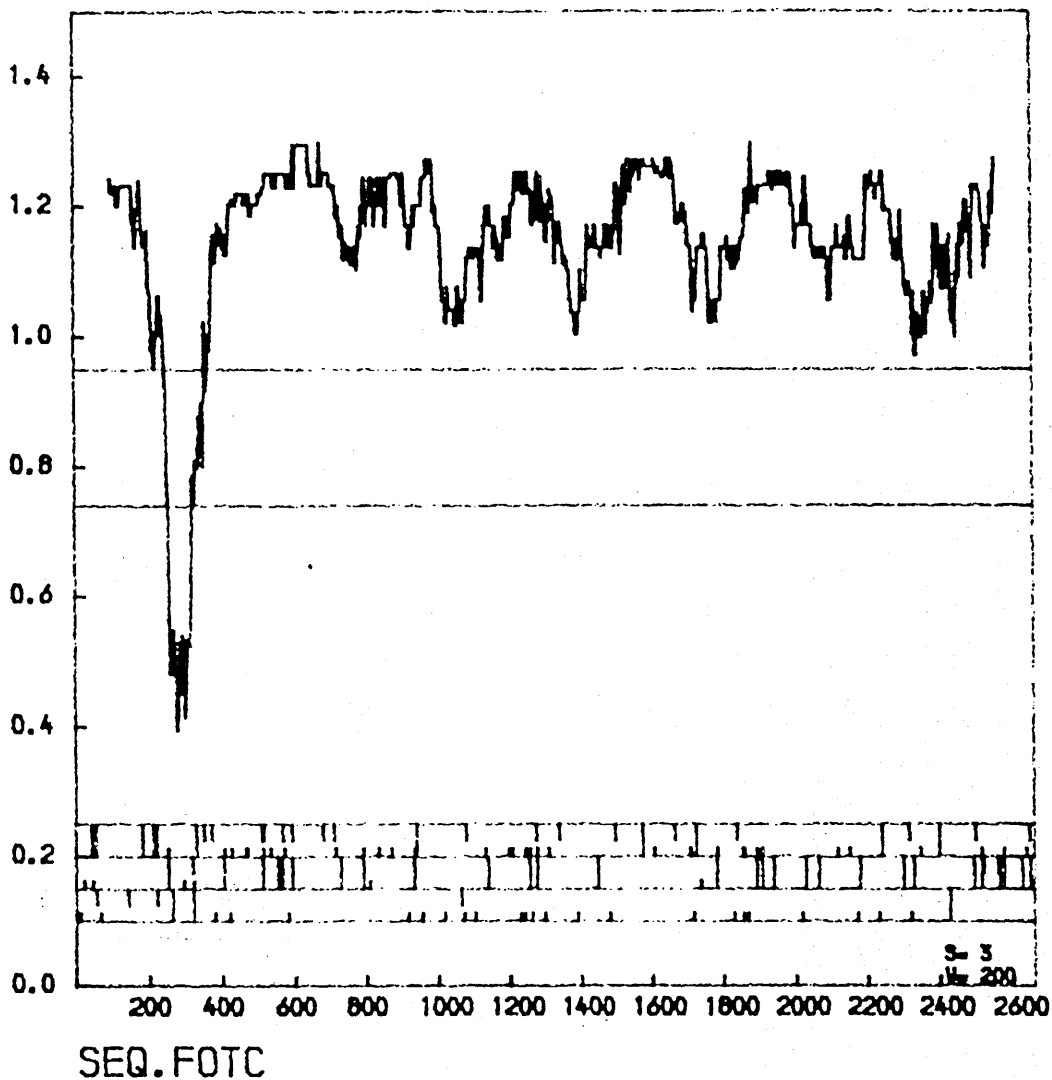
Fig 3.7 Graphical representation of Fickett analysis. Values plotted between the horizontal lines at 0.75 and 0.95 are not assigned as coding or non-coding sequence. Values above the 0.95 line represent coding sequence and values below the 0.75 line represent non-coding sequence as predicted by the Fickett analysis. The diagram below the graph represents the three open reading frames. Complete vertical lines indicate translational stop codons. Small vertical lines indicate translational start codons, descending lines represent an ATG codon and ascending lines represent a GTG codon.
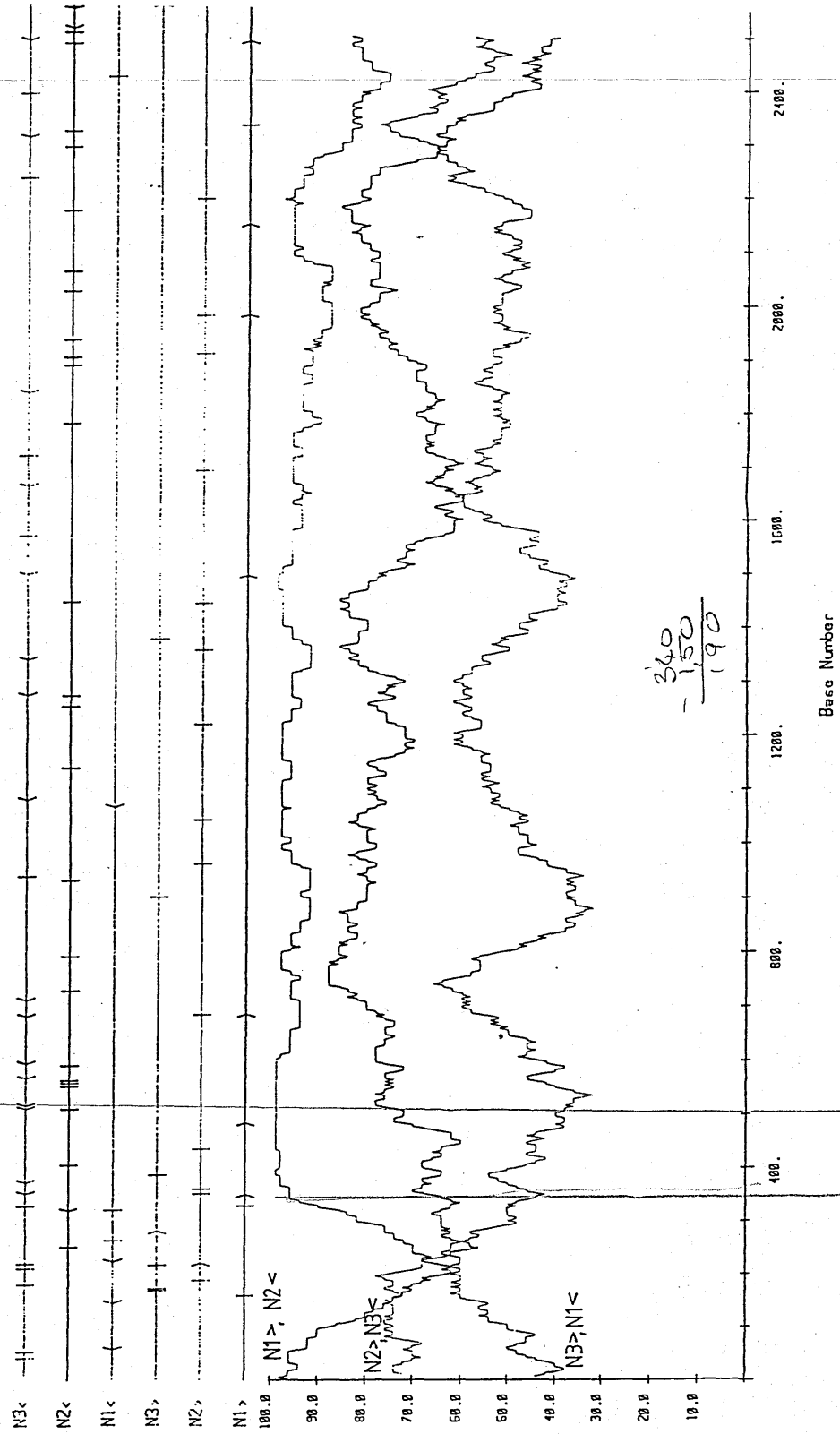
Fig 3.8 Frame analyses.

In the open reading frame analyses , arrowheads (> and <) indicate translational start codons and their orientation; vertical bars (I) represent translational stop codons. In the graphical analysis of G+C distribution the first triplet position and its associated reading frames are represented by a line marked N1> and N2<; the second triplet position and its associated reading frames are represented by a line marked N2> and N3<; the third triplet position and its associated reading frames are represented by a line marked N3> and

Frame analysis indicated the same open reading frames and also showed that the potential orf 3' to otc would be transcribed from the complementary strand (Fig 3.8). Orf1 spans the BamHI site and probably encodes the 3' end of the otcZ (Fig 1.4) gene product. otcZ151 requires sequences 5' to the BamHI site for complementation indicating that the coding sequence crosses this restriction site. There is no evidence of a transcriptional terminator downstream of the TGA codon at nt 190. It is predicted that orf2 encodes the otcI gene product. It is the only orf in the sequence with an in frame start and stop codon and it encodes a protein of 71.392KD which correlates very well with the predicted size of the protein, 70K, from in vitro transcription/translation assays (Doyle et al 1986). The only in frame ATG translational start codon between the stop codon at 190bp and the coding sequence of orf2, as predicted by computer analysis, occurs at 349bp. A comparison of the coding sequence of tetM and otc1 also suggests that the ATG at 349bp is the initiation codon due to the amino acid homology at the 5' end of both proteins (discussed in more detail below). A poor ribosome binding site (RBS) occurs 9bp from this ATG codon with a potentially stronger RBS 15bp from the start codon.

                    AAGGAATCCGAA
               329   340

However this is unusually distant from the initiation codon and it is unlikely to be utilized. It is possible that translation initiation occurs in the absence of a RBS. Three streptomycete genes are known to initiate translation in the absence of RBS's but in these cases transcription initiated at or very near the translation initiation site (Bibb and Janssen 1987, Horinouchi et al 1987).

The TGA translational stop codon at 2340bp is folowed by a 21bp inverted repeat spanning nucleotides 2346-2393bp (Fig 3.6). The theoretical free energy of formation (ΔG) of the hairpin-loop structure was calculated as -70kcal, according to the specifications of Tinoco et al (1973). The predicted hairpin -loop structure is followed by a short run of T residues, which is usually indicative of a terminator structure in the more intensely studied bacteria e.g. E.coli. Low resolution S1 mapping of the 3' end of the gene,

	U		C		A		G		
U	Phe	0	Ser	0	Tyr	0	Cys	0	U
	Phe	1	Ser	1	Tyr	1	Cys	0	C
	Leu	0	Ser	0	Ter	0	Ter	1	A
	Leu	0	Ser	1	Ter	0	Trp	1	G
C	Leu	0	Pro	0	His	0	Arg	0	U
	Leu	4	Pro	1	His	2	Arg	3	C
	Leu	0	Pro	0	Gln	1	Arg	0	A
	Leu	2	Pro	1	Gln	1	Arg	3	G
A	Ile	0	Thr	0	Asn	0	Ser	0	U
	Ile	3	Thr	1	Asn	2	Ser	2	C
	Ile	1	Thr	0	Lys	0	Arg	0	A
	Met	0	Thr	3	Lys	1	Arg	0	G
G	Val	0	Ala	0	Asp	1	Gly	1	U
	Val	3	Ala	3	Asp	4	Gly	3	C
	Val	1	Ala	0	Glu	0	Gly	0	A
	Val	2	Ala	2	Glu	6	Gly	0	G

Percentage base usage in each codon position –

U	C	A	G
3.2	9.5	6.9	13.8
9.0	6.9	10.1	7.4
1.1	18.0	2.1	12.2

Fig 3.9 Codon usage and percentage base usage in each position of a codon of orf1, proposed 3' end of otcZ.

	:	U		:	C		:	A		:	G		:	
U	:	Phe	0	:	Ser	0	:	Tyr	0	:	Cys	0	:	U
	:	Phe	20	:	Ser	9	:	Tyr	9	:	Cys	3	:	C
	:	Leu	0	:	Ser	0	:	Ter	0	:	Ter	1	:	A
	:	Leu	3	:	Ser	8	:	Ter	0	:	Trp	4	:	G
C	:	Leu	1	:	Pro	0	:	His	1	:	Arg	5	:	U
	:	Leu	19	:	Pro	16	:	His	22	:	Arg	23	:	C
	:	Leu	0	:	Pro	0	:	Gln	0	:	Arg	0	:	A
	:	Leu	50	:	Pro	24	:	Gln	13	:	Arg	30	:	G
A	:	Ile	1	:	Thr	0	:	Asn	1	:	Ser	1	:	U
	:	Ile	21	:	Thr	34	:	Asn	10	:	Ser	10	:	C
	:	Ile	0	:	Thr	0	:	Lys	1	:	Arg	0	:	A
	:	Met	6	:	Thr	17	:	Lys	12	:	Arg	3	:	G
G	:	Val	3	:	Ala	0	:	Asp	2	:	Gly	9	:	U
	:	Val	28	:	Ala	33	:	Asp	41	:	Gly	36	:	C
	:	Val	4	:	Ala	2	:	Glu	7	:	Gly	4	:	A
	:	Val	30	:	Ala	37	:	Glu	41	:	Gly	9	:	G

Percentage base usage in each codon position -

U	C	A	G
2.9	10.2	5.9	14.4
9.3	9.0	8.0	6.9
1.2	16.8	1.0	14.4

Fig 3.10 Codon usage and percentage base usage in each position of a codon of orf2, proposed otc1 gene.

approximately 250bp from the <u>Sst</u>I site i.e. at approximately nt 2400, which correlates well with the proposed terminator structure.

The <u>otc</u>I coding sequence is 72% G+C rich, the 3'end of <u>otc</u>Z is 68% G+C rich and the intervening sequence between these protein coding sequence is 66% G+C, as expected for streptomycete DNA which is on average 72% G+C rich. However the sequence from 2470-2590 has an unusually low G+C content of 54%. A second unusual feature of this sequence is that the A and T residues are grouped together in short runs of three or four residues (Fig 3.6). The significance of these features remains undetermined but similar clusters of T residues have been found 3' to the carboxy terminus of <u>tet</u>347 (Drocourt pers. comm. 1987).

The codon usage of both orf's is in good agreement with other streptomycete coding sequences. As expected, an extremely biased choice of synonomous codons is observed. The average % of GC residues found in the 1st, 2nd and 3rd positions of a codon in <u>otc</u>-1 are 74%, 48% and 97% respectively. These figures are reflected in the choice of synonomous codons (Fig 3.9, Fig 3.10).

The sequence has been analysed at the DNA level for the presence of direct and inverted repeats and the predicted amino acid sequence has been analysed for DNA binding motifs and mononucleotide binding motifs. No direct or inverted repeat strucures were found apart from the proposed terminator sequence at nt 2346-2393. The amino acid sequence was analysed according to the method of Dodd and Egan (1987) but no potential DNA-binding regions were discovered. A potential nucleotide binding site was located at amino acid residues 1-20. This conserved site usually occurs, as in this case, at the NH$_2$ terminal end of a protein (Moller and Amons 1985). The conserved nucleotide binding site of <u>otc</u>I exhibited most homology with the guanine nucleotide binding proteins and especially the translational elongation factors, indicating that the <u>otc</u>I gene product has the potential to bind GTP.

The predicted amino acid (a.a.) sequence of <u>otc</u>-1 has been compared to

51

DOTMAT: 7/20 MATCHES

Fig **3.11** A dotmatrix representation of homology at the 35% level, between the amino acid sequence of otc1 (x-axis) and tetM (y-axis)

DOTMAT, 7/10 MATCHES
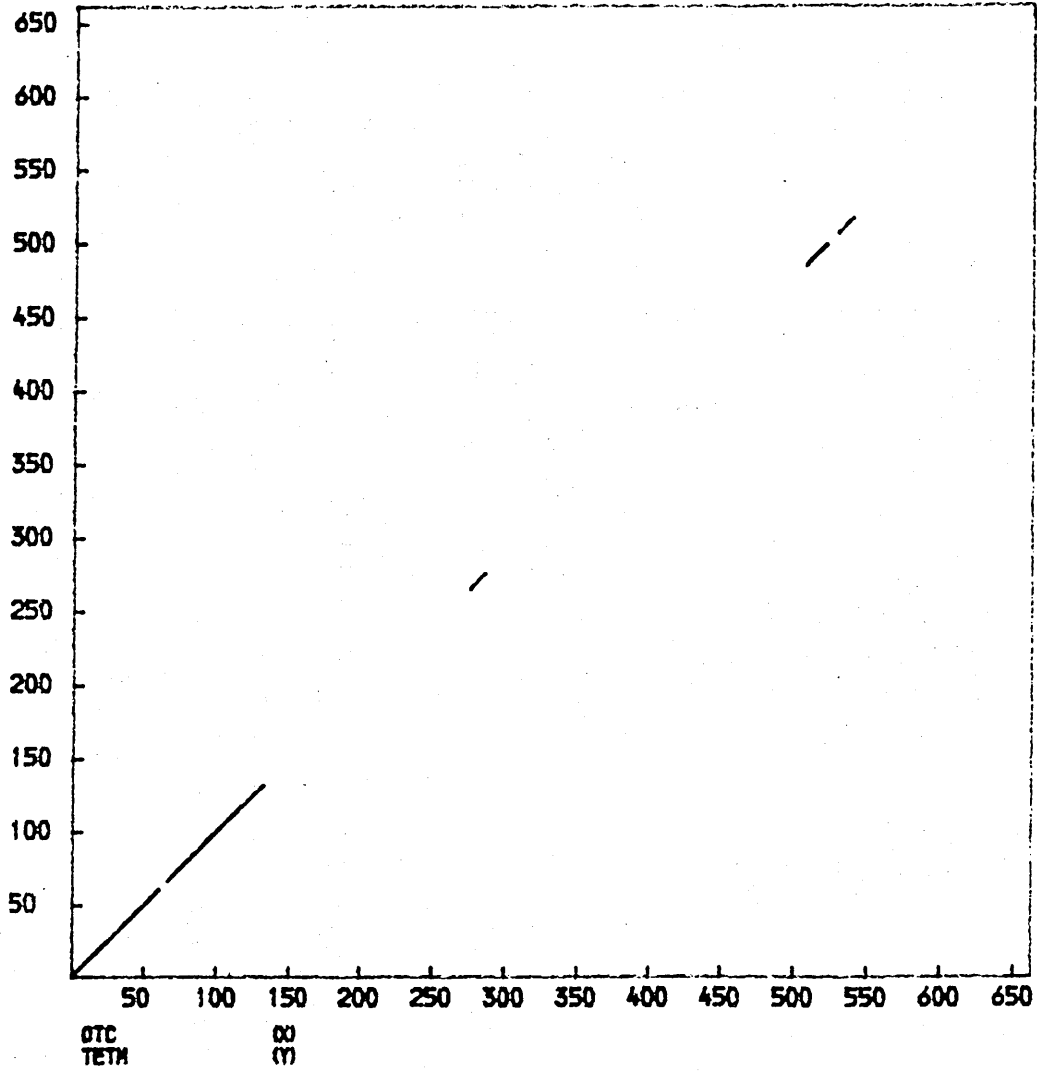
Fig 3.12 A dotmatrix representation of homology at the 70% level between the amino acid sequences otc1 (x-axis) and tetM (y-axis)

the amino acid of tetM from Streptococcus, tetC from pBR322 and tet347 (equivalent to otcII) from S.rimosus. No significant amino acid homology was observed between tetC and otc-1 or between tet347 and otcI, which is predictable considering that their mechanisms of resistance to tetracycline are quite different. The tetM resistance determinant confers resistance to the protein synthesizing machinery as does otc-1. A comparison of these two amino acid sequences reveals good homology at the 35% level which is concentrated predominately at the amino and carboxy termini (Fig 3.11). No further significant homology was discovered by considering the chemical equivalence of some amino acids. The NH$_2$ terminal regions of the proteins are very highly conserved. Fig 3.12 indicates substantial homology at the 70% level. The first 130 amino acids are 67% homologous. Three other regions are highly homologous when the sequences are aligned such that two deletions of 10 amino acids are created in the otcI sequence between a.a. 170-200 and 300-350. A deletion of one a.a. at 538-540 aligns the carboxy terminal ends of otcI and tetM to increase the homology between these two sequences. These deletions are represented in Fig 3.11 and Fig 3.12 by lines which are slightly displaced from the diagonal. This level of amino acid sequence homology is highly significant. The base compositions of the two genes are radically different (tetM - 37% GC : otc-1 - 72% GC). However the DNA sequences are 60% homologous over 400bp at the amino termini of the genes. This region shows 67% homology at the amino acid level. No other significant homology at the DNA level was apparent.

The predicted amino acid sequences of otcI and tetM have been compared with the amino acid sequence of EF-Tu, a translational elongation factor from E.coli (Arai et al 1980). Fig. 3.13 represents the homology between otcI and EF-Tu at the 35% level. The pattern is very similar to Fig 3.11 where homologous regions are concentrated at the NH$_2$ terminal ends of the proteins. All three proteins show homology over the first 150 amino acids (Fig 3.14). otcI and tetM are 65% homologous over this region and EF-Tu exhibits 36% homology with the other two sequences. If chemically equivalent amino acids are considered homologous, EF-Tu exhibits 48% homology with the NH$_2$ termini of otcI and tetM but there is little effect on the homology levels of the less or non-homologous regions. During peptide
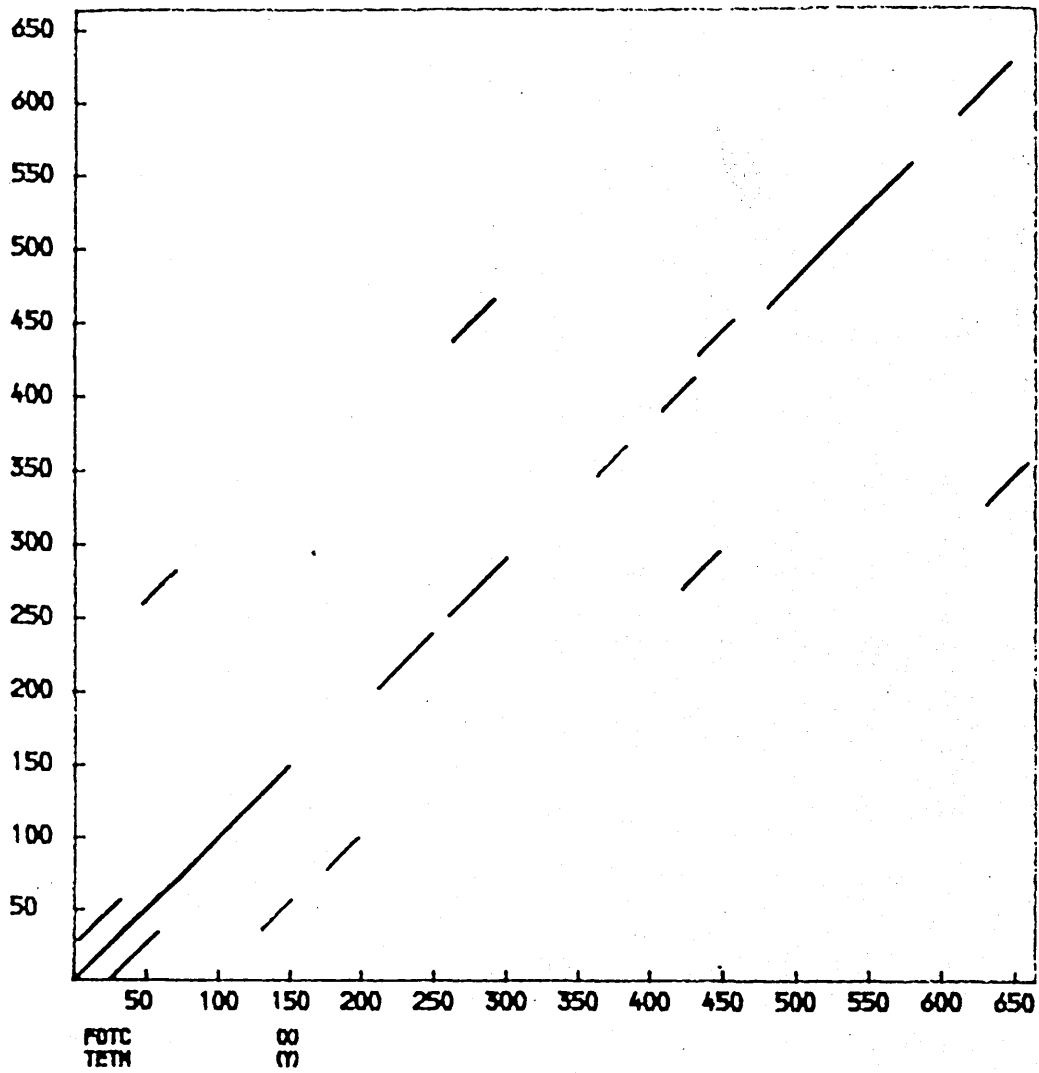
DOTMAT, 7/20 MATCHES
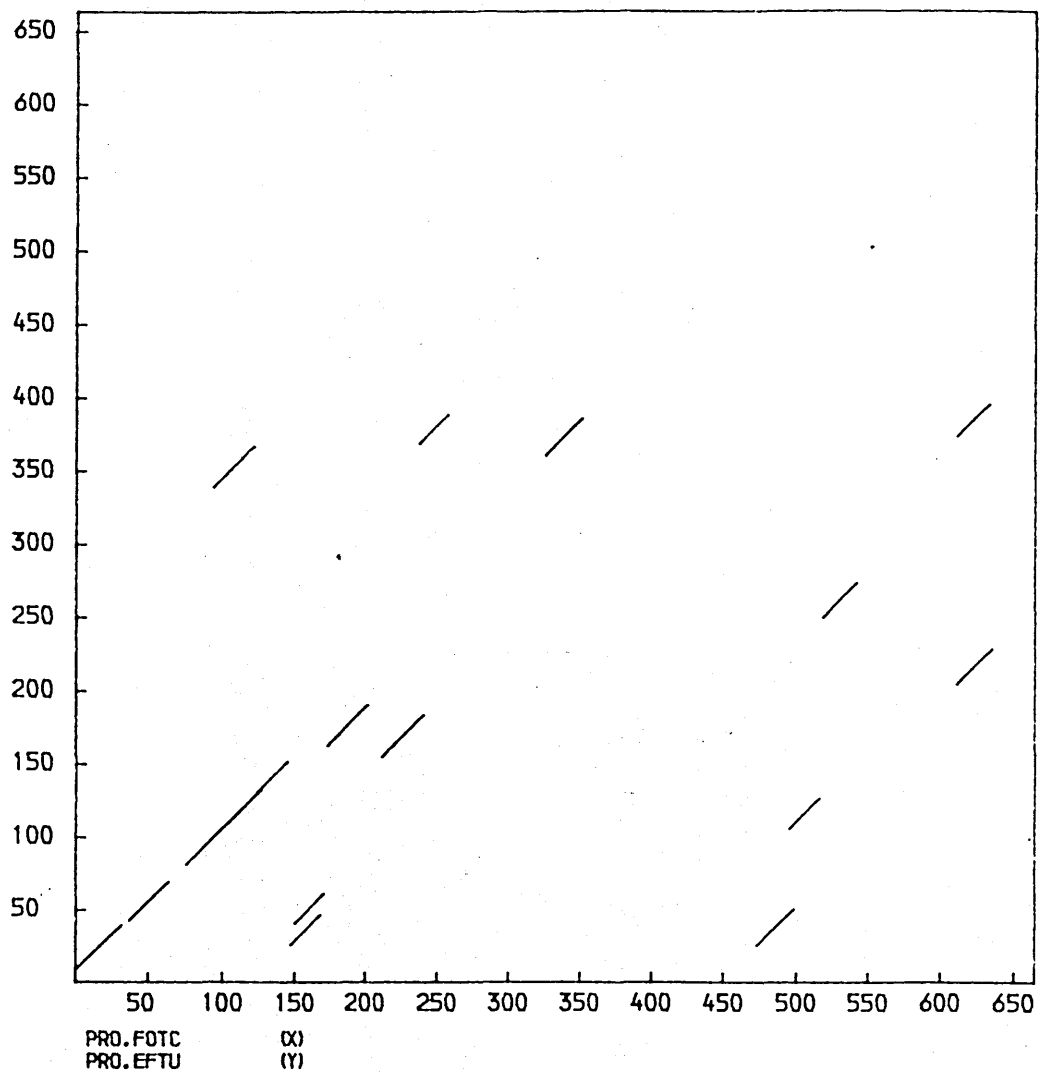
Fig 3.13   A dotmatrix representation of homology at the 35% level, between the amino acid sequence of <u>otc</u>I (x-axis) and EF-Tu (y-axis)

EF-Tu 9 K P H V N V G T I G H V D H G K T T L T A A I T T V L A K T

otc I 1 M N K L N L G I L A H V D A G K T S L T E R L L H R T G V I

tet M 1 M K I I N I G V L A H V D A G K T T L T E S L L Y N S G A I

EF-Tu Y G G A A R A F D Q I _ _ D N A P E E K A R G I T I N T S H

otc I D E V G S V D A G T T T T D S M E L E R Q R G I T I R S A V

tet M T E L G S V D R G T T K T D N T L L E R Q R G I T I Q T A I

EF-Tu V E Y D T P T R H Y A H V D C P G H A D Y V K N M I T G A A

otc I A T F V L D D L K V N L I D T P G H S D F I S E V E R A L G

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

EF-Tu Q M D G A I L V V A A T D G P M P Q T R E H I L L G R Q V G

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

tet M V L D G A I L L I S A K D G V Q A Q T R I L F H A L R K I G

EF-Tu V P Y I I V F L N K C D M V D D E E L L E L V E M E V R E L

otc I I P T L _ V F V N K I D R G G A R P D G V L R E I R D R L T

tet M I P T I _ F F I N K I D Q N G I D L S T V Y Q D I K E K L S

Fig 3.14   Amino acid sequence of EF-Tu (residues 9-158), otcI
(residues 1-150) and tetM (residues 1-150).   Identical amino acids
have been enclosed in boxes to indicate homology and the chemically
equivalent amino acids (K,R), (I,L) and (D,E) have also been
enclosed in boxes.

elongation EF-Tu binds GTP before combining with charged tRNA to form the ternary complex AA-tRNA.EF-Tu.GTP. The ternary complex binds to ribosomes at the vacant A site, the GTP is hydrolyzed, EF-Tu.GDP is released and AA-tRNA is oriented on the ribosome so that peptide bond formation can occur (Arai et al 1980). Tetracycline disrupts protein synthesis by interfering with the binding of the ternary complex, AA-tRNA.EF-TU.GTP, to the A site of the ribosome. otcI and tetM are approximately twice the size of EF-TU and have some other conserved regions not shared by EF-Tu.

5% homology at the amino acid level may be considered to occur randomly, and therefore homology at the 35% level indicates a possible evolutionary relationship (Doolittle 1981). The NH_2 ends of the otcI and tetM gene products are very highly conserved which indicates that these two proteins have the same mechanism of resistance and that the 130a.a. at the NH_2 terminal region are essential to the function of both proteins. The sequence homology between these two sequences may have occurred by divergent or convergent evolutionary processes. However all suggested cases of convergent evolution result in similarties of function without extensive amino acid homology (Bannister and Parker 1985) suggesting a case here of divergent evolution. The Benveniste and Davies (1973) hypothesis suggests that the plasmid and transposon resistance determinants are derived from the resistance determinants of antibiotic-producing species. This hypothesis has been tested by comparing the a.a. sequences of resistant determinants from antibiotic producing species with their functional counterparts on plasmids and transposons. The aminoglycoside phosphotransferase (aph) present in S.fradiae has been compared to the aph gene of bacterial transposons Tn903 and Tn5 (Thompson and Gray 1983). The a.a. sequences exhibit an overall homology of 36%-40% with more highly conserved regions in the carboxy termini. Less extensive a.a. sequence homology has been reported between the N-methyl transferases (erm) from S.erythraeus and the plasmid borne erm genes from S.aureus, S.sanguis and B.lichenformis (Uchiyama and Weisblum 1985). Similarly tet347 from S.rimosus and the plasmid borne tet resistant determinants in S.aureus and tetC from pBR322 show less extensive amino acid homology (Drocourt pers. comm. 1987). These reports support the hypothesis that resistant

determinants found on plasmids and transposons in pathogenic bacteria might have originated from the determinants of antibiotic producing bacteria where resistance is obligatory for survival. The homology between otcI and tetM further supports this hypothesis. The tetM protein is 23a.a. shorter than the corresponding otcI sequence, but the sequences can be aligned for maximal homology by deleting 21a.a. from the otcI sequence. The unusually high levels of homology at the NH_2 termini suggest this region has a functional role.

The homology between the OTC resistance genes otcI and tetM, and EF-Tu is very interesting. It indicates a functional relationship between the resistance genes and EF-Tu suggesting that these proteins have GTP binding sites and that they may even act as elongation factors. otcI and tetM proteins are approximately twice the size of the EF-Tu protein and have some conserved regions not shared by EF-Tu, indicating that these proteins may have other functional properties, probably involved in preventing tetracycline interfering with protein synthesis. Based on the homology between the proteins, a plausible hypothesis for the resistance mechanism of otcI and tetM is that they act as alternative tetracycline resistant elongation factors, possibly by substituting for the usual EF-Tu factors, which are inhibited by tetracycline. This would be the first example of an antibiotic resistance mechanism which is not mediated by inactivation of the drug, modification of the ribosomal target or efflux of the drug out of the cell.

CHAPTER 4

Promoter Probe Analysis of the _otc_1 Clone

A 2.65Kb BamHI/SstI clone encodes otcI (Butter et al 1986). This DNA fragment is situated at one end of the OTC biosynthetic early gene cluster (Fig 1.4). DNA sequencing data suggested that the otcI gene was transcribed from right to left (Fig 4.1, Section 3.6.2) and promoter probe vectors have been used to identify promoter activity in this clone. Subsequent dissection of the larger fragments exhibiting promoter activity by subcloning further defined the location of the presumptive promoters and indicated that transcription from these promoters is regulated. The S1 analysis data, presented and discussed in chapter 5, are consistent with the results discussed in this chapter and are discussed in conjunction with the data derived from the promoter probe constructs.

Two types of promoter-probe vectors are currently being used in Streptomyces. A low copy-number vector pARC1, which contains a promoterless gene directing synthesis of a brown pigment, has been used to identify promoter activity (Horinouchi and Beppu 1985). The advantages of this vector are the non-selective manner of identifying promoter activity and the low copy-number which partially recreates the natural environment of the promoter. However, pARC1 has only one restriction site for inserting fragments to test for promoter activity. The second vector system is based on a high copy-number vector pIJ101 (Ward et al 1986). Several useful features have been incorporated into the promoter-probe plasmids pIJ486 and pIJ487. A polylinker 5' to the promoterless neo gene from Tn5 facilitates the insertion and excision of fragments being tested. A terminator from E.coli phage fd has been incorporated 5' to the polylinker preventing any detectable transcriptional read through from vector promoters. A translational stop codon, 3' to the polylinker but in frame and 5' to the neo coding region has been included to prevent translational fusions which could alter the stability and/or the specific activity of the neo gene product. pIJ486 and pIJ487 differ only in having their polylinkers in opposite orientations. These vectors were chosen to analyse the potential promoter activity of subclones of the
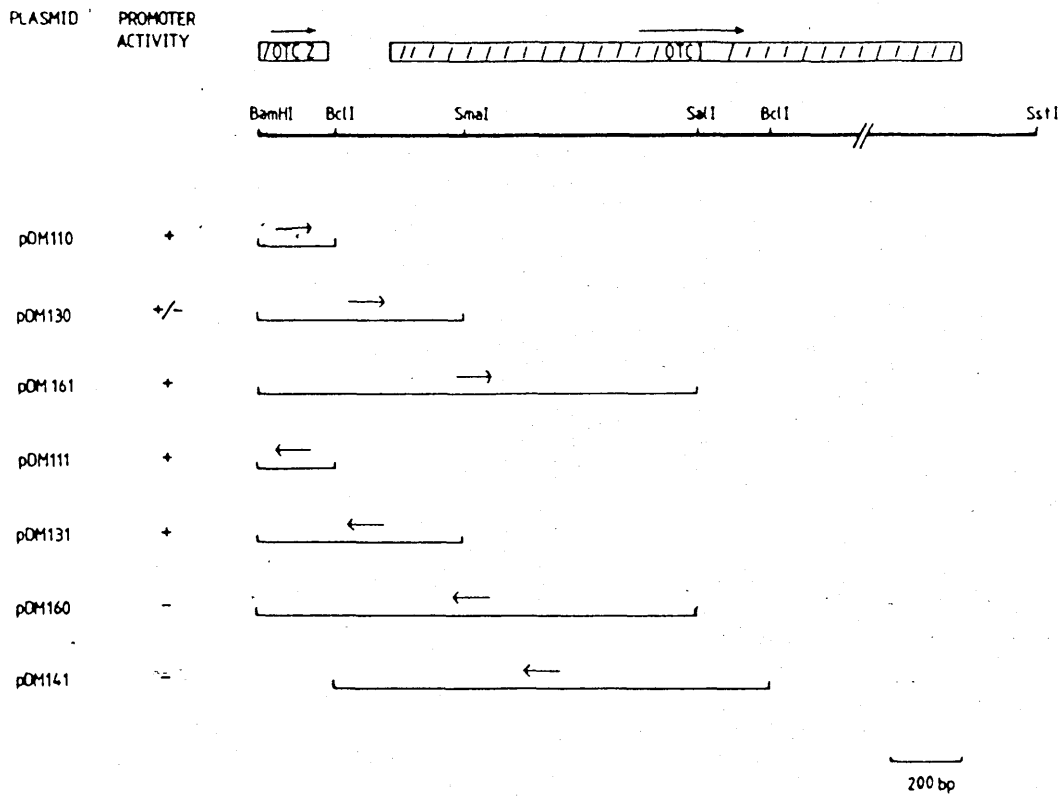
Fig 4.1 Diagram of promoter probe constructs based on the pIJ486/487 vectors. The arrows indicate the orientation of an insert with respect to the promoterless Tn5 neo gene. The otcI and otcZ coding sequences are denoted by hatched boxes and the arrows above the boxes indicate the direction of transcription.

BamHI/SstI fragment because they are well characterised and are more versatile as cloning vectors.

4.2 Results

4.2.1 Construction of promoter probe clones in pIJ486 and pIJ487

The fragments to be analysed in promoter probe vectors were initially subcloned in pUC18 and subsequently subcloned on HindIII/EcoRI fragments into pIJ486 and pIJ487. This approach was taken because the analysis of ligation products is quicker in E.coli than Streptomyces and the subsequent subcloning steps into Streptomyces vectors were simplified.

The 216bp BamHI/BclI, 573bp BamHI/SmaI, 1229bp BclI/BclI and 1234bp SalI/BamHI fragments from pGLW7 were subcloned into pUC18 to produce pDM101, pDM103, pDM104 and pDM106 respectively. The cloned fragments in pDM101 and pDM104 were orientated such that the intact BamHI site was nearest the HindIII site in the polylinker and the SalI site within the right end of the BclI/BclI fragment was nearest the EcoRI site (Fig 4.1).

HindIII/EcoRI fragments from these clones were subcloned into pIJ486 and pIJ487 vectors and transformed into TK24 to derive pDM110 and pDM111 from pDM101, pDM130 and pDM131 from pDM103, pDM141 from pDM104 and pDM160 and pDM161 from pDM106. The even numbered clones denote pIJ486 constructs and odd numbered clones pIJ487 constructs. All the Streptomyces constructs and their orientation with respect to the promoterless neo gene are diagrammed in Fig 4.1.

4.2.2 Assay of promoter activity from promoter probe constructs

Spores from TK24 containing each of the constructs including TK24 containing pIJ486 and pIJ487 as controls, were plated in a dilution series on emerson (Em) agar with thiostrepton (tsr) to assess the spore count. Approximately equal numbers of spores of each construct

PLASMID   Km CONCENTRATION (ug/ml)

PLASMID	0	100	200	300	400	500	600	700	800	900
pDM110	+++	+++	++	+	-	-	-	-	-	-
pDM130	+++	+++	++	+	-	-	-	-	-	-
pDM161	+++	+++	+++	+++	++	++	+	+	-	-
pDM111	+++	+++	+++	++	++	+	-	-	-	-
pDM131	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
pDM160	+++	-	-	-	-	-	-	-	-	-
pDM141	+++	-	-	-	-	-	-	-	-	-

Table 4.1 The table denotes resistance levels of promoter probe constructs to km at varying concentrations on Em agar plates with tsr at a concentration of 25ug/ml.

    +++ : uninhibited growth

    ++ : inhibited growth

    + : poor growth

were subsequently plated on Em plates with tsr at a concentration of 25ug/ml and kanamycin (km) at concentrations ranging from 100ug/ml to 900ug/ml. Several independent clones of each construct were tested for promoter activity.

No promoter activity was detected from pDM160 but pDM111 and pDM131 were both resistant to km at high levels (Table 4.1). The maximal level of resistance to kanamycin by pDM131 was not assessed because it showed no indication of being inhibited at 900ug/ml km. The inserts in pDM111 and pDM131 are subclones of pDM160 and because of the apparently anomalous results, a further 8 independent isolates of pDM160 were tested and the km concentration was lowered to 20ug/ml. All isolates were analysed by restriction enzyme mapping to ensure that no gross structural changes had occurred in the clones. No activity was detected from any of these further isolates of pDM160 or from the pDM141 construct.

Both pDM110 and pDM161 were resistant to km indicating promoter activity, but pDM161 was more resistant than pDM110 (Table 4.1). Independent isolates of pDM130 were found to be both resistant and sensitive to km. 18 independent isolates were tested on Em agar with tsr and km at a concentration of 20ug/ml. 9 isolates of pDM130 were sensitive and 9 isolates were resistant to km at this concentration. The 9 resistant isolates were tested at higher km concentrations and were resistant to km concentrations similar to pDM110. To test the stability of the resistant and sensitive phenotypes, plasmid DNA was isolated from these strains and used to transform S.lividans TK24. Secondary transformants were plated on Em agar with tsr and km. Sensitivity to km was stably maintained but secondary transformants from 3 of the isolates which were resistant to km became sensitive.

Plasmid DNA from sensitive and resistant isolates was resticted with PstI and SphI (Fig 4.2). Plasmid DNA from isolates 1, 4, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, and 18 showed no structural changes had occurred within the limits of resolution on a 1.2% (w/v) agarose gel. The above isolates include all of the resistant isolates and 4 of the
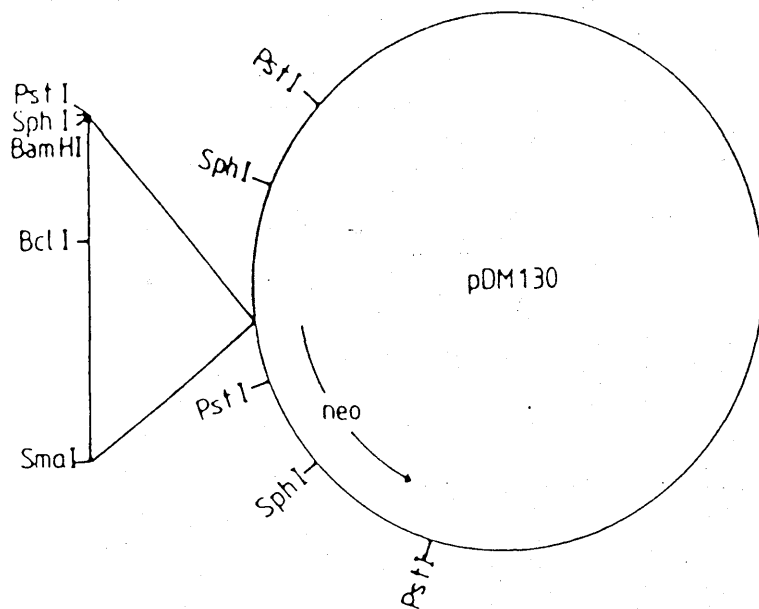
Fig 4.2

(A) Restriction map of pDM130, denoting the <u>Pst</u>I, <u>Sph</u>I, <u>Sma</u>I and <u>Bam</u>HI restriction sites.

(B) <u>Pst</u>I and <u>Sph</u>I restriction digests of independent isolates of pDM130 to confirm the status of the insert. pDM130 isolates 4,6,7,8,10,11,15,16 and 18 were resistant to km and 1,2,3,5,9,12,13,14 and 17 were sensitive.

(B) **Lane 1:** <u>Hin</u>dIII/<u>Eco</u>RI DNA marker

   Lane 2, 4, 6, 8, 10, 12, 14, 16 and 18: pDM130 isolates 1-9 digested with <u>Pst</u>I

   Lane 3, 5, 7, 9, 11, 13, 15, 17 and 19: pDM130 isolates 1-9 digested with <u>Sph</u>I

(C) As above except pDM130 isolates 10-18 replace isolates 1-9

(B)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
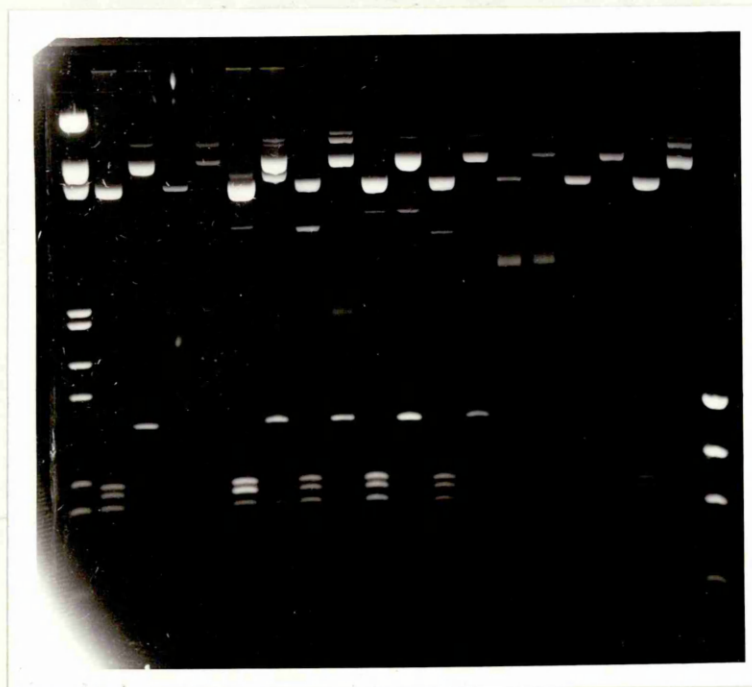
(C)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
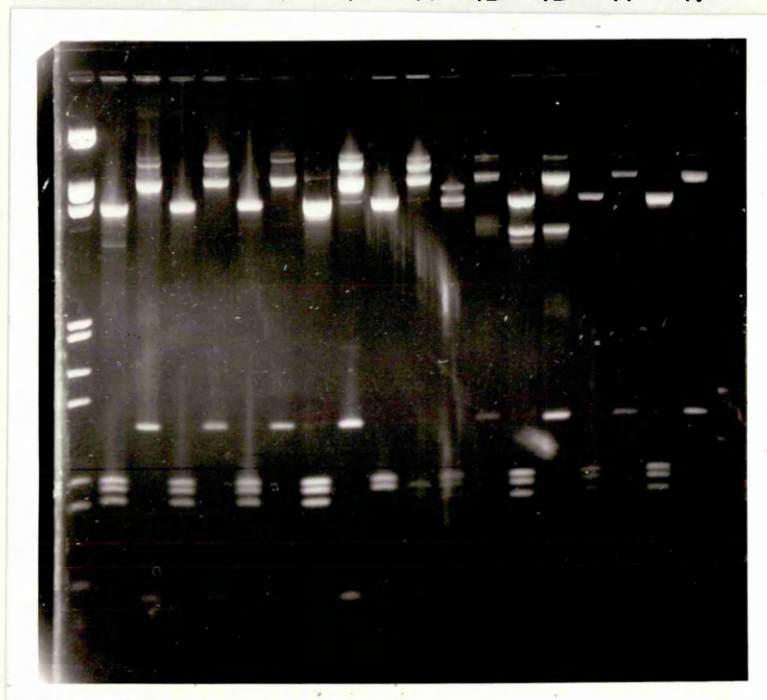
sensitive isolates. Plasmid DNA from sensitive isolates 3, 5, 9 and 14, form mixed populations of intact and deleted derivatives of pDM130. The intensity of the approximately 830bp PstI fragment (which includes the insert) varies with respect to the three plasmid derived PstI fragments, which indicates that the deletions are occurring within the insert, and several other smaller bands appeared (Fig 4.2). The 830bp PstI fragment in isolates 9 and 14 is very faint indicating that the majority of the plasmid population is composed of deletion derivatives. The deleted derivatives form a minority within the plasmid populations of isolates 3 and 5, especially 5 (Fig 4.2). Plasmid DNA from isolate 2 appears to have more complex rearrangments involving one of the plasmid derived PstI fragments as well. The most predominant deletion derivatives of plasmid DNA isolates 3, 5, 9 and 14 have deleted a fragment of approximately 300bp, smaller deletions of less than 100bp have also been observed in samples 3 and 5 (Fig 4.2).

4.3 Discussion

The smallest DNA fragment to exhibit promoter activity was the 216bp BamHI/BclI fragment in pDM110 and pDM111. The promoter activity in pDM110 will be referred to as the rightward transcriptional activity (towards the otcI structural gene) and the promoter activity from pDM111 as the leftward transcriptional activity (transcribing a mRNA of unknown function) (Fig 4.1). The presence of other promoters to the right of the BclI site cannot be ruled out because the promoter probe clones form a set of nested deletions. The S1 mapping analysis, presented in chapter 4, identified transcriptional start sites for the rightward transcript within the BamHI/BclI fragment and transcriptional start sites for the leftward transcript were clustered on either side of the BclI site, extending 10bp to the left of the BclI site and 40bp to the right of the BclI site. No other transcription starts were detected suggesting that promoter activity 5' to the otcI gene resides within and very close to BamHI/BclI fragment.

The km resistance levels of pDM111 and pDM131 are consistent with either a promoter to the left of the BclI site which has been partially disrupted by subcloning the BamHI/BclI fragment but still retains some promoter activity or with intact promoters on either side of the BclI site. Both possibilities would explain the increased resistance to km recorded with pDM131. S1 mapping results, presented and discussed in Chapter 4 support both of these hypotheses. 6 transcriptional start sites have been identified, one of which is 10bp to the left of the BclI site (named otcpA1) and therefore promoter motifs preceeding this start site are likely to have been partially disrupted. Part of the promoter has been replaced by vector polylinker sequences (Table 5.2), which disrupts a potential -10 prokaryotic consensus sequence and a potential consensus similar to the B.subtilis ctc promoter. However the km resistance levels indicate that this partially disrupted region is still functional as a promoter, so it is possible that the sequence just upstream of the transcription start is of more significance to promoter activity and/or that the polylinker sequence fortuitously provides the necessary additional sequence information for transcription initiation.

5 other transcriptional start sites involved in leftward transcription have been identified to the right of the BclI site, named otcpA2-6. These transcriptional starts could account for the increased km resistance levels recorded for pDM131 or the increase may simply be due to the undisrupted promoter sequence otcpA1, directing transcription from the transcriptional start to the left of the BclI site.

pDM160 clones were inhibited at concentrations as low as 20ug/ml of km. 10 independent isolates of pDM160 were tested for km resistance, all of which failed to grow. The plasmid DNA was analysed by restriction enzyme digestion and electrophoresis on agarose gels but no gross DNA rearrangements were found. This evidence suggests that pDM160 which is 661bp longer than pDM131 is somehow repressing the promoter activity recorded from pDM111 and pDM131. All of the extra

DNA in pDM160 is part of the coding region of otcI. pDM160 encodes 295 a.a. of the otcI protein, which is approximately 45% of the total protein. pDM131 encodes 75 a.a. and pDM111 is not within the coding region of otcI (see Chapter 3). The S1 mapping data discussed in chapter 4 supports these promoter probe results. The six transcription start sites identified for the leftward transcribing activity with RNA from S.lividans TK24 (pDM131) were not present with RNA from S.rimosus 4018 which contains the complete chromosomal copy of the otcI gene.

The mechanism of repression of the pDM160 construct could be due to a cis-acting regulatory sequence or a trans-acting repressor function. A cis-acting regulatory sequence seems unlikely because the target site would be a minimum of 340bp from the regulatory sequence which would be within the coding region of otcI. Enhancer sequences found 5' and 3' to eukaryotic promoters are known to positively regulate expression over large distance but there is no precedent for such sequences in prokaryotic organisms (Serfling et al 1985). The alternative mechanisms of repression would involve the otcI transcriptional product or the translational product (or at least the N terminal end of the polypeptide) repressing transcription from these promoters. There is no evidence that mRNA species which are translated can also act as antisense mRNA regulators (Green et al 1986), therefore the most likely mechanism of repression would be at the polypeptide level.

Both the function and length of the leftward transcript are unknown. The transcription unit overlaps the coding region of the 3' end of the otcZ gene which terminates 190bp to the right of the BamHI site. The distribution of bases in the otcZ coding region concurs with the asymmetric distribution of bases expected for Streptomyces coding regions (Thompson and Gray et al 1982), which implies that if the leftward transcript is translated its codon usage would be affected by that of otcZ. Analysis of the sequence indicates it is feasible that this transcript is translated . An ATG translational initiation codon occurs 144bp from the BamHI site and it is prece ded

by a potential ribosome binding site 6bp upstream. No translational stop codons occur in this frame before the BamHI site. An analysis of the codon usage of this potential coding region indicates that most synonomous codons generally used by the streptomycete coding regions are selected and the G/C base usage in the 1st, 2nd and 3rd codon positions would be 45%, 70% and 95%, respectively.

It is likely that this transcript plays some role in the regulation of the OTC biosynthetic pathway because it is regulated by DNA sequences in the coding region of otcI or by the otcI polypeptide itself. The question of whether the transcript is translated and functions at the protein level or plays a role as an anti-sense RNA regulator remains to be answered.

pDM110 and pDM130 are resistant to lower levels of km than pDM111 and pDM131, but the resistance level appears to increase with pDM161. S1 mapping with RNA derived from S.lividans TK24 containing pDM130 identified one transcription start point within or one base to the right of the BclI site, named otcIp1. This result is consistent with the fact that there is no increased resistance to km between pDM110 and pDM130. The increase in km^r of pDM161 may be explained if otcIp1 activity is repressed by the overlapping leftward transcriptional activity from otcpA1-6. It might be predicted that pDM161 will repress leftward transcription as does pDM160 and therefore any repression of otcIp1 activity by leftward transcription will be relieved in the pDM161 construct.

pDM130 clones display a curious instability with respect to km^r. Half of the original clones analysed were sensitive to km and a further 30% of the resistant clones became sensitive after re-transformation. No sensitive isolates have been observed reverting to km^r, suggesting that the change in phenotype is irreversible. Plasmid DNA from 4 of the 9 sensitve isolates showed no evidence of gross DNA rearrangments within the limits of detection on a 1.2% (w/v) agarose gel. This suggests that fine structural changes such as small deletions, point mutations or inversions may be responsible for the switch to km

sensitivity. Large deletions (approximately 100-300bp) have been observed in the plasmid isolates 2, 3, 5, 9 and 14, which are all sensitive to km. The plasmid populations of these isolates are mixed with respect to intact and deleted derivatives, therefore one would expect some resistance to km. As stated above these mixed populations of plasmid are all sensitive to km, suggesting that the intact insert fragment is rearranged in an unknown fashion probably similar to the rearrangements of the 4 sensitive isolates which exhibit no gross rearrangments.

Whatever rearrangments are occurring they appear not to affect the promoter activity of the leftward transcribing promoters, as the km^r phenotype of pDM131 displays no sign of instability. It is possible that the rearrangements are occurring due to the orientation of the insert in the vector. No gross rearrangments e.g. detectable deletions, were observed with plasmid DNA of pDM131 however relatively few isolates were analysed in comparison to isolates of pDM130 and it is possible that deletions were occurring but were not detected. Possible sites for the rearrangements may be deduced, if they are not orientation specific. Any rearrangments within 20bp of either side of the BclI site would be likely to disrupt the promoter activity of pDM131 unless a point mutation is occurring in the otcIp1 consensus sequence. Rearrangments of the otcIp1 consensus sequence would also affect pDM110, which exhibited no instability with its km^r phenotype. It would appear that the BclI/SmaI fragment is the only plausible site for rearrangments. However unless a terminator structure is created, a rearrangment is unlikely to have a direct affect on transcription. Alternatively the stability of the mRNA or polypeptide could be altered by a rearrangment in the 5' end of the mRNA. A 170bp deletion within the untranslated leader sequence of afsB resulted in failure to confer the $afsB^+$ phenotype to host strains (Horinouchi et al 1986) indicating that sequences 3' to the transcriptional start point have a role to play in the control of gene expression. Experiments to elucidate the cause of this effect have not been pursued. A direct approach to solving the nature of this mutation would be to 1) isolate an insert from a sensitive pDM130 clone and insert it into pIJ487, to

confirm that the leftward promoter activty is not affected 2) to
sequence the relevant part of the insert, looking for a small
deletion, insertion or point mutation present in the sensitive
isolates.

Results from the promoter-probe constructs in conjunction with S1
mapping data suggest that overlapping divergent promoters are located
at the BclI site 5' to to the otcI structural gene. Promoter-probe
constructs have indicated that the leftward transcriptional activity
is repressed by a factor to the right of the SmaI site, possibly by
the NH$_2$ terminal end of the otcI polypeptide. To test if the
repression is trans mediated it will be necessary to subclone the
BamHI/SalI fragment onto a compatible plasmid and transform this
plasmid into S.lividans containing pDM111 and pDM131. The effect of
the otcI polypeptide in trans on the promoter activity of pDM111 and
pDM131 could then be assayed.

The function of the leftward transcript is unknown but it is clear
that the expression of this transcript is regulated. A possible
function for this transcript indicated by the data, is that it acts as
an anti-sense mRNA regulator of the production pathway. This
transcript could bind to the otcZ mRNA and possibly other production
gene transcripts depending on the length of the message and disrupt
the expression of the OTC production genes. Anti-sense RNA has been
found to play an important role in gene regulation of prokaryotes.
The first E.coli antisense RNA regulation mechanism to be identified,
controlled the expression of OmpF expression (Mizuno et al 1984).
Several other examples have since been identified involving control
of plasmid replication, Tn10 transposition, expression of crp in
E.coli and expression of Q protein (Green et al 1986). All of these
anti-sense regulators are small transcripts which interact with the 5'
end of the message. Antisense RNA regulation has also been
demonstrated in artificial systems in both prokaryotes and eukaryotes.
An artificial anti-sense regulator which is complementary to lacZ mRNA
from the sequence encoding amino acid 8-1007, inhibits expression of lacZ
(Ellison et al 1985). However the importance of complementarity to

64

the leader sequence, in some cases, is demonstrated by a series of experiments on the inhibition of <u>omp</u>C (Coleman <u>et al</u> 1984). No natural antisense RNA regulator has been found which is complementary to the full length of the target transcript or interacts only with the 3' end of the target mRNA.

The size of the leftward trancript has not been established so it is possible that it terminates before the transcriptional start site of <u>otc</u>Z. It is also possible that <u>otc</u>Z is transcribed as part of a polycistronic transcript. In this case the leftward transcript would need to be transcribed the full length of such a polycistronic message to interact with the 5' end. It will be necssary to determine the transcrition termination point of the leftward transcript and the transcriptional start site of the <u>otc</u>Z transcript before the antisense regulator hypothesis can be further tested. If the function of the leftward transcript is mediated through antisense RNA regulation, the structure of the regulator would be unique amongst natural antisense RNA transcripts previously studied. The anti-sense RNA regulation hypothesis or a regulation hypothesis involving a possible translation product of the leftward transcript provides a mechanism of controlling the expression of the production pathway which would ensure that a cell was expressing <u>otc</u>I and therefore resistant to OTC, before the onset of OTC production.

CHAPTER 5

S1 Nuclease Mapping of Divergent Promoters 5' to the <u>otc</u>-1 Gene

5.1 Introduction

Relatively few sequences controlling transcription and translation of Streptomyces genes have been studied in detail. Expectations of discovering novel promoter features have arisen for two reasons. Streptomycetes express genes differentially throughout their complex life cycle and the expression of many of these genes will probably be controlled at the level of transcription. Streptomycete DNA has a 72% G+C content but most prokaryotic promoters have a high A+T content and the effect of an overall high G+C content on promoter structure and function is been investigated.

S1 mapping of the sequence upstream of the otcI structural gene has identified divergent transcriptional activity. The transcriptional start sites identified upstream of the structural gene will be discussed in relation to the promoter probe studies in Chapter 4 and the published data on sequence, expression and organization of Streptomyces promoters. The approximate location of the transcriptional stop site of the otcI gene has been identified by low resolution S1 mapping.

5.1.1 Streptomyces Promoter Sequences

Streptomycete promoter sequences have been reviewed recently by Hopwood et al (1986). The promoters were divided into three classes based on their homologies to other prokaryotic consensus promoter sequences. Recent publications have considerably increased the number of sequenced streptomycete promoters (Table 5.1) but they may still be divided into three distinct classes.

(1) those with homology with the E.coli consensus sequence

(2) sequences exhibiting homology with the B.subtilis ctc promoter

(3) a miscellaneous group, showing no significant homology with a recognized consensus sequence.

This categorization is somewhat arbitrary because the transcriptional

Table 5.1   The sequences around the -10 and -35 regions of a promoter have been displayed.   No attempt has been made to align theses sequences for maximum homology.   An * denotes promoter sequences identified only by S1 mapping, with no supporting evidence to indicate promoter actvity.

(A) Group 1 promoters

(B) Group 2 promoters.   The consensus sequence is underlined.

(C) Group 3 promoters.   Homology between AfsBp and aphp1 is denoted by a line between the sequences.   The 6bp homology between hygp and orfp2 and other sequences which are partly homologous, are underlined.

PROMOTER	-35 REGION	-10 REGION	+1	REFERENCE
gylp1	TTGACG	GAGACTCGC	U	Smith and Chater 1987
gylp2	TCGAAC	GTTAGAGTG	A	Smith & Chater 1987
0.9Kbp	TTGGCGG	GTAGTGTTC	U	Smith & Chater 1987
*dagp1	TTGTCA	TAGCAT·	A	Hopwood et al 1985
ermEp1	TGTGGGC	TGGTAGGAT·	G	Bibb & Janssen 1987
ermEp2	ATCTTGACG	GGGGAGGAT	C	Bibb & Janssen 1987
orfp1	CCTGCGA	CTAGCATCG·	A	Bibb & Janssen 1987
*kanpA	GTGTCGGAGG	GCGTACATT	G	Nakano & Ogawara 1987
pIJ101A	ATTGCGCA	TCAGACTGA	G	Buttner & Brown 1987
pIJ101B	CTTGACAC	CACAGTATG	?	Buttner & Brown 1987
pIJ101C	GATTGAGA	TGATCGCGG	G	Deng et al 1986
*orfJIIp	CTGACT	GACGAT	C	Hopwood et al 1986
SEP2	TTGACGG	TAAATTACC	A	Jaurin & Cohen 1985
SEP3	CTCTTGACA	TCATCATA	A	Jaurin & Cohen 1985
SEP6	ATATTGACA	GGTTATAT	G	Jaurin & Cohen 1985
E.coli	TTGACA	TATAAT		Hawley & McClure 1983
ctc:spoVG	AGG_TT	GG_ATTG_T		Tatti & Moran 1984
endoHp	TGACTGATTGAC	CCGGGGGCAGGGG	G	Westpheling et al 1985
vphp1	GCCCTGTCGCGC	CCGGCGGGAGCG	U	Janssen et al 1985
stap	AACCAGACCTCA	GGCCGGGCATAGCCT	A	Horinouchi et al 1987
kanp1	N.A.	CACCAGGGAGCGGC	T	Nakano et al 1987
orfIp1	TGGACGGTGGTG	ACCCGGGGATACGG	C	Voatli & Hutter 1987

PROMOTER	-35 REGION	-10 REGION	+1	REFERENCE
afsBp	GAGCGGGGGGCGG	TGCTGGACTTCTA	G	(Horinouchi et al 1986)
aphp1	AGGCGCGGAACGG	TGCCATGATGCCG	A	(Bibb & Janssen 1987)
aphp2	GGGGGATTCCCGGC	ACGCCATGTGACC	U	(Bibb & Janssen 1987)
pA1	TTTCGTCCCGGGC	GGACATTCTTGTT	G	(Bibb & Janssen 1987)
pA2	CGGCGCCAACGTG	GGCGGCATCATGG	A	(Bibb & Janssen 1987)
*pA3	ACTCGTGGTGCGG	CAACGTGCTGTCG	A	(Bibb & Janssen 1987)
*pA4	GCGTGCCACTCGT	TCCGGCGCAACGT	G	(Bibb & Janssen 1987)
*pA5	GCTGGTAGACGAA	TCCTTCGTTCACT	U	(Bibb & Janssen 1987)
tsrp1	CGGTCAGGGCAGC	TCGCGTAGGGTGT	A	(Janssen et al 1985)
tsrp2	CACGCAGCCCAGA	AGGCGAATACTTC	U	(Janssen et al 1985)
galp1	GTGACAGGGGGGT	TGTGTTATGTTTG	G	(Fornwald et al 1987)
*galp2	CTTTTCACTTCCG	CAAGCTGAAGTTC	G	(Fornwald et al 1987)
*amyp1	GAAAGCCTTGACG	TGAACGGCAGGCT	G	(Hoshiko et al 1987)
*amyp2	ACCGAAATGTCTT	TGACGCTGTTAAC	A	(Hoshiko et al 1987)
*amyp3	CGCAACTCACCGA	AAAGCCTTGACGC	U	(Hoshiko et al 1987)
sphp	CGGACGGCGGGCG	TCCGATAGGTTGC	A	(Vogtli & Hutter 1987)
orfip2	GGAGCGGGGTGTT	TATCGGAGATCGG	A	(Vogtli & Hutter 1987)
hygp	CCCGACGTGGCCG	TCGTCAACGCCTG	G	(Zalacain et al 1986)
orfp2	GATCCTCCCCGCA	CGTCAAGATCGAC	G	(Bibb & Janssen 1987)
orfp3	TTGCGGCGACGCG	CGTCCGAACTGCT	C	(Bibb & Janssen 1987)

analysis data are not fully comparable. For example, not all predicted promoters have been tested in promoter probe vectors. S1 analysis can occasionally result in artefacts due to the secondary structure of the region being analysed or mRNA processing may result in processed transcripts being identified. Therefore it is possible that some of the promoter sequences tabulated in Table 5.1, which have being identified by only one technique, are not functional in vivo. Any transcriptional start sites which have no additional evidence to suggest they are preceeded by functional promoters have been marked with an asterisk in Table 5.1.

The class 1 promoters, comprises of 15 of the 40 promoters sequenced to date (Table 5.1). A variety of gene functions are included in this group e.g. an antibiotic resistance gene, genes for glycerol utilization, plasmid encoded genes, an extracellular agarase gene and several genes of unknown function. These promoters exhibit significant homology with the most highly conserved bases in the E.coli consensus promoter sequence (Hawley and McClure 1983). Adenine occurs in 14 out of 15 cases at position 2 of the -10 region and thymine also occurs in 14/15 at position 6. The first three positions of the -35 region in prokaryotic consensus are TTG and these bases are conserved in 12/15, 13/15 and 15/15 cases respectively in these Streptomyces promoters. Mutation of the thymine at position 6 in the -10 region of ermEp1, essentially abolished all transcription from the promoter (Bibb and Janssen 1987). This T is the most highly conserved base in the E.coli consensus sequence (Hawley and McClure 1983) and would also appear essential to Streptomyces promoters in this group.

The class 2 promoters exhibit homology with the B.subtilis ctc promoter. The ctc promoter sequence is recognized by a minor form of sigma factor (sigma 37) in association with B.subtilis core RNA polymerase (Moran et al 1986). The five promoters in this group direct the transcription of genes encoding kanamycin (kan) resistance, viomycin (vph) resistance, streptothricin (sta) resistance, the endoglycosidase-H (endo-H) gene and an open reading frame (orfIp1) associated with the streptomycin (sph) resistance gene. The vphp1 and endo-H promoters have an eight consecutive base pair homology between the -12 and -21 region. The staP, kanP1 and orfIp1 have 5/8, 4/8 and

68

Three consecutive G residues, which occur between -13 and -17 depending on the promoter, are conserved among all five _Streptomyces_ promoters and also conserved in the _ctc_ promoter.

Methylation protection experiments on the _ctc_ promoter indicate 6 bases at -9,-14,-15,-16,-33 and -36 which are protected by E-sigma 37 (core RNAP-sigma37) (Moran _et al_ 1982) and base substitutions at these bases defined the G at -36, G at -14, G at -15 and G at -16 as most important in promoter recognition _in vitro_ by E-sigma 37 (Tatti and Moran 1984). RNAP from _S.rimosus_ transcribes _in vitro_ the base substitution mutants at -14, -15 and -16 at approximately 30%, 5% and 10% respectively, of the wild type _ctc_ promoter activity suggesting that these conserved bases are also important for promoter recognition in _S.rimosus_ (Craig Binnie pers. comm.). Transcription of the _spoVG_ promoter, from _B.subtilis_, is also directed by E-sigma37 (Moran _et al_ 1981) and the sequence conserved between these two promoters is displayed in Table 5.1. The -10 and -35 regions of these promoters are similar, unlike the 5 class 2 promoters where the most highly conserved sequences are situated between the -12 and -20 region. Sigma factor 37 is expressed in vegetatively growing cells but E-sigma 37 transcribes _in vitro_ _spoVG_ and _ctc_, which are expressed at the onset of sporulation and at the end of exponential growth phase (in non sporulating conditions), respectively. Expression of _spoVG_ als depends on the _spoO_ class of regulatory genes (Zuber and Losick 1983) indicating there is more than one factor involved in the control of expression.

RNAP heterogeneity has been demonstrated in _S.coelicolor_ (Westpheling _et al_ 1985). Two sigma factors were identified, sigma 49 and sigma 35, which confer specificity for transcription of the _B.subtilis_ _ctc_ and _veg_ promoters respectively. The _veg_ promoter is transcribed in _B.subtilis_ by E-sigma 43 which is functionally equivalent to the major _E.coli_ sigma factor; sigma 70. The _endo_-H promoter from _S.plicatus_ is also differentially transcribed by these sigma factors, utilizing E-sigma 49 but not E-sigma 35. The evidence suggests that a sigma factor with a similar recognition specificity to sigma 37 exists in _Streptomyces_ and in association with core RNA polymerase recognises

the five genes in this class. The conserved sequences of the class 2 promoters and the spoVG/ctc promoters (with the exception of the G residue at -36, -14, -15 and -16) suggest that there are also major differences between the promoter structures, which are possibly required for recognition of the promoter by other regulatory factors.

The remaining 20 promoters have been assigned to the third class. For eleven of these promoters the two most highly conserved bases in the -10 region of the E.coli consensus sequence are present but they display little homology at the -35 region. In E.coli, poor conservation of the -35 region has been correlated with promoters which are positively regulated by activator proteins (Raibaud and Schartz 1984). The regulation of these promoters, apart from those controlling the galactose operon (Fornwald et al 1986), has not been studied. This makes it difficult to infer any analogy with the E.coli data. The galactose operon has two promoters, galp1 which is regulated positively by galactose and galp2 which is expressed constitutively. Interestingly, the galp1 -35 region has an almost perfect E.coli consensus sequence but galp2 displays little homology to the consensus which suggests that the presence or absence of positive regulation has little effect on the sequence at -35 in this case. The vital importance of the consensus T in the -10 region of aphp1 was demonstrated by elimination of promoter activity after substitution of this T (Bibb and Janssen 1987). These data suggest that the sigma factor which is involved in specifing the transcription of aphp1 recognized this residue which is also the most conserved residue of the E.coli consensus sequence. The importance of the conserved A and T residues in the other promoters, especially those which are unusually close to the transcriptional start site, remains undetermined (Table 5.1). These promoters may also represent a new class or several new classes of promoters recognized by alternative sigma factors. The -35 regions of the 20 promoters have very little homology to each other and no consensus sequence is obvious. The -10 regions of hygp and orfp2 (associated with ermE) have 6bp of homology in the -10 to -16 region.

CGTCAA

70

promoters with a 4/6 bp match to this sequence. These promoters could form a separate class based on this homology but no homology was detected between the relevant -35 sequences. The 6 consensus promoter sequences which are recognized by minor sigma factors in B.subtilis (Johnson et al 1983) show no homology to the above sequence. If CGTCAA is a meaningful consensus, no analogous class of promoters has been discovered in B.subtilis. There is good homology between the afsBp and aphp1 promoters (Horinouchi et al 1986) but this homology does not extend to any of the other promoter sequences in this group.

To clarify the nature of class 3 promoters further investigations are required. For some pairs of promoters e.g. pA3, pA4 and amyp2, amyp3, the transcriptional start sites are within 6-7bp of each other and there is no unequivocal proof that both start sites are used in vivo, they may be experimental artefacts. It is possible that some of these sequences should not be regarded as promoter sequences at all. The regulation of the majority of these promoters has not been studied and such data could identify common motifs between the promoters which in turn could lead to investigation of further sequences with regulatory roles. Run off transcription experiments with reconstituted RNAP containing only sigma-35 would indicate directly if elements in the promoter sequence are part of an E.coli type consensus sequence or if they are fortuitously similar to the consensus. The relevance of the 6bp homology between hygp and orfp2 and the more extensive homology between afsBp and aphp1 may be established by the publication of further promoter sequences and more directly by base substitution in in vitro mutagenesis experiments.

5.1.2 Expression of Promoter Sequences in Streptomyces and E.coli

17 of the 40 promoters in Table 5.1 have been tested for promoter activity in E.coli. All six of the class 1 promoters tested are expressed in E.coli. The transcription start sites of pIJ101A, and pIJ101C are identical in Streptomyces and E.coli (Buttner and Brown 1987, Deng et al 1986). These data suggest that the initiation of transcription at class 1 promoters is due to a species of RNAP holoenzyme in Streptomyces that recognizes similar sequence

determinants to the consensus promoter of E.coli. The existence of such a species has been demonstrated in S.coelicolor (Westpheling et al 1985). The transcription start points of the three SEP were not detectable in Streptomyces and the promoters showed very low activity in Streptomyces (Jaurin and Cohen 1985). These promoters are highly expressed in E.coli and may be subject to regulation in Streptomyces. The SEP sequences have unusual features such as direct repeats and dyad symmetries, in the +10 to -60 region, which could be involved in the regulation of these promoters. Alternatively the SEP sequences, which were isolated in E.coli, may not have a functional role in Streptomyces, which would explain the low promoter activity.

Two of the class 2 promoters, vph and sta, have been tested in E.coli but are not expressed. The promoters of the genes in this group show no homology to the E.coli consensus sequence and are unlikely to be transcribed in E.coli. The B.subtilis ctc gene is also not expressed in E.coli (Tatti and Moran 1984).

Nine of the class 3 promoters (aphp1, aphp2, tsrp1, tsrp2, amyp1, amyp2, amyp3 sphp and hygp) have been tested but all were inactive in E.coli. Five of these promoters, aphp1, tsrp1, tsrp2, amyp1 and sphp, have some homology with the E.coli consensus sequence in the -10 region, but all the promoters have little or no homology with the -35 consensus sequence. However, as stated in Section 5.1.1 aphp1 is likely to be recognized by E-sigma 35 in Streptomyces which appears to be functionally equivalent to E-sigma 70 in E.coli. Either this homology is too poor to be recognized by E.coli RNAP or the promoters require alternative sigma or other factors not present in E.coli, to be transcribed. However the non expressed phenotype in E.coli maybe the result of other factors such as poor mRNA stability or poor translation; the aphp1 has no conventional ribosome binding site.

5.1.3 Expression of an E.coli promoter in S.lividans

The ampC B-lactamase gene of E.coli is expressed in S.lividans under the control of the ampC promoter (Jaurin and Cohen 1984). Base substitution up mutations of the promoter were made and the ampC gene was expressed in S.lividans at higher levels than from the wild type

```
        ampP32A                 ampP15G16    ampP11T
                                    G           T
                                    ↓           ↑
        A
        ↑
TGACAGTTGTCACGCTGATTGGTGTCGTTACAATCTAACGCATC
                                              +1
```
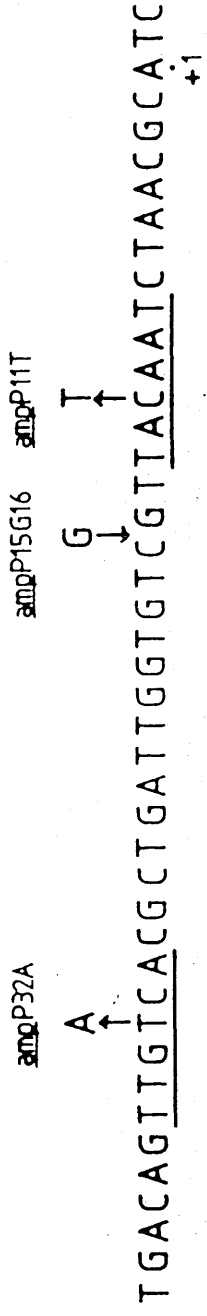
Relative Levels of ampC Specific mRNA

Bacterial Species	w.t.	ampP32A	ampP15G16	ampP11T
E.coli	1	21	16	7
S.lividans	1	7	30	2·5

Fig 5.1 Reproduced from Jaurin and Cohen (1984). The mutations ampP32A, ampP15G16 and ampP11T are displayed in the top half of the figure and their relative effects on promoter activity in E.coli and S.lividans in the lower half of the figure.

PROMOTER REGION

aph

aphp2

orfp1 orfp2 ermEp1 ermEp2
                        orfp3

pA1 pA4
     pA3
     pA2 aphP' pA5

ermE

kanp kanpA kanp1

kan

orfp1 orfp2 sphp

sph

50bp
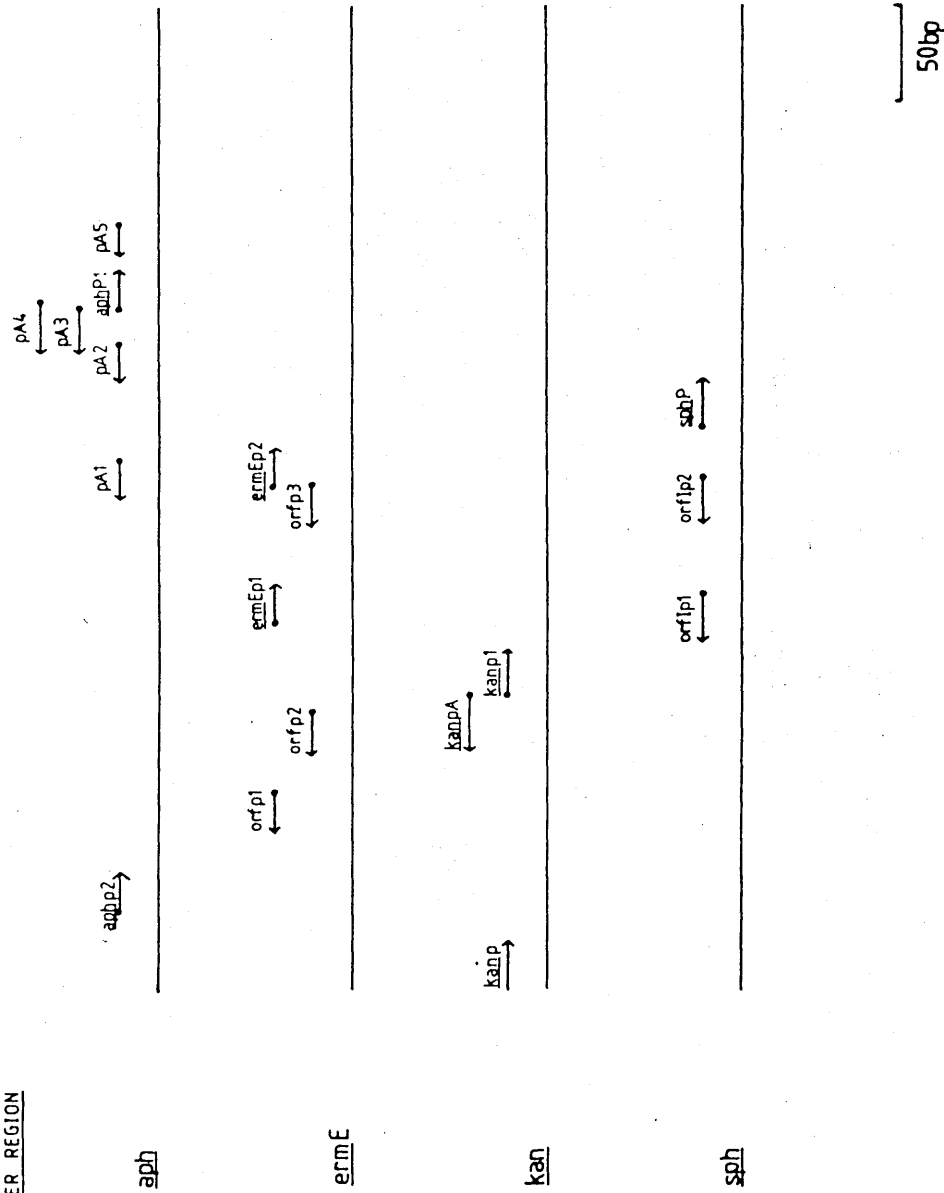
Fig 5.2 The transcription patterns of the aph, ermE, kan and sph
resistance genes are diagrammed to scale, indicating the
transcriptional complexity of these sequences.

promoter in S.lividans (Jaurin and Cohen 1984). Although for the three mutations studied there was increased levels of mRNA production in S.lividans and in E.coli, the relative effects varied between the two species (Fig 5.1). The evidence suggests that a species of Streptomyces RNAP, probably directed by sigma-35 (Westpheling et al 1985), recognizes the same promoter determinants as E.coli RNAP but the relative importance of these factors differ, reflected in the varying degrees of expression between S.lividans and E.coli.

5.1.4 Detailed analysis of the promoter regions of the erythromycin resistance gene (ermE), the neomycin phosphotransferase gene (aph), the kanamycin resistance gene (kan) and the hydroxystreptomycin phosphotransferase (sph) gene.

All four genes have complex transcription patterns, involving multiple transcripts (Fig 5.2). The promoters responsible for the expression of the resistance genes have been identified, kanp1, kanp2, ermEp1, ermEp2, aphp1 and sphp. The ermE1 and ermE2 transcripts which are expressed at the same level in S.erythraeus and in the heterologous host S.lividans, were not affected by the presence or absence of erythromycin (Bibb et al 1985). The kan1 transcript has been identified only in S.lividans and the kan2 transcript only in S.kanamyceticus cultured in antibiotic-producing medium (Nakano and Ogawara 1987). However kan2 transcripts have been identified before the onset of antibiotic production. The neomycin phosphotransferase gene is transcribed from only one promoter, aphp1, as the aph2 transcript terminates before the beginning of the structural gene (Bibb and Janssen 1987). The sph transcript has been identified in S.glaucescens and also in S.lividans (Vogtli and Hutter 1987). For both aphp1 and ermEp1 initiation of transcription is at the translational start site and therefore translation apparently occurs in the absence of a conventional ribosome binding site (Bibb and Janssen 1987).

The transcriptional analysis of these genes is further complicated by divergent and sometimes overlapping transcriptional activity (Fig 5.2). The function of these promoters is unknown but they may be coordinately regulated with promoters for the resistance genes.

The kanpA transcript is transcribed divergently with respect to the kanp1 and kanp2 transcripts. The kanpA transcript, like the kanp1 transcript, has been identified in S.lividans but not in S.kanamyceticus (Nakano and Ogawara 1987).

Five transcripts, pA1-5, have been identified around the aphp1 promoter (Bibb and Janssen 1987). The transcription start sites of pA2, pA3, and pA4 occur within 28bp of each other. The pA1 and pA2 promoter regions overlap the aphp1 promoter and produe more abundant transcripts than pA2, pA3 or pA4.

Three transcription start points have been identified for an open reading frame 185bp upstream of ermE (Bibb and Janssen 1987). orfp1 has been identified in both S.lividans and S.erythraeus and is expressed at marginally higher levels in S.lividans. orfp2 and orfp3 have been investigated and identified only in S.lividans.

Two transcription start points, orfIp1 and orfIp2, have been identified for an open reading frame 150bp upstream of sph in S.glaucescens (Vogtli and Hutter 1987). It has been suggested that there is a third promoter for this orf, within the sph transcript because of the strong full length protection of the probe after S1 nuclease protection (Vogtli and Hutter 1987).

The -10 and -35 regions of pA1-5, aphp1, aphp2, orfp2, orfp3, sphp, orfIp1, orfIp2 and kanp1 show little homology to the E.coli consensus sequence. OrfIp1 and kanp1 have been assigned to the group 2 promoters and all the others are assigned to the miscellaneous group 3 promoters. The transcriptional control mechanisms of the majority of these promoters are unknown but the variety and number of promoter signals indicates that the expression of these genes is highly regulated. Many different factors may be involved such as unknown sigma factors, positive regulators, repressor proteins and anti-sense RNA regulators. Further studies on the regulation of these promoters are required to establish the control mechanisms involved and to understand how they relate to each other, if at all.

B   Bc   D   M   1Kb   M   S Bs   Bc   2Kb   St
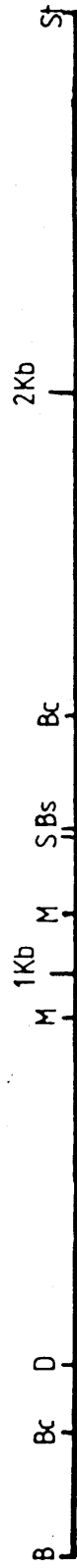
Probe 1

Probe 2

Probe 3

Probe 4

Probe 5

Probe 6

Fig 5.3 The diagram indicates which end and strand of a probe is radioactively labelled.

Probe 1: 1251bp BamHI/BstEII fragment, end labelled with kinase and [γ-^{32}P]ATP.

Probe 2: 923bp BamHI/MluI fragment, end labelled with kinase and [γ-^{32}P]ATP.

Probe 3: 327bp BamHI/DdeI fragment, end labelled with kinase and [γ-^{32}P]ATP.

Probe 4: 1554bp MluI/SstI fragment, end labelled with Klenow and [α-^{32}P]dCTP.

Probe 5: 327bp BamHI/DdeI fragment, end labelled with kinase and [γ-^{32}P]ATP.

Probe 6: 1212 BclI/SstI fragment, end labelled with Klenow and [α-^{32}P]dCTP and [α-^{32}P]dATP.

B – BamHI   Bc – BclI   Bs – BstEII   D – DdeI

M – MluI   S – SalI   St – SstI

5.1.5 Summary

Transcriptional studies on _Streptomyces_ have revealed two consensus sequences. The first is analgous to the _E.coli_ consensus and the second is typified by the _B.subtilis_ _ctc_ promoter. Two independent sigma factors specify transcription from these consensus promoters. The class 1 promoters interact with RNAP in a predictable manner. For example substitution of the most highly conserved base in the sequence, T at position 6 in the -10 region, eliminates transcriptional activity (Bibb and Janssen 1987). Analysis of a mutagenized _E.coli_ consensus promoter, at less highly conserved residues, in both _E.coli_ and _S.lividans_ produced similar trends of transcriptional efficency. The levels of expression differed between the two species (Jaurin and Cohen 1984) indicating that there may be more subtle differences in the interaction of RNAP with the _E.coli_ consensus sequence between the two species. Nearly half of the promoters studied have no obvious consensus sequence and the role of alternative sigma factors and other positively regulating factors needs to be investigated further.

The transcription patterns of the resistance determinants discussed, indicates complex regulation of the antibiotic resistance genes. However the nature of these regulation mechanisms have not been studied but the differential expression of transcripts in heterologous and host strains and overlapping promoter control sequences indicate that regulation is occurring. This complexity is not restricted to genes involved in antibiotic pathways, the amylase gene and glycerol operon also have multiple promoters (Hoshiko _et al_ 1987, Smith and Chater 1987).

5.2 Results

5.2.1 Low resolution S1 mapping of the 5' end of _otc_1

Probes 1 and 2 (Fig 5.3) were hybridised to RNA from _S.rimosus_ 4018 (cultured in TSB with exogenous OTC) at 70°C for 3hrs and treated with S1 nuclease according to the S1 mapping procedure in Section 2.15. The samples were run on a 1.2% (w/v) alkaline agarose gel (Fig 5.4)
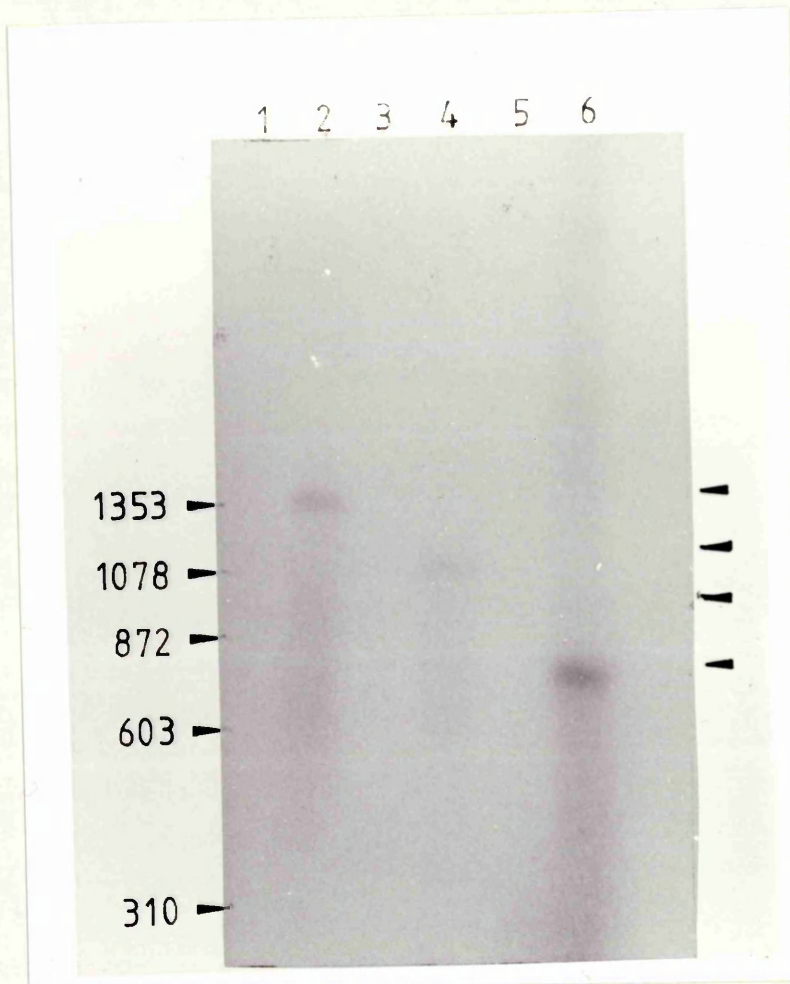
Fig 5.4 Low resolution S1 mapping of the start points and termination point of the otcI transcript

0.5pM of probe 4 (approx. 4×10^5cpm/pM) was hybridised with samples in lanes 1 and 2 at 68°C
0.5pM of probe 1 (approx. 1×10^5cpm/pM) was hybridised with samples in lanes 3 and 4 at 68°C
0.1pM of probe 2 (approx. 1×10^6cpm/pM) was hybridised with samples in lanes 5 and 6 at 70°C
Lane 1, 3 and 5: 40ug of tRNA
Lane 2, 4 and 6: 40ug of S.rimosus 4018 RNA (isolated from cells cultured in T.S.B. with exogenous OTC
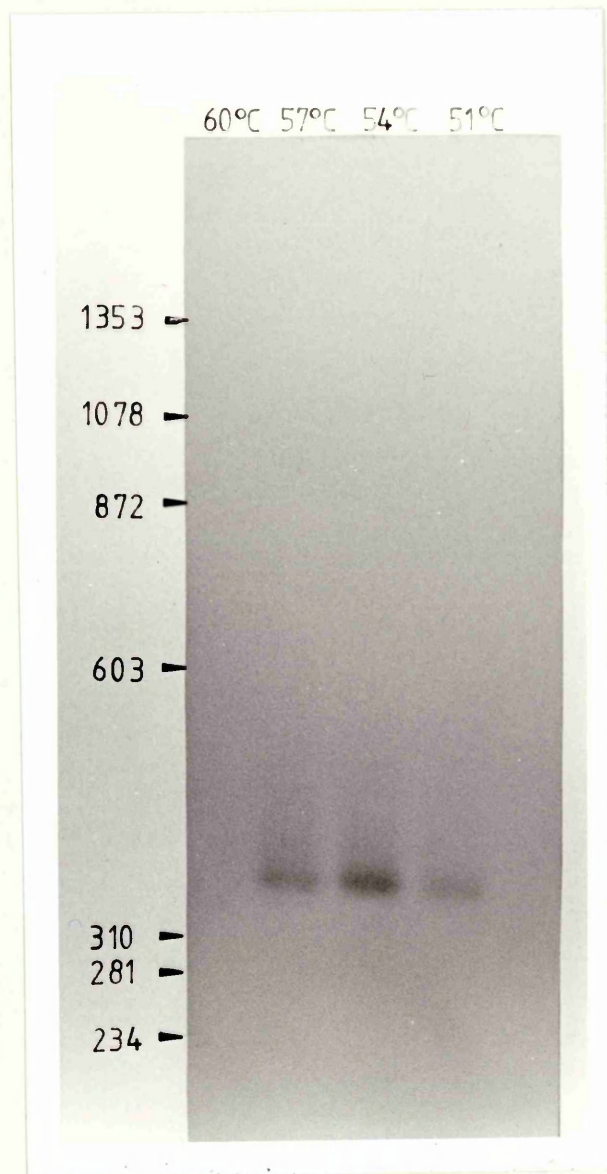
Fig 5.5 Determination of the T_m[DNA/DNA] of probe 3
A normal hybridisation reaction (section 2.15) was carried out in the presence of probe 3 and tRNA, at different temperatures

200-300bp to the right of the BamH1 site. A second protected fragment approximately 200bp longer, of lower intensity was identifed by probe 2 (Fig 5.4). The specific activity of probe 1 was lower than probe 2, which explains why the intensity of the common protected fragment is weaker with probe 1. The reduced specific activity probably also explains why the less intense and longer protected fragment was not identfied with probe 1. No protected fragments were detected in the control tracks (Fig 5.4). These data indicated two transcription start sites separated by 200-300bp. The predicted promoters for these transcripts were named otcIp1 and otcIp2, short and long transcripts respectively.

5.2.2 High resolution S1 mapping of otcIp1

The DNA/DNA melting temperature of probe 3 (Fig 5.3) was determined experimentally to be between 57°C and 60°C (Fig 5.5) using standard S1 hybridisation conditions. The theoretical T_m [DNA/DNA] of probe 3, which is 327bp long and 66.6% GC rich, is 58.6°C, in good agreement with the experimental value. Probe 3 was hybridised to RNA from S.lividans TK24 containing pDM131 and to RNA from S.rimosus 4018 (cultures grown in TSB in the presence and absence of exogenous OTC) at 62°C (2-5°C above the T_m [DNA/DNA]) for 3hrs and digested with S1 nuclease according to the procedure in Section 2.15. The samples were electrophoresed on a 6% (w/v) denaturing polyacrylamide gel. RNA from each source identified the same two transcriptional starts at nt 220 and nt 221 identifying the transcription start points of otcIp1 (Fig 5.6). RNA from S.rimosus 4018 identified a third transcriptional start at approximately nt 5-10, which defines the transcriptional start of otcIp2 more precisely (Fig 5.6).

5.2.3 High resolution S1 mapping of divergent promoter activity in the BamH1/Sma1 fragment.

Probe 5 was hybridised to RNA from S.lividans TK24 containing pDM131 and to RNA from S.rimosus (4018 cultures grown with exogenous OTC in TSB), at 62°C for 3hrs and digested with S1 nuclease aqcording to section 2.15. The samples were electrophoresed on a 6% (w/v)
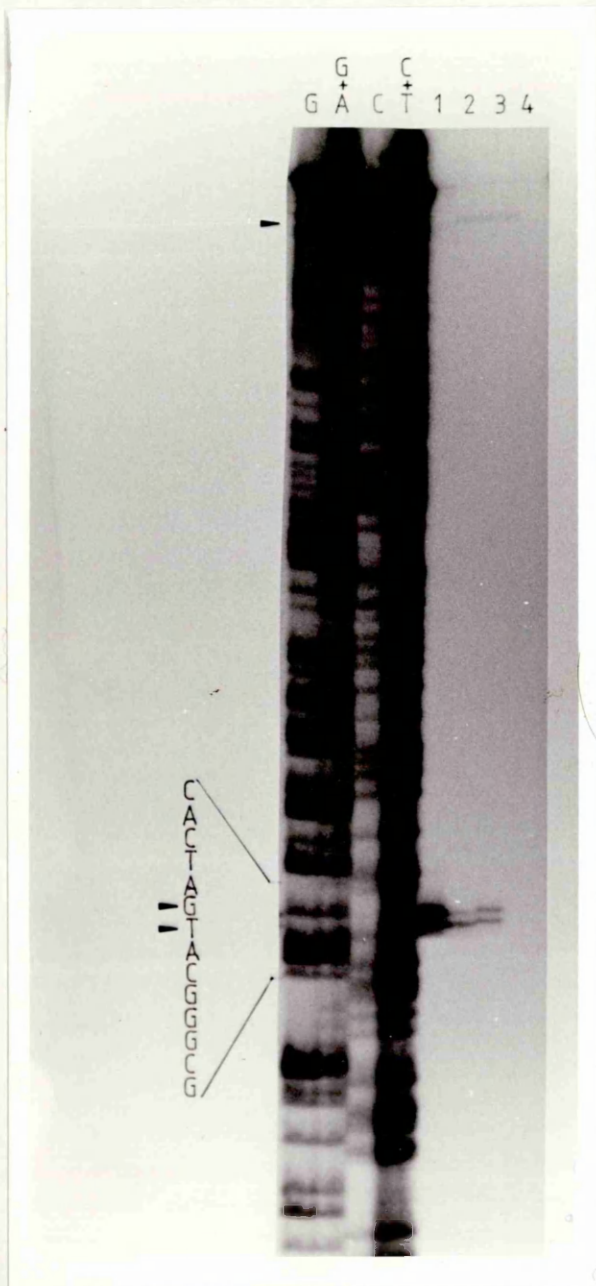
Fig 5.6 High resolution mapping of otcIp1 and otcIp2

Arrows indicate the start points coincident with the bands

0.1pM of probe 3 (approx. 2.5 x 10^6cpm/pM) was hybridised with samples in lanes 1-4 at 62°C

Lane 1: 40ug of S.lividans TK24 (containng pDM131) RNA

Lane 2: 40ug of S.rimosus 4018 RNA (isolated from cells cultured in T.S.B.)

Lane 3: 40ug of S.rimosus 4018 RNA (isolated from cells cultured in T.S.B. with exogenous OTC)

Lane 4: 40ug of tRNA

Fig 5.7 High resolution S1 mapping of <u>otcpA1-6</u>

Arrows indicate the start point coincident with the bands

0.05pM of probe 5 (approx. 7×10^6cpm/pM) was hybridised with samples in lanes 1-3 at 59°C

Lane 1: 40ug of <u>S.rimosus</u> RNA (isolated from cells cultures in T.S.B. with exogenous OTC)

Lane 2: 40ug of <u>S.lividans</u> TK24 (containing pIJ487) RNA

Lane 3: 40ug of <u>S.lividans</u> TK24 (containing pDM131) RNA

The hybridisation temperature was lowered to 59°C and the experiment repeated as above. Six protected bands were identified within a 40bp region with mRNA from S.lividans TK24 containing pDM131 (Fig 5.7). The most intense bands indicating transcriptional start sites were found at nt 211 and nt 225 (otcpA1 and otcpA3 promoters respectively). The protected band at nt 242 (otcpA5 promoter) was also relatively intense and the three weakest bands were at nt 222, nt 238 and nt 248 (otcpA2, otcpA4 and otcpA6 promoters respectively). No protected bands corresponding to those above were identified with control RNA from S.lividans TK24 containing pIJ487 or from S.rimosus 4018 RNA even after a five day exposure of the gel. A seventh protected band at nt 237 occurred with all RNA samples (Fig 5.7). The full length protection of the probe, caused by DNA/DNA hybrids, is particularly intense in the control track (Fig 5.7). However this was not a representative result because in all other experiments (data not shown) the full length protected band in control hybridizations was only as intense as the other samples.

5.2.4 Low Resolution S1 Mapping of the 3' end of the otcI transcript

Probes 4 and 6 (Fig 5.3) were used to map the 3' end of the otcI transcript. Probe 4 was hybridised to RNA from S.rimosus 4018 (cultures grown in TSB with exogenous OTC) at 68°C for 3hrs, digested with S1 nuclease as described in Section 2.15 and the samples electrophoresed on a 1.2% (w/v) alkaline agarose gel. Probe 4 protected a fragment of approximately 1300bp (Fig 5.4 a,b). Probe 6 was hybridised to RNA from S.rimosus 4018 and S.rimosus M15883 containing pPZ46 (Fig 1.4), at 68°C for 3hrs and digested with S1 nuclease as described in Section 2.15. The products were electrophoresed on a 1.2% (w/v) alkaline agarose gel and protected fragments of approximately 950bp were identified (Fig 5.8). The protected band from S.rimosus M15883 (pPZ46) sample was more intense, due to the high copy number of the plasmid in comparison to the single copy of otcI from S.rimosus 4018. These data indicated that the transcript terminates approximately 250bp before the SstI site.
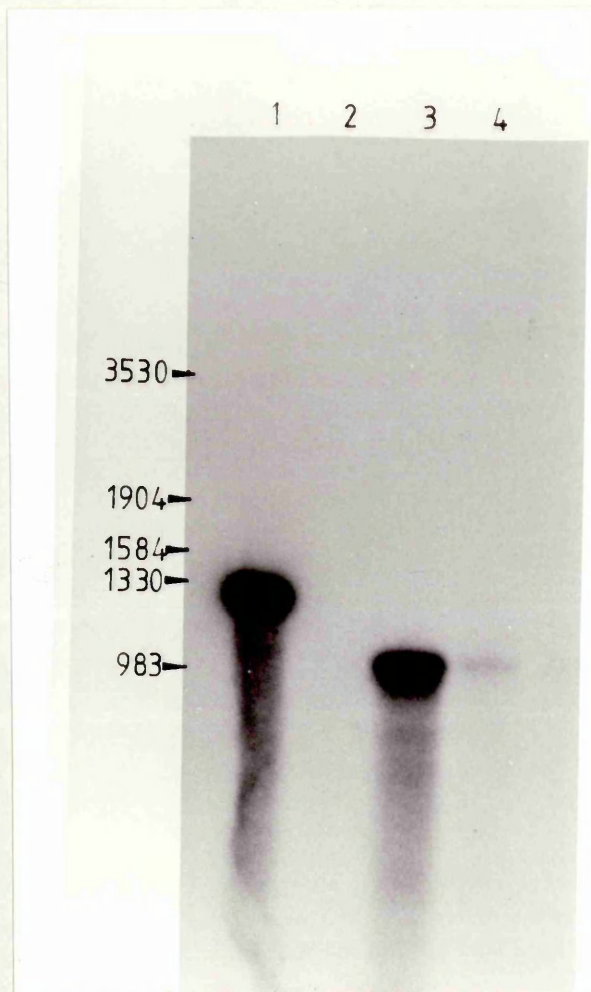
Fig 5.8 Low resolution S1 mapping of the termination point of the
otcI transcript

Probe 6 was hybridised to samples in lane 2, 3 and 4 at 58°C

Lane 1: Probe 6

Lane 2: 40ug of tRNA

Lane 3: 40ug of S.rimosus (containing pPZ46) RNA

Lane 4: 40ug of S.rimosus RNA

Divergent transcriptional start sites have been mapped upstream of the otcI structural gene. The promoters of the two start sites initiating transcription toward the otcI gene, have been named otcIp1 and otcIp2. 6 start sites were identified initiating divergent transcription, in relation to otcIp1 and otcIp2, and the potential promoters have been named otcpA1-6. Low resolution S1 mapping of the transcriptional termination site for otcI indicated a stop site approximately 250bp before the SstI site i.e. at approximately nt 2400. A translational stop codon within the otcI orf occurs at nt 2340 which is followed by a 23bp inverted repeat (Fig 3.5). The position of the inverted repeat in the DNA sequence coincides with the 3' end of the mRNA estimated from S1 mapping data. The transcription initiation pattern of this region is diagrammed in Fig 5.9.

otcIp1 and otcIp2 were initially identified by low resolution S1 mapping which indicated the appropriate probes to be used in high resolution S1 mapping. High resolution mapping identified two transcription start points for otcIp1 i.e. at nt 220 and nt 221 with RNA from S.rimosus 4018 and S.lividans TK24 containing pDM131. The sequence of the otcIp1 promoter (Table 5.2) has been assigned to the class 1 promoters discussed in the introduction. The -10 region has the major conserved residues of the prokaryotic consensus i.e. A at position 2 and a T at position 6 of the consensus. The -35 region incorporates 4/6 residues of the prokaryotic consensus sequence - TTGACA. This promoter has not been tested in E.coli for activity but if it behaves similarly to other promoters assigned to class 1, one would expect the promoter to be functional in E.coli. The transcriptional start site of otcIp1 supports the data from promoter probe constructs pDM110 and pDM130, which are inhibited by equivalent levels of km, indicating that the promoter control sequences are located within the BamHI/BclI insert.

A transcriptional start point for otcIp2 has only been identified with RNA from S.rimosus 4018. The transcript appears to be less abundant than the otcIpI transcript. The start point maps within 5-10bp of the BamHI site. A shorter probe will be required to resolve the actual

<u>otcIp1</u>

    AGCC<u>CTGACC</u>CTGCGTGGCCGCCCTTGC<u>TACC</u>GTGAT<u>CA</u>

<u>otcpA1</u>

    ATCCGCCCGGCGGGAACACA<u>CCC</u>G<u>CGGC</u>ATGATCACGG<u>T</u>

    gatctggggaattcgagctcggtacccgggGATCACGG<u>T</u>

<u>otcpA2</u>

    GTCGCTGTGATCCGCCCGGCGGGAACACACCCGCGGGG<u>C</u>

<u>otcpA3</u>

    TGACGTCGCTGTGATCCGCCCGGCGGGAACACACCCGC<u>G</u>

<u>otcpA4</u>

    GTTCCGCCCTCTCTGACGTCGCTGTGATCCGCCCGGC<u>GG</u>

<u>otcpA5</u>

    ATCGTTCCGCCCTCTCTGACGTCGCTGTGATCCGCCC<u>GG</u>

<u>otcpA6</u>

    TGCCGCATCGTTCCGCCCTCTCTGACGTCGCTGTGATC<u>C</u>

Table 5.2 The table displays potential promoter sequences upstream of
the transcriptional start sites. The sequence upstream of <u>otcIp3a</u> is
disrupted in pDM111, this sequence is diagrammed below <u>otcIp3a</u>.
Possible consensus sequences and the nt at the start site are
underlined. Lower case letters denote vector sequence from pDM111.

```
        D   P   A   T   T   T   V   H   D   V   L   Q   R   L   N   H   G   L   I   R
>  GGATCCGGCCACCACGACGGTGCACGACGTACTGCAACGGCTCAACCACGGCCTCATCCG
<  CCTAGGCCGGTGGTGCTGCCACGTGCTGCATGACGTTGCCGAGTTGGTGCCGGAGTAGGC
          •••••→
             otcIp2


        G   D   T   S   S   S   S   V   E   E   Y   R   A   E   I   E   E   A   G   F
>  CGGCGACACGTCGTCCAGCAGCGTCGAGGAGTACCGCGCGGAGATCGAGGAGGCCGGCTT
<  GCCGCTGTGCAGCAGGTCGTCGCAGCTCCTCATGGCGCGCCTCTAGCTCCTCCGGCCGAA



        R   V   R   Q   A   E   R   I   D   N   L   L   G   D   W   L   I   V   A   V
>  CCGGGTGCGGCAGGCGGAGCGCATCGACAACCTCCTCGGTGACTGGCTGATAGTCGCCGT
<  GGCCCACGCCGTCCGCCTCGCGTAGCTGTTGGAGGAGCCACTGACCGACTATCAGCGGCA


                                    otcpA1          otcpA2 otcpA3              otcpA4
                                    ←-10            ←     ←                     ←
     K   P   *-35
>  CAAGCCCTGACCCTGCGTGGCCGCCCTTGCTACCGTGATCATGCCCGCGGGTGTGTTCCC
<  GTTCGGGACTGGGACGCACCGGCGGGAACGATGGCACTAGTACGGGCGCCCACACAAGGG
                                          BcII
                                                    •→
                                             otcIp1

otcpA5  otcpA6
←       ←
>  GCCGGGCGGATCACAGCGACGTCAGAGAGGGCGGAACGATGCGGCAGGCGACCACGAATC
<  CGGCCCGCCTAGTGTCGCTGCAGTCTCTCCCGCCTTGCTACGCCGTCCGCTGGTGCTTAG



                                                            M   N   K   L
>  CGGCAGGAAAGGTCCGCATCACGACGCTAAGGAATCCGAACCCCCTCCATGAACAAGCTG
<  GCCGTCCTTTCCAGGCGTAGTGCTGCGATTCCTTAGGCTTGGGGGAGGTACTTGTTCGAC



        N   L   G   I   L   A   H   V   D   A   G   K   T   S   L   T   E   R   L   L
>  AATCTGGGCATCCTGGCCCACGTTGACGCCGGCAAGACCAGCCTCACCGAGCGCCTGCTG
<  TTAGACCCGTAGGACCGGGTGCAACTGCGGCCGTTCTGGTCGGAGTGGCTCGCGGACGAC



        H   R   T   G   V   I   D   E   V   G   S   V   D   A   G   T   T   T   T   D
>  CACCGCACCGGTGTGATCGACGAGGTCGGCAGCGTGGACGCCGGCACCACGACGACCGAC
<  GTGGCGTGGCCACACTAGCTGCTCCAGCCGTCGCACCTGCGGCCGTGGTGCTGCTGGCTG
```

Fig 5.9 DNA sequence from nt 1-480 (Fig 3.6)
Transcriptional start points are indicated by circles and the
direction of transcription by associated arrows. Presumptive -10
and -35 regions of otcIp1 and otcIp3a are underlined. An inverted
repeat is indicated by converging arrows. The amino terminal a.a.
sequence of otcI and carboxy terminal a.a. sequence of otcZ is
indicated.

with RNA from S.lividans TK24 (pDM131) but the transcription start site is very close to the BamHI site and it is quite possible that the promoter signals have been lost or at least disrupted by subcloning using the BamHI site. The promoter probe constructs do not include any sequence beyond the BamHI site. Therefore to investigate the promoter activity of sequences 5' to this start site, it will be necessary to subclone a fragment from pPZ46, which incorporates some sequence on either side of the BamHI site but not the otcIp1 promoter. It will be interesting to see if RNA derived from S.lividans containing such a construct, identifies a transcriptional start point in S1 mapping experiments. There is no sequence data 5' to the BamH1 site so it has not been possible to identify any sequences with potential promoter activity based on similarity to known consensus sequences. The preliminary S1 mapping of a transcriptional start site for otcIp2 suggests that promoter activity identified in constructs pDM110 and pDM130 is derived from otcIp1.

No transcriptional terminator has been identified for the otcZ gene. If the otcZ transcript overlaps the otcI transcript one would have expected to have identified full length protection of the relevant probes. However the RNA was harvested from cells cultured in T.S.B. medium, which does not support antibiotic production. Therefore the absence of the otcZ transcript may be due to the fact that it is not being expressed.

The divergent transcriptional activity identified by S1 mapping is far more complex. Possible promoters for each start site have been named otcpA1-6. This experiment was conducted initially under the same conditions used for high resolution mapping of otcIp1, because the probes were identical but differentially labelled. However no hydrids were identified and the hybridisation temperature was lowered by 3°C. Under these conditions the 6 protected bands were identified, suggesting that the BamHI/BclI DNA/RNA hybrid is less stable than the BclI/DdeI DNA/RNA hybrid. No bands corresponding to the 6 protected bands were duplicated in the control sample, indicating that the hybrids were real RNA/DNA hybrids. A computer search for indirect repeats with the potential to form hairpin-loops and therefore

arteractual bands, indicated that there were no stable structures which could be digested by S1 nuclease to account for any of the six protected bands. There is one band present in all samples (Fig 5.8) which may be accounted for by an inverted repeat with the potential to form a hairpin-loop with a ΔG=-17kcal. The band occurs at base 235 which is within the loop of this possible secondary structure. Because the band appeared in all tracks including the controls it has been disregarded.

Potential promoter sequences of all 6 transcription start sites have been displayed in Table 5.2. The sequence of otcpA1 may be grouped with the class 2 promoters. The sequence from -13 to -20 has a 7/8bp match with the endoH and vph promoters. The promoter probe data indicates that otcpA1 retains some activity in pDM111, which contains only the BamHI/BclI insert. However in this construction, BclI cuts at the -9 base of otcpA1. Streptomycete DNA sequences upstream of the -9 region have been replaced by the polylinker DNA sequence which exhibits some similarity to the conserved group 2 sequence (4/8bp identical) (Table 5.2). This region of identity is located 3bp nearer to the transcriptional start site but the km resistance levels indicate that this fragment still retains promoter activity. The larger promoter probe construct pDM131, incorporates all the start sites identified and exhibits considerably increased promoter activity (Fig 5.1). It is impossible to discern if this increased promoter activity is due to transcription from the undisrupted otcpA1 promoter or from the other 5 start sites or both.

The sequences upstream of the 5 other start sites show no homology to either the prokaryotic consensus or the class 2 consensus sequence and have been grouped with the class 3, miscellaneous promoters. They also show no homology to the tentative consensus sequences, established within the group between aphp1/afsBp and hygp/orfp2 (Table 5.1). These start sites are very close to each other, similar to pA2, pA3 and pA4 which direct the divergent transcription associated with the aph gene (Bibb and Janssen 1987). The transcription start sites of otcpA2 and otcpA3 are within 3bp of each other as are the start sites of otcpA4 and otcpA5 and these could be considered as alternative start sites from the same promoter. The other start sites

are sufficiently spaced apart to be considered as separate entities. mRNA processing could result in the 6 transcription start sites identified (S1 artefacts due to secondary structure have been ruled out) but one would not expect to find a class 2 type promoter sequence, otcpA1, upstream of the shortest transcript if it is simply the product of mRNA degradation. As discussed in the introduction and above, multiple closely spaced transcription starts have been observed previously but the function of such a complex arrangement of promoters is unknown. Multiple promoters may be useful in fine regulation of expression of a transcript. If further experiments demonstrate promoter activity associated with all or even some of these transcriptional start sites, at least two classes of promoter can be assumed to be regulating transcription. However otcpA1-6 transcriptional start sites have not been identified in hybridization experiments with RNA from S.rimosus 4018 which contains otcI mRNA. This result is consistent with the data from promoter probe construct, pDM160, which exhibits no promoter activity and includes 45% of the otcI polypeptide. A plausible explanation for these two results is that transcriptional activity from otcpA1-6 is repressed by factors within the otcI structural gene. This would introduce another level of complexity in the transcriptional control of this region. It is impossible to test, in promoter probe vectors, each potential promoter in isolation for promoter activity, because the sequences overlap extensively. However substitution of the bases at -13, -14 and -15 of otcpA1 should radically decrease the activity from this promoter (Craig Binnie pers. comm.) and it may then be possible to assay the contribution of otcpA1-6 to promoter activity in pDM131.

otcIpI and otcIp2 start sites have been identified with RNA from S.rimosus, which was cultured both in the presence and in the absence of exogenous OTC. otcIpA1-6 start sites have not been identified with RNA from S.rimosus,which was cultured in the presence and absence of exogenous OTC (data not shown). OTC in the medium would appear to have little or no effect on the expression of the otcI gene. However, this is not a good test of the effect of intracellular levels of OTC because this level is unknown whether the cells are cultured in the presence or absence of OTC. The effect of OTC on the expression of otcI needs to be investigated further.

The transcription pattern 5' to otcI is similar in many respects to the transcription patterns of the resistance genes discussed in section 5.1.4.. Transcription of otcI is directed by two promoters separated by 210-215bp. otcIp1 is expressed in both the host strain and in S.lividans and otcIp2 activity needs to be investigated in S.lividans. In addition divergent transcriptional activity from multiple transcriptional start sites occurs, overlapping with otcIp1. One respect in which otcI transcription pattern differs from the other resistance genes, is the close proximity of otcZ, a production gene (Fig 5.9). This creates a situation of overlapping transcripts. The function of the otcpA transcript has yet to be determined but it is highly probable that it is involved in the regulation of the OTC production pathway. The data resulting from promoter probe constructs and S1 mapping have defined certain points of the regulation of these transcription units. Transcription from otcpA1-6 is repressed in both promoter probe construct pDM160 and in S.rimosus 4018 cultures which are expressing otcI, indicating that the otcI polypeptide possibly controls the expression of the otcpA transcript. The activity of otcIp1 in pDM161 is increased, which implies that transcription from overlapping promoters otcpA1-6 is partially repressing the activity of otcIp1 in pDM110 and pDM130. As suggested in Chapter 3 the otcpA transcript may be involved as an anti-sense regulator of the production pathway genes. Regulation at the level of promoter consensus sequences must also be considered because two promoters have been identified, otcIp1 and otcpA1, which have the potential to be recognized by different sigma factors. The role of otcIp2 in the regulation of otcI expression needs to be investigated by the isolation and further S1 mapping of this promoter.

A more comprehensive understanding of the regulation of these genes requires consideration of the effect of OTC production on expression. It will be necessary to harvest RNA samples for S1 mapping experiments from different time points in the life cycle and to assay the cultures for OTC, in order to develop a clearer picture of this complex regulation mechanism.

Chapter 6

Concluding Remarks

Sequencing of the BamHI/SstI fragment encoding otcI identified the otcI open reading frame with an initiation codon at nt 349-352. The sequence also indicated that the orf for the otcZ structural gene terminated at nt 190. The transcription initiation pattern of pattern of the otcI gene presented in chapter 5 is derived from high resolution S1 mapping and is consistent with activity from promoter probe constructs. Two transcription start points separated by approximately 215bp direct the transcription of otcI. The sequence upstream of the start site at nt 219-220 (the otcIp1 promoter) resembles a streptomycete class 1 promoter which is the counterpart of the E.coli consensus promoter in Streptomyces. The sequence upstream of the start site at nt 5-10 (otcIp2) is unknown because it lies outside the sequenced BamHI/SstI fragment. The otcIp2 transcrit was slightly less abundant than the otcIp1. otcIp2 needs to be investigated further to identify sequences 5' to the transcription start site and also in promoter probe vectors. The transcription start site was not identified with RNA from S.lividans TK24 containing pDM131, which suggests that none of the promoter probe constructs contain this promoter. It will be interesting to identify which class of promoter otcIp2 resembles, in terms of the regulation of the otcI gene.

Divergent transcriptional activity was identified in promoter probe constructs and 6 clustered transcription start sites were identified at nt 211, 222, 225, 238, 242 and 248 (the otcpA1-6 promoters) with RNA from S.lividans TK24 containing pDM131. No transcription start sites were identified with RNA from S.rimosus 4018 which is consistent with the km sensitive phenotype of pDM160. It remains to be established that the transcription start sites otcpA1-6, are all preceeded by functional promoters. Because the start sites are closely clustered it is difficult to study potential promoter sequences in isolation. However it is possible to mutagenize the conserved sequence of otcpA1 and subsequently identify the effect on promoter activity from this region. The unusual, but not unprecedented, clustering of transcriptional start sites may play a role in the fine regulation of expression of this transcript. Two start sites within 5bp of each other, have been identified 5' to the

84

gal operon in E.coli (Queen and Rosenberg 1981). Initiation from P1 is activated by CRP-cAMP and initiation from P2 is induced by glycerol and inhibited by CRP-cAMP. The P2 transcript synthezises galE four times more efficiently than P1 mRNA and it has been suggested that the differential translation efficency of the P1 and P2 transcripts is directly related to the increase in size of the P2 transcript. The developmentally regulated gene, spoVG from B.subtilis, is also expressed from two overlapping promoters, which direct transcription from sites separated by 10bp (Johnson et al 1983). These promoters are recognized by two distinct sigma factors and in addition the expression of spoVG is regulated by the spo0 regulatory gene products. The complexities of this regulation system are not fully understood but they do show some similarity to the unusual transcription patterns which are being discoveed in Streptomyces species.

An incomplete picture of the regulation of the otcI gene has been established. The leftward divergent transcript is totally repressed in the presence of sequences within the N terminal end of otcI, most probably by the otcI polypeptide. When the leftward transcript is expressed it has a partially repressive affect on otcIp1, which may be explained by the effect of promoter sequences and transcription start sites overlapping. The role of otcIp2 has yet to be established.

The function of the leftward divergent transcript is unknown, but the fact that it is regulated by otcI indicates it may be involved in the regulation of OTC biosynthesis. There is no evidence that the transcript is translated but a potential ribosome binding site occurs at nt 150-156, 6bp upstream of a translational initiation codon. Translation of this transcript would create a situation of overlapping polypeptides with otcZ. An in vitro transcription/translation system may be used to look for a potential polypeptide product from this transcript. The leftward transcript also has the potential to act as an anti-sense RNA regulator of the expression of otcZ. This would be an unusual anti-sense RNA regulation mechanism because no anti-sense RNA regulators have been found which mediate control at the 3' end of a mRNA or extend along the full length of a mRNA.

To understand the regulation mechanisms it will be necessary first of all to identify the transcription end point of the leftward transcript and the transcription start site of otcZ. The role of the 6 transcription start sites otcpA1-6 must also be elucidated. The sequence of otcpA1 suggests that it is likelt to be transcribed by E-sigma49, but the role of sigma 49 in the temporal expression of genes has yet to be established. The sequences upstream of the 5 other transcriptional start sites have been assigned to the class 3 streptomycete promoters. Promoter probe constructs demonstrate *in vivo* promoter activity from this region, but further experiments are required to demonstrate if all or at least some of these start sites are preceeded by functional promoters. Clustered transcription start sites have been identified in other streptomycete species (Bibb and Janssen 1987) but these have also not been tested for activity in isolation and the function of clustered has yet to be determined.

All promoter probe constructs were analysed in S.lividans TK24 and S1 mapping experiments used RNA from S.lividans TK24 containing promter probe constructs or RNA from S.rimosus 4018 cultured in T.S.B.. Regulatory control of the region may be altered in the heterologous host S.lividans TK24 and T.S.B. medium does not support OTC biosynthesis, so again regulation may be altered from the usual mechanisms. To get a better understanding of the regulation of expression of these transcripts it will be necessary to harvest RNA from S.rimosus under controlled conditions. RNA samples should be harvested from S.rimosus 4018 grown in production medium (where the onset of OTC biosynthesis will be monitored) at different time points in the life cycle.

The organization of the promoters 5' to the otcI structural gene is very similar to the examples discussed in section 5.1.4. The promoter probe data in conjunction with the S1 analysis have indicated that these transcripts are coordinately regulated and have developed a picture of possible mechanisms of control. Further investigations of

86

concepts for the regulation of resistance genes and antibiotic biosynthetic genes in Streptomyces

The BamHI/SstI fragment has been sequenced. The predicted orf for otcI encodes a protein of 71.329Kd. The predicted amino acid sequence has been compared to the predicted amino acid sequence of tetM from Streptococcus. The 130 a.a. at the N terminal ends of both proteins are 66% homologous and three further small regions dispersed throughout the sequence show equivalent homology. A functional role has been proposed for these highly conserved regions. The overall homology between the two proteins contributes further evidence to the Benveniste and Davies (1973) hypothesis that antibiotic resistant determinants in non-producing species are derived from antibiotic producing species like the streptomycetes.

The highly conserved N terminal ends of otcI and tetM also exhibit similarities to the N terminal end of Ef-Tu from E.coli. This region includes the GTP binding site of EF-Tu which occurs at a.a. 4-20 in otcI and tetM and a.a. 12-28 in EF-Tu. Ef-Tu and otcI are 48% homologous over 150 a.a at the N terminal end when the chemical equivalence of amino acids is considered. This startling homology suggests that a functional relationship exists between these 3 proteins. It is possible that the tetracycline resistance genes act as an alternative elongation factor which protects the translation machinery from the inhibitory effect of tetracycline. To test if the otcI gene product plays a functional role in protein synthesis it will be necessary to reconstitute ribosomes in vitro, replacing the normal EF-Tu with the otcI polypeptide.

Bibliography.

Ambartsumyan NS, Mazo AM (1980) Elimination of the secondary structure effect in gel sequencing of nucleic acid. FEBS Lett. **114**:265-268

Antonov PP, Ivanov IG, Markov GG, Benign R (1978) Reassociation analysis of DNA in studying the genome size of Streptomyces. Studia Biophysica, Berlin **69**:67-74

Arai K, Clark BFC, Duffy L, Jones MD, Kaziro Y, Laursen RA, L'Italien J, Miller DL, Nagarkatti S, Nakamura S, Neilsen KM, Petersen TE, Takahashi K, Wade M (1980) Primary structure of elongation factor Tu from E.coli. Proc. Natl. Acad. Sci. USA **77**:1326-1330

Bannister JV, Parker MW (1985) The presence of a copper/zinc superoxide dismutase in the bacterium Photobacterium eriognathi: A likely case of gene transfer from eukaryotes to prokaryotes. Proc. Natl. Acad. Sci. USA **82**:149-152

Benveniste R, Davies J (1973) Aminoglycoside antibiotic-inactivating enzymes in actinomycete similar to those present in clinical isolates of antibiotic-resistant bacteria. Proc. Natl. Acad. Sci. USA **70**:2276-2280

Bibb MJ, Ward JM, Hopwood DA (1978) Transformation of plasmid DNA into Streptomyces at high frequency. Nature **274**:398-400

Bibb MJ, Schottel JL, Cohen SN (1980) A DNA cloning system for interspecies gene transfer in antibiotic producing Streptomyces. Nature **284**:526-531

Bibb MJ, Cohen SN (1982) Gene expression in Streptomyces: Construction and application of promoter probe plasmid vectors in Streptomyces lividans. Mol. Gen. Genet. **187**:265-277

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

Bibb MJ, Janssen GR, Ward JM (1986) Cloning and analysis of the promoter region of the erythromycin-resistance gene (ermE) of Streptomyces erythraeus. Gene **41**:E357-367

Bibb MJ, Janssen GR (1987) Unusual features of transcription and translation of antibiotic resistance genes in antibiotic producing Streptomyces. In Genetics of Industrial Microorganisms pp309-318 Eds. Alacevic M, Hranuel D, Toman P.

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

Burdett V, Inamine J, Rajagopalan S (1982) Heterogeneity of tetracycline resistance determinants in <u>Streptococcus</u>. J.Bacteriol. **149**:995-1004

Burdett V (1986) Streptococcal tetracycline resistance mediated at the level of protein synthesis. J. Bacteriol **165**:564-569

Butler MJ, Hunter IS, Sugden DA, Warren M (1986) Cloning and molecular architecture of genes involved in the biosynthesis of oxytetracycline. Abst. Fifth Int. Sym. Genet. Ind. Microorg. In Split, Yugoslavia p161

Buttner MJ, Joseph R, Jones GH, Ward JM, Fearnley IM, Bibb MJ (1986) Cloning and analysis of the gene encoding the extracellular agarase of <u>Streptomyces</u> <u>coelicolor</u>. Abstr Fifth International Symposium of the Genetics of Industrial Microorganisms. p161

Buttner MJ. Brown NL (1987) Two promoters from the Streptomyces plasmid pIJ101 and their expression in <u>E.coli</u>. Gene **51**:179-186

Chambers AE, Hunter IS (1984) Construction and use of a bifunctional streptomycete cosmid. Biochem. Soc. Trans. **12**:644-645

Chater KF, Merrick MJ (1979) <u>Streptomyces</u>. In. The biology of the Actinomycetes. Eds. Goodfellow M, Mordarski M, Williams ST. pp229-286

Chater KF, Bruton CJ, Springer W, Suarez JE (1981) Dispensable sequences and packaging constraints of DNA from the <u>Streptomyces</u> temperate phage ØC31. Gene **15**:249-256

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

Chater KF (1984) Morphological and physiological differentiation in <u>Streptomyces</u>. In. Microbial development. Eds. Losick R, Shapiro L. pp89-115

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

Coleman J, Green P, Inouye M (1984) The use of RNAs complementary to

specific mRNA to regulate the expression of individual bacterial genes. Cell 37:429-436

Cundliffe E, Thompson J (1979) Ribose methylation and resistance to thiostrepton. Nature 278:859-861

Davies J, Houk C, Yagisawa M, White IJ (1979) Occurrence and function of aminoglycoside modifying enzymes. In Genetics of Industrial Microorganisms. American Society for Microbiology, Washington DC. Eds. Sebek OK, Laskin AJ. pp166-169

Deng Z, Kieser T, Hopwood DA (1986) Expression of Streptomyces plasmid promoter in E.coli. Gene 43:295-300

Dodd IB, Egan JB (1987) Systematic method for the detection of potential lambda cro-like DNA-binding regions in proteins. J. Mol. Biol. 194:557-564

Doolittle RF (1981) Similar amino acid sequences: Chance or common ancestry? Science 214:149-159

Doyle D, Lee KY, Butler MJ, Hunter IS (1986) Molecular characterisation of oxytetracycline resistance in Streptomyces rimosus. Abstr. Fifth International Symposium on the Genetics of Industrial Microorganisms. pp136

Ellison MJ, Kelleher RJ, Rich A (1985) Thermal regulation of B-galactosidase synthesis using anti-sense RNA directed against the coding portion of the mRNA. J. Biol. Chem. 260:9085-9087

Enquist LW, Bradley SG (1971) Characterisation of deoxyribonucleic acid from Streptomyces venezuelae species. Dev. Ind. Microbiol. 12:225-236

Feitelson JS, Malpartida F, Hopwood DA (1985) Genetic and biochemical characterisation of the red gene cluster of Streptomyces coelicolor A3(2). J. Gen. Micro. 131:2431-2441

Fey G, Reiss M, Hersten H (1973) Interaction of tetracyclines with ribosomal subunits from E.coli. A fluorometric investigation. Biochem. 12:1160-1164

Fickett JW (1982) Recognition of protein coding regions in DNA sequences. Nucleic Acid Res. 10:5303-5318

Fisher SH, Bruton CJ, Chater KF (1987) The glucose kinase gene of Streptomyces coelicolor and its use in selecting spontaneous deletions for desired regions of the genome. Mol. Gen. Genet. 206:35-44

Fishman SE, Hershberger CL (1983) Amplified DNA in _Streptomyces fradiae_. J. Bacteriol **155**:459-466

Floss HG (1987) Hybrid antibiotics- the contribution of the new gene combinations. Biotechnology **5**:111-115

Fornwald JA, Schmitt FJ, Adams CW, Rosenberg M, Brawner ME (1987) Two promoters, one inducible and one constitutive, control transcription of the _Streptomyces lividans_ galactose operon. Proc. Natl. Acad. Sci. USA **84**:2130-2134

Friend EJ, Hopwood DA (1971) The linkage map of _Streptomyces rimosus_. J. Gen. Micro. **68**:187-197

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

Gray G, Selzer G, Buell G, Shaw P, Escanez S, Hofer S, Voegeli P, Thompson CJ (1984) Synthesis of bovine growth hormone by _Streptomyces lividans_. Gene **32**:21-30

Green PJ, Pines O, Inouye M (1986) The role of antisense RNA in gene regulation. Ann. Rev. Biochem. **55**:569-597

Grossman AD, Erickson JW, Gross CA (1984) The htpR gene product of _E.coli_ is a sigma factor for heat shock promoters. Cell **38**:383-390

Grossman AD, Losick R (1986) RNA polymerase heterogeneity in bacteria. In. Regulation of Gene Expression- 25 years on. Eds. Booth IC, Higgins CF. pp127-138

Haggerty DM, Schleif RF (1976) Location in bacteriophage Lambda DNA of cleavage sites of the site-specific endonuclease from _Bacillus amyloliquetaciens_ H. J. Virol. **18**:659-663

Hasegawa M, Hintermann G, Simonet JM, Crameri R, Piret J, Hutter R (1985) Certain chromosomal regions in _Streptomyces glaucesens_ tend to carry amplifications and deletions. Mol. Gen. Genet. **200**:375-384

Hawley DK, McClure WR (1983) Compilation and analysis of _E.coli_ promoter DNA sequences. Nucleic Acid Res. **11**:2237-2255

Helling RB, Goodman HN, Boyer HW (1974) Analysis of _EcoRI_ fragments of DNA from lambdoid bacteriophage and other viruses by agarose gel electrophoresis. J. Virol. **14**:1235-1244

creates targeted breakpoints for DNA sequencing. Gene **28**:351-359

Hershberger CL, Fishman SE (1985) Amplified DNA: Structure and significance. Microbiology (ASM) Ed. Leive L pp427-430

Hillen W, Schollmeier K Gatz C (1984) Control of expression of the transposon Tn10-encoded tetracycline resistance operon: 2. Interaction of RNA polymerase and tet repressor with the tet regulatory region. J. Mol. Biol. **172**:185-202

Hirschmann J, Wong P-K, Sei K, Keenes J, Kustu S (1985) Products of nitrogen regulatory genes ntrA and ntrC of enteric bacteria activate glnA transcription in vitro: Evidence that the ntrA product is a sigma factor. Proc. Natl. Acad. Sci. USA **82**:7525-7529

Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. Analytical Biochem **114**:193-197

Hopwood DA, Kieser T, Wright H, Bibb MJ (1983) Plasmids, recombination and chromosomal mapping in Streptomyces lividans 66. J. Gen. Micro. **129**:2257-2269

Hopwood DA, Malpartida F, Kieser HM, Ikeda H, Duncan J, Fujii I, Rudd BA, Floss HG, Omura S (1985) Production of hybrid antibiotics by genetic engineering. Nature **314**:642-644

Hopwood DA, Bibb MJ, Chater KF, Kieser T, Bruton CJ, Kieser HM, Lydiate DJ, Smith CP, Ward JM, Schrempf H (1985) Genetic manipulation od Streptomyces: A laboratory manual.

Hopwood DA, Bibb MJ, Chater KF, Janssen GR, Malpartida F, Smith CP (1986) Regulation of gene expression in antibiotic-producing Streptomyces. In. Regulation of gene expression- 25 years on. Eds Booth IR, Higgens CF. pp251-277

Horinouchi S Beppu T (1985) Construction and application of a promoter probe plasmid that allows chromogenic identification in Streptomyces lividans. J. Bacteriol. **162**:406-412

Horinouchi S, Suzuki H, Beppu T (1986) Nucleotide sequence of afsB, a pleiotrophic gene involved in secondary metabolism in Streptomyces coelicolor A3(2) and Streptomyces lividans. J. Bacteriol. **168**:257-269

Horinouchi S, Furuya K, Nishiyama M, Suzuki H, Beppu T (1987) Nucleotide sequence of the streptothricin acetyltransferase gene from Streptomyces lavendulae and its expression in heterologous hosts. J. Bacteriol. **169**:1929-1937

Hoshiko S, Makabe O, Nojiri C, Katsumata K, Satoh E, Nagaokra K (1987) Molecular cloning and characterisation of the <u>Streptomyces hygroscopicus</u> alpha-amylase gene. J. Bacteriol.**164**:1029-1036

Hunter IS, Friend EJ (1984) 'Restriction-deficient' mutants of industrial <u>Streptomyces</u>. Biochem. Soc. Trans. **12**:643-644

Janssen GR, Bibb MJ, Smith CP, Ward JM, Kieser T, Bibb MJ (1985) Isolation and analysis of <u>Streptomyces</u> promoters. In. Microbiology-1985 (A.S.M.) Ed. Leive L. pp392-396

Jaurin B, Cohen SN (1984) <u>Streptomyces lividans</u> RNA polymerase recognises and uses <u>E.coli</u> transcriptional signals. Gene **28**:83-91

Jaurin B, Cohen SN (1985) <u>Streptomyces</u> contain <u>E.coli</u>-type A+T rich promoters having novel structures. Gene **39**:191-201

Johnson WC, Moran CP, Losick R (1983) Two RNA polymerase sigma factors from <u>Bacillus subtilis</u> discriminate between overlapping promoters for a developmentally regulated gene. Nature **300**:800-804

Kieser T, Hopwood DA, Wright HM, Thompson CJ (1982) pIJ101, a multicopy broad host range <u>Streptomyces</u> plasmid: Functional analysis and development of DNA cloning vectors. Mol. Gen. Genet. **185**:223-238

Lake JA (1985) Evolving ribosome structure: Domain in archaebacteria, eubacteria, eocytes and eukaryotes. Ann. Rev. Biochem. **54**:507-530

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

Malpartida F, Hopwood DA (1984) Molecular cloning of the whole biosynthetic pathway of a <u>Streptomyces</u> antibiotic and its expression in a heterologous host. Nature **309**:462-464

Malpartida F, Hopwood DA (1986) Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in <u>Streptomyces coelicolor</u> A3(2). Mol. Gen. Genet. **205**:66-73

Malpartida F, Hallam SE, Kieser HM, Motamedi H, Hutchinson CR, Butler MJ, Sugden DA, Warren M, McKillops C, Bailey CR, Humphreys GO, Hopwood DA (1987) Homology between <u>Streptomyces</u> genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes. Nature 325:818-821

tetM tetracycline resistance determinant of the <u>Streptococcal</u> conjugative shuttle transposon Tn1545. Nucleic Acid Res. **14**:7047-7058

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560-564

Mendez B, Tachibana C, Levy SB (1980) Heterogeneity of tetracycline resistance determinants. Plasmid **3**:99-108

Mikulik K, Karnetovei J, Quyen N, Blumauerova M, Komersoua I, Vanek Z (1971) Interaction of tetracycline with protein synthesising system of <u>Streptomyces aureofaciens</u>. J. Antibiotics **24**:801-809

Mikulik K, Jiranova A, Janda I, Weiser J (1983) Susceptibility of ribosomes of the tetracycline-producing strain of <u>Streptomyces aureofaciens</u> to tetracyclines. FEBS. Lett. **152**:125-130

Mizuno T, Chou M-Y, Inouye M (1984) A unique mechanism regulating gene expression: Translational inhibition by a complementary RNA transcript (micRNA). Proc. Natl. Acad. Sci. USA **81**:1966-1970

Moazed D, Noller HF (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. Nature **327**:389-394

Moller W, Amons R (1985) Phosphate binding sequences in nucleotide binding proteins. FEBS. Lett. **186**:1-7

Moran CP, Lang N, Banner CDB, Haldenwang WG, Losick R (1981) Promoter for a developmentally regulated gene in <u>Bacillus subtilis</u>. Cell **25**:783-791

Moran CP Jr, Johnson WC, Losick R (1982) Close contacts between sigma-37 RNA polymerase and a <u>B.subtilis</u> chromosomal promoter. J. Mol. Biol. **162**:709-713

Motamedi H, Hutchinson CR (1987) Anthracycline antitumour antibiotics. Organisation of the genes for tetracenomycin C production in <u>Streptomyces glaucesens</u>. In. Genetics of industrial microorganisms. Eds. Alacevic M, Hranueli D, Toman P. pp355-364

Murakami T, Anzai H, Imai S, Satoh A, Nagaoka K, Thompson CJ (1986) The bialaphos biosynthetic genes of <u>Streptomyces hygroscopicus</u>: Molecular cloning and characterisation of the gene cluster. Mol. Gen. Genet. **205**:42-50

Muto A, Asawa S (1987) The guanine and cytosine content of genomic

Nakano MM, Mashiko H, Ogawara H (1984) Cloning of the kanamycin resistance gene from a kanamycin producing Streptomyces species. J. Bacteriol. 157:79-83

Nakano MM, Ogawara H (1987) Isolation and characterisation of ribosome resistance gene from Streptomyces kanamyceticus. In. Genetics of Industrial Microorganisms. Eds. Alacevic M, Hranuel D, Toman P. pp177-184

Nomura M, Mizushima S, Ozaki M, Traub P, Lowry CV (1969) Structure and function of ribosomes and their molecular components. Cold Spring Harbor Sym. Mol. Biol. 34:49-61

Ohnuki T, Katoh T, Imanaka T, Aiba S (1985) Molecular cloning of tetracycline resistance genes from Stretomyces rimosus in Streptomyces griseus and characterisation of the cloned genes. J. Bacteriol. 161:1010-1016

Ohnuki T, Imanaka T, Aiba S (1985) Self cloning in Streptomyces griseus of a str gene cluster for streptomycin biosynthesis and streptomycin resistance. J. Bacteriol. 164:85-94

Piendl W, Flett F, Altenbucher J, Platt J, Cullum J (1986) Analysis and comparison of the amplifiable DNA regions in Streptomyces lividans 66 and Streptomyces coelicolor A3(2). Abstr. Fifth International Symposium on the Genetics of Industrial Microorganisms. p42

Pulido D, Vara JA, Jimenez A (1986) Cloning and expression in biologically active form of the gene for human interferon alpha 2 in Streptomyces lividans. Gene 45:167-176

Queen C, Rosenberg M (1981) Differential translation efficiency explains discoordinate expression of the galactose operon. Cell 25:241-249

Raibaud O, Schwartz M (1984) Positive control of transcription initiation in bacteria. Ann. Rev. Genet. 18:173-206

Rhodes PM, Winskill N, Friend EJ, Warren M (1981) Biochemical and genetic characterisation of Streptomyces rimosus mutants impaired in oxytetracycline biosynthesis. J. Gen. Micro. 124:329-338

Rhodes PF, Hunter IS, Friend EJ, Warren M (1984) Recombinant DNA methods for the oxytetracycline producer Streptomyces rimosus. Biochem. Soc. Trans. 12:586-587

95

sequences in strains of <u>Streptomyces</u> produced by an interspecific protoplast fusion. Mol. Gen. Genet. **182**:336-340

Sanger F, Nicklen S, Coulson AR (1977a) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463-6467

Sanger T, Air GN, Barell BG, Brown NL, Coulson AR, Fiddes JC, Hutchinson CA III, Slocombe PM, Smith M (1977b) Nucleotide sequence of bacteriophage ØX174 DNA. Nature **265**:687-691

Scherer GFE, Walkinshaw MD, Arnott S, Morre DJ (1980) The ribosome binding sites recognised by <u>E.coli</u> ribosomes have regions with signal character in both the leader and protein coding segments. Nucleic Acid Res. **8**:3895-3907

Schrempf H (1983) Deletion and amplification of DNA sequences in melanin-negative variants of <u>Streptomyces reticuli</u>. Mol. Gen. Genet. **189**:501-505

Seno ET, Bruton CJ, Chater KF (1984) The glycerol utilization operon of <u>Streptomyces coelicolor</u>: Genetic mapping of <u>gly</u> mutations and the analysis of cloned <u>gly</u> DNA. Mol. Gen. Genet. **193**:119-128

Serfling E, Jasin M, Schaffner W (1985) Enhancers and eukaryotic gene transcription. Trends in Genetics **1**:224-230

Singer CE, Ames BN (1970) Sunlight ultraviolet and bacterial DNA base ratios. Science **170**:822-826

Skinner RH, Cundliffe E (1980) Resistance to the antibiotics viomycin and capriomycin in the <u>Streptomyces</u> species which produce them. J. Gen. Micro. **120**:95-104

Skinner RH, Cundliffe E (1982) Dimethylation of adenine and the resistance of <u>Streptomyces erythraeus</u> to erythromycin. J. Gen. Micro **128**:2411-2416

Smith CP, Chater KF (1987) Physiology, genetics and molecular biology of glycerol utilization in <u>Streptomyces coelicolor</u>. In. Genetics of industrial microorganisms. Eds. Alacevic M, Hranuel D, Toman P. pp7-17

Stanzak R, Matsushima P, Baltz RH, Rao RN (1986) Cloning and expression in <u>Streptomyces lividans</u> of clustered erythromycin biosynthesis genes from <u>Streptomyces erythraeus</u>. Biotechnology **4**:229-232

polymerase from <u>Bacillus subtilis</u>. J. Mol. Biol. :285-297

Thompson CJ, Ward JM, Hopwood DA (1982) Cloning of antibiotic resistance and nutritional genes in <u>Streptomyces</u>. J. Bacteriol. **151**:668-677

Thompson CJ, Gray GS (1983) Nucleotide sequence of a streptomycete aminoglycoside phosphotransferase gene and its relationship to phosphotransferases encoded by resistance plasmids. Proc. Natl. Acad. Sci. USA **80**:5190-5194

Tinoco I, Borer PN, Dengler B, Levine MD (1973) Improved estimation of secondary structure in Ribonucleic acids. Nature New Biol. **246**:40-41

Truitt CL, Ray GL, Trempy JE, Da-Jian Z, Halenwang WG (1985) Isolation of <u>Bacillus subtilis</u> mutants altered in expression of a gene transcribed <u>in vitro</u> by a minor form of RNA polymerase (E-sigma37). J. Bacteriol. **161**:515-522

Uchiyama H, Weisblum B (1985) N-methyltransferase of <u>Streptomyces erythraens</u> that confers resistance to the macrolide-lincosamide-streptoyramin B antibiotics: Amino acid sequence and its homology to cognate R-factor enzymes from pathogenic bacilli and cocci. Gene **38**:103-111

Vogtli M, Hutter R (1987) Characterisation of the hydroxystreptomycin phosphotransferase gene (<u>sph</u>) of <u>Streptomyces glaucescens</u>: Nucleotide sequence and promoter analysis. Mol. Gen. Genet. **208**:195-203

Ward JM, Janssen GR, Kieser T, Bibb MJ, Butler MJ, Bibb MJ (1986) Construction and characterisatioin of a series of multicopy promoter probe vectors for <u>Streptomyces</u> using the aminoglycoside phosphotransferase gene from Tn5 as indicator. Mol. Gen. Genet. **203**:468-478

Westpheling J, Ranes M, Losick R (1985) RNA polymerase heterogeneity in <u>Streptomyces coelicolor</u>. Nature **313**:22-27

Williams G, Smith I (1979) Chromosomal mutations causing resistance to tetracycline in <u>Bacillus subtilis</u>. Mol. Gen. Genet. **177**:23-29

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**:103-119

Young M, Cullum J (1987) A plausible mechanism for large-scale

chromosomal DNA amplification in streptomycetes.   FEBS Lett.
212:10-14

Zalacain M, Gonzalez A, Guerrero MC, Mattaliano RJ, Malpartida T,
Jimenez A (1986)   Nucleotide sequence of the hygromycin B
phosphotransferase gene from <u>Streptomyces hygroscopicus</u>.   Nucleic
Acid Res. **14:**1565-1581

Zuber P, Losick R (1983)   Use of <u>lacZ</u> fusion to study the role of the
<u>spo</u>0 genes of <u>B.subtilis</u> in developmental regulation.   Cell
35:275-283

APPENDIX: Restriction sites found in sequence of _otc_ resistance gene

From   1 to   2657 there is/are   2 occurrence(s) of AatII
Sequence(s): /GACGTC/

   258/1 2128/1

From   1 to   2657 there is/are   1 occurrence(s) of AccI
Sequence(s): /GTATAC/GTACAC/GTCTAC/GTCGAC/

   1234/4

From   1 to   2657 there is/are   7 occurrence(s) of AcyI
Sequence(s): /GACGCC/GACGTC/GGCGCC/GGCGTC/

   258/2 385/1   457/1   658/4   826/1 2100/1 2128/2

From   1 to   2657 there is/are   0 occurrence(s) of AflII
Sequence(s): /CTTAAG/

From   1 to   2657 there is/are   4 occurrence(s) of AflIII
Sequence(s): /ACACGT/ACATGT/ACGCGT/ACGTGT/

   66/1   923/3   954/1 1103/3

From   1 to   2657 there is/are   0 occurrence(s) of AhaIII
Sequence(s): /TTTAAA/

From   1 to   2657 there is/are   7 occurrence(s) of AluI
Sequence(s): /AGCT/

   356/1   488/1   902/1 1700/1 2029/1 2542/1 2653/1

From   1 to   2657 there is/are   2 occurrence(s) of ApaI
Sequence(s): /GGGCCC/

   842/1 2284/1

From   1 to   2657 there is/are   7 occurrence(s) of ApaLI
Sequence(s): /GTGCAC/

   20/1 1015/1 1069/1 1099/1 1849/1 2014/1 2167/1

From   1 to   2657 there is/are   5 occurrence(s) of AvaI
Sequence(s): /CCCGAG/CCCGGG/CTCGAG/CTCGGG/

   573/2   613/4 1319/2 2278/1 2375/2

From   1 to   2657 there is/are 16 occurrence(s) of AvaII
Sequence(s): /GGACC/GGTCC/

   311/2   518/2   633/2   689/1 1034/2 1059/1 1146/1 1399/2
   1512/1 1563/2 1679/1 1755/1 1977/2 2350/1 2385/2 2423/1

From   1 to   2657 there is/are   0 occurrence(s) of AvaIII
Sequence(s): /ATGCAT/

From   1 to   2657 there is/are   0 occurrence(s) of AvrII
Sequence(s): /CCTAGG/

From   1 to   2657 there is/are   1 occurrence(s) of BalI
Sequence(s): /TGGCCA/

    527/1

From   1 to   2657 there is/are   1 occurrence(s) of BamHI
Sequence(s): /GGATCC/

    1/1

From   1 to   2657 there is/are   13 occurrence(s) of BbvI
Sequence(s): /GCAGC/GCTGC/

    78/1   417/2   449/1   498/1   768/2   903/2   1089/2   1122/2
1224/2   1281/2   1479/2   1491/2   1601/1

From   1 to   2657 there is/are   3 occurrence(s) of BbvII
Sequence(s): /GAAGAC/GTCTTC/

    721/2   1288/2   2131/2

From   1 to   2657 there is/are   2 occurrence(s) of BclI
Sequence(s): /TGATCA/

    216/1   1445/1

From   1 to   2657 there is/are   4 occurrence(s) of BglI
Sequence(s): /GCC-------GGC/

    799/1   2138/1   2262/1   2277/1

From   1 to   2657 there is/are   0 occurrence(s) of BglII
Sequence(s): /AGATCT/

From   1 to   2657 there is/are   7 occurrence(s) of BinI
Sequence(s): /GGATC/GATCC/

    1/1   2/2   248/1   777/2   1593/2   2489/1   2523/2

From   1 to   2657 there is/are   1 occurrence(s) of BsmI
Sequence(s): /GAATGC/GCATTC/

    707/2

From   1 to   2657 there is/are   3 occurrence(s) of BspMI
Sequence(s): /ACCTGC/GCAGGT/

    1178/1   1525/1   1884/2

From   1 to   2657 there is/are   0 occurrence(s) of BspMII
Sequence(s): /TCCGGA/

From   1 to   2657 there is/are   6 occurrence(s) of BssHII
Sequence(s): /GCGCGC/

    749/1   1003/1   1154/1   1156/1   1969/1   2628/1

From   1 to   2657 there is/are   3 occurrence(s) of BstEII
Sequence(s): /GGT-ACC/

    1251/1   1375/1   1887/1

From   1 to   2657 there is/are   0 occurrence(s) of BstXI

From   1 to   2657 there is/are   28 occurrence(s) of CauII
Sequence(s): /CCCGG/CCGGG/

    121/2   242/2   573/1   574/2   741/2   864/1   876/1   891/2
   1020/1   1080/2   1248/2   1319/1   1320/2   1331/2   1380/2   1385/2
   1461/1   1767/1   2057/1   2067/1   2104/1   2191/1   2232/2   2375/1
   2376/2   2475/2   2526/2   2598/2

From   1 to   2657 there is/are   14 occurrence(s) of CfrI
Sequence(s): /CGGCCA/CGGCCG/TGGCCA/TGGCCG/

      6/1   198/4   527/3   647/2   908/4   989/4   1245/2   1304/1
   1550/2   1759/2   1943/2   2117/2   2222/4   2269/2

From   1 to   2657 there is/are   18 occurrence(s) of Cfr10I
Sequence(s): /ACCGGC/ACCGGT/GCCGGC/GCCGGT/

    113/3   388/3   427/2   460/3   785/1   829/3   1045/1   1243/1
   1552/3   1681/1   1757/1   1918/1   2002/4   2022/4   2240/1   2341/1
   2403/3   2466/4

From   1 to   2657 there is/are   1 occurrence(s) of ClaI
Sequence(s): /ATCGAT/

    2520/1

From   1 to   2657 there is/are   4 occurrence(s) of DdeI
Sequence(s): /CT-AG/

    327/1   2062/1   2427/1   2620/1

From   1 to   2657 there is/are   4 occurrence(s) of DraII
Sequence(s): /AGG-CCC/AGG-CCT/GGG-CCC/GGG-CCT/

    1052/2   1352/4   1511/1   2283/3

From   1 to 2657 there is/are 1 occurrence(s) of DraIII
Sequence(s): /CAC----GTG/

   1018/1

From   1 to 2657 there is/are 0 occurrence(s) of EcoRI
Sequence(s): /GAATTC/

From   1 to 2657 there is/are 0 occurrence(s) of EcoRI*
Sequence(s): /AATT/

From   1 to 2657 there is/are 11 occurrence(s) of EcoRII
Sequence(s): /CCAGG/CCTGG/

   372/2   540/2   636/2   942/2   1056/2   1434/2   1818/2   2139/2
   2202/1   2220/2   2414/2

From   1 to 2657 there is/are 0 occurrence(s) of EcoRV
Sequence(s): /GATATC/

From   1 to 2657 there is/are 0 occurrence(s) of Eco47III
Sequence(s): /AGCGCT/

From   1 to 2657 there is/are 2 occurrence(s) of EspI
Sequence(s): /GCT--AGC/

   2061/1 2619/1

From   1 to 2657 there is/are 6 occurrence(s) of FinI
Sequence(s): /GGGAC/GTCCC/

   755/2   1334/1   1400/2   1473/1   1678/1   2126/1

From   1 to 2657 there is/are 40 occurrence(s) of Fnu4HI
Sequence(s): /GC-GC/

    60/1   78/1   127/1   200/1   261/1   417/1   449/1   495/1
   498/1   503/1   666/1   746/1   768/1   799/1   837/1   903/1
   991/1   1089/1   1119/1   1122/1   1160/1   1182/1   1224/1   1281/1
  1407/1   1425/1   1479/1   1486/1   1491/1   1530/1   1549/1   1501/1
  1863/1   1942/1   2053/1   2259/1   2268/1   2271/1   2274/1   2602/1

From   1 to 2657 there is/are 11 occurrence(s) of GdiII
Sequence(s): /CGGCCG/TGGCCG/

   198/2   647/1   908/2   989/2   1245/1   1550/1   1759/1   1943/1
   2117/1   2222/2   2269/1

From   1 to 2657 there is/are 6 occurrence(s) of GsuI
Sequence(s): /CTCCAG/CTGGAG/

   490/2   1435/2   2041/2   2140/2   2415/2   2549/2

```
From      1 to   2657 there is/are   4 occurrence(s) of ...
Sequence(s): /AGGCCA/AGGCCT/TGGCCA/TGGCCT/

   527/3   1820/4   1832/1   1961/1

From      1 to   2657 there is/are   4 occurrence(s) of HaeII
Sequence(s): /AGCGCT/AGCGCC/GGCGCT/GGCGCC/

   410/2   609/3   1457/2   1689/3

From      1 to   2657 there is/are  36 occurrence(s) of HaeIII
Sequence(s): /GGCC/

     7/1      50/1    112/1    199/1    375/1    528/1    577/1    648/1
   843/1     862/1    909/1    990/1   1054/1   1128/1   1167/1   1246/1
  1305/1    1354/1   1383/1   1451/1   1551/1   1760/1   1821/1   1833/1
  1944/1    1962/1   2001/1   2118/1   2223/1   2243/1   2270/1   2276/1
  2285/1    2406/1   2431/1   2465/1

From      1 to   2657 there is/are  17 occurrence(s) of HgaI
Sequence(s): /GACGC/GCGTC/

    81/2    323/1    385/1    457/1    659/2    753/2    826/1    922/1
  1105/2   1114/1   1221/1   1232/2   1268/1   1713/1   1750/2   1987/1
  2100/1

From      1 to   2657 there is/are  13 occurrence(s) of HgiAI
Sequence(s): /GAGCAC/GAGCTC/GTGCAC/GTGCTC/

    20/3    619/4   1015/3   1069/3   1099/3   1258/1   1294/1   1849/3
  2014/3   2167/3   2227/1   2323/4   2652/2

From      1 to   2657 there is/are   6 occurrence(s) of HgiCI
Sequence(s): /GGCACC/GGCGCC/GGTACC/GGTGCC/

   463/1   1204/1   1388/1   2046/4   2344/1   2400/4

From      1 to   2657 there is/are   1 occurrence(s) of HgiEII

  1000/1

From      1 to   2657 there is/are   5 occurrence(s) of HgiJII
Sequence(s): /GAGCCC/GAGCTC/GGGCCC/GGGCTC/

   842/3   1764/1   1806/4   2284/3   2652/2

From      1 to   2657 there is/are  45 occurrence(s) of HhaI
Sequence(s): /GCGC/
    96/1    139/1    411/1    501/1    610/1    629/1    696/1    749/1
   751/1    848/1    856/1    994/1   1003/1   1005/1   1039/1   1117/1
  1154/1   1156/1   1158/1   1279/1   1342/1   1410/1   1423/1   1458/1
  1469/1   1494/1   1556/1   1635/1   1690/1   1733/1   1748/1   1771/1
  1898/1   1940/1   1969/1   1971/1   1990/1   2096/1   2331/1   2355/1
  2362/1   2381/1   2536/1   2628/1   2630/1
```

From   1 to   2657 there is/are   4 occurrence(s) of HindII
Sequence(s): /GTTAAC/GTCAAC/GTTGAC/GTCGAC/

   382/3   556/2   727/2   1234/4

From   1 to   2657 there is/are   0 occurrence(s) of HindIII
Sequence(s): /AAGCTT/

From   1 to   2657 there is/are   7 occurrence(s) of HinfI
Sequence(s): /GA-TC/

   296/1   332/1   360/1   478/1   1213/1   1615/1   2481/1

From   1 to   2657 there is/are   6 occurrence(s) of HinGUII
Sequence(s): /GGATG/CATCC/

   55/2   369/2   513/2   678/2   1077/2   2313/2

From   1 to   2657 there is/are   0 occurrence(s) of HpaI
Sequence(s): /GTTAAC/

From   1 to   2657 there is/are   52 occurrence(s) of HpaII
Sequence(s): /CCGG/

      5/1   114/1   121/1   242/1   300/1   389/1   428/1   461/1
    516/1   574/1   741/1   786/1   830/1   865/1   877/1   891/1
   1021/1   1046/1   1080/1   1193/1   1244/1   1248/1   1265/1   1320/1
   1331/1   1380/1   1385/1   1453/1   1462/1   1553/1   1682/1   1758/1
   1768/1   1919/1   1933/1   2003/1   2023/1   2058/1   2068/1   2105/1
   2192/1   2232/1   2241/1   2342/1   2348/1   2376/1   2388/1   2404/1
   2467/1   2475/1   2526/1   2598/1

From   1 to   2657 there is/are   14 occurrence(s) of HphI
Sequence(s): /GGTGA/TCACC/

   158/1   404/2   509/2   791/2   898/1   932/2   1218/1   1253/2
   1274/2   1375/1   1889/2   1904/2   2586/1   2592/1

From   1 to   2657 there is/are   0 occurrence(s) of KpnI
Sequence(s): /GGTACC/

From   1 to   2657 there is/are   0 occurrence(s) of MaeI
Sequence(s): /CTAG/

From   1 to   2657 there is/are   11 occurrence(s) of MaeII
Sequence(s): /ACGT/

   27/1   68/1   259/1   380/1   532/1   956/1   1364/1   1783/1
   2129/1   2321/1   2635/1

From   1 to   2657 there is/are   6 occurrence(s) of MaeIII
Sequence(s): /GT-AC/

   159/1   1219/1   1252/1   1376/1   1888/1   1903/1

From   1 to   2657 there is/are   4 occurrence(s) of MboII
Sequence(s): /GAAGA/TCTTC/

    722/2   1289/2   1418/2   2132/2

From   1 to   2657 there is/are   2 occurrence(s) of MlaI
Sequence(s): /TTCGAA/

    2035/1   2547/1

From   1 to   2657 there is/are   2 occurrence(s) of MluI
Sequence(s): /ACGCGT/

    923/1   1103/1

From   1 to   2657 there is/are   3 occurrence(s) of MmeI
Sequence(s): /TCCAAC/TCCGAC/GTTGGA/GTCGGA/

    583/2   1752/4   2251/4

From   1 to   2657 there is/are   37 occurrence(s) of MnlI
Sequence(s): /CCTC/GAGG/

    52/1   86/2   107/2   110/2   151/1   154/1   267/2   344/1
    402/1   442/2   561/1   598/2   655/2   930/1   949/2   982/2
   1051/2   1165/2   1392/1   1465/2   1517/1   1540/2   1569/1   1579/2
   1596/1   1687/2   1840/2   2011/2   2044/2   2122/2   2143/2   2163/1
   2209/2   2281/2   2418/2   2426/1   2487/2

From   1 to   2657 there is/are   2 occurrence(s) of MstI
Sequence(s): /TGCGCA/

    695/1   1493/1

From   1 to   2657 there is/are   6 occurrence(s) of NaeI
Sequence(s): /GCCGGC/

    113/1   388/1   460/1   829/1   1552/1   2403/1

From   1 to   2657 there is/are   0 occurrence(s) of NarI
Sequence(s): /GGCGCC/

From   1 to   2657 there is/are   0 occurrence(s) of NcoI
Sequence(s): /CCATGG/

From   1 to   2657 there is/are   0 occurrence(s) of NdeI
Sequence(s): /CATATG/

From   1 to   2657 there is/are   0 occurrence(s) of NheI
Sequence(s): /GCTAGC/

```
From      1 to    2657 there is/are    4 occurrence(s) of NlaIII
Sequence(s): /CATG/

   220/1   348/1   1497/1  2496/1
```

```
From      1 to    2657 there is/are   17 occurrence(s) of NlaIV
Sequence(s): /GG---CC/

     1/1    463/1    689/1    842/1   1033/1   1204/1   1352/1   1388/1
  1399/1   1512/1   1678/1   2046/1   2187/1   2283/1   2284/1   2344/1
  2400/1
```

```
From      1 to    2657 there is/are    1 occurrence(s) of NotI
Sequence(s): /GCGGCCGC/

   2268/1
```

```
From      1 to    2657 there is/are    0 occurrence(s) of NruI
Sequence(s): /TCGCGA/
```

```
From      1 to    2657 there is/are    7 occurrence(s) of NspBII
Sequence(s): /CAGCGG/CAGCTG/CCGCGG/CCGCTG/

    58/3    225/3    800/3    838/3   1120/4   2182/1   2272/3
```

```
From      1 to    2657 there is/are    1 occurrence(s) of NspCI
Sequence(s): /ACATGC/ACATGT/GCATGC/GCATGT/

   1496/3
```

```
From      1 to    2657 there is/are    1 occurrence(s) of PflMI
Sequence(s): /CCA-------TGG/

   2642/1
```

```
From      1 to    2657 there is/are    4 occurrence(s) of PpaI
Sequence(s): /GAGACC/GGTCTC/

   642/2   1188/2   1315/1   1438/1
```

```
From      1 to    2657 there is/are    1 occurrence(s) of PpuMI
Sequence(s): /AGGACCC/AGGACCT/GGGACCC/GGGACCT/AGGTCCC/AGGTCCT/GGGTCCC/GGGTCCT/

   1511/1
```

```
From      1 to    2657 there is/are    0 occurrence(s) of PstI
Sequence(s): /CTGCAG/
```

```
From      1 to    2657 there is/are    2 occurrence(s) of PvuI
Sequence(s): /CGATCG/

    851/1   1520/1
```

From   1 to   2657 there is/are   0 occurrence(s) of PvuII
Sequence(s): /CAGCTG/

From   1 to   2657 there is/are 16 occurrence(s) of RsaI
Sequence(s): /GTAC/

```
    29/1     91/1    958/1    975/1   1026/1   1176/1   1356/1   1627/1
  1652/1   1875/1   1916/1   2006/1   2071/1   2310/1   2411/1   2460/1
```

From   1 to   2657 there is/are   4 occurrence(s) of RsrII
Sequence(s): /CGGACCG/CGGTCCG/

```
   517/2   1754/1   2349/1   2384/2
```

From   1 to   2657 there is/are   1 occurrence(s) of SalI
Sequence(s): /GTCGAC/

```
  1234/1
```

From   1 to   2657 there is/are   1 occurrence(s) of SauI
Sequence(s): /CCT-AGG/

```
  2426/1
```

From   1 to   2657 there is/are 17 occurrence(s) of Sau3A
Sequence(s): /GATC/

```
     2/1    103/1    217/1    249/1    435/1    547/1    735/1    777/1
   852/1   1446/1   1521/1   1593/1   1695/1   1779/1   2490/1   2523/1
  2607/1
```

From   1 to   2657 there is/are 31 occurrence(s) of Sau96I
Sequence(s): /GG-CC/

```
   311/1    375/1    518/1    576/1    633/1    689/1    842/1    843/1
   862/1   1034/1   1053/1   1059/1   1127/1   1146/1   1353/1   1382/1
  1399/1   1512/1   1563/1   1679/1   1755/1   1977/1   2000/1   2243/1
  2276/1   2284/1   2285/1   2350/1   2385/1   2423/1   2464/1
```

From   1 to   2657 there is/are   0 occurrence(s) of ScaI
Sequence(s): /AGTACT/

From   1 to   2657 there is/are 39 occurrence(s) of ScrFI
Sequence(s): /CC-GG/

```
   121/1    242/1    372/1    540/1    573/1    574/1    636/1    741/1
   864/1    876/1    891/1    942/1   1020/1   1056/1   1080/1   1248/1
  1319/1   1320/1   1331/1   1380/1   1385/1   1434/1   1461/1   1767/1
  1818/1   2057/1   2067/1   2104/1   2139/1   2191/1   2202/1   2220/1
  2232/1   2375/1   2376/1   2414/1   2475/1   2526/1   2598/1
```

From 1 to 2657 there is/are 22 occurrence(s) of SduI
Sequence(s): /GGGCCC/GGGCTC/GGGCAC/GAGCCC/GAGCTC/GAGCAC/GTGCCC/GTGCTC/GTGCAC/

```
  20/9    619/8    842/1   1015/9   1069/9   1099/9   1203/3   1258/6
1294/6   1387/3   1764/4   1806/2   1849/9   2014/9   2047/7   2110/7
2145/3   2167/9   2227/6   2284/1   2323/8   2652/5
```

From 1 to 2657 there is/are 30 occurrence(s) of SecI
Sequence(s): /CC--GG/

```
  46/1     58/1    154/1    225/1    524/1    572/1    573/1    596/1
 741/1    800/1    818/1    838/1   1149/1   1319/1   1331/1   1596/1
1673/1   1738/1   1817/1   2114/1   2120/1   2201/1   2202/1   2232/1
2272/1   2279/1   2375/1   2376/1   2433/1   2485/1
```

From 1 to 2657 there is/are 12 occurrence(s) of SfaNI
Sequence(s): /GATGC/GCATC/

```
 141/2    278/1    316/2    368/2    506/2    677/2    684/1   1076/2
1271/2   1661/2   2151/1   2518/2
```

From 1 to 2657 there is/are 1 occurrence(s) of SfiI
Sequence(s): /GGCC-----GGCC/

```
2276/1
```

From 1 to 2657 there is/are 3 occurrence(s) of SmaI
Sequence(s): /CCCGGG/

```
 573/1   1319/1   2375/1
```

From 1 to 2657 there is/are 0 occurrence(s) of SnaBI
Sequence(s): /TACGTA/

From 1 to 2657 there is/are 0 occurrence(s) of SnaI
Sequence(s): /GTATAC/

From 1 to 2657 there is/are 0 occurrence(s) of SpeI
Sequence(s): /ACTAGT/

From 1 to 2657 there is/are 1 occurrence(s) of SphI
Sequence(s): /GCATGC/

```
1496/1
```

From 1 to 2657 there is/are 4 occurrence(s) of SplI
Sequence(s): /CGTACG/

```
1365/1   1626/1   1651/1   2459/1
```

From 1 to 2657 there is/are 0 occurrence(s) of SspI
Sequence(s): /AATATT/

From   1 to   2657 there is/are   1 occurrence(s) of SstI
Sequence(s): /GAGCTC/

   2652/1

From   1 to   2657 there is/are   5 occurrence(s) of SstII
Sequence(s): /CCGCGG/

    58/1   225/1   800/1   838/1   2272/1

From   1 to   2657 there is/are   0 occurrence(s) of StuI
Sequence(s): /AGGCCT/

From   1 to   2657 there is/are   0 occurrence(s) of StyI
Sequence(s): /CCAAGG/CCATGG/CCTAGG/CCTTGG/

From   1 to   2657 there is/are   22 occurrence(s) of TaqI
Sequence(s): /TCGA/

    84/1   105/1   144/1   437/1   481/1   566/1   602/1   623/1
   653/1   737/1   1235/1   1292/1   1394/1   1519/1   1523/1   1640/1
  1664/1   1697/1   1838/1   2036/1   2521/1   2548/1

From   1 to   2657 there is/are   13 occurrence(s) of TaqII
Sequence(s): /CACCCA/CACCGA/GACCCA/GACCGA/TGGGTC/TGGGTC/TCGGTG/TCGGTC/

   156/7   405/2   474/4   870/2   896/7   963/4   1216/7   1254/2
  1373/7   1741/6   1890/2   2199/3   2584/7

From   1 to   2657 there is/are   56 occurrence(s) of ThaI
Sequence(s): /CGCG/

    59/1   95/1   97/1   226/1   502/1   630/1   750/1   752/1
   781/1   801/1   839/1   847/1   849/1   855/1   924/1   978/1
   993/1   1002/1   1004/1   1006/1   1038/1   1104/1   1116/1   1155/1
  1157/1   1159/1   1173/1   1210/1   1231/1   1278/1   1341/1   1409/1
  1422/1   1488/1   1548/1   1557/1   1611/1   1709/1   1749/1   1772/1
  1941/1   1968/1   1970/1   1989/1   2085/1   2097/1   2273/1   2302/1
  2332/1   2354/1   2361/1   2363/1   2382/1   2535/1   2627/1   2629/1

From   1 to   2657 there is/are   1 occurrence(s) of Tth_I
Sequence(s): /GAC----GTC/

   439/1

From   1 to   2657 there is/are   0 occurrence(s) of Tth_II
Sequence(s): /CAAACA/CAAGCA/TGTTTG/TGCTTG/

From   1 to   2657 there is/are   0 occurrence(s) of XbaI
Sequence(s): /TCTAGA/

From   1 to   2657 there is/are   0 occurrence(s) of XhoI
Sequence(s): /CTCGAG/

From   1 to   2657 there is/are   3 occurrence(s) of XhoII
Sequence(s): /AGATCC/AGATCT/GGATCC/GGATCT/

    1/3   776/1   1592/1 .

From   1 to   2657 there is/are   7 occurrence(s) of XmaIII
Sequence(s): /CGGCCG/

   647/1   1245/1   1550/1   1759/1   1943/1   2117/1   2269/1

From   1 to   2657 there is/are   0 occurrence(s) of XmnI
Sequence(s): /GAA------TTC/