Analysis of the transfer region of the *Streptomyces* plasmid SCP2*

A thesis submitted for the degree of Doctor of Philosophy

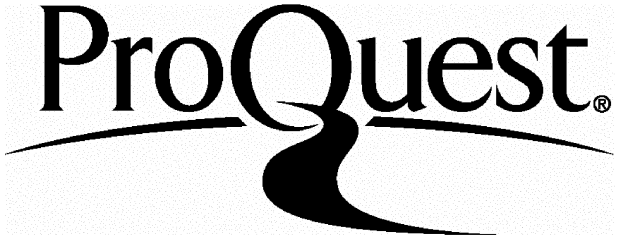
of the University of London

Majib Ur Rehman

Department of Biology

University College London

1995



ABSTRACT

pIJ903 is a bifunctional derivative of the 31.4kb low copy number, conjugative *Streptomyces coelicolor* plasmid SCP2*. pIJ903 was used in order to determine the complete nucleotide sequence of a critical part (3.35kb) of the transfer region (*Sac*I (23) - *Pst*I (18)). Three complete putative open reading frames with the codon usage typical of *Streptomyces* genes have been identified: *traA* (479 amino acids [aa]), *traB* (138 aa) and *traX* (291aa). The deduced sequences of none of the putative proteins showed any overall similarity to known protein sequences.

traA is essential for SCP2* intermycelial plasmid transfer and pock formation and it encodes a protein of molecular weight 50,389 Daltons. The TraA protein contains a nucleotide binding sequence and shows a faint similarity to DNA helicases, suggesting that it, like Tra of pIJ101 and TraSA of pSAM2 might be a DNA helicase. *traB* encodes a putative protein of MW 14,703 Daltons with a potential leucine zipper motif, suggesting that it may be a DNA binding protein. *traX* would encode a protein of 30,587 Daltons but no function can be attributed to this protein as yet.

The *traB* gene has been cloned into pT7-7 creating pQR524, a construct which produces a 16,100 Dalton fusion protein *in vivo* in *E.coli*. This TraB fusion protein has been used as an antigen in order to raise polyclonal rabbit antibodies against it. The TraB antibodies were shown to bind to TraB in *E.coli* but no TraB was detected in *Streptomyces* crude protein extracts.

The DNA sequence proposed in this work is slightly different to that previously published. The divergence in the sequences causes a shift of the translational start codon of *traA*, such that *traA* would be translated into a protein 36 amino acids longer at its N-terminus. Other than the difference at the N-terminus the TraA proteins are very similar. Sequence analysis of the *traX* and *traA* indicated that these two genes may be translationally coupled, with the stop codon of *traX* overlapping the start codon of *traA*. *In vitro* protein analysis of the *traX* and *traA* genes in *E.coli* suggests that these genes synthesise a protein of approximately 80,000 Daltons (the sum of TraX and TraA). Further analysis suggests that the 80,000 Dalton protein produced *in vitro* could be a fusion protein, maybe the result of ribosomal frameshifting.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank Dr Hilary Richards and Dr John Ward for their continuous help throughout this project. I would also like to thank Dr Tobias Kieser of the John Innes Institute for providing strains and help with raising the antibodies.

May I also thank everyone in the lab namely; Hana, Mushtaq, Les, Carol, Sue and a very special thanks to Shradha for putting up with me for so long.

Thanks also to Saleem Mohammed, Sejal Patel, Robert Wardle, Steve Athwall and Kamran Salim for their constructive comments, criticisms and encouragement.

Finally, thanks to my sisters Shain and Sajda for helping me prepare this manuscript.

DEDICATION

I would like to dedicate this thesis to my parents who have supported and encouraged me throughout my education.

TABLE OF CONTENTS

LIST OF FIGURES

PAGE NO.

LIST OF TABLES

CHAPTER ONE: Introduction

1.0 Introduction

1.1 The Streptomycetes

1.1.1 Growth and Morphology

Streptomyces are Gram positive, obligately aerobic, mycelial bacteria that are widely distributed in the environment, particularly soil, where they form a large and essential part of the natural population. The Streptomycetes undergo a complex cycle of morphological differentiation resembling that of filamentous fungi (Figure 1.1). Vegetative growth is via the formation of a network of substrate mycelium (branching, multinucleate hyphae) that penetrate and degrade complex organic debris by the secretion of hydrolytic enzymes (Chater, 1984). Differentiation occurs on solid surfaces under conditions of nutrient limitation, with substrate mycelium giving rise to specialised, spore bearing aerial hyphae (Chater, 1984). The aerial hyphae actually form by cannibalising the lysing vegetative mycelium (Chater, 1984). During this stage the organism is particularly sensitive to invading micro-organisms which are attracted to the lysing mycelia. The enormous diversity of antibiotics that can be isolated from these bacteria may be the result of the evolution of strong chemical defence mechanisms to protect this vulnerable stage of the life cycle (Chater & Hopwood, 1983).

1.1.2. Importance of *Streptomyces*

The main impetus for the intensive study of the *Streptomyces* has been the vast array of secondary metabolites produced by this genus, including many products of widespread use in human and veterinary medicine and agriculture. The *Streptomyces* hold a position of considerable medical and

provided the potential for a more rational approach to increasing levels of antibiotic yields in fermentations through an understanding of antibiotic biosynthesis and its regulation, and the possibility of creating new antibiotics to combat the spread of natural resistance. Most of the cloning systems available at this time had been developed for *E.coli*, an organism taxonomically and genetically different from *Streptomyces*. There was thus an obvious need to develop dedicated cloning systems for the isolation, manipulation and analysis of *Streptomyces* genes, and it has been research to this aim that has lead to the identification and characterisation of many of the *Streptomyces* plasmids known today.

In the last decade several advances have aided the discovery and characterisation of *Streptomyces* plasmids and their subsequent use as cloning vectors. Possibly of most importance was the development of reliable, rapid and effective procedures for the isolation of plasmid DNA from *Streptomyces* (Kieser, 1984). This, together with the discovery of the means to introduce plasmid DNA into *Streptomyces* protoplasts at high efficiency (Bibb *et al.*, 1978), and the significant observation, that conjugative *Streptomyces* plasmids cause pocks (Bibb *et al.*, 1977), providing a convenient genetic marker for the recognition and manipulation of plasmids, have all contributed to a rapid development of cloning techniques involving *Streptomyces* plasmids (Hopwood *et al.*, 1986).

1.2.2 Variety & distribution of *Streptomyces* plasmids

Naturally occurring plasmids have been reported in many *Streptomyces* strains. Since the first physical characterisation of a *Streptomyces* plasmid (Schrempf *et al.*, 1977), a profusion of plasmids have been identified. These plasmids run the spectrum from low to high copy number (one to several hundred), large to small size (>200kb to <4kb) and

nearly all can conjugate, except for naturally occurring deletions of conjugative forms (Tomich, 1988; reviewed by Hopwood *et al.*, 1986). Most *Streptomyces* plasmids carry genes which allow very efficient (up to 100%) plasmid transfer and moderately efficient exchange of chromosomal genes (Kieser & Hopwood, 1991). No natural, covalently closed circular plasmid has been proven to carry genes for other properties such as antibiotic biosynthesis or antibiotic/metal resistance commonly associated with plasmids of Gram negative bacteria (Kieser & Hopwood, 1991).

The first *Streptomyces* plasmid to be identified was SCP1 in *Streptomyces coelicolor* A3(2) (Vivian, 1971), which has recently been characterised in more detail (Kinashi & Shimaji-Murayma, 1991) and shown to be a giant linear plasmid (Kinashi *et al.*, 1987). However the first to be isolated physically was a covalently closed circular (CCC) molecule of 31kb (Schrempf *et al.*, 1977), distinct from SCP1, later found to be identical to a second sex plasmid of *Streptomyces coelicolor,* SCP2. This plasmid has been identified through the occurrence of spontaneous variants (SCP2*) capable of promoting increased fertility in *S.coelicolor* A3(2) (Bibb *et al.*, 1977).

Thus far all characterised *Streptomyces* plasmids are double stranded DNA molecules and most are covalently closed circular (CCC) DNAs, but there are also double stranded linear plasmids and integrating elements (plasmids that exist integrated in the host chromosome and can also be found as autonomous plasmids). Hopwood *et al.* (1986) reported that CCC DNA plasmids were found in about 30% of wild type *Streptomyces* strains examined. *Streptomyces* plasmids rarely seem to code for properties affecting host phenotype other than conjugation. Methylenomycin production and resistance by SCP1 (*S.coelicolor*) and pSV1 (*S.violaceus-ruber*) may be the only well established examples of *Streptomyces* plasmids coding for properties other than conjugation (Hopwood *et al.*, 1986).

1.2.3 *Streptomyces* plasmids as cloning vectors

Most streptomycete plasmid cloning vectors have been developed from three different naturally occurring *Streptomyces* plasmid molecules. These are: 1) those based on the *S.lividans* plasmid pIJ101, which are generally small, have a high copy number and a broad host range (Kieser *et al.* 1982); 2) those derived from SLP1.2, which are generally a medium size, with an intermediate copy number and a limited host range (Bibb *et al.*, 1981) and 3) those derived from the *S.coelicolor* plasmid SCP2*, which are mostly large, have a low copy number and a fairly broad host range (Lydiate *et al.*, 1985).

It is not practical in the scope of this work to review the large variety of *Streptomyces* based cloning vectors currently available. Reviews of the subject can be found in Hopwood *et al.* (1983, 1986 & 1987). The use of *Streptomyces* cloning vectors for gene cloning and expression are further reviewed by Bibb & Cohen (1982); Chater *et al.*(1982); Chater (1990) and Tomich (1988).

1.3 Plasmid transfer

Self transmissible (conjugative) plasmids are found widely in Gram negative and Gram positive bacteria. The transfer mechanisms and associated DNA processing and fertility functions of those found in Gram negative bacteria are well characterised in comparison to those of the Gram positives. Cell to cell contact is suggested as being a universal requirement of conjugal transfer, but the mechanisms by which this occurs in Gram positive bacteria appear quite distinct from those found in Gram negative bacteria (Dunny, 1990).

In conjugative plasmids of Gram negative bacteria, such as F, and other plasmids where more detailed analyses have been carried out, conjugation can be divided on both physiological and genetic bases into two parts: the recognition of recipient cells by donor cells that leads to mating pair formation, and the subsequent transfer of DNA (Willetts, 1988). It is not practical in the context of this Chapter to review in detail the transfer mechanisms of Gram negative bacteria, comprehensive reviews can be found in Willetts & Skurray, (1987); Willetts, (1988); Clewell, (1993) however, a brief overview of the important features of plasmid transfer are described below.

Not all plasmids of Gram negative bacteria encode self-transfer functions, but those that do, appear to have similar mechanisms of gene transfer. Unlike certain other conjugative plasmids in which expression of the transfer genes (*tra*) is repressed, the *tra* genes of F are always derepressed. Therefore F^+ (donor) cells always have pili on their outer surface. Briefly, the mechanism of conjugal transfer involves the binding of the F pilus, specifically to a protein in the outer membrane (OmpA) of the F^- cell (recipient), thus initiating transfer replication and the process of gene transfer by conjugation. A nick is made in the F plasmid at the origin of transfer (*oriT*), followed by a rolling circle mechanism of replication in which the intact strand is used as the template and the 3' end generated by the nick is used as primer. This action causes the 5' end of the nicked single strand to be displaced and transferred to the F^- cell, within which the single stranded molecule of DNA is duplicated by a chromosomally encoded DNA polymerase and re circularized.

More than 25 genes within the *tra* region have been proposed to be involved in conjugal transfer of F and F-like plasmids (Willetts & Skurray, 1987). Most of the transfer genes can be grouped according to the

conjugation stage or transfer-related property that requires their expression. These groups of genes include: **1)** Regulation of transfer; expression of the transfer functions is dependant upon the regulatory product, TraJ, expression of which is repressed by the *finP* product when a *finO* product is available. In F plasmids, the FinO protein is not synthesised due to a disruption (by an Insertion Element- IS3) in the *finO* gene, therefore the *tra* genes are derepressed (Ippen-Ihler & Skurray, 1993). **2)** Pilus biosynthesis and assembly; donors carrying F and F-like plasmids produce pili, which protrude from the surface of the cell and initiate mating contact with the recipient cells (Willetts, 1988). Three *tra* gene products are required for the synthesis of the subunits (pilin) of the pilus, and a further 11 *tra* gene products are required for the assembly of the pili (Willetts & Skurray, 1987). **3)** Conjugal DNA metabolism; four *tra* gene products may be involved in the DNA related events in the donors, Which include the site specific single stranded nick at *oriT*, 5' → 3' displacement of the nicked strand and its subsequent displacement into the recipient cell.

In addition to the above gene products, further transfer region gene products have been implicated in mating pair stabilisation prior to gene transfer and surface exclusion to prevent the entry into the cell of plasmids of the same incompatibility group (Inc) (Willetts & Skurray, 1987). More comprehensive reviews of the roles of the individual *tra* genes during transfer can be found in Willetts & Skurray, (1987) and Ippen-Ihler & Skurray, (1993).

As previously stated, cell to cell contact is a universal requirement of conjugal transfer, though the mechanisms by which mating pair formation takes place, can vary. Conjugal transfer of DNA seems to be less frequently studied among Gram positive bacteria than among Gram negative bacteria however, some of the fundamental aspects of the process seem to differ between the two classes. It has been suggested that in for

example *Enterococcus faecalis*, pili do not play a role in mating pair formation, rather plasmid containing donor cells form clumps with cells that lack the plasmid and plasmid transfer is suggested as taking place within these clumps (Clewell, 1981). Clumping results from the interaction between a proteinaceous microfibrillar substance referred to as the aggregation substance (AS) on the surface of the plasmid containing donor and a binding substance (BS) on the surface of the plasmid free recipient (Clewell, 1993). Both donor and recipient cells of *Enterococcus* have been suggested as producing the binding substance, but the aggregation substance is produced only when a plasmid bearing donor is in close proximity of a cell that lacks that particular plasmid (recipient). The recipient cell produces a chromosomally encoded small peptide molecule referred to as a sex pheromone (Clewell, 1993). Plasmid free recipients produce multiple sex pheromones, each exhibiting specificity for a given plasmid type. The pheromones produced by the recipient cells, diffuse through the medium and enter the donor cell, inducing the plasmid encoded gene to synthesise the aggregation substance. Although the donor cells also carry chromosomal genes encoding the sex pheromones, sex pheromones are not synthesised due to a repressor product encoded by the plasmid. When a plasmid is acquired, secretion of the related pheromone is prevented however, unrelated pheromones continue to be produced (Clewell, 1993). Thus far pheromone responding plasmids have been confined to the enterococci and their close relatives.

Transfer of *Streptomyces* plasmids from plasmid harbouring donor strains to plasmid free recipients occurs efficiently on solid media, but has not been characterised in liquid culture, and is generally thought not to occur in liquid media. Very little is currently known about the underlying mechanisms of plasmid transfer in *Streptomyces*. However the very size of conjugative *Streptomyces* plasmids, and more so the sizes of the regions

localised as containing the *tra* genes ~2kb for pIJ101 (Kieser *et al.*, 1982); ~5kb for pSAM2 (Smokvina *et al.*, 1991) and ~9kb for SCP2* (Bibb *et al.*, 1977), suggests that conjugal transfer in *Streptomyces* may be genetically and structurally simpler than in plasmids of other bacteria or that the majority of the genes involved in transfer are located on the chromosome. These compare to the size of transfer regions of plasmids of Gram negative bacteria which can occupy many kilobases. For example, the transfer region of the F plasmid of *E.coli*, occupies around 33kb (from a plasmid size of 100kb), though much of this codes for genes for the assembly of the mating pili (Willets & Skurray, 1987). The simplicity of the system in *Streptomyces* is consistent with the inability to detect, by electron microscopy, the presence of pili or other mating structures in conjugating hosts (Akagawa, 1987). This may also reflect the lack of a need for sex pili to ensure prolonged close contact of the non-motile *Streptomyces* hyphae. It follows from this fact, and from the fact that *Streptomyces* plasmids carry no or relatively few genes affecting the phenotype of the host, that much of the DNA carried by the *Streptomyces* plasmids is of unknown function except for DNA encoding genes involved with transfer and fertility.

It is not known whether DNA transfer in *Streptomyces* is single stranded or double stranded. Replication of the *Streptomyces* plasmid, pIJ101, has been proposed to be through a single stranded intermediate, probably by a rolling circle mechanism (Hopwood & Kieser, 1993; Zaman *et al.*, 1992). It is likely also, that plasmid transfer in *Streptomyces* is replicative.

1.3.1 Pock formation in *Streptomyces*

As suggested, conjugation functions of many *Streptomyces* plasmids require very little genetic information on the plasmid, no structures such as the sex pili have been implicated as necessary to form a mating pair,

and a single *tra* gene is essential and sufficient for transfer. Nevertheless, conjugation is very efficient, nearly 100% of the recipient progeny receive plasmids when donor and recipient strains are mated on solid media (Bibb & Hopwood, 1981; Kieser *et al.*, 1982).

Most *Streptomyces* plasmids studied give rise to 'pocks' (Bibb & Hopwood, 1981; Bibb *et al.*, 1977), a property which has been used as an indicator for plasmid transfer (Kendall & Cohen, 1987). The pock phenotype can be used instead of antibiotic selection in transformation experiments: one plasmid carrying transformant can be seen in a lawn of 10^8 plasmid free individuals (Bibb *et al.*, 1978). Pock formation is a phenomenon unique to conjugative plasmids of actinomycetes and the phenotype has been used in the identification of cryptic plasmids devoid of any identifiable phenotypic markers. Pocks are observed when a plasmid-containing 'donor' spore develops within a confluent lawn of a strain lacking the plasmid (recipient). Under these conditions, transfer of the plasmid results in a 'pock' where the growth and/or development of the newly infected recipient culture is retarded in a region of up to 1-2mm around the plasmid-containing donor (Hopwood & Kieser, 1993). Growth inhibition is often accompanied by an obvious change in physiology, most often as a precocious production of pigmented antibiotics observed in the pocks of *S.coelicolor* and *S.lividans* (Hopwood & Kieser, 1993). Bibb *et al.* (1977) first observed the zones of retarded growth, they noticed that the inhibition zones followed the contours of the plasmid containing colony at constant width, leading them to suggest that their formation was dependent upon hyphal contact rather than a diffusable inhibitor. Pock forming *Streptomyces* plasmids are described as exhibiting 'lethal zygosis' (Ltz) on the plasmid free strains, due to the similarity to lethal zygosis in *E.coli* (Skurray & Reeves, 1973), in which F^- cells are killed by simultaneous conjugation with several Hfr individuals (Bibb *et al.*, 1977).

However in *Streptomyces* this phenomenon probably does not lead to death of the recipient strain (Bibb *et al.*, 1977) since it is possible to isolate plasmid containing cells from within the inhibition zones.

Hopwood *et al.* (1985a) proposed that transfer of *Streptomyces* plasmids was accomplished through two steps: intermycelial transfer mediated by transfer functions and intramycelial transfer directed by spread functions. It has also been suggested that pock formation is a manifestation of intramycelial transfer of plasmid (spreading) within the mycelium of the recipient culture after primary transfer (intermycelial) from the donor (Kendall & Cohen, 1987).

1.3.2 Fertility

With the exception of two plasmids isolated from *Streptomyces lividans* SLP3 and SLP4 (Hopwood *et al.*, 1983) all conjugative *Streptomyces* plasmids tested so far, are fertility factors, able to promote the transfer of chromosomal markers from donors to recipient (also referred to as chromosomal mobilising ability [cma] Holloway, 1979). *Streptomyces* plasmids may be responsible for most (if not all) of the fertility naturally observed in members of this genus (Hopwood *et al.*, 1986). The observation that SLP3 and SLP4 are conjugative but lack fertility properties suggests that fertility is not an obligate consequence of plasmid transfer and that other *Streptomyces* plasmids may posses a specific fertility function. It was also shown that different *Streptomyces* plasmids promote fertility at characteristically different levels irrespective of their normal transfer efficiency (Hopwood *et al.*, 1985a). No evidence has been found to suggest that autonomous plasmids such as SCP2* and pIJ101 integrate into the chromosome to cause fertility as do plasmids of Gram negative bacteria. The mechanism by which non-integrating *Streptomyces* plasmids promote the

transfer of chromosomal sequences remains unknown (Hopwood *et al.*, 1985a).

1.4 Analyses of the transfer regions of *Streptomyces* plasmids.

Many self transmissible *Streptomyces* plasmids have been identified, but transfer regions have only been localised on a relatively few of them, including the non-integrating plasmids, pIJ101 (Kieser *et al.*, 1982; Kendall & Cohen, 1987), pSN22 (Kataoka *et al.*, 1991a), SCP2* (Bibb *et al.*, 1977) and integrating elements pSAM2 (Smokvina, 1991; Hagege *et al.*, 1993) and SLP1 (Omer & Cohen, 1984). However, the transfer region has been sequenced only in the case of pIJ101 (Kendall & Cohen, 1988) and pSAM2 (Hagege *et al.*, 1993), of which the *Streptomyces lividans* plasmid pIJ101 (for which the complete sequence of the plasmid is available) has been well characterised (Kendall & Cohen, 1987 &1988; Stein *et al.*, 1989; Stein & Cohen, 1990; Zaman *et al.*, 1992; Tai & Cohen, 1993). The transfer region of pSN22 has also been sequenced (Hagege *et al.*, 1993; Hopwood & Kieser, 1993) but as yet unpublished (Kataoka *et al.*, Unpublished). The most studied of these plasmids pIJ101, pSN22 and pSAM2 are discussed further in this section, whilst the relatively poorly understood mechanisms of SCP2* are summarised in a later section.

1.4.1 pIJ101

pIJ101 is an 8.9kb CCC DNA plasmid isolated from *Streptomyces lividans* (Kieser *et al.*, 1982). It has a high copy number (40 - 300 copies/chromosome), broad host range, is self transmissible by conjugation, is pock forming and is able to promote chromosomal recombination at high frequency in *S.lividans* and *S.coelicolor* (Kieser *et al.*, 1982). More is currently known about the biology of the pIJ101 plasmid than

any other *Streptomyces* plasmid. The complete plasmid has been sequenced (Kendall & Cohen, 1988), and some of the transfer properties of the plasmid have been further analysed (Stein *et al.*, 1989; Stein & Cohen, 1990; Zaman *et al.*, 1992; Tai & Cohen, 1993).

In vitro insertion and deletion studies localised the transfer region (*tra*) to a 1.4kb fragment, and the spread function (*spd*) to a 1.7kb fragment (Kieser *et al.*, 1982). Disruption of the transfer region was shown to abolish pocking, since, presumably no transfer was able to take place. On the other hand, disruption of the spread region caused small pocks to be formed, suggesting that transfer was taking place, but the plasmid was unable to migrate within the recipient mycelium (Kieser *et al.*, 1982). Kieser *et al.* (1982) also showed by replica plating, that transfer of the plasmid was confined to the area of the pock. They further suggested that both transfer and spread probably required *cis* - acting functions, since they could not be induced to transfer or spread in the presence of normal transmissible, pock forming plasmids.

Apart from the previously identified *tra* and *spd,* genetic analysis identified four loci (*kilA*, *kilB*, *korA* & *korB*) involved in transfer and pock formation (Kendall & Cohen, 1987). The *kil* genes so called because of their lethal effects, could not be introduced into *S. lividans* in the absence of a suitable *kor* (kil override) gene in either *cis* or *trans* (Kendall & Cohen, 1987). Both *kil* genes are involved in plasmid transfer, since disruption of *kilA* abolished pocking while disruptions to *kilB* reduced the size of the pocks observed (Kendall & Cohen, 1987). The *korA* gene was able to override the lethality of both *kilA* and *kilB*, whereas *korB* overrode *kilB* only (Kendall & Cohen, 1987). As previously stated, disruptions to *tra* and *spd* produced no pocks and small pocks respectively, thus *kilA* and *tra* are involved in

intermycelial transfer, whilst *kilB* and *spd* are involved with the intramycelial spread of the transferred plasmid.

Kendall & Cohen, (1988) sequenced pIJ101 and showed that the genetically defined *kilA* and *tra* loci were encoded by a single (621 amino acid[aa]) ORF (*tra*). They further showed that the predicted Tra protein probably had no association with membranes (due to the absence of significant hydrophobic domains), or with DNA binding proteins, but did show the presence of a putative ATP binding domain, suggesting that the Tra protein may be involved in an active ATP requiring process. Transcription of the *tra* promoter was suggested as extending through to the *spd* locus. Analysis of the *spd* region, indicated the presence of two ORFs, designated *spdA* and *spdB* (Kendall & Cohen, 1988). The presence of strong hydrophobic domains, each sufficiently long to span the host membrane, in both *spd* ORFs were shown, thus suggesting that they may be associated with membranes.

Kendall & Cohen, (1987) showed the *korA* gene was able to override lethal effects of both *kil* genes, whereas *korB* could only override the effects of *kilB*. Stein *et al.* (1989) revealed the nature of *kor* mediated repression to be at the transcriptional level. The KorB protein repressed transcription from its own promoter and that of the *kilB* gene. *KorA* and *tra* were found to have overlapping divergent promoters (identified by Buttner & Brown, 1987) from which transcription was repressed by the *korA* product. The mechanism by which the *kilB* phenotype is repressed by *korA* is not transcriptional, and is as yet unknown (Stein *et al.*, 1989). Sequence analysis by Kendall & Cohen (1988), showed that the predicted amino acid sequence of the *korA* and *korB* ORFs, showed no homology with any other repressor proteins. However, the amino acid sequence of both genes contained helix-turn-helix motifs, similar to the DNA binding domains of many other prokaryotic repressor proteins (Kendall & Cohen, 1988; Harrison, 1991; Pabo

& Sauer, 1984). The promoter regions of *korB* and *kilB*, which are repressed by *korB*, were shown to be very similar (Kendall & Cohen, 1988). The binding of the KorB protein, expressed in *E.coli* (Stein & Cohen, 1990; Zaman *et al.*, 1992; Tai & Cohen, 1993) has been studied, and using DNase I footprinting the recognition sites of the KorB protein located to the *kilB* and *korB* promoter regions (Zaman *et al.*, 1992) suggesting that KorB controls the expression of the *kilB* gene at the transcriptional level.

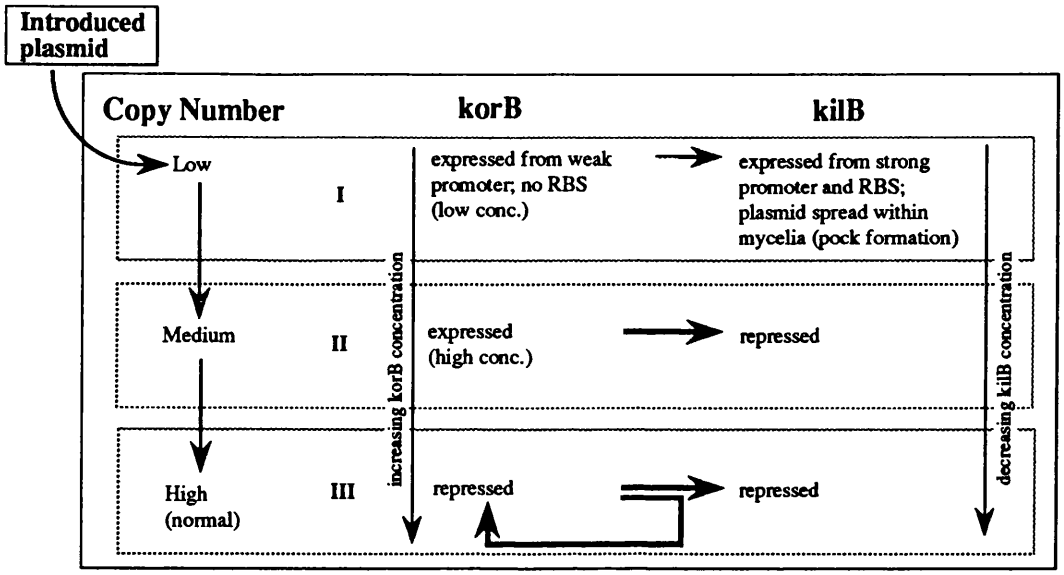
A model for KorB regulation of *korB* and *kilB* genes has been proposed by Zaman *et al.* (1992) and is represented diagramatically in Figure 1.2. This model assumes that the *kilB* promoter is more efficient than that of *korB*, and that the affinity of KorB protein for the *kilB* operator is stronger than that for the *korB* operator.

I) Upon transfer of pIJ101 to a new host, the plasmid is present only at a low copy number and only a limited amount of KorB protein is synthesised in the cell. The low levels of KorB in the cells results in the derepression of the *kilB* gene, and an increase in KilB protein concentration. KilB may act to retard cell growth until an appropriate copy number is reached, before allowing the plasmid to spread throughout the recipient mycelium (resulting in pocks). **II)** As plasmid copy number increases to the desired level, the concentration of KorB also increases. The increase in KorB causes the protein to actively repress *kilB* transcription by binding to the *kilB* operator. **III)** Once KorB repressor concentration increases beyond a certain limit and all the *kilB* operator sites have occupied, KorB is then able to bind to its own operator and negatively regulate KorB production. **IV)** Since KorB has a relatively low affinity for its own promoter, some read through expression takes place ensuring a basal level of KorB in the cell. A drop in KorB concentration below the basal level causes its dissociation from its own

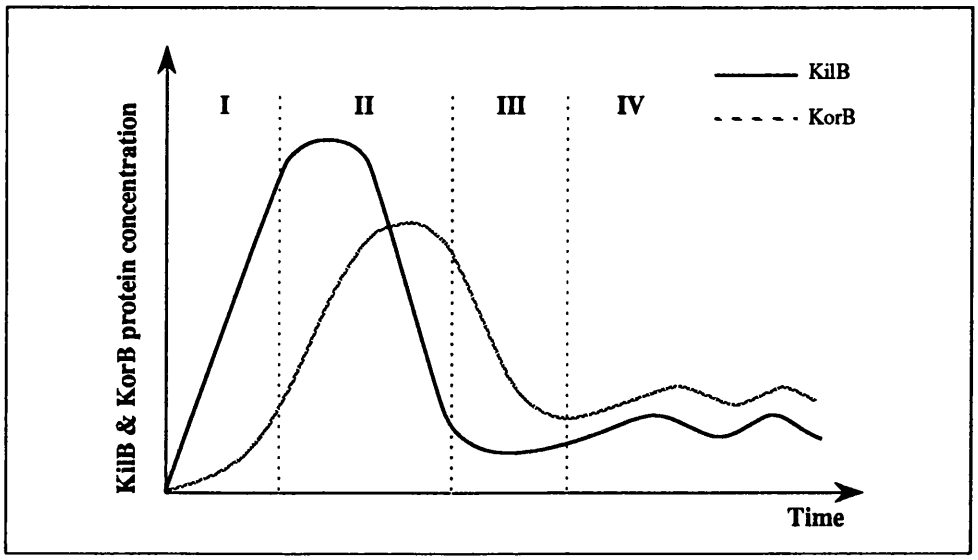
FIGURE 1.2 Model for KorB regulation of *korB* and *kilB* genes.

[Figure reproduced from Zaman *et al.* (1992)]

A)



B)



operator and further KorB expression. Thus a constant level of KorB is maintained within the cell which controls the KilB protein concentration.

1.4.2 pSN22

pSN22 is an 11kb, self-transmissible, pock forming, multicopy CCC DNA plasmid isolated from *Streptomyces nigrifaciens* (Kataoka *et al.*, 1991a). *In vitro* insertion and deletion studies localised the transfer and pock regions to a ~7kb fragment (Kataoka *et al.*, 1991a). Mutational analysis showed that the 7kb fragment contained five genes (*spdA, spdB, traA, traB, traR*) postulated as being involved in plasmid transfer and pock formation (Kataoka *et al.*, 1991a). Three of these *traA*, *traB* & *traR* were shown to be essential for pock formation. Disruption of *traB* completely abolished transfer with no pocks being observed, no other mutation had such a devastating effect, and thus the *traB* gene product was suggested as being the major protein involved in plasmid transfer (Kataoka *et al.*, 1991a & b). *traA* mutants did not form pocks, but the plasmids were still transmissible, a phenotype not observed in any other *Streptomyces* plasmids (Kataoka *et al.*, 1991a). The *traR* gene was suggested as being a regulatory factor involved in expression of *tra* genes, and thus analogous to the *kil-kor* system associated with other *Streptomyces* plasmids (Kendall & Cohen, 1987), since *traR* could not be inactivated in plasmids which had an undisrupted *traB* gene (Kataoka *et al.*, 1991a & b). Disruption of the *spd* genes (*spdA*, *spdB*) reduced pock size and thus were hypothesised as facilitating the movement of plasmids within the *Streptomyces* mycelium similar to the *spd* genes of pIJ101 (Kataoka *et al.*, 1991a).

1.4.3 pSAM2

pSAM2 is an 11kb integrating element isolated from *Streptomyces ambofaciens* (Pernodet *et al.*, 1984). This element is self-transmissible, pock forming and able to mobilise chromosomal markers (Smokvina *et al.*, 1991). A functional map of pSAM2 was deduced from phenotypes exhibited in *S.lividans* by numerous deletion or insertion derivatives (Smokvina *et al.*, 1991) and the transfer and pock formation regions sequenced (Hagege *et al.*, 1993).

The transfer and pock forming regions were localised to a ~3kb (*NruI*(8)-*SmaI*(2)) fragment (Smokvina *et al.*, 1991). Sequence analysis identified seven putative ORFs (Hagege *et al.*, 1993). The *tra* of pIJ101 (Kendall & Cohen, 1987) and *traB* of pSN22 (Kataoka *et al.*, 1991) are essential for plasmid transfer, a similar gene *traSA* was also identified in pSAM2, which was shown to be essential for conjugal transfer and pock formation (Hagege *et al.*, 1993). Sequence similarities were also found between *tra* (pIJ101) and *traSA* (pSAM2), both putative proteins Tra (621aa) and TraSA (306aa) contained probable nucleotide binding domains, suggesting the possibility of a common function, possibly in the use of energy from ATP hydrolysis to drive plasmid transfer (Hagege *et al.*, 1993). Similar ATP binding enzymes involved in plasmid transfer of Gram-negative bacteria have been identified, for example the *KilBI* of RK2 (Motallebi-Veshareh *et al.*, 1992).

Functional similarities were also found with the spread genes of pIJ101 (*spdA, spdB*) and pSN22 (*spdA, spdB*) which were proposed to be involved with intramycelial transfer since disruption reduced the size of the pocks formed. *spdA*, *spdB*, *spdC* and *spdD* gene mutants from pSAM2, were also affected in their pock sizes, but additionally in their transfer abilities, a phenomenon not observed with the *spd* gene mutants of pIJ101, but similar to

kilB mutants of pIJ101, which were shown to be affected in plasmid transfer & pock size (Kendall and Cohen, 1987; Hagege *et al.*, 1993). The putative Spd proteins of pIJ101 and pSAM2 could be membrane associated, but no sequence homology between them was observed (Hagege *et al.*, 1993).

Kendall & Cohen (1987), identified a *kil-kor* system associated with the transfer region of pIJ101, where a *kil* gene could not be cloned in the absence of a suitable *kor* (kil override) gene. A similar system has been shown in pSAM2, were the major transfer gene *traSA* is suggested as being a *kil* locus, since disruption of the *Bam*HI (11) - *Sac*I (14) region was lethal, but if the *Bst*EII (4) - *Nru*I (8) region was deleted simultaneously, *Streptomyces* transformants could be obtained (Smokvina *et al.*, 1991). Similar to the observation made by Kataoka *et al.*(1991a), they showed that *traR* could not be inactivated in plasmids which expressed *traB*. It is possible that the *korSA* gene could encode a protein with amino acid sequence similarities to *korA* of pIJ101 (Hagege *et al.*, 1993), suggesting that it too could be a transcriptional regulatory protein, controlling the expression of *traSA* (Hagege *et al.*, 1993), since it also contains a helix-turn-helix motif found in many prokaryotic repressor proteins (Hagege *et al.*, 1993; Harrison, 1991). Hagege *et al.*(1993) also showed that *traSA* showed significant sequence similarity to *tra* of pIJ101, even though it was half the size. They also showed that *korSA*, *traR* and *korA* (from pSAM2, pSN22 & pIJ101 respectively) observed sequence homology.

In summary, functional similarities in the properties of the *tra* and *spd* genes of pIJ101, pSN22 and pSAM2 can be observed. Mutations to the *tra* and *spd* functions generally abolished transfer and pock formation or lead to reduced pock sizes respectively. All three plasmids have been suggested as possessing a *kil-kor* type regulatory system. It not being possible to introduce a *kil* loci into a *Streptomyces* host in the absence of a suitable *kor*

gene. In most cases the *kor* genes, themselves have generally been shown to possess DNA binding motifs of one form or another. Some weak amino acid sequence similarities between the various proteins of pIJ101, pSN22 and pSAM2 have also been seen.

Differences however, in the transfer and pock formation genes of the three *Streptomyces* plasmids are also obvious. The sizes of some of the proteins, predicted from the amino acid sequences, suggested as being similar have been shown to vary quite considerably (eg. Tra of pIJ101 [621 amino acids] and TraSA of pSAM2 [306 amino acids]). Some of the transfer and spread genes themselves have been identified for some plasmids but not for others. A summary of the transfer and spread genes of the above plasmids can be found in Figure 1.3.

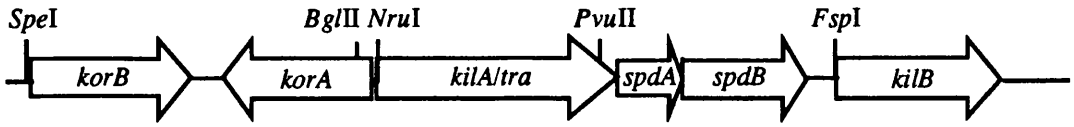
1.5 *Streptomyces coeliolor* plasmid SCP2*

SCP2* (Lydiate *et al.*, 1985) is a derivative of the sex plasmid SCP2 (Bibb *et al.*, 1977) isolated from *Streptomyces coelicolor*. Both plasmids are 31.4kb in size and physically indistinguishable from one another. However the high fertility variant SCP2*, exhibits a much more pronounced lethal zygosis reaction, than SCP2. SCP2* is a large, broad host range plasmid of very low copy number (1-4). Several genetic functions have been approximately mapped, and a restriction map for 13 enzymes has been determined (Larson & Hershberger, 1986; Lydiate *et al.*, 1985). A physical and functional map of SCP2* is shown in Figure 1.4.

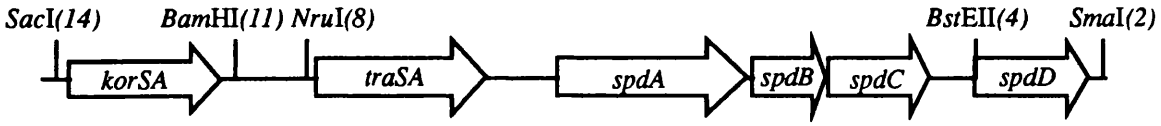
Most of the cloning vectors derived from SCP2* have a low copy number, allowing cloned chromosomal genes to be maintained at a dosage at most only slightly above normal. These plasmids are useful for cloning large DNA fragments, even plasmids with large DNA inserts (up to 35kb such as those of antibiotic biosynthetic genes) are stably inherited and no

FIGURE 1.3 Summary of the functional locations of the *tra* and *spd* genes of pIJ101, pSAM2 and pSN22

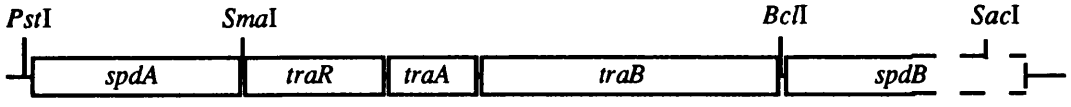
pIJ101



pSAM2



pSN22



Selected restriction sites are indicated (numbers in parentheses refer to sites mentioned in the text). Directions of the genes (where known) are indicated. Diagram not drawn to scale.
[Figure adapted from: Kendall & Cohen, 1987 & 1988; Kataoka *et al.*, 1991a & b; Hagege *et al.*, 1993].

structural rearrangements have been reported in any of the clones (Lydiate *et al.*, 1985).

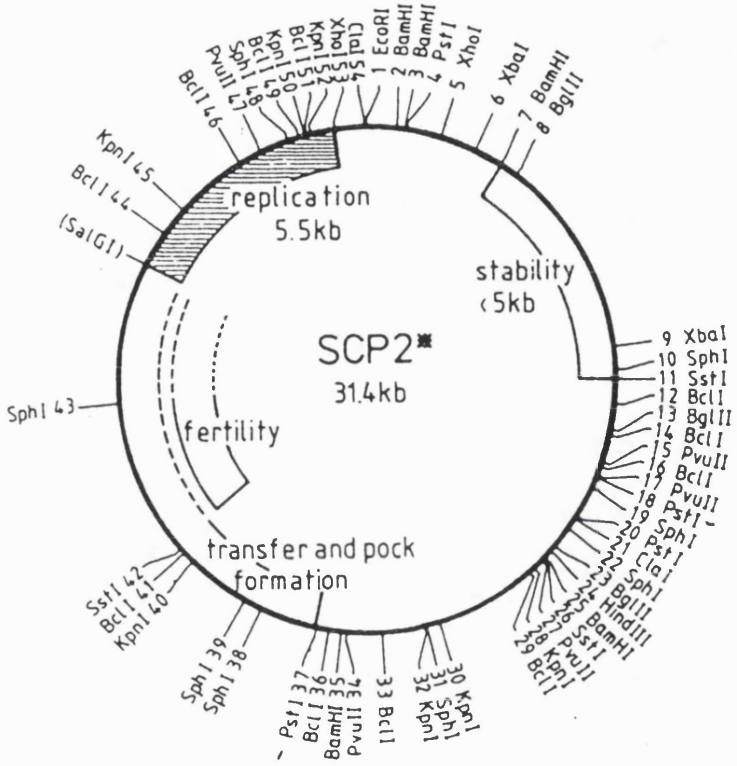
1.5.1 Construction of pIJ903.

The phenotype of SCP2* derivatives deleted *in vitro* revealed that plasmid transfer, replication, pock formation fertility functions were all present on a *Pst*I fragment (*Pst*I A), and a second *Pst*I fragment (*Pst*I B) contained a possible partition function involved in promoting stable inheritance of the plasmid (Bibb *et al.*, 1980). The SCP2* derivative pSCP103 was constructed which consisted of the *Pst*I fragments A and B of SCP2* joined in their native configuration, carrying all the known biological functions of the SCP2* parent.

The 1.76kb thiostrepton resistance gene (*tsr*) from pIJ30 (Thompson *et al.*, 1982) was introduced as a *Bam*HI fragment into linearized (partial digestion with *Bcl*I) pSCP103, creating pIJ919 (Lydiate *et al.*, 1985). The 25.8kb bifunctional vector, pIJ903 (Lydiate *et al.*, 1985) was able to transform *E.coli* to ampicillin resistance and *S.lividans* to thiostrepton resistance. pIJ903 contains a single *Bam*HI cloning site, this constructed by the *in vitro* deletion of the four small *Bam*HI-*Xho*I fragments from pIJ919 and their subsequent replacement by the *Bam*HI-*Sal*I fragment of pBR327.

Shuttle plasmids with *E.coli* and *Streptomyces* replicons allow for easy DNA isolation and manipulation in *E.coli*, while phenotypic analysis can be performed when the plasmids are returned to *Streptomyces*. Particular attention has been given to the construction of pIJ903 (Figure 1.5), since this is the cloning vector used throughout this study in order to further examine the transfer region of SCP2*.

FIGURE 1.4 Restriction enzyme and functional map of the 31.4kb *Streptomyces coelicolor* plasmid SCP2*



The minimal replicon, *rep* was shown to be located to a 5.5kb fragment between a *Sal*I site and *Eco*RI (1). The 8.9kb region between *Pst*I (37) and clockwise to *Sal*I is the region shown to be involved with transfer and pock formation of the plasmid and its derivatives. All functional locations shown above were derived from Lydiate *et al.* (1985); Larson & Hershberger, (1986).

Restriction enzyme sites are numbered, using the unique *Eco*RI site as site 1. There are many sites for *Sal*I, *Sac*II and *Sma*I, but none for *Eco*RV. Unique sites are *Eco*RI (1) and *Hin*dIII (24).

[Figure reproduced from Hopwood *et al.*(1985b)]

FIGURE 1.5 Construction of pIJ903: a *Streptomyces/E.coli* shuttle vector

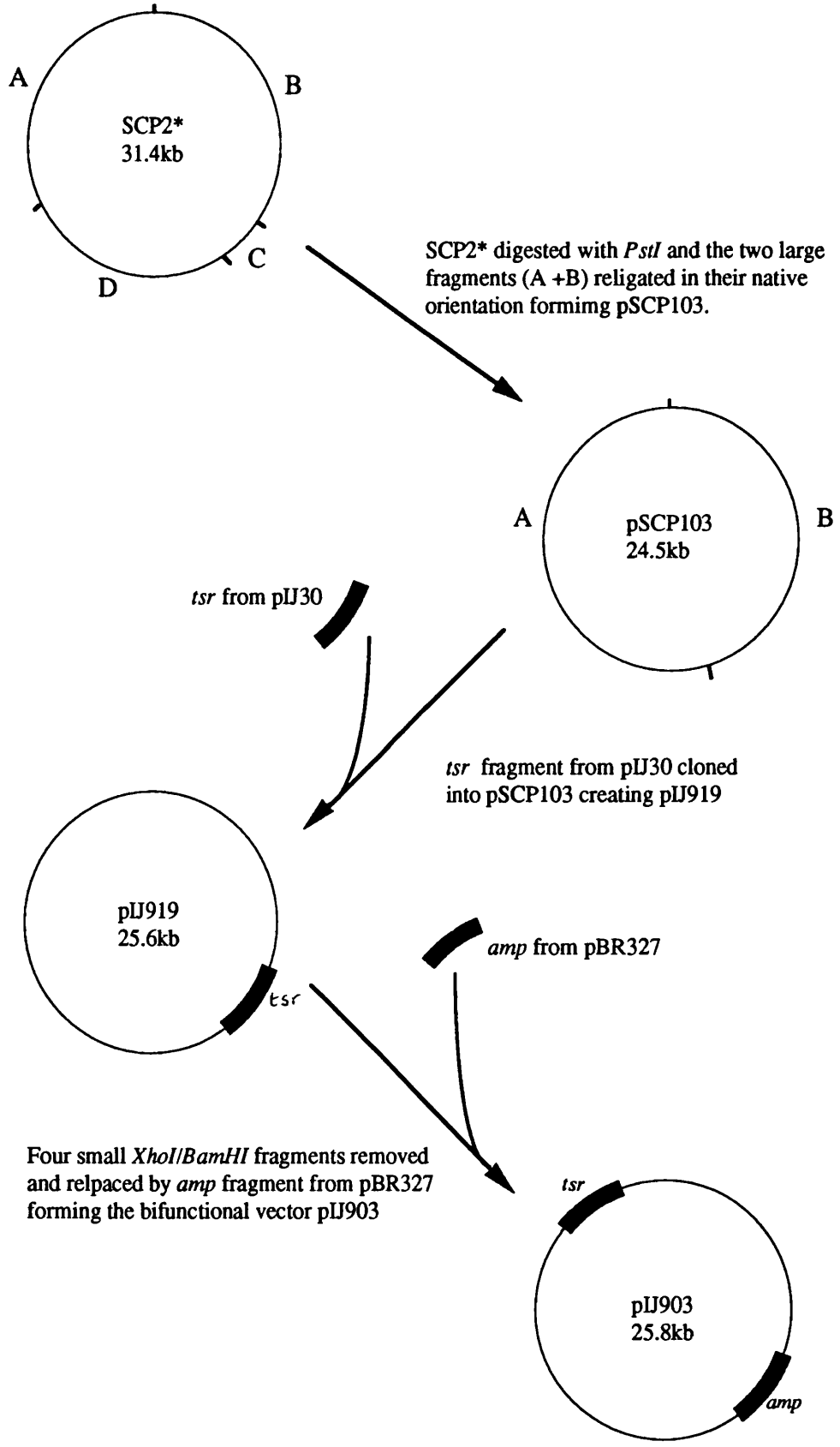
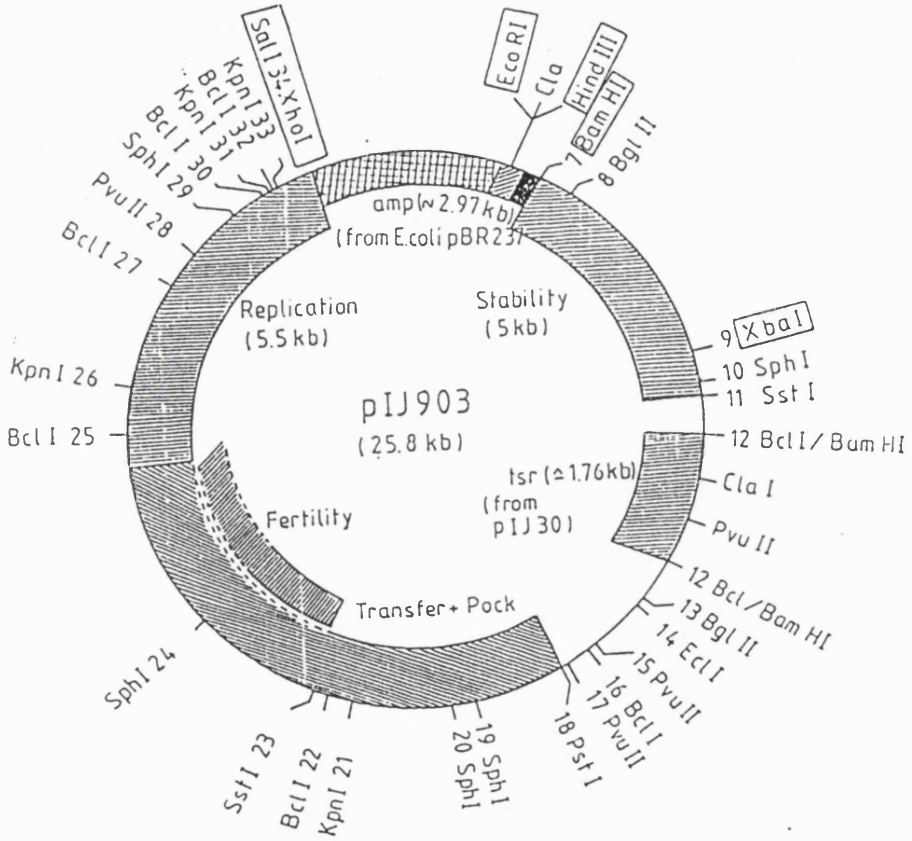


FIGURE 1.6 Restriction endonuclease cleavage and functional map of pIJ903



pIJ903 (25.8kb) is a bifunctional cloning vector containing all the known functions of the parent plasmid: SCP2*, but additionally able to transform *Streptomyces* to thiostrepton resistance and *E.coli* to ampicillin resistance.

The restriction sites are also shown, there is a single *Bam*HI cloning site and as for SCP2* many sites for *Sal*I, *Sac*II and *Sma*I.

[Figure reproduced from Hopwood *et al.*, 1985b]

1.6 Aims of the project

At the outset of this project very little was known about the location and organisation of the transfer and pock functions of SCP2*. The main aim of this work was to obtain the complete nucleotide sequence of the transfer region of the *Streptomyces* plasmid SCP2*, using its derivative pIJ903 (Figure 1.6). This bifunctional vector, which retains all the biological properties of SCP2*, may be shuttled between *S.lividans* and *E.coli* without structural rearrangements. pIJ903 also has a higher copy number in *E.coli* (about 20 copies per host chromosome) and therefore making it easier to isolate and manipulate DNA in *E.coli* for subsequent analysis in *Streptomyces*. The DNA sequence can then be used to identify open reading frames (ORFs), and map the locations of possible genes. Further analysis of the primary amino acid sequence of the genes may determine possible functions of the encoded proteins, bringing us closer to an understanding of the mechanism of plasmid transfer in *Streptomyces*.

It is hoped that this investigation will provide new knowledge which can subsequently be used in the long term to improve and design new *Streptomyces* cloning vehicles. The availability of such vectors will no doubt prove vital in the progression of gene cloning in *Streptomyces*. Any further information about the biology and genetics of *Streptomyces* plasmids is of paramount importance, for such material can be utilised by industry for strain and yield improvement, modification of existing or production of new secondary metabolites, by what is commercially and economically a most important genus.

CHAPTER TWO: MATERIALS & METHODS

2.1 Materials

All chemical reagents used in this work were supplied by BDH, Promega or Sigma (unless specified otherwise) and were of analytical grade or equivalent. Bacteriological reagents were obtained from Oxoid or Difco.

Antibiotics were purchased from Sigma apart from Thiostrepton, which was a gift from S.J.Lucania of E.R.Squibb and Sons Inc, NJ, USA.

Restriction and DNA modifying enzymes were obtained from various sources and all radiolabelled isotopes were supplied by Amersham and NEN (Dupont).

2.2 Strains & Plasmids

E.coli and *Streptomyces* host strains and plasmids used are listed in Tables 2.1 & 2.2. *Streptomyces coelicolor* strains M124, A700, M144 and *Streptomyces* plasmids pIJ8017, pIJ8019 were obtained from T. Kieser of the John Innes Institute, Norwich, UK.

2.3 Culture Media

2.3.1 *E.coli* Culture Media

All media used for cultivating *E.coli* were from Sambrook *et al.*, (1989) and are listed below.

TABLE 2.1 *Streptomyces* & *E.coli* Host Strains

E.coli host strains

Strain	Genotype	Reference
JM107	F' *traD36 lacI^{q}Δ (lacZ) M15 proAB IΔ(lac-proAB) endA1 hsdR17 mcrA*	Yanisch-Perron *et al.* (1985)
BL21(DE3)	F$^-$ *hsdS gal* [λD69Ø (lac*UV5p-T7* gene 1)]	Studier & Moffatt (1986)
GM242	*dam recA leuB proA2 his$^-$ metB lacY galK ara14 thi$^-$*	Lydiate et al. (1985)

Streptomyces host strains

S.lividans		
TK64	*pro$^-$ str^R*	Hopwood *et al.*(1985b)
TK23	*spc^R*	Hopwood *et al.*(1985b)
S.coelicolor		
M124	*pro$^-$ Arg$^-$ cys$^-$*	Hopwood *et al.*(1985b)
A700	*pro$^-$ Arg$^-$ cys$^-$*	Hopwood *et al.*(1985b)
M144	-	Hopwood *et al.*(1985b)

TABLE 2.2 *Streptomyces* & *E.coli* Plasmids

E.coli plasmids

Plasmid	Phenotype	Reference
pUC18/19	amp^R, LacZ'	Yanisch-Perron *et al.*(1985)
M13mp18/19	amp^R, LacZ'	Yanisch-Perron *et al.*(1985)
pT7/7	amp^R	Tabor (1990)
pIJ903	amp^R	Lydiate *et al.*(1985)

Streptomyces Plasmids

Plasmid	Phenotype	Reference
$SCP2^+$	-	Bibb *et al.* (1977)
$SCP2^*$	-	Bibb *et al.* (1977)
pIJ903	tsr^R	Lydiate *et al.* (1985)
pIJ101	-	Kieser *et al.* (1982)
pIJ8017	tsr^R, vio^R	Brolle *et al.* (1993)
pIJ8019	tsr^R, vio^R	Brolle *et al.* (1993)

L Broth

10g/l Bacto tryptone

5g/l Bacto yeast extract

10g/l NaCl

1g/l Glucose

pH 7.0 15g/l agar added for L Agar.

7g/l agar added for Soft Agar.

2xYT Medium

16g/l Bacto tryptone

10g/l Bacto yeast extract

5g/l NaCl

pH 7.0

M9 Minimal Medium

100ml/l 10x M9 Salts

12g/l Agar

at time of use add:

2ml/l $MgSO_4$ (1M)

10ml/l Glucose (20%)

0.1ml/l $CaCl_2$ (1M)

10x M9 Salts

60g/l $Na_2HPO_4 \cdot 12H_2O$

30g/l KH_2PO_4

5g/l NaCl

10g/l NH_4Cl

NAXI Plates

Oxoid Nutrient agar base (25g/l) with the following supplements

100µg/ml ampicillin

40µg/ml IPTG

80µg/ml X-gal

2.3.2 *Streptomyces* Culture Media

Media used for cultivating *Streptomyces* were principally from Hopwood *et al.* (1985b) with the changes listed below.

R2YE

103g/l Sucrose[A]

0.25g/l K_2SO_4

10.12g/l $MgCl_2$

10g/l Glucose[B]

0.1g/l Difco Casaminoacids

15g/l Agar

at time of use add:

10ml KH_2PO_4(0.5%)

80ml $CaCl_2$(3.68%)

15ml L-Proline(20%)

100ml TES (5.73% pH7.2)

2ml Trace elements solution

5ml NaOH (1M)

50ml Difco Yeast Extract(10%)[C]

Trace Elements

40mg/l $ZnCl_2$

200mg/l $FeCl_3.6H_2O$

10mg/l $CuCl_2.2H_2O$

10mg/l $MnCl_2.4H_2O$

10mg/l $Na_2B_4O_7.10H_2O$

10mg/l $(NH_4)_6MO_7O_{24}.4H_2O$

[A] Sucrose omitted

[B] Glucose replaced by Mannitol for better sporulation (Brolle *et al.*, 1993)

[C] 0.1% final concentration yeast extract used

TSB

30g/l Tryptone Soya Broth (Oxoid)

15g/l TSB and 20g/l agar were used for 1/2 strength TSB agar.

YEME

3g/l Difco yeast extract

5g/l Difco bacto peptone

3g/l Oxoid malt extract

10g/l Glucose

after autoclaving

2ml/l of 2.5M $MgCl_2.6H_2O$ was added

SGM

10g/l Oxoid malt extract

10g/l Oxoid bacto peptone

20ml Glycerol pH 7.0 (NaOH)

for SGM agar add 2% agar

2.3.3 Media Supplements

Table 2.3 shows working concentrations of the various antibiotics used. ampicillin, spectinomycin, streptomycin & viomycin in water, chloramphenicol in ethanol, rifampicin in methanol and thiostrepton in DMSO were prepared as stock solutions. Those prepared in water were filter sterilised (0.22μM, Acrodisc filters), all antibiotics were stored at -20°C except for rifampicin which was stored at 4°C.

X-gal stock solutions were prepared in N.N'-dimethylformamide, IPTG in water as described by Sambrook *et al.* (1989) and stored at -20 °C.

2.4 Culture Conditions

2.4.1 *E.coli* growth & storage

All *E.coli* samples were grown at 37°C unless indicated otherwise. Storage was in the form of 20% glycerol stocks at -70°C (long term storage) or on agar plates at 4°C (short term). Liquid cultures were normally grown in 20ml universal bottles containing 5ml of the appropiate media (or alternatively in conical flasks) in a cooled orbital incubater (Gallenkamp) at 200rpm.

2.4.2 *Streptomyces* growth & storage

All *Streptomyces* cultures were incubated at 30°C, and stored as spore suspensions in 20% glycerol at -20°C.

Spore suspensions were prepared (Hopwood *et al.*, 1985b) by flooding a well sporulating plate with 10ml of sterile water, the spores were then scraped off using

TABLE 2.3 Media Supplements

Supplement/ (stock soln. mg/ml)	Final conc.(μg/μl) *E.coli* solid	*E.coli* liquid	*Streptomyces* solid	*Streptomyces* liquid
Ampicillin (50)	200	50	-	-
Chloramphenicol (34)	100	25 (170)[A]	-	-
Rifampicin (20)	-	200	-	-
Thiostrepton (50)	-	-	50	5
Streptomycin (10)	-	-	20	20
Spectinomycin (50)	-	-	100	100
Viomycin (30)	-	-	30	30
X-gal (20)	80	-	-	-
IPTG (20)	40	-	-	-
(1M)[B]	-	1mM[B]	-	-

[A] concentration of chloramphenicol used for plasmid amplification in JM107.

[B] for induction of the *lac* promoter during protein expression.

an inoculating loop. Following this the spore suspensions were then centrifuged (MSE Centaur 2) at 4000rpm, for 10mins, and the resulting spore pellet resuspended in 2ml of 20% glycerol and stored at -20°C. If required the spore suspension was filtered by passing it through a sterile 5ml gilson tip plugged with cotton wool, the filtrate then stored at -20°C.

2.4.3 Pock formation

A dense lawn (~10^8 spores) of the recipient strain was spread on to an agar plate and allowed to dry for a few minutes. The donor spores were then either spread directly onto the recipient lawn (a sufficient number of spores to give 20-200 potential pocks) or alternatively the donor spores were streaked onto the lawn of recipient spores. The plates were then dried for 15 mins and incubated at 30°C. Plates were observed frequently for the presence of pocks (which were seen as small areas of repressed growth surrounded by a ring of pigment production on the reverse of the plate).

The donor spores were also spread or streaked onto plates without the recipient strain, this served as a control, since we could determine the number of donor spores that had in fact been plated and thus determine the potential number of pocks expected.

2.4.4 Testing For Plasmid Transfer

In cases were pocks were not clearly visible but plasmid transfer was thought to have taken place, the occurance of plasmid transfer was determined by taking a loopfull of the prospective transconjugants and plating them onto an agar plate with the relevant antibiotics to select for recipient and plasmid. Thus growth on these plates indicated that plasmid transfer had taken place since the transconjugants had acquired resistance genes from both the donor and the recipient strains.

2.5 *E.coli* Transformations

2.5.1 Preparation of competent cells

Competent cells were prepared using a method described by Morrison (1979), essentially a single *E.coli* colony was picked from an M9 minimal medium plate and used to inoculate 5ml of L broth and grown at 37°C overnight. 4ml of this overnight culture was then used to inoculate 200ml of L broth (supplemented with 20mM $MgCl_2$) and grown at 37°C untill an $O.D._{590}$ of ~0.4 was reached (~2-3hrs). The cells were then placed on ice for 15mins, following which they were harvested by centrifugation (4000rpm - 10mins at 4°C). The cell pellet was then resuspended in 40ml of ice-cold 75mM $CaCl_2$, 15% glycerol and placed on ice for a further 10 mins. The cells were again harvested as previously and the pellet resuspended in 10ml of the same buffer. The competent cells were then aliquoted (0.5ml) in to microcentrifuge tubes and frozen in a ethanol/dry-ice bath and stored at -70°C untill required.

2.5.2 Transformation of competent cells

Transformations were carried out by thawing an aliquot of competent cells on ice for 15 mins. The cells were then mixed with the DNA (maximum of 5%v/v) in T.E and returned to the ice for a further 30 mins. The transformation mix was then heat shocked at 37°C for 10 mins followed by 5 mins on ice. After the heat shock treatment the transformation mix was added to 5ml of L broth and incubated at 37°C for 2 hours. Following this 100 - 200µl was plated out onto an appropriate agar plate and incubated overnight.

2.5.3 Transfection of competent cells

Transfection of DNA was carried out using a method modified from Sambrook *et al.* (1989). Essentially the method was similar to that for transformations except that after the heat shock treatment 100µl of the transfection mix was added to 5ml of soft agar (45°C), 50µl IPTG (20mg/ml), 100µl X-gal (20mg/ml) and 200µl of

E.coli (grown to exponential phase ~ 6hrs). The mix was plated onto an appropriate agar base and incubated at 37°C overnight.

2.6 Plasmid DNA Isolations

2.6.1 Small scale DNA isolations

Small scale plasmid DNA isolations were carried out using a method similar to that described by Birnhoim & Doly (1979). Essentially a single colony was used to inoculate 5ml of L broth and incubated overnight at 37°C (with the appropiate antibiotic). 1.5ml of the overnight culture was then spun in a microcentrifuge (MSE microcentaur) at 13k rpm for 30secs, and the supernatant discarded. The cell pellet was then resuspended in 100µl of lysis buffer (25mM Tris pH8.0, 0.9% glucose, 10mM EDTA, 2mg/ml lysozyme) and incubated at room temperature for 5 mins. 200µl of alkaline SDS (0.2M NaOH, 1% SDS) was then added and the contents mixed, followed by the addition of 150µl of ice cold 4M Sodium Acetate (pH6.0) and the tube incubated on ice for 5 mins. The sample was then spun (13k rpm - 5 mins) and the supernatant transferred to a fresh tube. The supernatant was subsequently extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 v/v) followed by an extraction with chloroform:isoamyl alcohol (24:1 v/v). The DNA was then precipitated by adding 2 volumes of ice cold ethanol and storing on ice for 15 mins, followed by a spin (13k rpm - 10 mins) to pellet the DNA. The DNA was then washed once with 70% ethanol and the pellet dried and resuspended in 30µl of TER (10mM Tris pH7.6, 1mM EDTA, 30µg/ml RNase A).

2.6.2 Large scale DNA isolations

Two methods were routinely employed for the large scale isolation of plasmid DNA, both are described below.

Method A

A single colony was picked and grown overnight in L broth as above, this overnight culture was then used to inoculate 50ml of L broth and grown until an

OD_{590} of ~0.6 was reached. This culture was then used to inoculate 500ml of L broth and grown for a further 3hrs (OD_{590} ~0.4) followed by plasmid amplification by the addition of chloramphenicol to a final concentration of 170µg/ml and grown overnight.

The cells were harvested by spinning (4000rpm, 4°C for 10mins) and the resulting cell pellet resuspended in 50ml of STE (0.1M NaCl, 10mM Tris pH 8.0, 1mM EDTA) and the cell pelleted as above. The cell pellet was then resuspended in 20ml of lysis buffer and incubated at room temperature for 5 mins, followed by the addition of 25ml of alkaline SDS. After mixing 20ml of ice cold 4M Sodium Acetate (pH6.0) was added and the suspension incubated on ice for 10mins followed by centrifugation (8000rpm, 4°C, 15 mins) using a Sorvall RC5B centrifuge.

Following this the supernatant was removed and 50ml of 50%(w/v) PEG 6000 was added and left at -20°C for 1 hour, after which the sample was again spun as before. The supernatant was discarded and the pellet resuspended in 5ml of TE (10mM Tris pH7.6, 1mM EDTA), 0.5ml of 1M NaCl and precipitated by adding 10ml of ethanol and stored at -20°C (24 hrs) or -70°C (1hr) as was convenient.

The DNA pellet was recovered by spinning (10k rpm, 4°C, 20 mins), the supernatant removed and the pellet resuspended in 8ml of TE. 8.4g of Caesium Chloride, 800µl of Ethidium Bromide (10mg/ml) were added and the mixture placed in a Dupont-Sorvall ultra-centrifuge tube (sealed using a Dupont crimper) and centrifuged (Sorvall OTD-65B Ultra-centrifuge) at 45k rpm for 48hrs.

The bands were extracted as described in Sambrook *et al.*(1989) and the Ethidium Bromide removed by several extractions with Butanol, followed by dialysis with TE overnight. The purity and concentration of the DNA was determined as described in section 2.6.5, the DNA then precipitated and resuspended in TE (to a final concentration of ~1mg/ml).

Method B

An alternative method used to isolate plasmid DNA without the need to amplify DNA using chloramphenicol involved the use of the QIAGEN plasmid DNA isolation kit supplied by Hybaid.

The system was used as described by the manufacturer, essentially the bacteria were grown and harvested as described in the previous method. The resulting pellet was resuspened in buffer P1 (100µg/ml RNase A, 50mM Tris, 10mM EDTA pH8.0) followed by the addition of buffer P2 (0.2M NaOH, 1% SDS). After incubation at room temperature for 5 mins, buffer P3 (2.55M KaC pH4.8) was added and the sample centrifuged to remove cell debris. The supernatant was applied to a QIAGEN column (QIAGEN tip 100) previously equilibrated with Buffer QBT (750mM NaCl, 50mM MOPS, 15% EtoH, 0.15% Triton X-100, pH7.0), the column was then washed with Buffer QC (1M NaCl, 50mM MOPS, 15% EtoH, pH7.0). The DNA then eluted with Buffer QF (1.25M NaCl, 50mM MOPS, 15% EtoH, pH8.2) and precipitated using iso-propanol and resuspended in a suitable volume of TE (pH 8.0).

2.6.3 Rapid Screening for Recombinants

Rapid screening for recombinants was carried out using a method similar to that described by Chowdhury (1991). Prospective recombinant colonies were picked and used to inoculate 5ml of L broth and grown at 37°Cwith vigorous shaking for 6 hours. 0.7ml of the culture was then extracted with phenol:chloroform, followed by precipitation with iso-propanol. The resulting pellet was washed with 70% ethanol and resuspended in TER. An aliquot of this DNA was then used to set up a digest with the appropiate restriction enzyme.

2.6.4 Isolation of Single Stranded DNA

A fresh exponential phase culture of host cells (JM107) was diluted (1:100) in 2xYT and infected with a M13 plaque and grown at 37°C for 5-8 hours.

1.5ml of this cell culture was then spun for 2mins and the supernatant transferred to a fresh tube. 200µl of 25%PEG/2.5M NaCl was added and the tube stored on ice for 20mins, after spinning for 5mins the supernatant was discarded and the pellet resuspended in 200µl of TE buffer. The DNA was then extracted once with phenol:chloroform and once with chloroform followed by precipitation with ethanol and resuspended in 20µl of TE buffer.

2.6.5 Estimation of DNA concentration

DNA concentrations were determined spectrophotometrically by measuring the absorbance of the DNA sample at 260nm (An O.D $_{260nm}$ of 1.0 corresponds to 50µg/ml dsDNA or 20µg/ml ssDNA).

DNA purity was determined by taking an additional absorbance reading at 280nm (an O.D $_{260/280nm}$ of 1.8 or above corresponds to clean DNA) as described by Sambrook *et al.* (1989).

2.7 DNA Manipulations

2.7.1 Restriction Digests

All restriction digests were performed according to the suppliers instuctions and using the supplied buffers. Restriction digests were terminated by the addition of stop mix (4M Urea, 50%(w/v) Sucrose, 0.05M EDTA, 0.1% (w/v) Bromophenol blue) when the digest was to be loaded onto an agarose gel. In instances where further processing of the digest was required the reaction was terminated by incubating at 65°C for 15 mins (or as necessary) and the DNA purified using the purification protocol described in section (2.7.3).

2.7.2 DNA Ligations

DNA ligations were performed by incubating the DNA (in TE) with 3µl of 10x T4 Ligase buffer (200mM Tris pH7.6, 50mM $MgCl_2$, 50mM DTT), 3µl of 10mM ATP, water to 30µl and 10units of T_4DNA ligase. The ligation mix was

incubated (4°C overnight or at room temperature for 1 hour) and transformed into competent cells (2.5.2/3).

2.7.3. DNA Purification

Two methods were used for the general purification of nucleic acids.

Method A

The first involved extracting the DNA with an equal volume of phenol:chloroform, followed by precipitation with ethanol and washing with 70% ethanol. Following which the DNA was dried and resuspended in an appropiate volume of TE.

Method B

The alternative method employed the Wizard DNA Clean Up system supplied by Promega. This was used by adding 1ml of the Wizard DNA clean-up resin (in 6M Guanidine Thiocyanate) to the DNA to be purified (200-500µl) and passing the solution through a column. The DNA was then washed in wash buffer (80% iso-propanol) and eluted in 50µl of TE.

2.7.4. Dephosphorylation of DNA

Dephosphorylation of DNA was carried out by using Calf intestinal phosphtase (CIP). The DNA to be dephosphorylated was incubated in CIP buffer (50mM Tris pH 9,1mM $MgCl_2$, 0.1mM $ZnCl_2$, 1mM Spermidine) with 1unit of enzyme per µg of DNA and incubated at 37°C for 1 hour. The reaction was terminated by incubating at 70°C for 15 mins, and the DNA subsequently purified by extracting with phenol:chloroform and precipitating with ethanol (2:7:3).

2.7.5 DNA Modification

Filling of recessed 3' ends;

3' recessed ends were made blunt by digesting ~5µg of DNA with the appropiate enzyme, upon completion of the digestion reaction, 1µl of a solution containing all four dNTPs (2mM) was added and 3 units of Klenow fragment of *E.coli* DNA polymerase 1 for each µg of DNA. The reaction was then incubated at room temperature for 20 mins, after which the reaction was terminated by extracting with phenol:chloroform and precipitating with ethanol.

Removing protruding 3'ends;

After digesting the DNA with the appropiate enzyme, 2µl of dNTPs (all four at 2mM) and 2 units of T_4 DNA polymerase were added and the reaction incubated for 20 mins at 15°C. The reaction was terminated and the DNA purified as described previously (2.7.3).

2.7.6 End Labelling of DNA

5' end labelling of DNA was carried out using T_4 Polynucleotide kinase (PNK) and ^{32}P-λ-ATP. The DNA to be labelled was treated with CIP as described (2.7.4), in order to remove the overhanging 5' phosphate. ~1µg of DNA was incubated at 37°C for 1 hour in T_4PNK buffer (50mM Tris pH7.6, 10mM $Mgcl_2$, 1mM DTT, 1mM EDTA, 1mM spermidine), 100µCi of ^{32}P-ATP (3000Ci/mmol) and 10 units of T_4PNK. The reaction was terminated by heating to 65°C for 15 mins. The unincorporated label was removed from the labelled DNA using a NAP-5 column (Pharmacia) according to the manufacturers instructions. The labelled DNA being eluted in TE buffer (pH8.0).

2.8 Agarose Gel Electrophoresis

Gel electrophoresis was carried out using mini gels (50ml) for standard analysis and large gels (200ml) for preparative electrophoresis. Gels (0.7-1.7%

agarose) were prepared in TBE (45mM Tris, 45mM Boric acid, 5mM EDTA) and run at 1-10v/cm as appropiate. DNA bands were stained in Ethidium Bromide (0.5µg/ml) during electrophoresis and visualized under UV light.

2.8.1 Recovery of DNA Fragments

DNA was recovered from agarose gels using one of two methods, either electro-elution into dialysis tubing (Sambrook *et al.*, 1989) or by the use of low melting point agarose gels.

Electro-elution :

Electro elution involved cutting the band of interest from the gel (prepared & run in 0.5 x TBE), and placing the agarose slice in a dialysis tube[A] filled with buffer (0.5 x TBE) and carring out electrophoresis in the same buffer. The gel slice was then discarded and the buffer containing the DNA from the agarose gel extracted twice with phenol:chloroform. 0.2 volumes of 10M Ammonium Acetate and 2 volumes of ethanol were added to the aqueous phase and the DNA precipitated. The DNA was resuspended in TE buffer and reprecipitated by adding 3M Sodium Acetate and ethanol.

LMP gels :

Low melting point agarose (BRL - Ultra pure) gels were prepared in 0.5 x TBE and run in the same buffer (at 4°C). The band of interest was excised and placed in a microcentrifuge tube and incubated in an 65°C water bath. The molten agarose was then extracted using the Wizard DNA clean up kit (2.7.3).

[A] Dialysis tubing was prepared (Sambrook *et al* ., 1989) by boiling the tubing (cut into ~15cm lengths) in 2% Sodium bicarbonate and 1mM EDTA. After washing the tubing several times it was autoclaved in 1mM EDTA and stored at 4°C .

2.8.2 Estimation of size of fragments

The size of linear DNA fragments was estimated by comparing their mobility with those of known molecular weight. The standard molecular weight markers included λ/*Hind*III or λ/*BstE*II for fragments in the 20 - 1Kb range and ϕ174/*Hae*III or ϕ174/*Hinc*II for fragments in the 1.5 - 0.1Kb range.

2.9 DNA Sequencing

2.9.1 Sequencing strategy

The strategy for sequencing of the transfer region was based on a method described by Bankier *et al.* (1987), and similar to that used to sequence the small *Streptomyces* plasmid pIJ101 (Kendall & Cohen, 1988).

Sub-fragments of the transfer region of pIJ903 were cloned into appropriate vectors and subjected to DNA sequencing as described (2.9.2). Since this procedure alone would not produce the complete sequence, several 'shot-gun' libraries each using a different restriction enzyme were constructed. The sub-fragments cloned were then sequenced (2.9.2) to provide overlaps between the larger fragments.

2.9.2 Template preparation

All double stranded plasmid DNA templates used for sequencing were prepared using the Wizard minipreps DNA purification system (Promega). The kit was used as described by the manufacture except for a few minor modifications. Essentially the protocol was similar to that described earlier (2.6.1). 3ml of an overnight culture was pelleted by centrifugation and resuspended in resuspension buffer (50mM Tris pH 7.5, 10mM EDTA,100μg/ml RNase), lysed by the addition of lysis buffer (0.2M NaOH,1%SDS), and neutralized by the addition of 2.55M potassium acetate. After spinning to remove the cell debris the supernatant was mixed with the Wizard minipreps DNA purification resin (in 7M Guanidine Hydrochloride) and passed through a column. The column was then washed with wash buffer (200mM NaCl, 20mM Tris, 5mM EDTA, 50% ethanol) and the DNA eluted in 50μl of TE. It was

found that the quality of DNA isolated could be improved by washing the cell pellet with STE buffer before extracting the DNA. By eluting the DNA in 50µl of TE twice, the quantity could also be increased. The DNA yield from the above system was found to be between 5-10µg of DNA.

For double stranded plasmid DNA the DNA was denatured using a combination of methods described by Chen & Seeburg (1985) and Hattori & Sakaki (1986). To the DNA to be denatured (~5µg) was added 0.1 volumes of denaturing mix (2M NaOH, 2mM EDTA) and the sample incubated at 37°C for 20 mins. On completion of this incubation period 0.1 volumes of 3M Sodium Acetate was added and the DNA precipitated with 3 volumes of ethanol. The DNA then pelleted and washed with 70% Ethanol.

2.9.3 Sequencing reactions

Sequencing of DNA was carried using the standard dideoxy chain termination DNA sequencing method described by Sanger *et al.* (1977) and using the Sequenase Version 2.0 DNA sequencing kit supplied by United States Biochemical.

All the subsequent DNA sequencing reactions were performed in micro-centrifuge tubes (typically 0.5ml) and were split into three stages: annealing of primer to template, labelling & termination.

Annealing:

The DNA template (~5µg for denatured double stranded plasmid DNA & ~ 2µg for single stranded DNA) was resuspened in 6µl of water, 2µl of Sequenase reaction buffer (40mM Tris pH 7.5, 20mM $MgCl_2$, 50mM NaCl) and 2µl of the appropiate sequencing primer (Table 2.4). Annealing of primer to template was carried out by incubating at 37°C for 30 mins (for dsDNA) or by heating to 65°C for 2 mins and then allowing to cool to room temperature (for ssDNA). On completion of the incubation period the tubes were placed on ice for at least 5 mins.

Labelling:

For standard labelling reactions (reading up to 300 bases from primer) 1µl of 100mM DTT, 2µl dGTP labelling mix (7.5µM dGTP, 7.5µM dCTP, 7.5µM dTTP), 1µl [α-^{35}S] dATP (10µCi/µl 1000Ci/mmol), 1µl Mn buffer (0.15M Sodium Isocitrate, 0.1M $MnCl_2$ - only added if reading close to the primer) and 2µl of Sequenase Version 2.0 T7 DNA polymerase (diluted 1:4 in Glycerol enzyme dilution buffer - 20mM Tris, 2mM DTT, 0.1mM EDTA, 50% glycerol) were added and reaction incubated at 17-20°C for 3 mins.

Termination:

After incubating the labelling mix for 3 mins, 3.5µl was transferred to four separate tubes each containing 3µl of the appropiate termination mix ddT, ddG, ddC or ddA (80µM dITP, 80µM dATP, 80µM dCTP, 80µM dTTP, 50mM NaCl & 8µM ddATP,ddTTP, ddGTP or ddCTP) and incubation continued for a further 5 mins. Following this incubation period 4µl of stop mix (95% Formamide, 20mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol) was added and the samples stored at -20°C. Prior to loading the sequencing reactions were heated at 95°C for 5 mins and 2-3µl loaded onto the gel.

2.9.4 Gel Preparation

Denaturing polyacrylamide gels were prepared by adding 15ml (for a 6% gel) or 20ml (8%) of 40% Acrylamide stock solution[A], 50g Urea, 10ml of 10xTBE and water to 100ml. After filtering the gel mix (Whatman No.4) 660µl of 10% AMPS and 60µl of TEMED were added and the mix poured between two siliconized (with Dimethyldichlorosilane solution) glass plates (39 x 33cm & 42 x 33cm) and allowed to polymerize. The spacers used were 0.4mm thick and a sharks tooth comb (0.4mm, 5.7mm width) was used to form the wells.

TABLE 2.4 DNA Sequencing primers

Primer	sequence	for use with
pUC sequencing primer (-40) 17mer	- 5`d[GTTTTCCCAGTCACGAC] 3`	pUC/M13
pUC reverse sequencing primer (-24) 16mer	- 5`d[AACAGCTATGACCATG] 3`	pUC/M13
T7 promoter primer 23mer	-5`d[TAATACGACTCACTATAGGGAGA] 3`	pT7-7

All the above primers were supplied by New England Biolabs.

Gels were run in 1xTBE (90mM Tris, 90mM Boric acid,10mM EDTA) buffer at 70w in a BRL series 2 DNA sequencing apparatus. On completion the gels were fixed for 10 mins (0.4mm gels) or 30 mins (0.4-1.2mm gels) in 5% Acetic acid, 15% Methanol. After rinsing several times in distilled water the fixed gels were placed on two sheets of Whatman 3MM paper, covered in plastic wrap and dried for 2 hours at 80°C in a gel drier (Bio-Rad).

The dried sequencing gel was then exposed to autoradiographic film (35 x 43cm, Hyperfilm-MP Amersham) at room temperature overnight or as necessrary and developed using an X-ograph X2 automatic developer.

A 40% Acrylamide stock solution was prepared by adding 38g Acrylamide, 2g N,N Methyl Bis-Acrylamide in water. The solution was then de-ionized (`Amberlite` MB-1 monobed mixed resin) and filtered through Whatman filter paper (Bankier *et al.*, 1987)

2.9.5 Sequence Analysis

DNA Sequences were entered into a computer using the Beckman Gelmate 1000 sonic digitising system and assembled in conjunction with Microgenie sequence analysis software (Queen & Korn,1984).

The assembled sequence was analysed using 'DNA Strider" (Marck, 1988). Protein analysis were carried out using the programs of the Genetics Computer Group package (GCG) version 7.0, of the University of Wisconsin (Devereux *et al.*, 1984).

Homologies to known proteins in the Genbank, EMBL and SwissProt collections were identified by using the FASTA and WORDSEARCH programs. Structural motifs in protein sequences were identified using PROFILESCAN, and the program MOTIFS was used to search for protein motifs as defined by Bairoch (1992). Comparisons of two or more protein/DNA sequences were performed using the programs; BESTFIT, GAP or PILEUP.

2.10 Polyacrylamide Gel Electrophoresis

2.10.1 Gel Preparation

Polyacrylamide gel electrophoresis was carried on 15% gels as described by Laemmli (1970), using the BRL vertical gel electrophoresis system.

Resolving gels (40ml) were prepared by adding 20ml of 30% acrylamide to 9.2ml H_2O, 10ml 1.5M Tris (pH8.8), 0.4ml 10% SDS, 0.4ml 10% Ammonium Persulphate and the gel polymerized by the addition of 50µl of TEMED (Sambrook *et al.*, 1989).

Stacking gel (10ml) was prepared by adding 1.7ml of 30% acrylamide to 6.8ml H_2O, 1.25ml 1M Tris (pH6.8), 0.1ml 10% SDS, 0.1ml 10% Ammonium Persulphate and polymerized by the addition of 20µl of TEMED (Sambrook *et al.*, 1989).

Protein samples were boiled in an equal volume of loading buffer (5% β-Mercaptoethanol, 3.4% SDS, 15% Glycerol, 0.01% Bromophenol blue, 47mM Tris

pH6.8) for 5 mins prior to loading. Electrophoresis was performed at 40mA (4-6hours) in Tris-gylcine electrophoresis buffer (25mM Tris, 250mM gylcine, 0.1% SDS). On completion, gels were dismantled and stained in Coomassie stain (0.25% coomassie brilliant blue R250, 45% Methanol, 10% Acetic acid) for ~30mins and then destained in several changes of destain (45% Methanol, 10% Acetic acid) until bands were clearly visable.

Molecular weight markers (Sigma broad range 200-6.5kDa) were also run on protein gels in order to estimate the size of the protein bands.

2.10.2 Estimation of Protein Concentration

Protein concentrations were determined spectrophotometrically using the method described by Bradford (1976). BSA was used as a standard and absorbance readings were taken at 595nm.

2.11 *In Vivo* Protein Expression

2.11.1 Protein Expression in *E.coli*

pUC based clones were grown in L broth for 2 hours (from single colonies) and then IPTG was added to a final concentration of 1mM and growth allowed to continue overnight. However for pT7-7 based clones a single colony was grown overnight in L broth, the overnight culture was then diluted 1/40 in fresh L broth and incubation continued for a further 2 hours ($O.D._{590}$ of ~0.4), at which point IPTG was added to a final concentrartion of 1mM and the sample incubated for a further 4hours (Studier *et al.*,1990).

The induced samples were then spun and the cell pellets resuspended in protein extraction buffer (50mM HEPES pH 8.0, 1mM PMSF), normally in 10-20% of the original culture volume.

2.11.2 Protein Extraction

Both *E.coli* and *Streptomyces* proteins were extracted by sonication (MSE Soniprep). Sonication was carried out at amplitude 14 for 10 x 10sec blasts with 10 sec cooling intervals, the sonicated samples were then stored at -20°C. To prepare pellet and supernatant fractions the sonicated samples were spun in a microfuge for 2 mins at top speed, the supernatant was then transferred to a fresh tube, the pellet resuspended in buffer and spun again, the supernatant discarded and the pellet resuspended in buffer.

2.11.3 Protein Sample Preparation

E.coli protein samples were prepared by inducing the cells (2.11.1) and the proteins extracted as described (2.11.2), the resulting proteins were then analysed on SDS polyacrylamide gels.

However for *Streptomyces* grown on solid media a different approach was taken. For samples grown in liquid medium the procedure for preparing protein extracts was similar to that for *E.coli.*

To obtain *Streptomyces* samples from solid media, *Streptomyces* were grown on sterile nitrocellulose membranes (Bio-Rad Nitrocellulose disks 80mm). The membranes were placed on slightly damp agar in 90mm Sterilin plates, followed by plating of the spore sample and incubation at 30°C.

Proteins samples were prepared by removing the nitrocellulose membrane to a petri dish and scraping the *Streptomyces* off the membrane and into buffer using an inoculating loop. After spinning for 2 mins in a microfuge the pellet was resuspended in a suitable volume of protein extraction buffer (50mM HEPES pH8.0, 1mM PMSF), and the proteins extracted as described (2.11.2) before analysing on SDS polyacylamide gels.

2.12 *In Vitro* Protein Expression

The S30 *E.coli in vitro* Coupled Transcription/ Translation (CTT) kit based on a principle described by Zubay(1973). The kit supplied by Promega was used for all *in vitro* protein expression work.

2.12.1 Template Preparation & Reactions

Plasmid DNA used to perform the CTT reactions was isolated using the method described in section 2.9.2. The reactions were performed by mixing to the DNA (~5µg in 14µl) 25µl of the Amino acid mix (minus Methione), 15µl of S30 *E.coli* Extract and 1µl of L-[^{35}S] Methionine (10µCi/µl 1000Ci/mmol) and the reaction incubated at 37°C for 2 hours.

The protein samples were then precipitated using 4 volumes of Acetone (15 mins on ice) and spun at 13K rpm for 5 mins. The resulting protein pellet was dried and resuspended in SDS protein sample buffer.

2.12.2 Gel preparation & Autoradiography

Protein gels were prepared and run as described in section 2.10.2 except that pre-stained molecular weight markers (Bio-Rad Low range) were loaded. After completion of the run the gels were dried as described in section 2.9.4 and exposed to autoradiographic film.

2.13 Protein Analysis

2.13.1 Antigen Preparation

Polyacrylamide gels were prepared and run as described (2.10.2) except that whilst preparing the gel no comb was inserted thus forming one continous well. On completion the gel was stained in 0.3M Copper Chloride (Lee *et al.*, 1987) until bands appeared and washed several times in water. The band of interest was then excised using a scalpel and fragmented by repeatedly passing back and fourth through a 10ml syringe (Harlow & Lane, 1988).

An alternative method (Diano *et al.*,1987; Knudsen,1985) was to transfer the proteins onto nitrocellulose (2.13.2) and then stain the nitrocellulose for total protein using Ponceu S (2% Ponceu S in 30% TCA, 30% sulphosalicylic acid) as described by Harlow & Lane (1988). After washing with several changes of water the band of interest was excised with a scalpel and sonicated in Phosphate Buffered Saline (PBS) until the resulting solution could be passed through a 21G hyperdermic needle.

2.13.2 Electro-Blotting (S.Wilson - personal communication)

The transfer of proteins from polyacrylamide to nitrocellulose membrane was set up by cutting nine sheets of filter paper (LKB) and one sheet of nitrocellulose membrane (Schleicher & Schuell) to the same size as the gel. Three sheets of filter paper were then soaked breifly in Anode Buffer 1 (20% Methanol, 300mM Tris pH 10.4), this was followed by three sheets soaked in Anode Buffer 2 (20% Methanol, 25mM Tris pH 10.4), the nitrocellulose sheet soaked in Anode Buffer 2 , followed by the gel and finally three sheets of filter paper soaked in Cathode Buffer (20% methanol, 25mM Tris pH 7.6). This sandwich was then placed in the blotting apparatus (LKB Multiphor 2 electrophoresis unit) and the proteins transfered from gel to nitrocellulose at 200mA for 1 hour.

Once completed the system was dismantled and the nitrocellulose sheet placed between several sheets of filter paper and allowed to dry.

2.13.3 Immunoassay (S.Wilson - personal communication)

The dried nitrocellulose membrane was incubated in Tris Buffered Saline (TBS - 150mM NaCl, 25mM Tris pH7.2) and 2%w/v powdered low fat milk. After incubating for 1 hour the primary Antibody was added (1:1000 -1:2000 dilution) and incubation continued overnight.

Following the incubation with primary antibody the membrane was washed several times in TBS (~3x10 min washes) and placed in TBS and powdered

milk as previously and incubated with Anti Rabbit IgG (whole molecule) coupled to Horseradish Peroxidase (Sigma - immunochemicals), for 1 hour. The filter was again washed as above in TBS, and after the final wash the filter developed.

Two methods were used for the detection of proteins, a colorimetric reaction and chemiluminescense (ECL).

The colorimetric assay was carried out by placing the nitrocellulose filter in developing solution (30ml water, 5ml 4 Chloro-1-Napthol (5mg/ml in MeOH), and 30µl Hydrogen Peroxide) for 5-10mins or until bands were clearly visable. The colour enhancement was terminated by washing the filter in water and then drying between several sheets of filter paper.

Chemiluminescense was carried out using the ECL kit supplied by Amersham as described in their protocols booklet. Essentially the two active solutions were mixed prior to filter development (in equal quantities) and the filter placed in the solution for 1 min. After which the filter was exposed to autoradiographic film (30 sec, 1 min, 2 min exposures) and the film developed as described (2.9.4).

2.13.4 Gel Retardation Assay

Gel retardation was carried out as described by (Hennighausen & Lubon,1987) and using a Band Shift kit supplied by Pharmacia. Essentially 40µg of crude protein extract was preincubated with 4µg poly(dI-dC).poly(dI-dC) (in 10mM Tris, 1mM EDTA), 2µl NP-40, 10µl binding buffer (0.1M Tris pH7.5, 0.5M NaCl, 5mM DTT) for 15mins at room temperature. 1-5ng of end labelled DNA (~10,000cpm) was then added and the reaction incubated for a further 15mins at room temperature, on completion 4µl of loading dye (250mM Tris pH7.5, 0.2% Bromophenol blue, 0.2% Xylene cyanol, 40% glycerol) was added and the samples run on a 3% non denaturing polyacrylamide gel.

3% non denaturing polyacrylamide gels were prepared by adding 5ml of 30% acrylamide to 10ml of 5xTBE, 35ml H_2O, and polymerized by the addition of 700µl of 10% Ammonium Persulphate and 50µl of TEMED. Gels were pre-run in

1xTBE (90mM Tris, 90mM Boric acid, 10mM EDTA) for 1 hour before loading and running the samples. Gels were then dried and exposed to X-ray film as described previously (2.9.4).

2.14 Abbreviations

AMPS: Ammonium persulphate

BSA: Bovine serum albumin

DMSO: Dimethyl sulphoxide

DTT: Dithiothreitol

EDTA: Diaminoethanetetra-acetic acid disodium salt

HEPES: N-2-Hydroxyethylpiperazine-N'-2-ethanesulphonic acid

IPTG: Isopropyl β-D-thiogalactoside

MOPS: 3-[N-Morpholino]propanesulfonic acid

PBS: Phosphate Buffered Saline

PEG: Polyethylene glycol

PMSF: Phenylmethylsulfonyl fluoride

SDS: Sodium dodecyl sulphate

TCA: Trichloroacetic Acid

TEMED: N,N,N',N'-Tetramethylethylenediamine

TES: N-tris[Hydroxymethyl]methyl-2-aminoethanesulfonic acid

X-gal: 5-Bromo-4-chloro-3-indoyl-β-D-galactopyranoside

CHAPTER THREE: Sequence analysis of a critical part of the transfer region of the *Streptomyces* plasmid SCP2*.

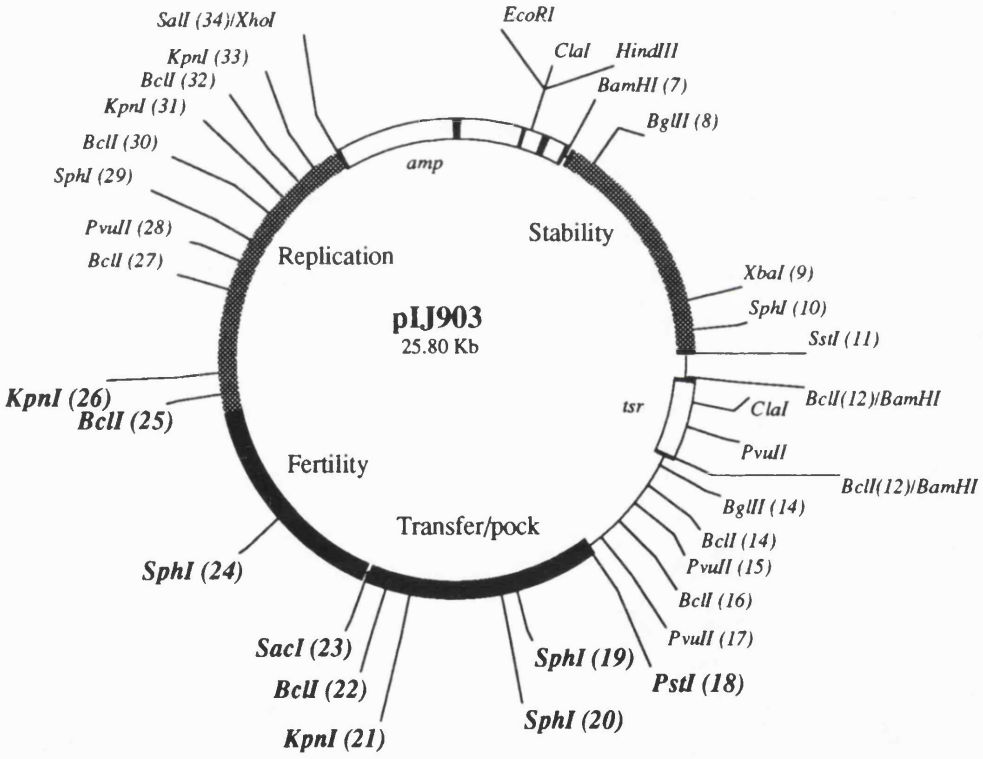
3.0 Introduction

Previously it has been suggested that much is currently known about the plasmid transfer mechanisms of Gram-negative bacteria (such as *E.coli*), however very little is known in comparison, about the transfer mechanisms of Gram-positive organisms such as *Streptomyces.* p1J101, the small, high copy number, broad host range *Streptomyces* plasmid (Kieser *et al.*, 1982) is the only *Streptomyces* plasmid for which the transfer region has been precisely located and the DNA sequence for the whole plasmid is available (Kendall and Cohen, 1988).

The topic of this study, SCP2* (Bibb & Hopwood, 1981), is very poorly understood, largely due to the difficulty experienced when working with this large, unit copy number *Streptomyces coelicolor* plasmid. At the start of this study the regions involved in plasmid transfer had only been approximately located (to the ~9kb *Pst*I (18) - *Kpn*I (26) fragment - Fig. 3.1) by *in vitro* deletion studies (Lydiate *et al.*, 1985).

The following section describes the sequencing of a critical part of the transfer region of the *Streptomyces* plasmid SCP2* using the standard dideoxy chain termination sequencing method (Sanger *et al.*, 1977). p1J903 (Figure 3.1), the *Streptomyces/E.coli* shuttle vector (Lydiate *et al.*, 1985) is used for all the initial sub- cloning of the transfer region. pIJ903 contains all the essential regions of SCP2* (plus the ampicillin resistance gene and the origin of replication from pBR327 - Fig.1.4). The strategy employed to sequence the transfer region is similar to that previously employed by

FIGURE 3.1 Map of *Streptomyces/E.coli* shuttle vector pIJ903.



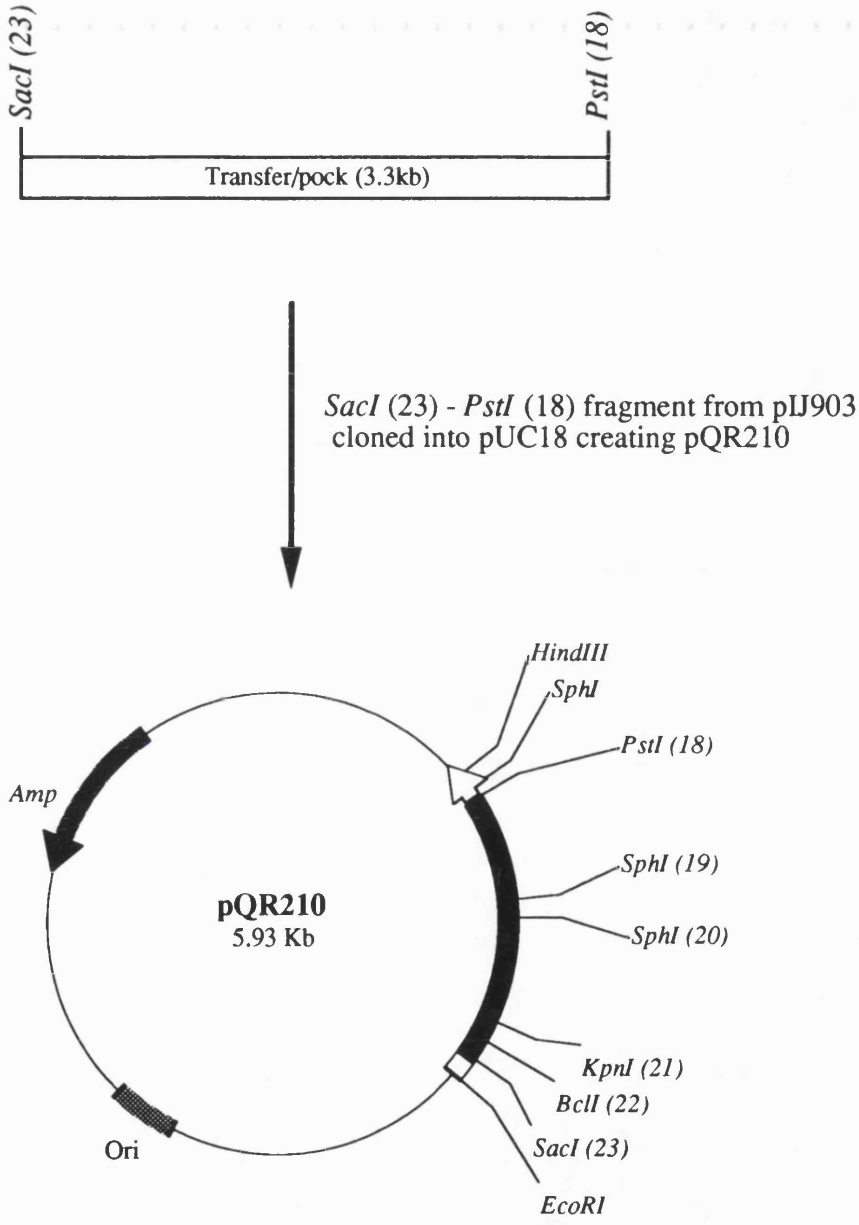
Kendall & Cohen, (1988) to obtain the DNA sequence of pIJ101. Essentially, the transfer region was sub-cloned into pUC vectors and then used to create several 'shot-gun' libraries using various frequently cutting enzymes. These constructs were then sequenced from both ends and the DNA sequence aligned with the aid of a computer.

3.1 Sub-cloning of the transfer region

A large scale plasmid DNA preparation of pIJ903 in *E.coli* was carried out as described (2.6.2). 20µg of pIJ903 DNA was then digested to completion with *Sac*I and *Pst*I and the 3.3kb *Sac*I (23) - *Pst*I (18) fragment (see Figure 3.1) containing part of the transfer region was extracted from a LMP agarose gel (2.8.1). A small fraction of this fragment was then digested with *Sph*I in order to confirm that the correct fragment had in fact been isolated, three fragments (1.7kb *Sac*I (23) - *Sph*I (20), 0.4kb *Sph*I (20) - *Sph*I (19) & 1.2kb *Sph*I (19) - *Pst*I (18)) were seen on an agarose gel as expected. The cleaned fragment was then used to set up a ligation (2.7.2) with pUC18 digested with *Sac*I and *Pst*I and the ligation mix used to transform (2.5.2) competent *E.coli* JM107 and plated out on NAXI plates (2.3.1). Individual colonies were then screened to locate a clone that contained the 3.3kb (*Sac*I (23) - *Pst*I (18)) fragment from pIJ903, this construct was named pQR 210 (Figure 3.2).

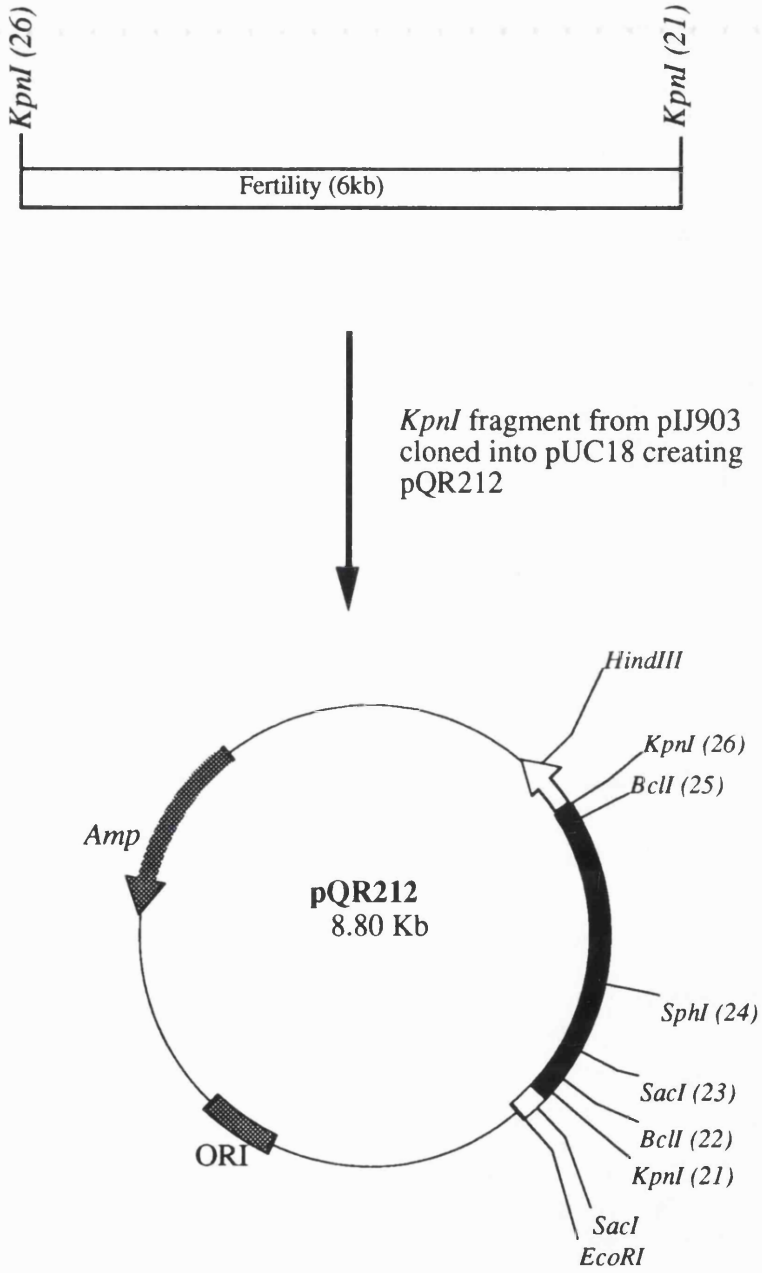
Similarly pIJ903 DNA was digested to completion with *Kpn*I, and the reaction run on a 1% LMP agarose gel. The ~6kb *Kpn*I (26) - *Kpn*I (21) fragment containing the transfer/pock region was extracted as described (2.8.1). As above, a small portion of the extracted DNA was digested with *Sph*I and run on an agarose gel. The presence of two bands (3.5kb *Kpn*I (21) - *Sph*I (24) & 4.2kb *Kpn*I (26) - *Sph*I (24)) when the gel was viewed under ultra violet light, indicated that the correct fragment had in fact been isolated.

FIGURE 3.2 Construction of pQR210



The 3.3kb *Sac*I (23) to *Pst*I (18) fragment from pIJ903 was subcloned into pUC18 creating pQR210. The construction of pQR210 is described in detail in the text. The numbers of the restriction sites refer to the map of pIJ903 shown in Figure 3.1.

FIGURE 3.3 Construction of pQR212



The 6kb *Kpn*I (26) to *Kpn*I (21) fragment from pIJ903 was subcloned into pUC18 creating pQR212. The construction of pQR212 is described in detail in the text. The numbers of the restriction sites refer to the map of pIJ903 shown in Figure 3.1.

A small portion of this *Kpn*I DNA fragment was used to set up a ligation with pUC18 (digested with *Kpn*I), which was then used to transform some competent cells (2.5.2). Recombinant colonies were screened (2.6.1) in order to find a clone that contained the ~6kb *Kpn*I fragment from p1J903, this was named pQR212 (Figure 3.3). In order to determine the orientation of the *Kpn*I fragment in pQR212 a small portion of the DNA was digested with *Sac*I and run on a 1% agarose gel. The presence of two bands (0.4kb and 6kb) indicated that the orientation of the fragment was as shown in Figure 3.3.

Thus the clones pQR210 and pQR212 contain the entire region predicted to be associated with plasmid transfer (~9Kb). For the purpose of sequencing, smaller segments of the transfer region were also sub cloned into pUC18 or pUC19 vectors. Figure 3.4 illustrates some of these constructs which span the whole of the 9kb transfer/pock region of p1J903, (a more detailed description of the individual clones is given in Table 3.1).

3.2 Construction of 'shotgun' libraries

Most of the sub-clones created above could only be used to sequence the ends since they were too large to be sequenced in their entirety, and it was not possible to sub-clone the region further due to the lack of known restriction sites within this region (only 9 restriction endonuclease sites mapped within the ~9kb transfer/pock region). An alternative approach was thus necessary in order to complete the sequence of the entire transfer region. A 'shot gun' strategy was used whereby the transfer region was digested with enzymes known to cut *Streptomyces* DNA frequently and the resulting pool of fragments cloned into the appropriate sites of pUC18/19 or M13mp18/19.

Approximately 5µg of the 3.3kb fragment (*Pst*I (18) - *Sac*I (23)) was digested with *Sau*3AI (or *Sal*I, *Taq*I, *Sma*I, *Alu*I, *Hae*III) and a small fraction of the restriction digest analysed on a 1.2% agarose gel.

TABLE 3.1 Regions sub cloned from the transfer/fertility regions of the *Streptomyces* plasmid pIJ903

Clone	Origin
pQR210	*Pst*I(18) - *Sac*I(23) fragment from pIJ903 cloned into pUC18
pQR211	as pQR210 but in pUC19
pQR212	*Kpn*I(21)- *Kpn*I(26) fragment from pIJ903 cloned into pUC18
pQR213	*Sac*I(23) - *Kpn*I(26) fragment from pIJ903 cloned into pUC18
pQR214	*Sph*I(20) - *Sph*I(24) fragment from pIJ903 cloned into pUC19
pQR500	*Sac*I(23) - *Sph*I(20) fragment from pIJ903 cloned into pUC19
pQR501	*Pst*I(18) - *Sph*I(19) fragment from pIJ903 cloned into pUC19
pQR503	*Sph*I(20) - *Sph*I(19) fragment from pIJ903 cloned into pUC19
pQR504	*Sac*I(23) - *Sph*I(24) fragment from pIJ903 cloned into pUC19
pQR505	*Sph*I(24) - *Kpn*I(26) fragment from pIJ903 cloned into pUC19
pQR506	*Sac*I(23) - *Kpn*I(21) fragment from pIJ903 cloned into pUC18
pQR507	*Sac*I(23) - *Kpn*I(21) fragment from pIJ903 cloned into pUC19
pQR508	*Sph*I(20) - *Kpn*I(21) fragment from pIJ903 cloned into pUC19
pQR509	*Kpn*I(26) - *Bcl*I(25) fragment from pIJ903 cloned into pUC19
pQR510	*Bcl*I(25) - *Sac*I(23) fragment from pIJ903 cloned into pUC19

FIGURE 3.4 Map of the clones obtained from the transfer region.

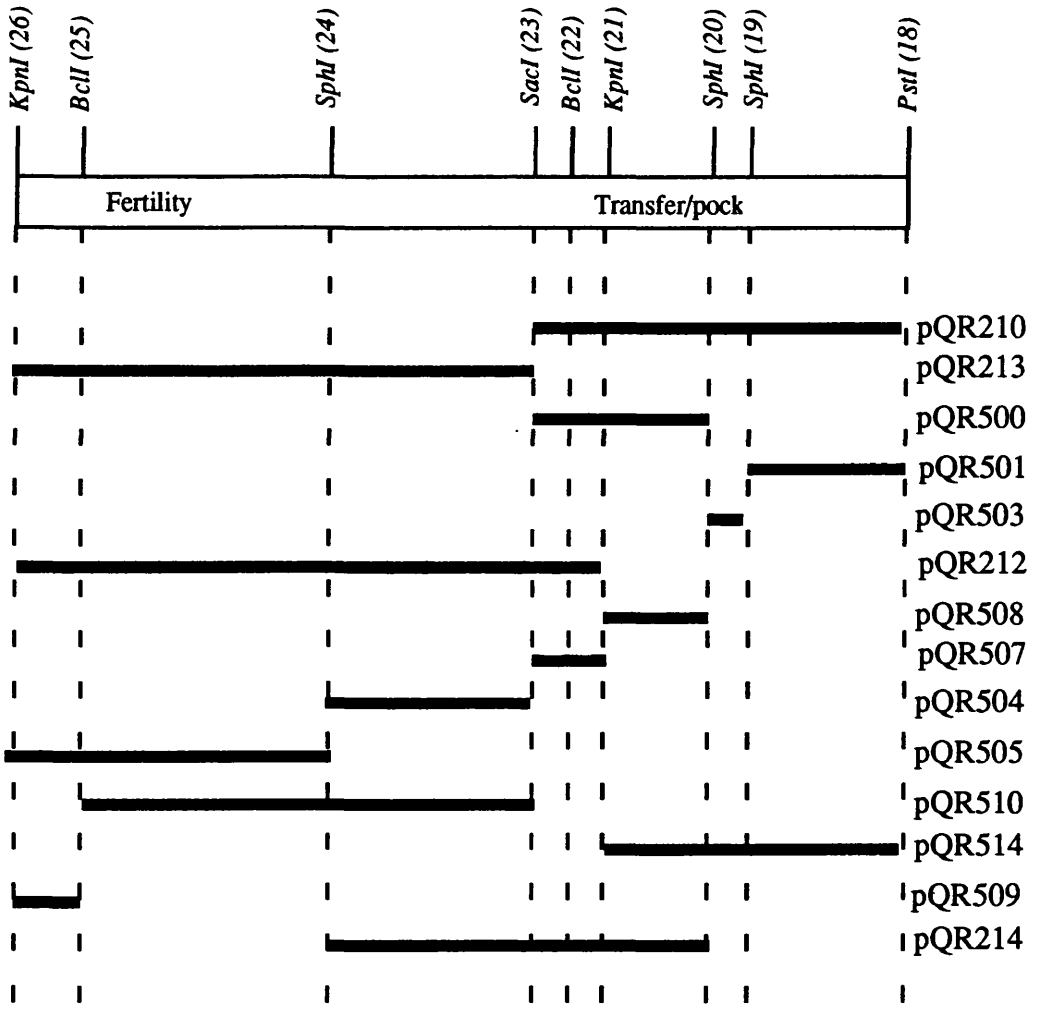


Figure 3.5 shows an example of a DNA gel illustrating the large number of fragments produced when various enzymes are used for the shotgun digestion. The remainder was used to set up a ligation, followed by transformation into competent cells. Recombinant colonies were then screened (2.6.3) in order to find clones that had taken up any inserts (Figure 3.6 shows an example of a large scale screening experiment).

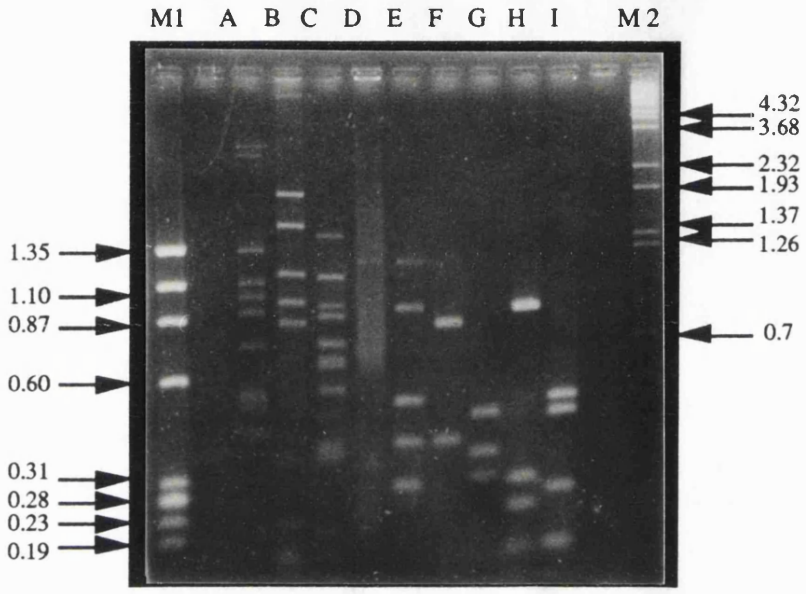
A similar procedure was carried out with various enzymes and fragments resulting in over 50 clones varying in size from 200bp to 1.6kb, all containing fragments originating from within the transfer/pock region of the *Streptomyces* plasmid pIJ903. These libraries together with the clones constructed (Table 3.1) were then sequenced in order to obtain the complete nucleotide sequence of the transfer region of the plasmid pIJ903.

3.3 Sequencing the transfer region

All the pUC based constructs were sequenced using the forward and reverse sequencing primers (2.9.3), those constructs which were smaller than 400bp (such as pQR503, pQR506 & any created using the 'shot-gun' strategy) were sequenced end to end. Single stranded DNA prepared from M13 based clones was sequenced using the forward sequencing primer only. All clones were sequenced on average three times, with some of the difficult regions being sequenced up to six times. In order to read as much sequence as possible from the gels, double loading was routinely carried out on wedged sequencing gels, and in many cases 5% polyacrylamide sequencing gels were used to run the sequencing reactions.

Developed autorads were used to read the sequence, (sequences read 2-3 times) and the information entered into a computer using a Beckman Gelmate Sonic digitising gel reader.

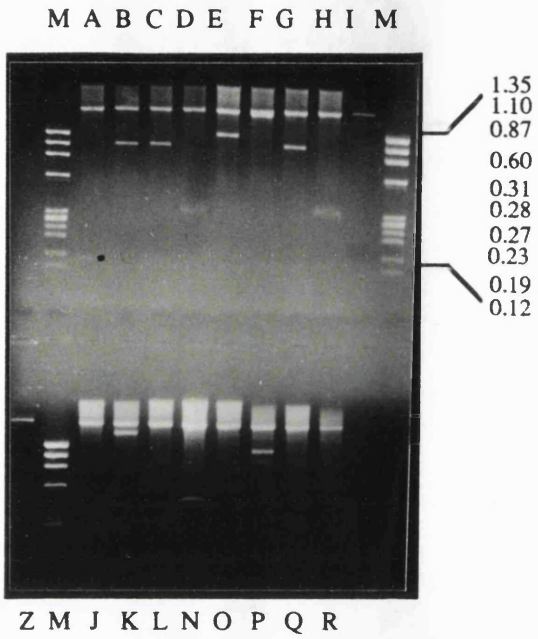
FIGURE 3.5 DNA gel showing the fragments produced by the shotgun cloning experiments.



Lanes A-E: *Sac*I (23) - *Pst*I (18) fragment digested with *Alu*I, *Sal*I, *Sau*3AI *Sma*I and *Hae*III respectively. Lanes F+G: *Pst*I (18) - *Sph*I (19) fragment digested with *Taq*I and *Sma*I respectively. Lanes H+I: *Sac*I(23) - *Sph*I (20) fragment digested with *Taq*I and *Sma*I .

Above is an example of a DNA gel showing the large array of fragments produced in the shotgun cloning experiments. Essentially ~5µg of DNA was digested and the resulting digest separated on a 1.2% agarose gel. Molecular weight DNA markers (M1 - Φ174/*Hae*III, M2 - λ/*Bst*EII) are shown, the sizes of the individual bands are indicated (kb) by arrows.

FIGURE 3.6 DNA gel showing the rapid screenng method employed to identify recombinants.



Prospective recombinant colonies were picked from NAXI plates and screened using the method described in 2.6.3. The DNA was then digested with *Eco*RI and *Hind*III and the digest separated on a 1.4% agarose gel.
Lanes A-I and J-R refer to individual recombinant colonies picked for screening. M is the molecular weight DNA marker ϕ174/*Hae*III, the size of the bands is indicated in kb. Z is control DNA (pUC18 vector) linearized with *Eco*RI. Inserts can be found in all the samples except for Lanes J, L, O, Q & R.

During the latter part of the sequencing project the emphasis turned from sequencing the entire transfer/fertility region to concentrating on the smaller yet critical transfer (*Sac*I (23) - *Pst*I (18)) region .

3.4 Modifications required for sequencing *Streptomyces* DNA

Due to the unusually high G+C content of *Streptomyces* DNA several modifications were required to the standard sequencing protocols in order to eliminate some of the problems generally associated with high G+C content templates. One of the main problems associated with such templates is band compression due to the high stability of secondary structures formed by the G+C rich template.

Methods of overcoming gel artefacts produced as a result of secondary structure formation include the use of dITP (deoxy-inosine triphosphate) in place of the standard dGTP (deoxy-guanine triphosphate) in the sequencing reactions. Alternatively, 40% formamide was incorporated in the sequencing gel for eliminating very strong compression s, however due the toxicity of formamide its use was restricted to a last resort policy. It was also found that increasing the termination reaction temperatures from 37°C to 50°C seemed to decrease the amount of secondary structures formed and thus reducing the amount of band compression's observed. The stability of the T7 DNA polymerase at the elevated temperature was not a problem since the enzyme was diluted in an increased concentration of glycerol, the effect of the extra glycerol (which would normally distort standard sequencing gels) was overcome by running the sequencing reactions on a sequencing gel that had been prepared with a Glycerol Tolerant Electrophoresis buffer (GTE) in place of Tris-Borate Electrophoresis (TBE) buffer (2.9.4).

3.5 Sequence alignment

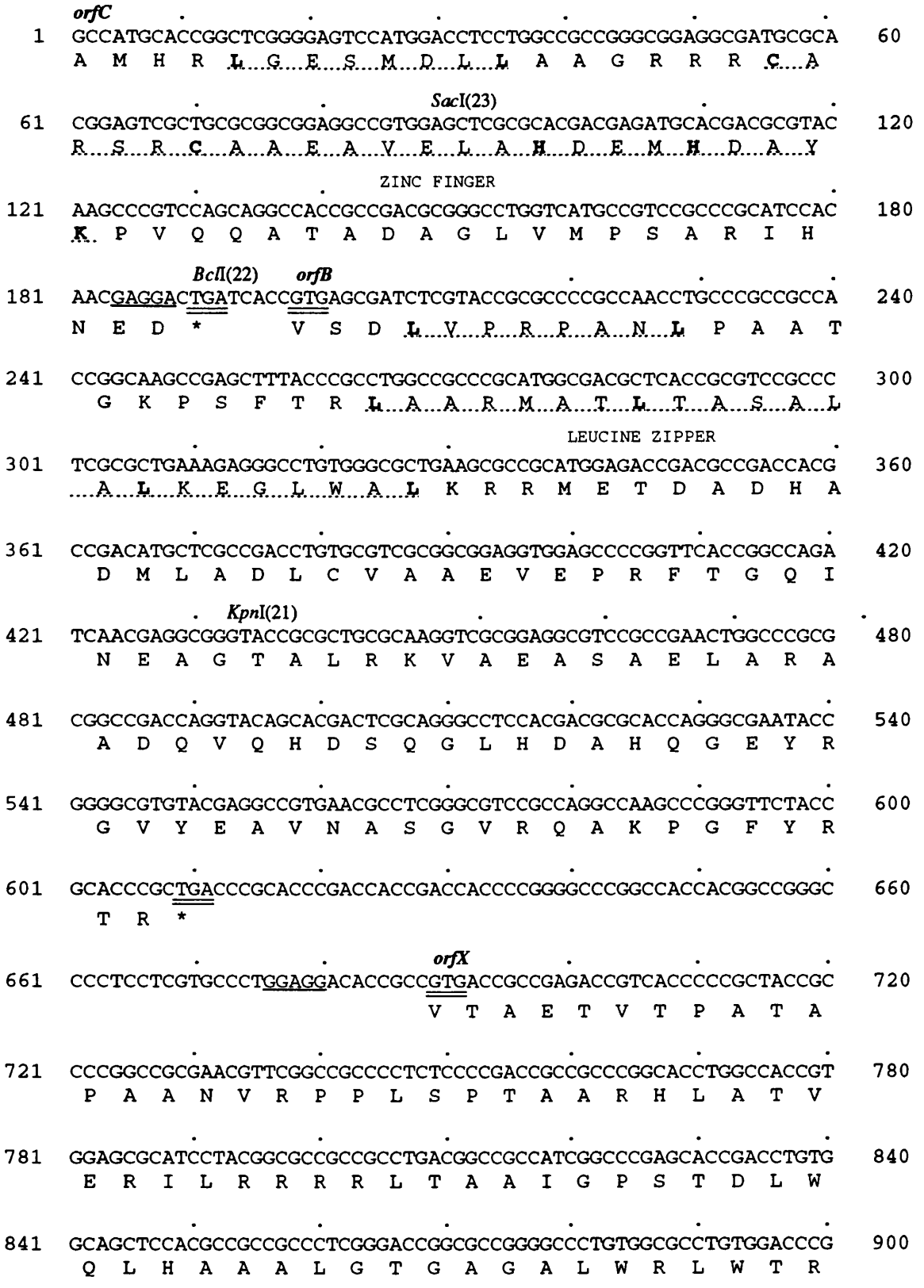
The completed sequence was aligned using Microgenie DNA sequencing software (Queen & Korn, 1984), and extends from the *Pst*I (18) site through and beyond the *Sac*I (23) site. The complete DNA sequence together with the putative open reading frames, selected restriction sites and protein motifs is presented in Figure 3.7. The complete DNA sequence consists of 3352bp with an overall G+C of 75.1% which is similar to the expected value of approximately 74% (Bibb *et al.*, 1984) for *Streptomyces* DNA sequences, the nucleotide composition of the region sequenced is shown in Table 3.2. A more comprehensive restriction map of the sequenced region is also shown in Appendix A.

TABLE 3.2 Nucleotide composition of the transfer region of pIJ903

Base	Number	% of Total
A	436	13.1%
G	1125	33.6%
C	1394	41.5%
T	395	11.8%
G+C	2519	75.1%
A+T	831	24.9%

Total number of nucleotides - 3352

FIGURE 3.7 Nucleotide sequence of the transfer region of the *Streptomyces* plasmid pIJ903.



```
901  CACCCGCGACCGTGAGGAAGGCTTCGGCGGCCGCCTCCTGACGTCCTGCTACAAGGCCGT  960
       T  R  D  R  E  E  G  F  G  G  R  L  L  T  S  C  Y  K  A  V

961  TCCGGTCCTCGGGCTGTCGACCGCGTACGGCGTCGCCCTGGCCGTCCCCGGCACCGCCTG  1020
       P  V  L  G  L  S  T  A  Y  G  V  A  L  A  V  P  G  T  A  W

1021 GTGGGAGGTCGCCGCGTCCGTCGCCGTCGCCACCGCGTCGGCAGTCGCCGTGCCCCTGAC  1080
       W  E  V  A  A  S  V  A  V  A  T  A  S  A  V  A  V  P  L  T

1081 CCGCTCCCGGGGCCTGCGCCGGGCCGCCGAGAACCTGCCCGCCGTCGTCGCCGAGCAGGC  1140
       R  S  R  G  L  R  R  A  A  E  N  L  P  A  V  V  A  E  Q  A

1141 ACCGGCCGAGGAGCCCGAACAGGACGGCTACGAGGGCGACGTCGCCCGGCACTGGGCAGC  1200
       P  A  E  E  P  E  Q  D  G  Y  E  G  D  V  A  R  H  W  A  A

1201 CTCCCCGGCGACCGGCACCACGCGCCTGGCCCACGTCCGCCAGTACCACCCGGGCCGCCC  1260
       S  P  A  T  G  T  T  R  L  A  H  V  R  Q  Y  H  P  G  R  P

1261 CGACTTCGAGGCCGTGATCCTCGCGCAGCCCGGCGAGGCCGTCCCCGGCAGCCTGGACCG  1320
       D  F  E  A  V  I  L  A  Q  P  G  E  A  V  P  G  S  L  D  R

1321 CCGGGCCGTCGCCGCCGTCTACGACGTCCCCGAGGAGGCCGTGCGGCTCGCCCCGGTCCC  1380
       R  A  V  A  A  V  Y  D  V  P  E  E  A  V  R  L  A  P  V  P

1381 CGGACACGGACCCGGCCGGATCGCCGTCTGTGTCGCCCCCGTCGAACTCCTGGCCCGCCA  1440
       G  H  G  P  G  R  I  A  V  C  V  A  P  V  E  L  L  A  R  Q

1441 GCAGCACCAGGAGCCCGGCGACCAGCTCGCCGCCCTGTGGGACGCGAAGGTGTCCGCCCC  1500
       Q  H  Q  E  P  G  D  Q  L  A  A  L  W  D  A  K  V  S  A  P

                          SalI
1501 GAAGGCGTCGCCCCGGGCGTGGAGCTGGTCGACCACCGCATCGAGGCGGACCGCATCGTG  1560
       K  A  S  P  R  A  W  S  W  S  T  T  A  S  R  R  T  A  S  *

     orfA
1561 ATGCGCGTGGAGGCCCCGGACACGCAGCTCCTCGCCCTCCCGCGACTCCCCCTCGCCCGG  1620
     M  R  V  E  A  P  D  T  Q  L  L  A  L  P  R  L  P  L  A  R

1621 GCCCTCGGGGTGGAGGACCCGGAACTCCTGATGATCGAGACGGACGGCATGGCGCACGGC  1680
       A  L  G  V  E  D  P  E  L  L  M  I  E  T  D  G  M  A  H  G

1681 GTCGTCACCCTCTACCGGGAGCACCCGCTGCTGACCGTCCGCGAGGCGACCCCGGACGAC  1740
       V  V  T  L  Y  R  E  H  P  L  L  T  V  R  E  A  T  P  D  D

                          SphI(20)
1741 CTGGTGATGGACCGGCACGGCCGTATCGCCATCGGCATGCGCCACGACGGCCGGCCCGCG  1800
       L  V  M  D  R  H  G  R  I  A  I  G  M  R  H  D  G  R  P  A

1801 CGCTGGCCGCTCTATGACCCGGAGCTGGGGGCCCTGACGGACCTCCTGGTCGGTGCCCCA  1860
       R  W  P  L  Y  D  P  E  L  G  A  L  T  D  L  L  V  G....A....P
```

```
1861  GGCTCGGGCAAGTCCGTGACCCTGCTGACCCTGATCGCCGCCGAGCGGACGAACGGCATC  1920
      G  S  G  K  S  V  T  L  L  T  L  I  A  A  E  R  T  N  G  I
         P-LOOP

1921  GTCTCCGTGGTCGCCGACGCGCAGGACGGCATGTCGCTCCCGGAGGCGGAGGGCCGGGTG  1980
      V  S  V  V  A  D  A  Q  D  G  M  S  L  P  E  A  E  G  R  V

1981  TTCCACTTCGGGGCCGGGCAGGCGGAGGCGGCCGCCACCCTCGCCGCGTACTCCGCCGTC  2040
      F  H  F  G  A  G  Q  A  E  A  A  A  T  L  A  A  Y  S  A  V

2041  GCCTCCTACCGCCAGGAGGTGAGCGCCGCGAACGGCTGGGGGTCCTTCACCCTGGGCAAG  2100
      A  S  Y  R  Q  E  V  S  A  A  N  G  W  G  S  F  T  L  G  K

                                                  SphI(19)
2101  CCGTGGCGCCTGGCCATCCTGACCATGGACGAGATCAACCGCATGCTGGCCATGGACTCC  2160
      P  W  R  L  A  I  L  T  M  D  E  I  N  R  M  L  A  M  D  S
                                DEAD BOX
2161  GGGCTGCCCGCACCGTTCCGCATGTGGGTTGCCGGGATGCTCGGGGCCGGACAGATCACG  2220
      G  L  P  A  P  F  R  M  W  V  A  G  M  L  G  A  G  Q  I  T

2221  TGGCGGAAGGTCGGCATGGGCGTCCGGATCGCCGGTCAGTCCATCCACCTCGCCGACCTT  2280
      W  R  K  V  G  M  G  V  R  I  A  G  Q  S  I  H  L  A  D  L

2281  GCGGACTCCGAGAAGATCCGCGCGAACGCCAAGAACGGGTCCGTCTGGCTCGGGCGGGTC  2340
      A  D  S  E  K  I  R  A  N  A  K  N  G  S  V  W  L  G  R  V

2341  AACTCCTCCATGACGCGCTCCATGGCGTCCGACATGACCACCGGCGCCGTGGAGATCACG  2400
      N  S  S  M  T  R  S  M  A  S  D  M  T  T  G  A  V  E  I  T

2401  CCCGTCCCGAAGTACTTCGGGGTGACCGGCACGGCCGAGCTGGATGCCGCATGGTCCGGC  2460
      P  V  P  K  Y  F  G  V  T  G  T  A  E  L  D  A  A  W  S  G

2461  GACGAGGCCCCCACCGGTCCGGTCACCGCCGGGACGGCCTGGCTCATCCAGTCCGGTCAG  2520
      D  E  A  P  T  G  P  V  T  A  G  T  A  W  L  I  Q  S  G  Q

2521  CCGTATCTCTCCCGGGTCTGGCGGGCCGTCAAGGAGGCCCGGACGTATCCGAGCCTGATC  2580
      P  Y  L  S  R  V  W  R  A  V  K  E  A  R  T  Y  P  S  L  I

2581  GGCCTCATGGAGTCCTCGCCCCTGGTCGACTTCACCCCGCCCGAGGCCGACGTCTTCCGC  2640
      G  L  M  E  S  S  P  L  V  D  F  T  P  P  E  A  D  V  F  R

2641  CGTGCTTACGACGAGGCCCTGACGATCGCCGTGGCGCTCCTGGAGGGCGACGAGGACGGC  2700
      R  A  Y  D  E  A  L  T  I  A  V  A  L  L  E  G  D  E  D  G

                SacII
2701  GGAGGCGCACCGGCCCCCGCGGCTCCCCGCCGGACCGAGCAGGCCGCCGACGCGGGCTCC  2760
      G  G  A  P  A  P  A  A  P  R  R  T  E  Q  A  A  D  A  G  S

2761  TCCCGTGTCCCGACGCCGCCCGCTACCGCGCGGACCCTCCAGGTCCTCGTTCTGGACGCC  2820
      S  R  V  P  T  P  P  A  T  A  R  T  L  Q  V  L  V  L  D  A

2821  CTGGCCGCCGGGCAGCTCCGCACCCGCGAGATCCGCAAGGCGGTAGGCGCCGGCACCGAC  2880
      L  A  A  G  Q  L  R  T  R  E  I  R  K  A  V  G  A  G  T  D
```

```
2881  GGCGGCCCGTCCTCCGGGTCCGTCGACAACGCCCTCCAGGCCCTCCAGGACGCCGGGCTC  2940
      G  G  P  S  S  G  S  V  D  N  A  L  Q  A  L  Q  D  A  G  L

2941  GTCGCCCGCGTCACCCACGGCGTGTGGGCCCGCACCGACCAGGCCGACCCCTCCTGACCC  3000
      V  A  R  V  T  H  G  V  W  A  R  T  D  Q  A  D  P  S  *

3001  CGCACCCCTCGCCGGGCCCCTGACGCGCCTTACGACGGCCGCCGGGGGCCCGGCCCTCTC  3060

3061  TTCCCCGCCGGAGGACCCCGTGACCGCCCCCAGCGTCGCCGTACAGGCCCCCGAGCGGGC  3120

3121  CGACACCAGACCAGCACCGCCCGCGTCCCTGGAGGACCGCCTGACGGCCGTGAGCGCCCT  3180

3181  CATGGACGACCGGCTCACGGTCGCCGCCCTGGCCGTCGCCGTGAACGCCGCCGTGAGCGT  3240

3241  CCCCGACGTCGGGCCCGTCGACCTGGCCGACGTCGTCCGCGTCCCCGCCGACGTCCAGCC  3300

                                                  PstI(18)
3301  CTACCGGCCGCCCGTCGACCTCTACCCGACCCCGGTCGCCGCCGTTCTGCAG          3352
```

DNA sequence and ORFs of the transfer region are shown. Translations for the putative *traB*, *traX*, *traA* and the incomplete *traC* genes are shown below the nucleotide sequence (start & stop codons are double underlined). Putative ribosome binding sites (RBS) are underlined (solid line), as are probable protein motifs (dotted line) identified using the program MOTIF (Devereux *et al.*, 1984). The locations of selected restriction sites are also indicated.

The region was sequenced as described in the Materials & Methods and in this section using the dideoxy chain termination method (Sanger *et al.*, 1977).

3.6 Sequence analysis

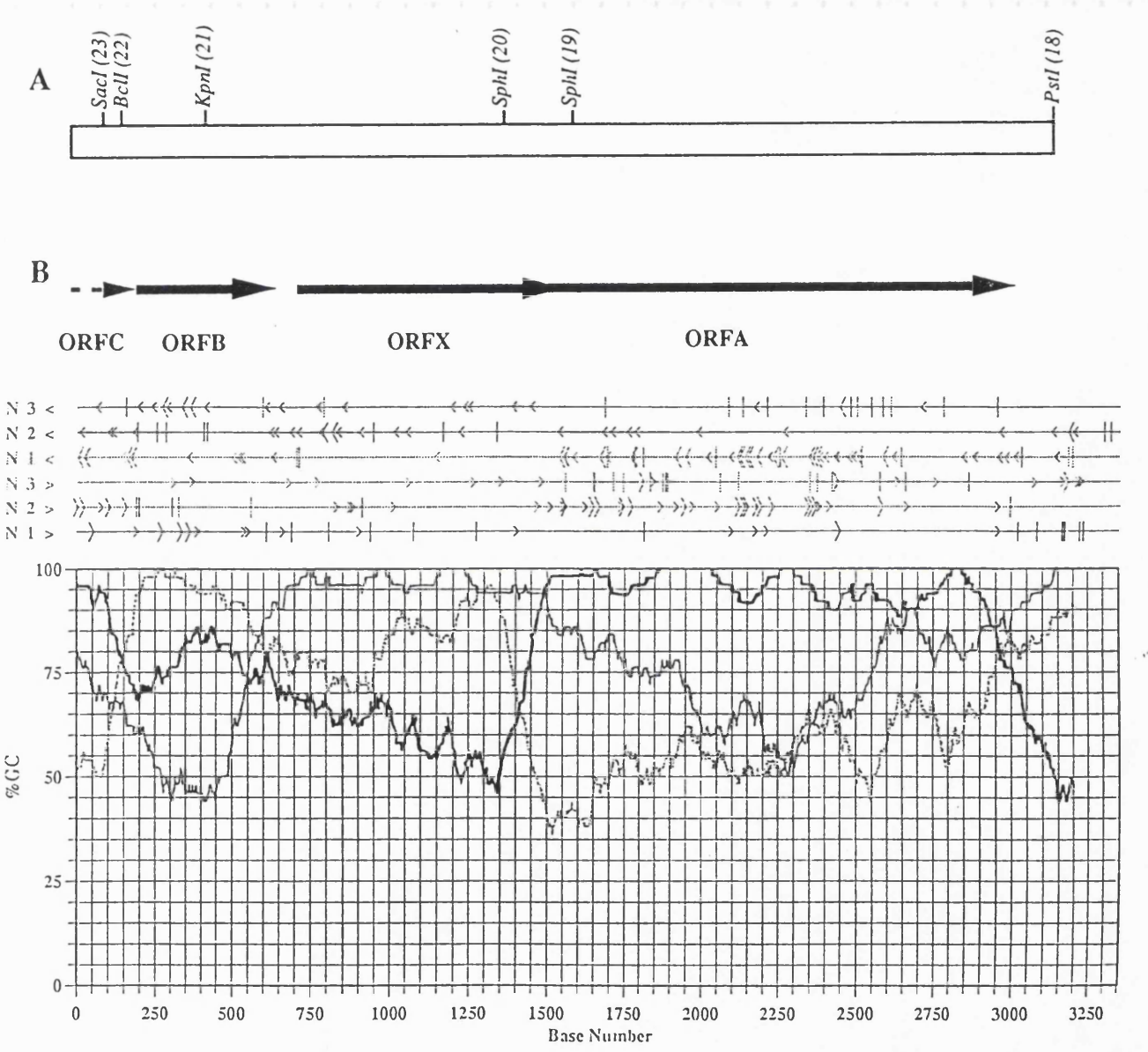
Due to the unusually high GC% content of *Streptomyces* DNA, protein coding regions tend to have a non-random codon usage, such that coding sequences have about 70% G+C in the first position of the codon, 50% G+C in the second and 90% G+C in the third position of the triplet (Bibb *et al.*, 1984). The FRAME analysis (Bibb *et al.*, 1984) uses this principle to plot graphically the G+C% composition at each of the three positions of the codon, allowing one to determine the probable protein encoding regions of the DNA at a glance.

Figure 3.8 shows such a plot for the sequence of the transfer region of pIJ903. The thick line (-) is the G+C composition at every first position of a triplet, the thin line (-) at every second position and the dashed line (- -) at every third position of the triplet over a window of 50 bases.

Regions of the sequence that do not encode proteins would lie close together due to the random use of bases. However regions encoding proteins would be far apart, such that, the third position of the codon would have the highest G+C and the second the lowest, with the first lying in between the other two lines. Analysing the FRAME plot suggests that there are three protein coding regions located around base positions 200 - 600, 700 - 1550 and 1550 - 3000 which are designated ORFs B, X, and A. Stop (I) and start (> ATG, > GTG) codons are also shown above the plot, the six possible reading frames (three in each direction) are represented as N_1, N_2, N_3 (>) and N_1, N_2, N_3 (<).

Closer inspection of the 'FRAME" plot and the sequence for suitable start and stop codons, suggests that ORF B starts at nucleotide (nt) 198 with a GTG translational start codon and would terminate at the translational stop codon TGA located at nt 611 giving rise to an ORF of 138 amino acids. Similarly ORF X extends from nt 689 (GTG) through to nt 1559

FIGURE 3.8 FRAME plot of the sequenced transfer region of pIJ903.



A - Part of the transfer region sequenced in this study.

B - Location of the putative ORFs within the transfer region.

C - Location of the start (ATG > & GTG >) and stop (l) codons in all six reading frames.

D - Graph showing %GC at the 1st (-), 2nd (-) & 3rd (--) positions of the codons over a window of 50 bases.

(TGA) giving rise to an ORF of 291 amino acids. The exact start codon for ORF A is rather more difficult to determine due the presence of many in frame start codons in the area indicated by the FRAME plot as being the likely N-terminus of the reading frame, however a translational stop codon (TGA) is located at nt 2995. There is also an incomplete open reading frame located upstream of the ORF B. The TGA stop codon for this incomplete ORF, designated as ORF C is located at nucleotide 193.

Thus the FRAME analysis suggests the presence of three putative open reading frames (ORF A, ORF B & ORF X), though the FRAME plot is a very a powerful tool for identifying protein coding regions in *Streptomyces* DNA (Kendall & Cohen, 1988), the N-terminal start sites are only approximate rather than exact and thus identifying the true protein start codon can be quite difficult in the absence of substantiating protein sequence data (Kendall & Cohen, 1988), particularly since GTG start codons appear quite frequently in *Streptomyces* DNA sequences (Buttner *et al.*, 1987).

Where necessary the choice of start codon was made using a combination of the FRAME plot and inspection of the sequences preceding the probable start codon, the sequence just upstream of the putative start codon was compared for complementarity to the 3' end of the 16S rRNA of *S.lividans* (5'- GAUCACCUCCUUUCU -3') suggesting the presence of a putative ribosome binding site (Kendall & Cohen, 1988). Strohl (1992) analysed the putative Shine-Dalgarno sequences of many *Streptomyces* genes and suggested the conserved sequence to be [A/G]GGAGG. Table 3.3 shows the sequences preceding the probable start sites of all the open reading frames; note ORF A has more than one possible start site, the three most probable ones as suggested by the FRAME plot are shown in Table 3.3, others are located either too far upstream or downstream of the predicted N-terminal region by the FRAME plot. The ATG (nt 1561) is suggested as being the most likely translational start site due to its proximity to the ribosome binding site

and the FRAME plot, however without further protein sequence data the exact site cannot be specified.

From the information available thus far, it can be concluded that there are three putative ORFs in the transfer region of the *Streptomyces* plasmid SCP2*, all of which are transcribed in the same direction. The molecular weights (as estimated from the nucleotide sequence), number of codons and overall G+C% (of each ORF) are presented in Table 3.4. The program ISOELECTRIC (Devereux *et al.*, 1984) was used in order to determine the isoelectric point of each of the three proteins (Table 3.4).

TABLE 3.3 Sequences preceding the initiation codons

ORF	Sequence
ORF B	CCCGCATCCACAACGAGGACTGATCACC**GTG**
ORF X	CCCTCCTCGTGCCCTGGAGGACACCGCC**GTG**
ORF A	ATCGAGGCGGACCGCATC**GTG**ATGCGC**GTG**

Putative initiation codons are indicated in bold.

Probable Ribosome-binding sites are underlined (bases exhibiting complimentarity to the 3` end of 16S rRNA - 5`-GAUCACCUCCUUUCU-3`)

As can be seen the largest open reading frame ORF A encodes a putative protein of 50,389 Daltons (479 codons) with an overall %GC (total) of 73%, the smallest is ORF B (138 codons) with a putative protein of molecular weight 14,703 Daltons. ORF X is the third, with the 291 codons

coding for a possible protein of 30,587 Daltons. All three ORFs show a total GC% within the range expected for *Streptomyces* protein coding regions.

TABLE 3.4 ORFs identified in the transfer region of SCP2*

ORFs	Molecular weight	No.of codons	%GC	pI
ORF A	50,389	479	73.4	5.07
ORF B		138	73.2	7.6
ORF X		291	77.8	10.3

The G+C% at each position of the coding triplet is specific to *Streptomyces* , with the first being 70% G+C, the second being 50% and 90% G+C for the third position of the codon. Table 3.5 shows the G+C% at each of the three positions in the triplet for all three ORFs. Analysis of Table 3.5 suggests that the r-is nothing untoward about the G+C% at each of the three positions and thus further confirms that these ORFs are likely to be protein coding regions.

If the base composition outside the predicted open reading frames was analysed it would be expected to be uniformly distributed within each nucleotide triplet. The maintenance of such a biased codon usage throughout each coding region and the subsequent disappearance of the triplet periodicity outside of the reading frames strengthens the assignment of initiation and termination codons for the proposed open reading frames.

Codon frequency tables can also serve as a method for confirming protein coding regions in *Streptomyces* DNA. As a consequence of the high overall GC% found in *Streptomyces* DNA, the codon usage patterns are also biased. Table 3.7 indicates the codon usage patterns of the three ORFs identified in the transfer region of the *Streptomyces* plasmid pIJ903. Further analysis of this table clearly indicates the non random nature of the codon usage patterns, since 20 out of the 61 possible codons are not used at all in ORF A and 6 further codons are only used once. A similar pattern is observed in the other two open reading frames where 26 codons for ORF B and 24 codons for ORF X are not used, and 9 codons in ORF B and 6 codons in ORF X are only used once. In non - protein coding regions one would expect to see a random distribution of codons which is clearly not the case in the above examples.

TABLE 3.5 % G+C at the 1st, 2nd and 3rd codon positions of the ORFs of the transfer region of pIJ903.

ORF	No. of codons	%G+C at codon position:		
		1st	2nd	3rd
ORF A	479	73	54	95
ORF B	138	74	50	93
ORF X	291	77	61	94

Due to this unusually high G+C content of *Streptomyces* DNA certain codons are also used more frequently than others. Codons used rarely or to a very limited extent in *Streptomyces* (but seen frequently in *E.coli*)

TABLE 3.6 Codon usage table for ORF A, ORF B AND ORF X

		ORF A	ORF B	ORF X			ORF A	ORF B	ORF X
Phe	TTT	-	1	-	Ser	TCT	-	-	-
	TTC	7	2	2		TCC	23	2	6
Leu	TTA	-	-	-		TCA	-	-	-
	TTG	-	-	-		TCG	3	2	6
	CTT	1	-	-	Pro	CCT	-	-	-
	CTC	25	5	9		CCC	11	4	15
	CTA	-	-	1		CCA	1	-	-
	CTG	22	8	15		CCG	19	2	10
Ile	ATT	-	-	-	Thr	ACT	-	-	-
	ATC	16	1	4		ACC	21	7	19
	ATA	-	-	-		ACA	-	-	-
Met	ATG	16	3	-		ACG	12	1	3
Val	GTT	2	-	2	Ala	GCT	3	-	1
	GTC	22	3	20		GCG	48	15	42
	GTA	1	2	-		GCA	3	-	5
	GTG	11	4	6		GCG	17	12	9
Tyr	TAT	3	-	-	Cys	TGT	-	-	1
	TAC	5	3	5		TGC	-	1	1
***	TAA	-	-	-	***	TGA	1	1	1
	TAG	-	-	-	Trp	TGG	10	1	9
His	CAT	-	-	-	Arg	CGT	3	-	1
	CAC	7	4	7		CGC	17	10	15
Gln	CAA	-	-	-		CGA	1	-	-
	CAG	14	6	9		CGG	14	2	11
Asn	AAT	-	-	-	Ser	AGT	-	-	-
	AAC	7	3	2		AGC	2	2	3
Lys	AAA	-	1	-	Arg	AGA	-	-	-
	AAG	8	4	3		AGG	-	-	1
Asp	GAT	1	1	-	Gly	GGT	4	1	-
	GAC	29	7	9		GGC	27	7	16
Glu	GAA	1	2	3		GGA	2	-	2
	GAG	24	7	14		GGG	15	1	3

*** Stop Codons

include GAT (Asp), CGT (Gly), GTT (Val) and GAA (Glu) (Bernan *et al.*, 1985). In contrast CGG (Arg) is found frequently in *Streptomyces* but rarely in *E.coli* (Bernan *et al.*, 1985). Therefore further examination of the codon usage patterns of the three open reading frames identified, can reveal whether there are any unusual features. GAT (Asp), CGT (Gly), GTT (Val) and GAA (Glu) are used very infrequently in all three ORFs, whereas CGG (Arg) is used frequently in both ORF A, ORF X but to a lesser extent in ORF B.

A similar observation can be made when comparing the amino acid usage charts of the three ORFs (Table 3.7). Certain amino acids seem to be favoured more so than others. Alanine for example seems to be a very popular amino acid appearing 155 times in total out of a total of 908 amino acids, thus equating to an overall percentage of 17%. On the other hand amino acids such as Phenylalanine and Asparagine seem to be less popular.

Codon usage tables, amino acid usage charts and GC% at the three positions of the triplet all serve as a method of confirming the presence of protein coding regions due to the non random nature of codons, amino acids and GC%. Similar comparisons for non protein coding regions would be expected to have a totally random appearance.

A search of the regions preceeding the putative ORFs revealed no homology to sequences associated with Streptomycete promoters (Strohl, 1992). The entire sequenced region was also subjected to a search for inverted repeats and regions of dyad symmetry, no repeats of any significance were determined.

3.7 Analysis of open reading frames

3.7.1 Search for protein homology

During the latter part of this project the DNA sequence of part of the transfer region of SCP2* was published (Brolle *et al.*, 1993). The three proteins identified from the sequence presented in this thesis were compared

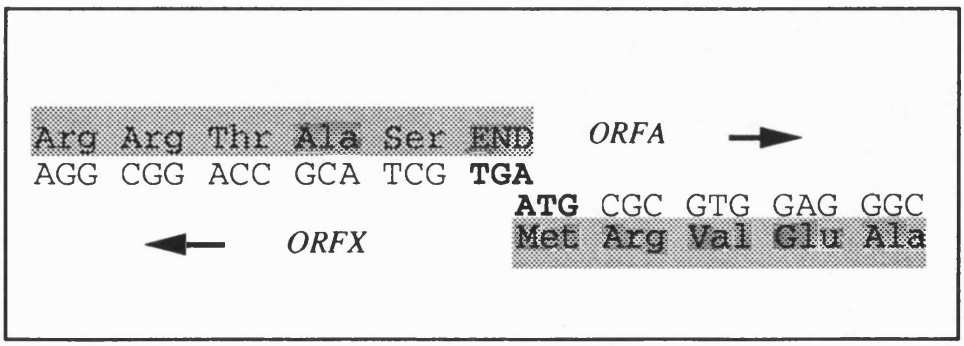
TABLE 3.7 Amino acid composition of the three ORFs

	ORF A		ORF B		ORF X	
	n	n (%)	n	n (%)	n	n (%)
Hydrophobic						
Glycine	48	10.0	9	6.5	21	7.2
Alanine	71	14.8	27	19.6	57	19.6
Valine	36	7.5	9	6.5	28	9.6
Leucine	48	10.0	13	9.4	25	8.6
Isoleucine	16	3.3	1	0.7	4	1.4
Methionine	16	3.3	3	2.2	-	-
Phenylalanine	7	1.5	3	2.2	2	0.7
Tryptophan	10	2.1	1	0.7	9	3.1
Proline	31	6.5	6	4.3	25	8.6
Neutral						
Serine	28	5.8	6	4.3	15	5.2
Threonine	33	6.9	8	5.8	22	7.6
Asparagine	7	1.5	3	2.2	2	0.7
Glutamine	14	2.9	6	4.3	9	3.1
Cysteine	-	-	1	0.7	2	0.7
Hydrophilic						
Aspartic acid	30	6.3	8	5.8	9	3.1
Glutamic acid	25	5.2	9	6.5	17	5.8
Lysine	8	1.7	5	3.6	3	1.0
Histidine	7	1.5	4	2.9	7	2.4
Arginine	35	7.3	12	8.7	28	9.6
Tyrosine	8	1.7	3	2.2	5	1.7
Total	479		138		291	

to those published by Brolle *et al.* (1993). The predicted amino acid sequence for both TraB and TraX showed 100% homology to the published sequence whereas TraA was found to be significantly different at its N-terminus. The ORF identified in this study was 36 amino acids longer (479 codons) than that proposed by Brolle *et al.* (1993). This increase in the number of codons resulted in an increase in the size of the predicted protein from 46.5kDa (Brolle *et al.*,1993) to 50.4kDa (this work). It is interesting to note that the proposed translational stop codon (TGA) for the *traX* gene overlaps the proposed ATG start codon of the *traA* ORF. This overlapping stop-start motif suggests the possibility of translational re-initiation and could indicate translational coupling (Aksoy *et al.*, 1984).

Figure 3.9 shows the predicted sequence across the junction of *traX* and *traA* illustrating the translational coupling effect suggested above. Such arrangemants of ORFs are fairly common amongst genes within an operon. Examples of translational coupling found in transfer regions of other systems include the TraI of the F plasmid (Bradshaw *et al.*, 1990), *tra* region of the conjugative Staphlococcal plasmid pSK41 (Firth *et al.*, 1993) and within the *klbB* region of the plasmid RK2 (Thomson *et al.*, 1993). The sequencing discrepancies resulting in the different predicted TraA proteins are further investigated in the next section: Chapter four.

FIGURE 3.9 Sequence across the predicted junction of ORF X & ORF A.



Having determined the probable coding capacity of the three ORFs identified within the transfer region of pIJ903, the predicted amino acid sequences were analysed for any sequence or structural similarity to previously characterised proteins of known function using the GCG sequence analysis software (2.9.5).

The programs FASTA and WORDSEARCH (Devereux *et al.*, 1984) identified no significant homology with any protein sequences in any of the databases. However a very weak similarity was observed between *E.coli* DNA Helicase and the N-terminal end of the TraA protein (NB. this similarity was not found when the published sequence of Brolle *et al.* for TraA was used to search the database). The overall similarity is weak (Figure 3.10), such a weak similarity can almost be considered as random but it is noteworthy in this instance due to the possible function of the TraA protein (discussed later). A similarly weak homology was also observed with the Tra protein of pIJ101. The aligned sequences of *traA* (SCP2*) and *Tra* (pIJ101) are shown in Figure 3.11. The WORDSEARCH profile of *traX* revealed a very weak similarity to the Rep protein of pIJ101 as shown in Figure 3.12.

Hydrophobicity plots were also drawn for all the proteins (Figure 3.13). No strong hydrophobic domains, suggesting association with membranes were found in any of the proteins, other than a short region of about 10 - 12 hydrophobic residues found in TraX.

3.7.2 Search for conserved protein 'Motifs'

Since the search for sequence similarities with previously characterised proteins of known function revealed no significant indications as to the probable functions of the identified proteins, it was decided to search the three ORFs and the incomplete *traC* ORF for homology to specific sequence motifs (Bairoch, 1992) using the programs MOTIFS and PROFILESCAN (Devereux *et al.*, 1984).

FIGURE 3.10 Comparison of predicted amino acid sequence of TraA with that of a E.coli DNA Helicase.

```
                        10         20         30         40
A                MRVEAPDTQLLALPRLPLARALGVEDPELLMIETDGMAHGVVTLYREH
                                   |||:  ::::||:::::| ::: | : |:
B  QHNQAYTEAMLTEYADFFRQVESSPLNPAQARAVVNGEHSLLVLAGAGSGKTSVLVARAG
       180        190        200        210        220        230

   50         60         70         80         90         100
   PLLTVREATPDDLVMDRHGRIAIGMRHDGRPARWPLYDPELGALTDL-LVGAPGSGKSVT
    ||:  ||:|:::::   || | : :::  ::| :  | :  :: :| |   : ::|:|:
   WLLARGEASPEQILLLAFGRKAAEEMDERIRERLHTEDITARTFHALALHIIQQGSKKVP
       240        250        260        270        280        290

   110        120        130        140        150        160
   LLTLIAAERTNGIVSVVADAQDGMSLPEAEGRVFHFGAGQAEAAATLAAYSAVASYRQEV
   ::: ::::
   IVSKLENDTAARHELFIAEWRKQCSEKKAQAKGWRQWLTEEMQWSVPEGNFWDDEKLQRR
       300        310        320        330        340        350
```

The aligned sequence of part of TraA (A - top strand) and a DNA Helicase (B - bottom strand) is shown above. The alignment was done by using the FASTA program (2.9.5). Identical amino acids are indicated by bars; double dots indicate similar amino acids.

This interestingly revealed the presence of a P-loop motif located between nucleotide numbers 1861 and 1875 (underlined in Figure 3.7). The [GxxxxGKS] sequence also referred to as the 'A' consensus sequence (Walker *et al.*, 1982) common to ATP/GTP binding proteins (Saraste *et al.*, 1990), suggests that ATP may be required for the TraA protein product to function. Brolle *et al.* (1993) also reported the presence of a possible 'DEAD' box motif ([V/I]LDEADx[M/L]LxxGF) common to DNA Helicases (Linder *et al.*, 1989), this is also observed (nt: 2124 - Figure 3.7) in the *traA* ORF identified here when some mismatches are allowed for in the MOTIFS search. Though the match is poor (Figure 3.14) the highly conserved -DE- pair is present (Linder *et al.*, 1989).

FIGURE 3.11 Comparison of predicted amino acid sequence of TraA with that of Tra of pIJ101.

```
  5 APDTQLLALPRLPLA.RALGVEDPELLMI.............ETDGMAHGV 41    A
    ||     :|  : : | | :::  || |                | : | :
  2 APTLAPFAARWDAEADRRMALRTPEHLKALMDAQKGYNSARSTAATAKSQ 51    B
             .          .          .          .          .
 42 VTLYR...EHPLLTVREA..TPDDLVMDRHGRIAIGMR..HDGRPAR... 81
     |  |    :|| ||| |  |:|     :::     ::  : : ||
 52 RTAARAASKNPLSTVRRAARTADKAARTHRDQAKTKLKAARKDYPATLRA 101
             .          .          .          .          .
 82 WPLYDPELGALTDLLVGAPGSGKSVTL.....LTLIAAERTNGIVS...V 123
    :::     : |: :  ::| | :     :     | ||||:    :::   |
102 RAVQAHAMHAVPGAGISALGWDQAGAWPVAGSLALIAANVAALVIGRRKV 151
             .          .          .          .          .


             .          .          .          .          .
394 QAADAGSSRVPTPPATARTLQVLVLDALAAGQLRTREIRKAVGAGTDGGP 443
    : | |:   ||  |  | |    |  |:: |     ::       :|  |
543 ETAPAAE..VPAQPTKAPTNREKVAAAIGTGATTVADV......ATVTGI 584
             .          .          .
444 SSGSVDNALQALQDAGLVARVTHGVWARTDQADP 477
      |||  |:  | ||| | |   | :    | :
585 NKGSVSKAVKQLLDAGEVLRSEDGSLSVVTQVGE 618
```

The aligned sequences of parts of TraA (A- top strand) and Tra of pIJ101 (B- bottom strand) are shown above. The alignment was done by using the BESTFIT program (2.9.5). Identical amino acids are indicated by bars; double dots indicate similar amino acids.

Analysis of the predicted amino acid sequence of the *traB* gene product showed a perfect match to a structure referred to as the Leucine Zipper (Landschlulz *et al.*, 1988). This motif ([LxxxxxxLxxxxxxLxxxxxxL] located between nucleotides 264 - 329 in Figure 3.7) is known to be involoved in protein-protein interactions of several eukaryotic proteins (Landschlulz *et al.*, 1988), and is in fact found in the *klBC* gene of the *kil-kor* regulon of the plasmid RK2 (Thomson *et al.*, 1993). It may be posible to extend this motif if a gap is allowed for at the N-terminus, since a [LxxxxxxL] is found at nt:207 (see figure 3.7 & 3.14). However no other DNA binding domains such as the

FIGURE 3.12 Comparison of predicted amino acid sequence of TraX with that of Rep of pIJ101.

```
  2 TAETVTPATAPAANVRPPL.........SPTAARHLATVERILRRRRLTAA 43    A
137 TEVTVGQINGWHPHIHAIVLVGGRTEGERSAKQIVATFEPT......GAA 180   B

 44 IGPS.TDLWQLHAAALGTGAGALWRLWTRTRDREEGFGGRLLTSCYKAVP 92
181 LDEWQGHWRSVWTAALRKVNPAF......TPDDRHGVDFKRLETERDAND 224

 93 VLGL..STAYGVALAVPGTAWWEVA.ASVAVATASAVAV..PLTRSRGLR 137
225 LAEYIAKTQDGKAPAL......ELARADLKTATGGNVAPFELLGRIGDL. 267

138 RAAENLPAVVAEQAPAEEPEQDGYEGDVARHWAASPATGTTR.LAHVRQY 186
268 ..TGGMTEDDAAGVGSLEWNLSRWH.EYERATRGRRAIEWTRYLRQMLGL 314

187 HPGRPDFEAVILAQPGEAVPGSLDRRAVAAVYD.........VPEEAVRL 227
315 DGGDTEADDLDLLLAADADGGEL.RAGVAVTEDGWHAVTRRALDLEATRA 363

228 APVPGHGPGRIAVCVAPVELLARQQHQE.....PGDQLAALWDAKVSAPK 272
364 AEGKDGNEDPAAVGERVREVLALADAADTVVVLTAGEVAEAY.ADMLAAL 412

273 ASPRAWSWSTTASRRTA 289
413 AQRRE...EATARRRRE 426
```

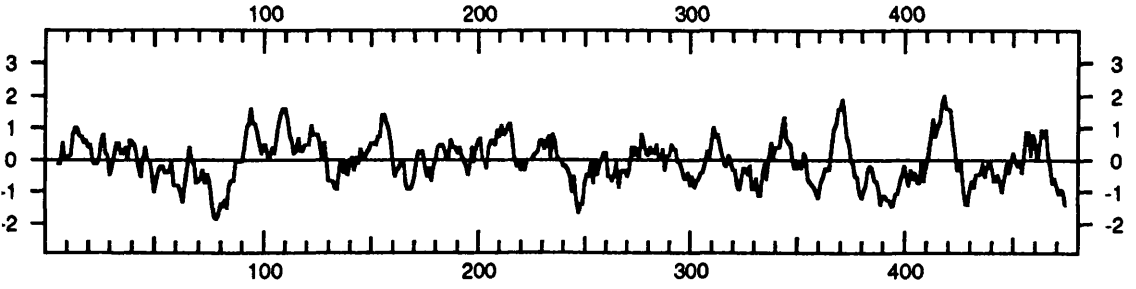
The aligned sequence of TraX (A- top strand) and Rep of pIJ101 (B- bottom strand) are shown above. The alignment was done by using the BESTFIT program (2.9.5). Identical amino acids are indicated by bars; double dots indicate similar amino acids.

Helix-turn-Helix or Zinc finger motifs were found in the predicted *traB* sequence. No conserved motifs were found when a similar search was carried out on the *traX* ORF using either MOTIFS or PROFILESCAN.

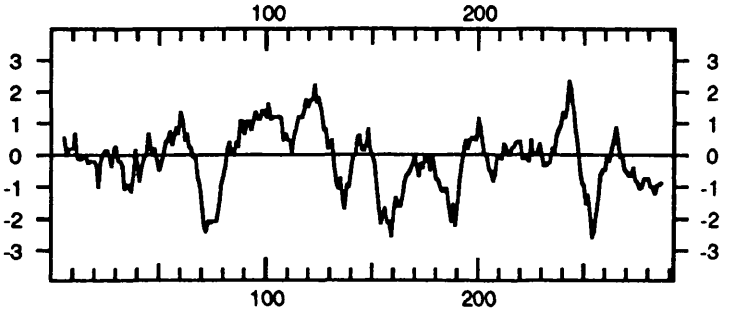
The incomplete *traC* ORF was subjected to a similar search for protein motifs. No significant matches were observed, though a similarity to a Zinc finger (C2H2 type) domain (Klug & Rhodes, 1987) was noted. The Zinc

FIGURE 3.13 Hydrophobicity plots of the TraA, B & X proteins.

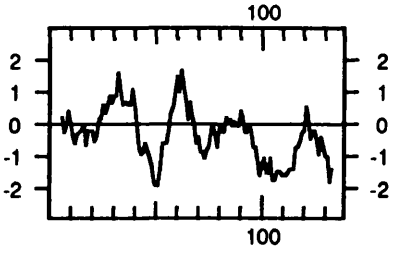
TraA



TraX



TraB



The hydrophobicity plots drawn for all three proteins are shown above. The plots were obtained using DNA Strider (Marck, 1988). No strongly hydrophobic domains sufficient in length to span a membrane were identified, other than a short hydrophobic segment 10 - 12 residues in length found in TraX (located between 80 - 110 bp on the above diagram).

finger domains (Cx{4}Cx{12}Hx{3}H) are found in numerous DNA binding proteins (Klug & Rhodes, 1987), however with only part of the sequence available the citing of such a domain must be made with some caution. A summary of the protein motifs identified within the ORFs in the sequenced region are summarized in Figure 3.14.

FIGURE 3.14 **Protein motifs identified within the transfer region of pIJ903.**

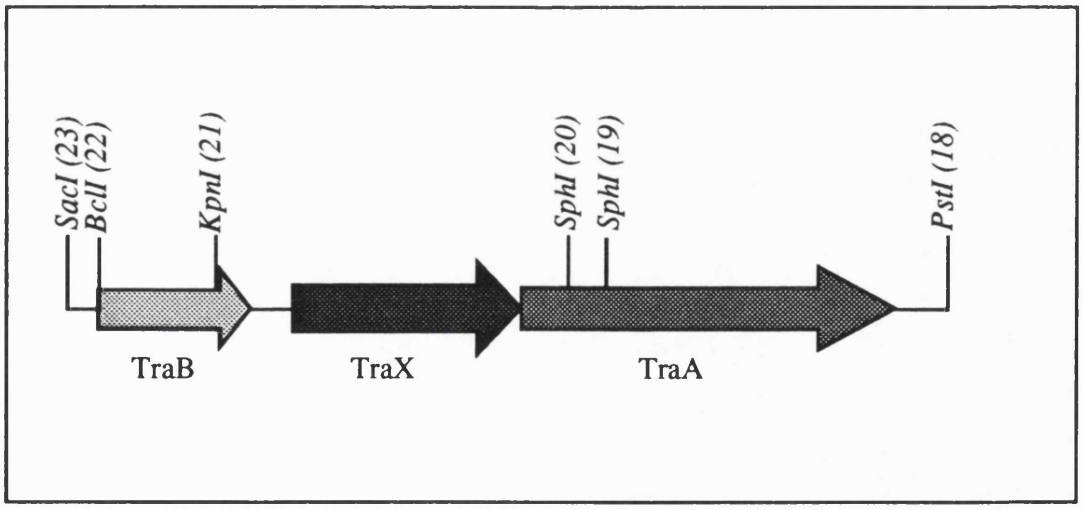
p-Loop	**G**XXXX**GKS**
traA	**G**APGS**GKS**
DEAD box	V/I**LDEAD**XM/L**L**XX**GF**
traA	TM**DE**INR**ML**AMDS
Leucine Zipper	**L**XXXXXX**L**XXXXXX**L**XXXXXX**L**
traB	**L**VPRPAN**L**[11x]**L**AARMAT**L**TASALA**L**KEGLWA**L**
Zinc Finger	**C**XXXX**C**XXXXXXXXXXXX**H**XXX**H**
traC	**C**ARSR**C**AAEAVELA**H**DEM**H**DAY**K**

3.8 Summary

A critical part (3352bp - *Sac*I(23) to *Pst*I(18)) of the transfer region of the *Streptomyces* plasmid pIJ903 (a derivative of SCP2*) has been sequenced and the resulting primary sequence analysed.

FRAME analysis identified three putative open reading frames, all of which are transcribed in the same direction. The predicted amino acid sequence of these three ORFs has been determined and suggests that proteins of 50.4kDa (ORF A), 14.7kDa (ORF B) and 30.6kDa (ORF X) would be encoded by them. The locations of these ORFs within the sequenced transfer region are summarised in Figure3.15.

FIGURE 3.15 Location of the genes identified within the transfer region of the *Streptomyces* plasmid SCP2*.



The *traA* sequence proposed in this work is a contradiction to that published (Brolle *et al.*, 1993), the differences causing a change in the predicted size of the protein and a change in the N-terminal end, leading to the possibility of the *traX* and *traA* ORFs being translationally coupled.

The predicted amino acid sequences of the three open reading frames were also analysed for sequence homology to previously characterised proteins of known function. Only *traA* showing any similarity, though very weak, with a DNA Helicase and the *Tra* gene product of pIJ101. No significant homologies were found with the other two sequences. Further analysis for sequence motifs revealed a match to the P-loop motif (nucleotide binding fold domain of ATP requiring enzymes) and a DEAD box motif (common to DNA Helicases) in the sequence for *traA*. The possibility of a leucine zipper motif (involved in regulation by DNA binding) in the *traB* ORF has been suggested. However this is far from being a specific pattern and since no other DNA binding domains were identified, *traB* cannot be termed as being a DNA binding protein without further analysis. *TraX* on the other hand failed to show any significant match to any of the motifs in the data banks.

CHAPTER FOUR: Protein analysis of the transfer region of SCP2*.

4.0 Introduction

The previous section shed light on a contradiction between the amino acid sequence of the *traA* ORF proposed here and that published by Brolle *et al.* (1993). As mentioned, the divergence in the sequences caused the two proteins to be significantly different at their N-terminal ends, but with similar C-terminal ends. The *traA* ORF identified during this work translated to a protein 36 amino acids longer (at the N-terminus), than that published. The sequencing discrepancies causing these differences are further investigated in this section. The difference in lengths of the two reading frames proposed for *traA* also lead to differing molecular weights, 46.5kDa being the published value, whereas a molecular weight of 50.4kDa is estimated by this work.

The change in the position of the translational start codon also resulted in the possibility of the *traX* and *traA* open reading frames being translationally coupled, such that the termination codon (TGA) of *traX* overlaps the start codon (ATG) of *traA* .

```
traX     Ala Ser End
         GCA TCG TGATG CGC GTG
                 Met Arg Val      traA
```

Translational coupling is fairly wide spread amongst bacteria, first being suggested in the *trp* operon of *E.coli* between the *trpD* and *trpE* genes (Oppenheim & Yanofsky, 1980) where it was suggested that translation of the upstream gene (*trpD*) was required for the efficient translation of the downstream gene (*trpE*) (Oppenheim & Yanofsky, 1980) .

An important point to note is that the *traA* gene product seems to be one of the major proteins involved in plasmid transfer, since disruption of this gene results in no plasmid transfer and the complete abolition of pocks (Brolle *et al.*, 1993). Thus determining the true translational start of *traA* would be of fundamental importance if further work is to be carried out on this major transfer protein of SCP2*. This section concentrates on further sequence analysis of the disputed region in order to confirm the sequence and analyses some different approaches of studying the predicted protein products encoded by the transfer region.

4.1 Sequence comparisons

The differences in the two postulated sequences are located to nucleotide positions (nt)* 1616, 1651, 2067 & 2554 as shown below, the sequence published by Brolle *et al.* (1993) is underlined and the upper sequence is the sequence determined in this work. The differences are shown in bold type.

nt. 1616: CCCCTC GCCCGG

CCCCTC**G**GCCCGG

nt. 1651: ATCGAG ACGGAC

ATCGAG**A**ACGGAC

nt. 2067: GTGAGCG**C**CGCGA

GTGAGCG**G**CGCGA

nt. 2554: TCAAG**A**AGGCCCGG

TCAAG**G**AGGCCCGG

* numbers refer to the nucleotide positions in Figure 3.7.

The effect of the DNA sequence differences on the predicted *traA* ORFs are illustrated in Figures 4.1a &b. As can be seen the two additional bases

proposed in the published sequence are located towards the N-terminal end of the ORF, causing the start site (as determined by the FRAME plot & the presence of a putative ribosome binding site) to shift by an additional 36 amino acids. The two base substitutions are located in the central region of the ORF (shown in Figure 4.1b), causing the substitution of amino acids from Alanine to Glycine (nt. 2067) and Glutamic acid to Lysine (nt. 2554). It is interesting to note that the differences in the DNA sequence do not cause a major difference in the predicted amino acid sequence other than the N-terminal ends and two substitutions in the amino acids.

4.2 Construction of pQR516

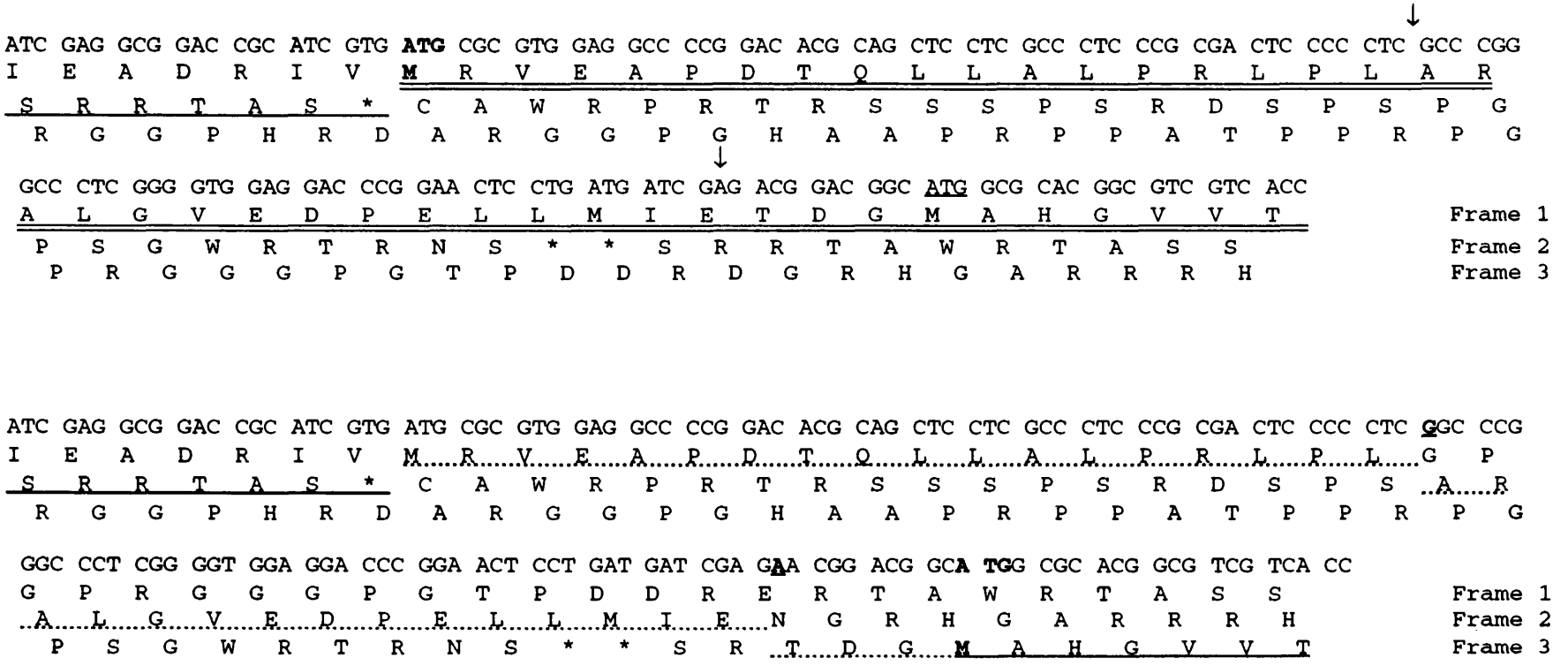
The major difference in the TraA proteins occurred as a result of the base differences located at nucleotide positions 1616 and 1651. Analysis of the restriction sites in this region identified a 250bp *Sph*I (nt:1782) and *Sal*I (nt:1530) fragment containing these two differences of interest. 10μg of pQR500 DNA was digested to completion with the above enzymes and the resulting digest resolved on a LMP agarose gel. The 250bp fragment was extracted from the gel (2.8.1) and cloned into pUC18 creating the 2.9kb construct designated pQR516 (Figure 4.2).

4.3 Sequence analysis of the disputed region

pQR516 (Figure 4.2) was sequenced using both forward and reverse sequencing primers (2.9) in order to obtain the DNA sequence of both strands of DNA. Selected regions of the sequence are presented in Figure 4.3. As can be observed the additional bases proposed by Brolle *et al.* (1993) (underlined in Figure 4.3) are clearly absent from both strands of the DNA.

The other discrepancies reported earlier (base substitutions), were also investigated by resequencing clones in which the sequence had been

FIGURE 4.1a DNA sequence comparison across the junctions of the *traX* & *traA* ORFs.



Top sequence: is that proposed by this work. The **ATG** start codon for *traA* is shown in bold type, and the amino acid sequence (Frame 1) is shown (double underlined). The start codon proposed by Brolle *et al.* is underlined and the positions of the sequencing discrepancies are indicated by arrows.
Bottom sequence: is that proposed by Brolle *et al.* The **ATG** start codon is shown in bold face, the amino acid sequence (Frame 3) is underlined and the sequencing discrepancies are in bold face/underlined. The dotted line indicates the amino acid sequence proposed by this work and the change in reading frames caused by the additional nucleotides in the sequence postulated by Brolle *et al.* can be seen. The amino acid sequence underlined by a single line is the C-terminal end of *traX*.

FIGURE 4.1b Amino acid sequence comparison of the two proposed SCP2* *traA* ORFs.

```
  1 MRVEAPDTQLLALPRLPLARALGVEDPELLMIETDGMAHGVVTLYREHPLLTVREATPDD 60
                                        ||||||||||||||||||||||
                                        MAHGVVTLYREHPLLTVREATPDD

 61 LVMDRHGRIAIGMRHDGRPARWPLYDPELGALTDLLVGAPGSGKSVTLLTLIAAERTNGI 120
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    LVMDRHGRIAIGMRHDGRPARWPLYDPELGALTDLLVGAPGSGKSVTLLTLIAAERTNGI

121 VSVVADAQDGMSLPEAEGRVFHFGAGQAEAAATLAAYSAVASYRQEVSAANGWGSFTLGK 180
    |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
    VSVVADAQDGMSLPEAEGRVFHFGAGQAEAAATLAAYSAVASYRQEVSGANGWGSFTLGK

181 PWRLAILTMDEINRMLAMDSGLPAPFRMWVAGMLGAGQITWRKVGMGVRIAGQSIHLADL 240
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    PWRLAILTMDEINRMLAMDSGLPAPFRMWVAGMLGAGQITWRKVGMGVRIAGQSIHLADL

241 ADSEKIRANAKNGSVWLGRVNSSMTRSMASDMTTGAVEITPVPKYFGVTGTAELDAAWSG 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    ADSEKIRANAKNGSVWLGRVNSSMTRSMASDMTTGAVEITPVPKYFGVTGTAELDAAWSG

301 DEAPTGPVTAGTAWLIQSGQPYLSRVWRAVKEARTYPSLIGLMESSPLVDFTPPEADVFR 360
    |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
    DEAPTGPVTAGTAWLIQSGQPYLSRVWRAVKKARTYPSLIGLMESSPLVDFTPPEADVFR

361 RAYDEALTIAVALLEGDEDGGGAPAPAAPRRTEQAADAGSSRVPTPPATARTLQVLVLDA 420
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    RAYDEALTIAVALLEGDEDGGGAPAPAAPRRTEQAADAGSSRVPTPPATARTLQVLVLDA

421 LAAGQLRTREIRKAVGAGTDGGPSSGSVDNALQALQDAGLVARVTHGVWARTDQADPS*  479
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    LAAGQLRTREIRKAVGAGTDGGPSSGSVDNALQALQDAGLVARVTHGVWARTDQADPS*
```

The complete amino acid sequence of both the postulated *traA* ORFs is presented above (aligned using BESTFIT). The upper sequence is the one proposed by this work and the lower one is that postulated by Brolle *et al.* (1993).The differing size and amino acids are indicated in bold type.

FIGURE 4.2 Construction of pQR516.

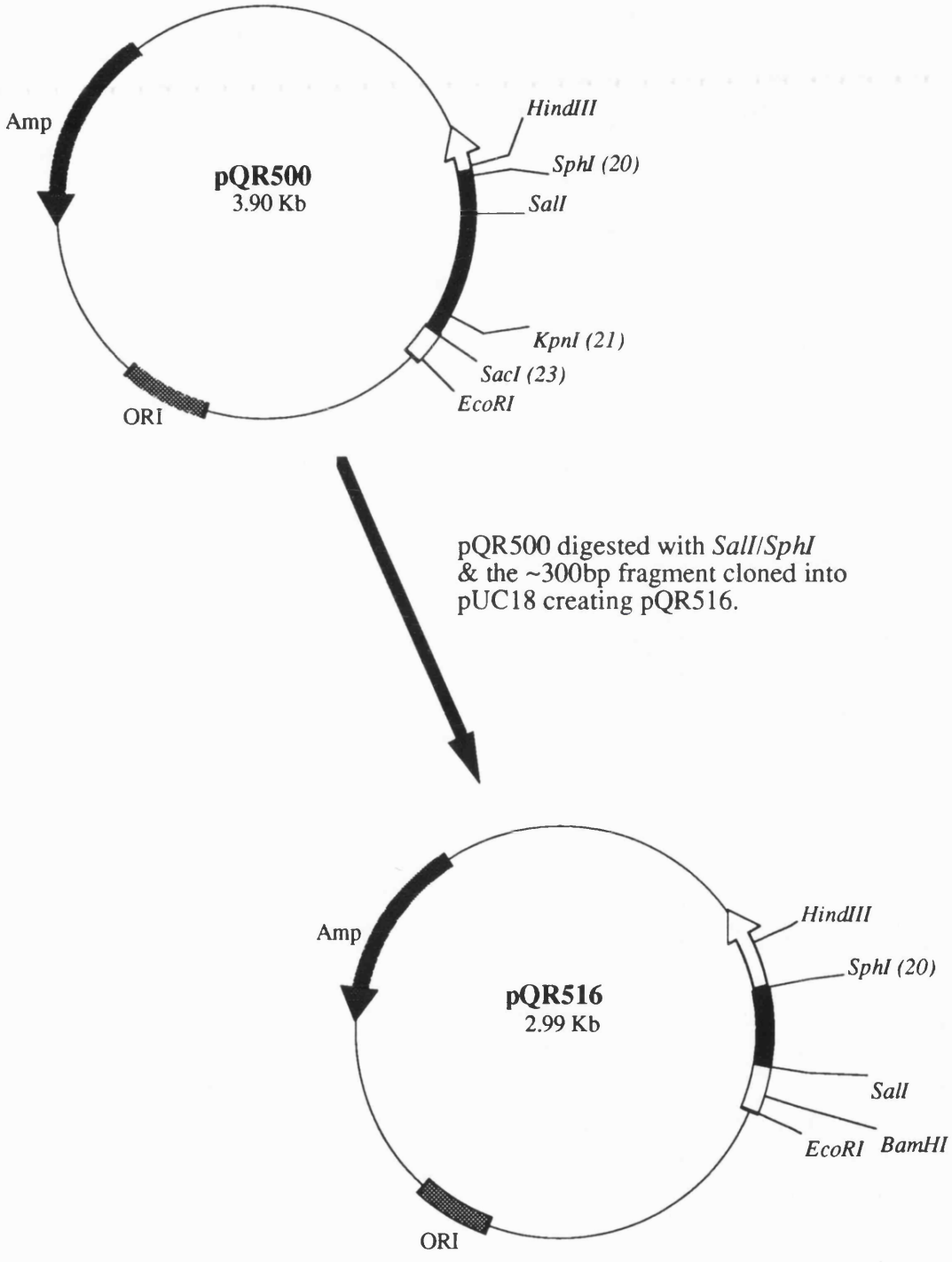
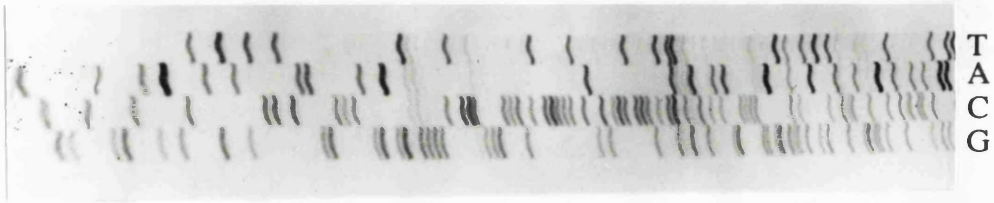


FIGURE 4.3 Sequence of pQR516



GGCA GAGCT·····CCCG CTCCCCC
GGCA**A**GAGCT·····CCCG**G**CTCCCCC

CCCCCTCG CCC·····TCGAGA CGC
CCCCCTCG**G**CCC·····TCGAGA**A**CGG

Auto radiograms of the DNA sequence of selected regions of pQR516 are shown. The sequence was obtained using forward & reverse sequencing primers, by the chain termination sequencing method (2.9). The sequence postulated by Brolle and co-workers is underlined.

previously located. Figure 4.4 shows the sequence of clones pQR503 (containing the *Sph*I (19) to *Sph*I (20) fragment - Table 3.1) and Taq 01/01 (containing a 463bp *Sph*I (19) to *Taq*I (nt:2610) fragment), from which again it can be observed that the sequence proposed by this work is likely to be the correct sequence for the transfer region of the *Streptomyces* plasmid SCP2*. These regions and those from pQR516 were sequenced many times (> 6 times) using different clones and sequencing different strands, each time the sequence read agreed with that proposed by this work.

4.4 Restriction enzyme analysis of the sequenced region

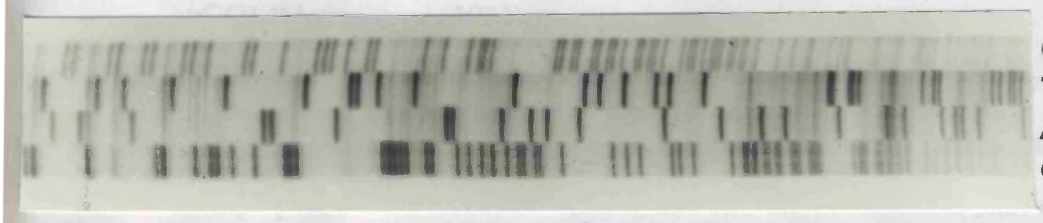
The complete *Sac*I to *Pst*I region was analysed for restriction sites and compared to the sites identified in the published sequence, in order to look for differences in the number or location of sites and to see if the difference in the sequence had caused the addition or deletion of any sites.

This indicated that there was in fact an additional *Ava*I restriction site created, by the sequence proposed in this work. However when further analysed it was obvious that simple restriction enzyme analysis of the fragment could not be used to verify our sequence since there were two *Ava*I restriction sites located 8bp away from each other in the *Sal*I to *Sph*I fragment of pQR516. Running such a digest on a polyacrylamide gel would not allow one to see the presence of the 8bp fragment. An alternative approach would be to digest the DNA with *Ava*I and end label the fragments, followed by analysis on a polyacrylamide gel. However this also has its problems since there is no guarantee that the small 8bp fragment would not be lost during the labelling/cleaning steps of this reaction.

4.4 Construction of pQR514

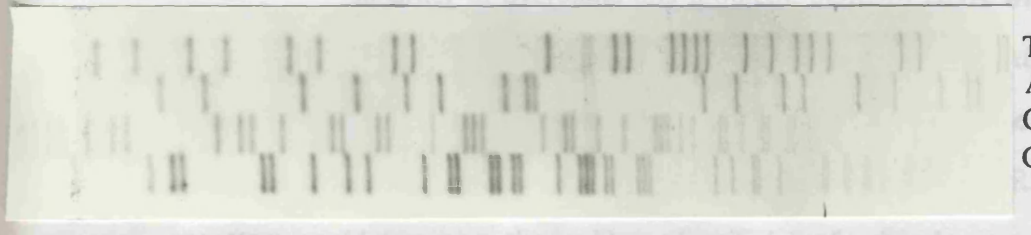
In order to examine the *traA* gene product, it was decided to use the S30 *E.coli in vitro* Coupled Transcription/Translation (CTT) kit (Zubay,1973).

FIGURE 4.4 Sequence of pQR503 & Taq 01/01.



GGGGGTCGGCAAGCGCCGCGAGTGGAGG.

GGGGGTCGGCAAGCGCGGCGAGTGGAGG.



TATGCAGGCCCGGAGGAACTGCCGG(

TATGCAGGCCCGGAAGAACTGCCGG(

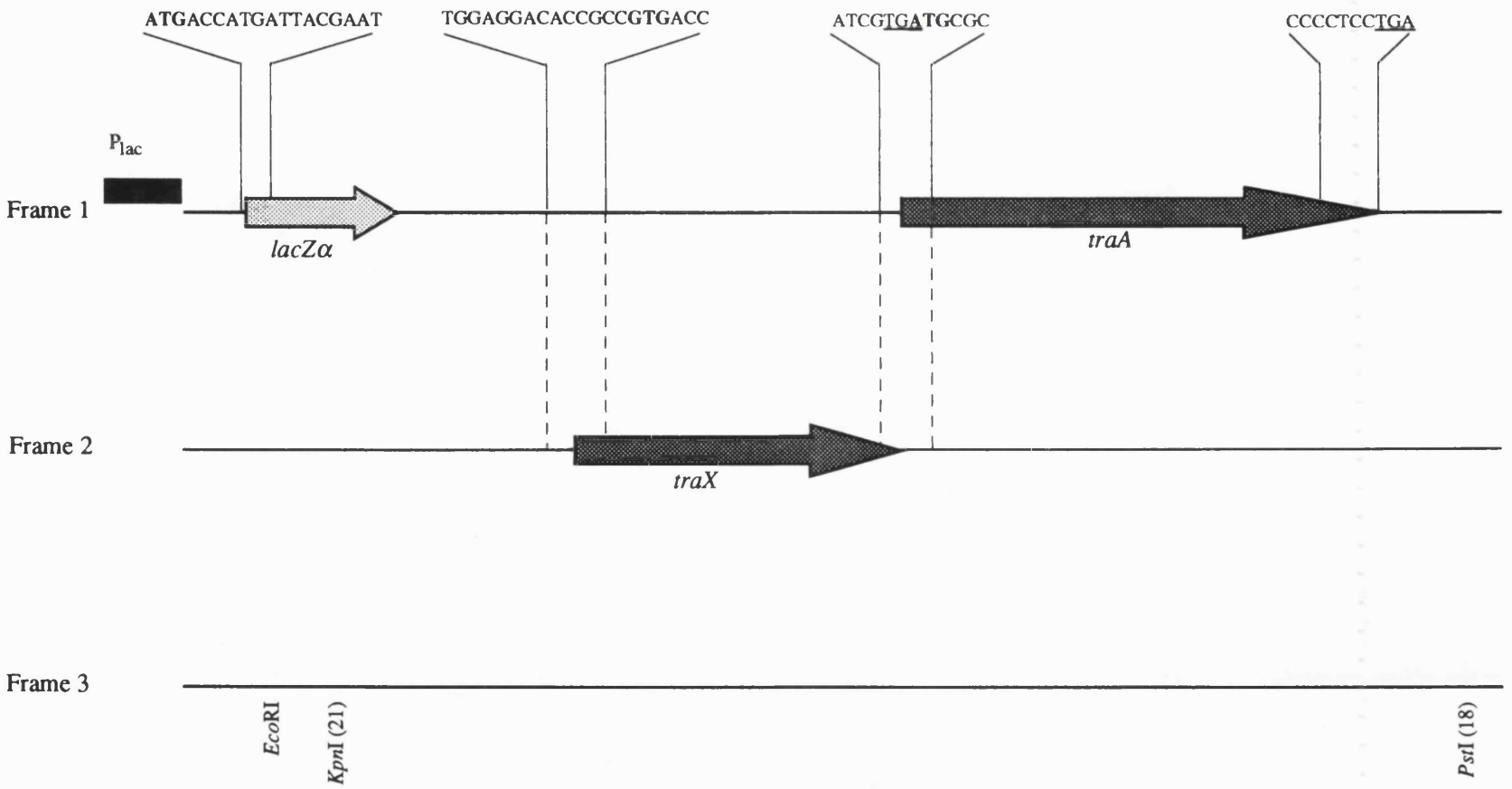
Autoradiograms of the DNA sequence of selected regions of pQR503 [nt: 1780-2147 (top)] & Taq 01/01 [nt: 2147- 2610 (bottom)] are shown. The sequence was obtained using the dideoxy chain termination sequencing method as described (2.9). The differences between the two sequences are highlighted.The sequence postulated by Brolle and co-workers is underlined.

A similar *E.coli* based system had been used previously to successfully synthesise the KorB protein from the *Streptomyces* plasmid pIJ101 (Zaman *et al.*, 1992) and analyse the products of the transfer region of the *Staphylococcus aureus* plasmid pGO1 (Morton *et al.*, 1993).

To achieve this the transfer region of SCP2* had to be sub-cloned into an appropriate vector, such that a suitable promoter was present upstream of the proposed *traA* ribosome binding site and the translational start codon. The previously constructed clone pQR210 (Figure 3.2) was examined to see whether this could be used as the template for the reactions. However the construct was not suitable since, the cloning of the *Sac*I - *Pst*I fragment into pUC18 had created a fusion of reading frames across the junction with *LacZ*. This juxtaposition between lac from *lacZα* and the *Streptomyces* DNA could possibly artificially force the system to use a non-natural start. Thus the use of this construct could possibly lead to the synthesis of a fusion protein of ~50kDa (size estimated from the predicted amino acid sequence). Since this product was similar in size to the protein under investigation it was decided to avoid the use of this construct.

In order to overcome this problem, pQR210 DNA was digested with *Kpn*I and *Pst*I , releasing the ~3.0kb fragment containing the *traX* and *traA* genes. This fragment was recovered from a low melting point agarose gel (2.8.1) and sub-cloned into the *Kpn*I - *Pst*I sites of pUC18, creating pQR514. This construct would overcome the problem of artificially forcing the system to use a non-natural start since any read through from the *lacα* start codon would be terminated early due to the presence (within the first 50 codons) of termination codons in all three forward reading frames (see Figure 4.5 & 4.9). pQR514 was then grown overnight and used to isolate plasmid DNA using the method described in 2.12.1. 2-5 μg of this DNA was then used as the template in the coupled transcription/translation reactions (2.12.1), and the samples resolved on a 15% SDS polyacrylamide gel. Examination of the autoradiogram (Figure 4.6) failed to reveal the presence of the ~50kDa protein expected from the *traA* ORF.

FIGURE 4.5 DNA Sequence across the junction of pQR514

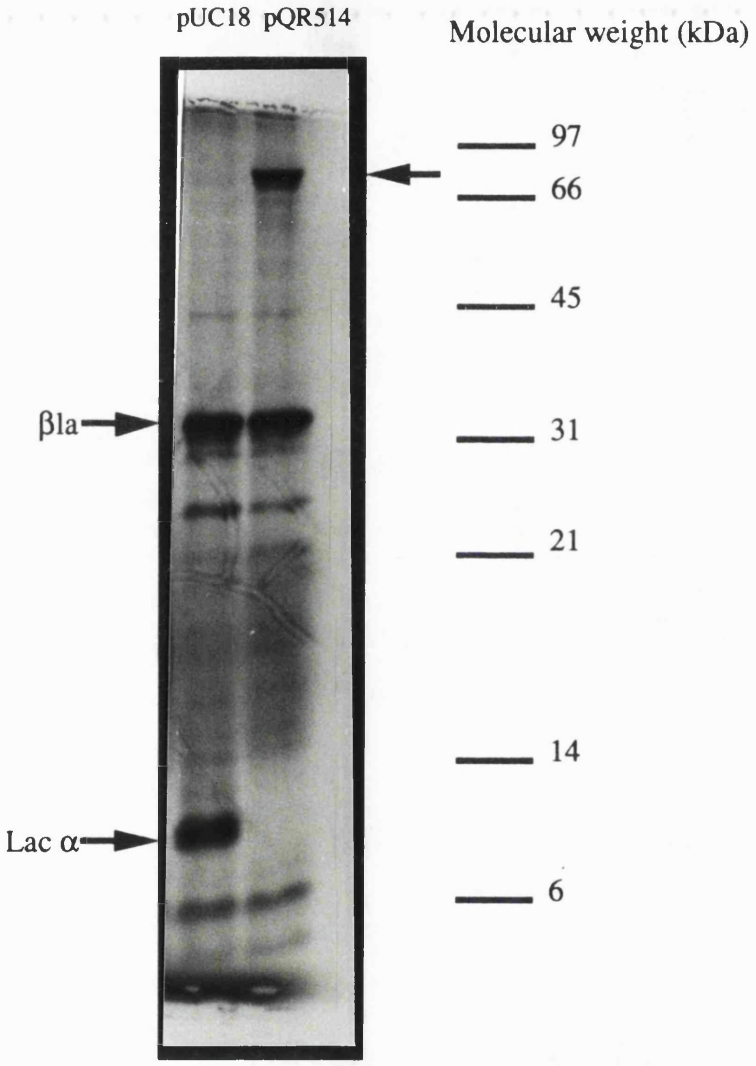


The read through from the lac promoter in all three reading frames is shown. The DNA sequence of selected regions is also shown as are putative start codons (bold face) and stop codons (underlined). Translational termination codons are located in all three reading frames between the premature termination codon of *lacZα* and the start codon of *traX*.

However careful examination of the gel revealed the presence of an unexpected protein band (Figure 4.6) which was absent from the pUC18 control. The presence of this band was initially somewhat confusing, on repeating the reactions and analysing the protein products on a 10% SDS polyacrylamide gel (Figure 4.7), the protein band estimated to be about 80kDa, was again observed. To confirm that the protein being observed was from the cloned *Streptomyces* DNA in pQR514, the construct was linearized using *Eco*RI, and this linearized template was used in the CTT reactions. As can be seen from Figure 4.8 the additional protein band observed previously disappeared when the template was linearized at the *Eco*RI site. The presence of this additional band migrating as an 80kDa protein could not have been due to any of the other reading frames within the cloned segment, due to the presence of many stop codons (Figure 4.9). The largest ORF (located in Frame 1 - nt: 1100 to 1600 - Figure 4.9), other than the *traX* and *traA* ORFs, would not be large enough to encode a protein product of the size observed. Therefore since no other ORF could account for the 80kDa protein observed, and it had been shown that the protein was from the cloned *Streptomyces* DNA, it was thought that the observed 80kDa protein was the result of a fusion between the *traX* (30kDa) and *traA* (50kDa) genes, which were proposed earlier to be translationally coupled.

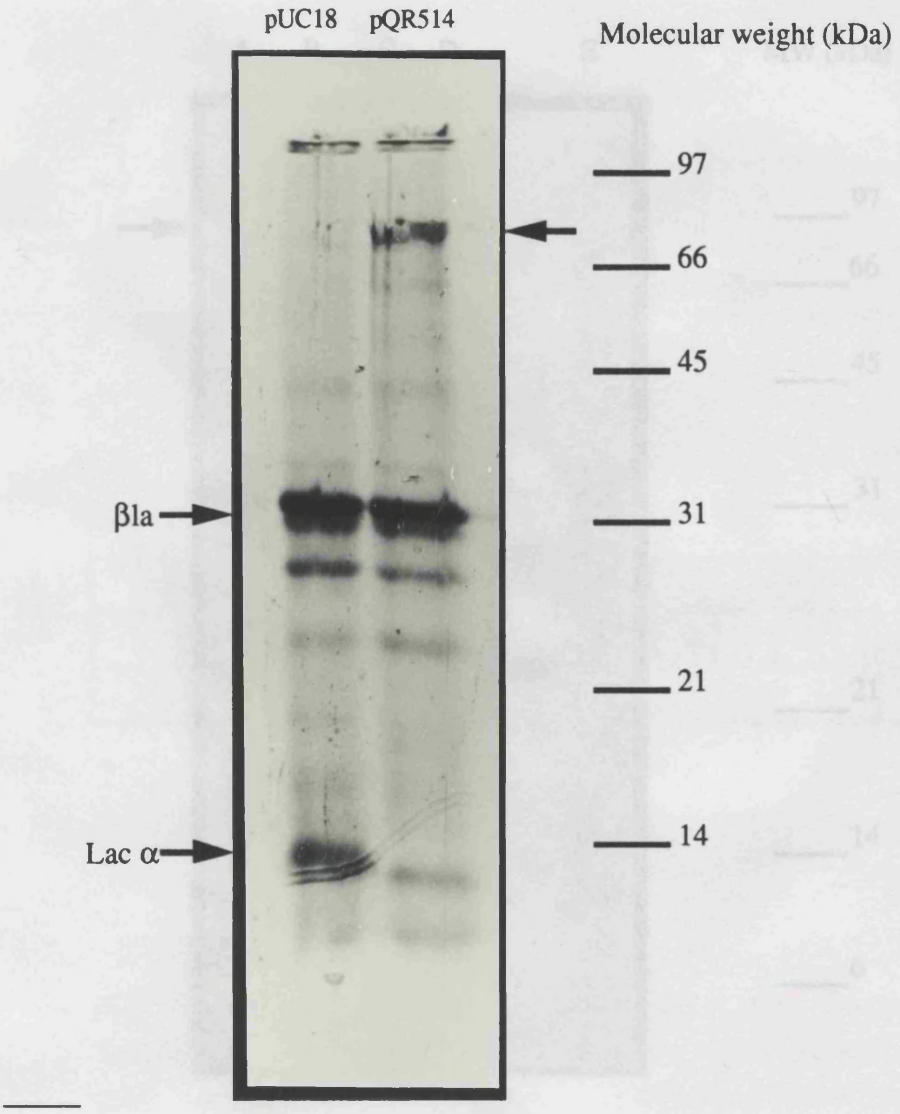
In order to further examine the protein being synthesised, a unique *Sac*II restriction site was identified approximately 300bp (~100 amino acids) at nt: 2720 from the stop codon of *traA* (see Figure 3.7 & 4.9). It was predicted that if pQR514 was digested with *Sac*II, and the resulting template used in the CTT reactions, the protein product synthesised would be reduced in size from ~80kDa to ~70kDa due to the absence of the ~100 amino acids from the C-terminal end of *traA*. This reduction in size would not only confirm that the 80kDa protein being observed was from the cloned region in pQR514 but also strengthen the hypothesis that it was the result of translational fusion between the *traX* and *traA* gene products. Figure 4.8 shows the results of such an experiment and as can be

FIGURE 4.6 Autoradiogram of *in vitro* protein extracts of pUC18 & pQR514 resolved by 15% SDS PAGE



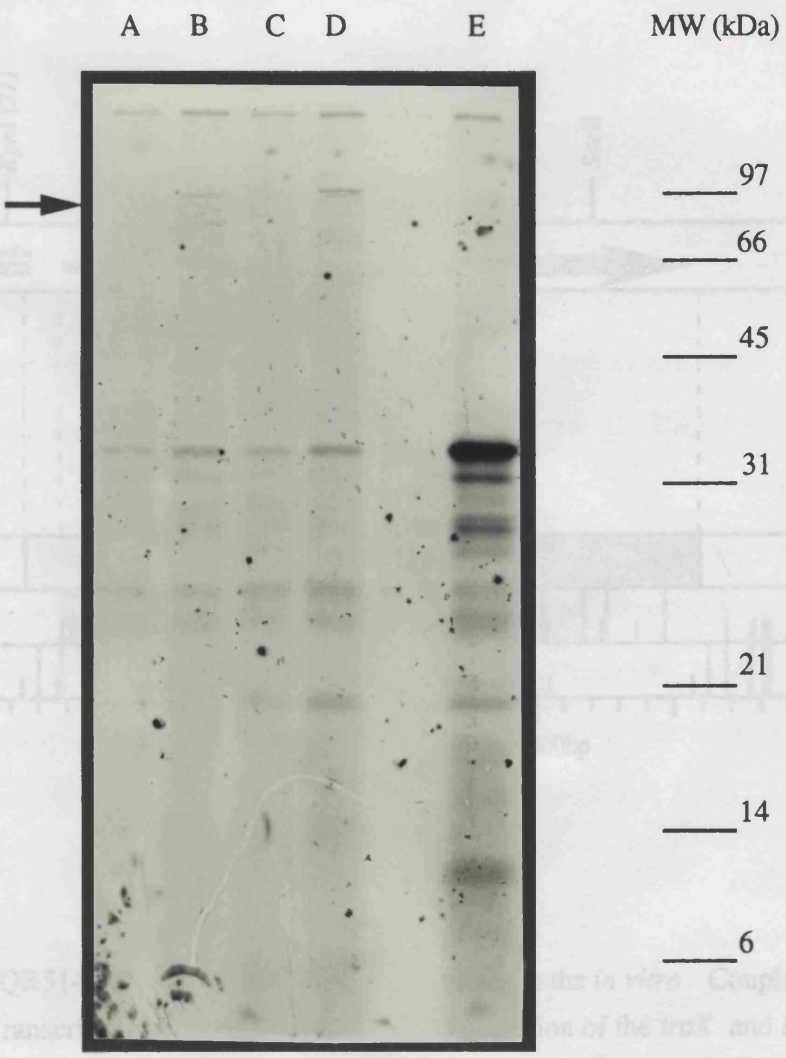
^{35}S [Met] labelled protein extracts of pUC18 & pQR514 (synthesised (2:12) using the S30 - *E.coli in vitro* Coupled Transcription /Translation kit) were resolved by 15% SDS PAGE. The expected *traA* gene product (~50kDa) was not observed in the pQR514 sample. However an additional band of ~80kDa, clearly absent from the pUC18 control was observed.
The ~12kDa (*LacZ`α*) and ~31kDa (βla) gene products are indicated by arrows.

FIGURE 4.7 Auto radiogram of *in vitro* protein extracts of pUC18 & pQR514 resolved by 10% SDS PAGE



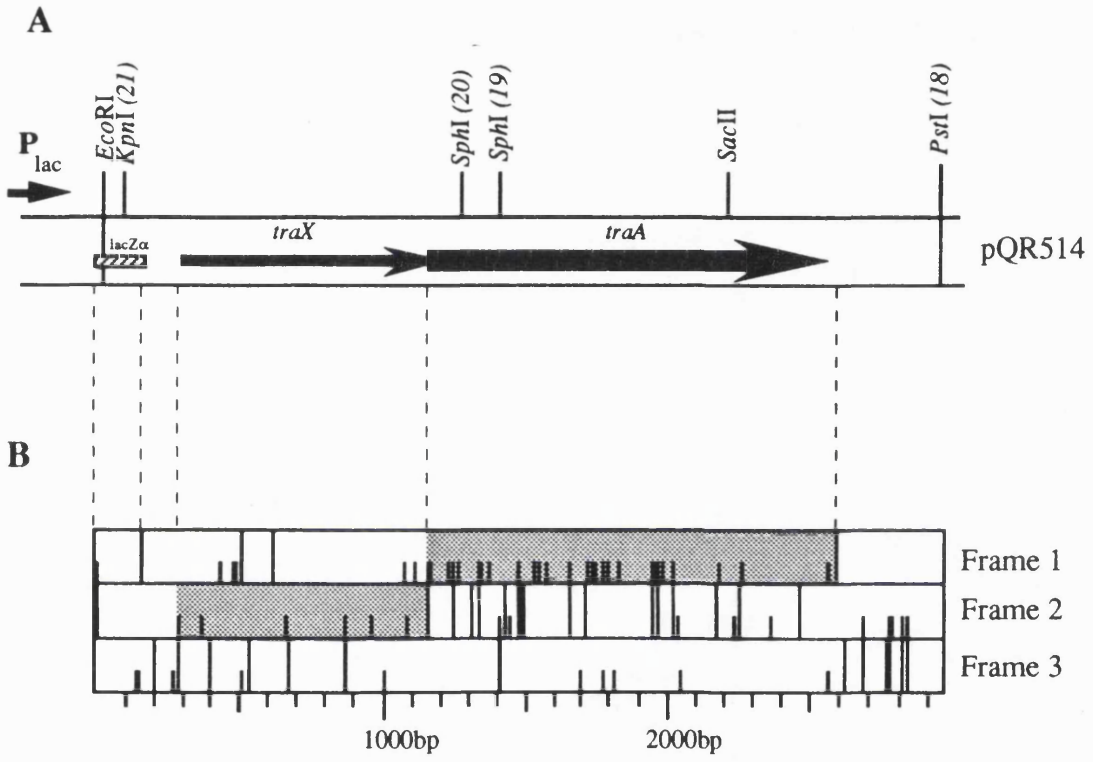
^{35}S [Met] labelled protein extracts of pUC18 & pQR514 (synthesized (2:12) using the S30 - *E.coli in vitro* Coupled Transcription /Translation kit) were resolved by 10% SDS PAGE. The additional band present in the pQR514 sample was estimated to be about 80kDa (as compared to the Molecular weight standards).The ~12kDa (*LacZ`α*) and ~31kDa (βla) gene products are indicated by arrows.

FIGURE 4.8 Auto radiogram of in vitro protein extracts of linearized pQR514 resolved by 15% SDS PAGE.



Lane A: pQR514 (*Sac*II) Lane B: pQR514 (*Pst*I)
Lane C: pQR514 (*Eco*RI) Lane D: pQR514 (uncut) Lane E: pUC18
^{35}S [Met] labelled *in vitro* protein extracts of pUC18 & pQR514 (linearized) resolved by 15% SDS PAGE. The 80kDa protein band observed in uncut pQR514 disappears when linearized (*Eco*RI) template is used, but can still be seen when pQR514 (*Pst*I) is used which cuts downstream of the *traA* ORF.

FIGURE 4.9 Open reading frame map of pQR514



A- pQR514 constructed for use as the template in the *in vitro* Coupled Transcription/Translation reactions. The location of the *traX* and *traA* genes with respect to *Lac* α and the lac promoter (P_{lac}) are indicated.

B- Start/stop codon map of the region of pQR514 shown in A above, obtained using ORF MAP (DNA Strider; Marck, 1988). The three forward reading frames are represented with stop codons (as tall bars) and start codons (short bars). The positions of *traX* (Frame 2), *traA* (Frame 1) and part of *lacZα* (Frame 1) are also represented.

seen the 80kDa protein disappeared when *Eco*RI digested template was used. It is possible that when *Sac*II digested template was used a very faint band of the size expected may have been present. However this band is not visible in Figure 4.8 (the approximate position of this band is indicated). This area could not be pursued further due to restrictions on time.

4.5 Summary

The sequence of the transfer region of SCP2* has been further investigated and evidence supporting the sequence proposed (Chapter Three) has been presented. Analysis of the protein products encoded by the transfer region have shown the presence of an unexpected protein band of ~80kDa, which has been shown to be encoded by the cloned *Streptomyces* DNA, since digestion with *Eco*RI but not *Pst*I causes the protein to disappear. This work proposes that the 80kDa protein observed could be the product of translational fusion between the *traX* (50kDa) and *traA* (80kDa) gene products. These two genes could be translationally coupled, if the DNA sequence presented in this work is correct and the assumption of the native translational start site is correct.

CHAPTER FIVE: Cloning & Expression of the *traB* gene.

5.0 Introduction

The sequence analysis carried in Chapter Three identified three possible open reading frames (ORF's) *traA*, *traB* and *traX*. The smallest, *traB* which may be a regulatory protein and thus of particular interest, warranted further investigation. The mutagenic studies carried out on *traB* (Brolle *et al.*, 1993) using the transposon Tn4560 showed reduced plasmid transfer ($2x10^{-5}$) and the complete abolition of pocks. This suggested that the *traB* gene may be involved in plasmid transfer rather than plasmid spread. Analysis of the predicted *traB* ORF (section 3.7) indicated that the 137 amino acids of *traB* would encode a putative protein of 14.7kDa. Further analysis suggested the presence of a possible leucine zipper motif, which has been shown to be involved in DNA binding (Landschulz *et al.*, 1988; Vinson *et al.*, 1989)), thus raising the question; is the product encoded by ORF B involved in a possible regulatory role ?

The aim of the work performed here was to clone the *traB* open reading frame into an appropriate vector and to overexpress the protein *in vivo* in *E.coli* as a fusion protein in sufficient quantities to be able to raise polyclonal antibodies against it. These would hopefully provide an assay for studying the role of TraB in *Streptomyces* plasmid transfer. pUC18 (for map see Appendix B) with its multiple cloning site (MCS) and inducible lac promoter upstream of the MCS, was chosen as the vector of choice for the expression of the foreign gene product as a fusion protein.

5.1 Construction of pQR521

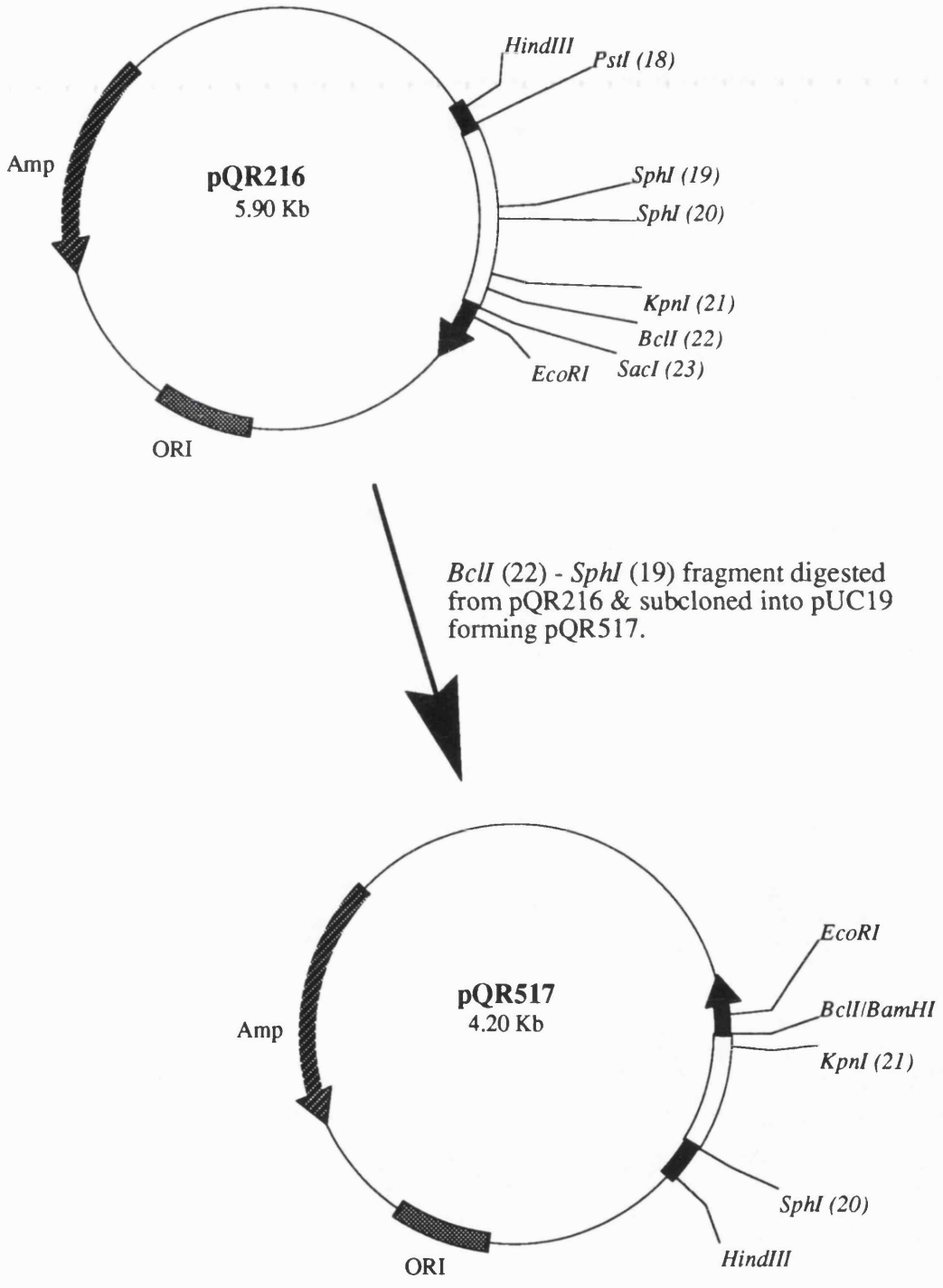
10μg of pQR216 [as pQR210 (Figure 3.2) but transformed in *E.coli* GM242 (Table 2.1)] DNA was digested to completion with *Bcl*1 and *Sph*1 resulting in the release of three fragments, ~ 0.4kb (*Sph*I(19) to *Sph*I(20)), ~1.5kb (*Sph*I(19) to *Bcl*I(22) fragment containing ORF B) and ~4.0kb (*Sph*I(20) to *Bcl*I(22) remainder of

the transfer region & the vector). The reaction was run on a 1% Low melting point agarose gel and the 1.5kb band containing the *traB* ORF was extracted from the gel as described in 2.8.1. After checking the recovery on a 1.2% agarose gel, about 1μg of the *BclI-SphI* fragment was ligated to ~0.5μg of pUC19 DNA (cut with *Bam*HI & *Sph*I). The ligation mix was then used to transform some competent *E.coli* cells (2.5.2) and the transformation mix plated out overnight at 37°C on NAXI plates (2.3.1).

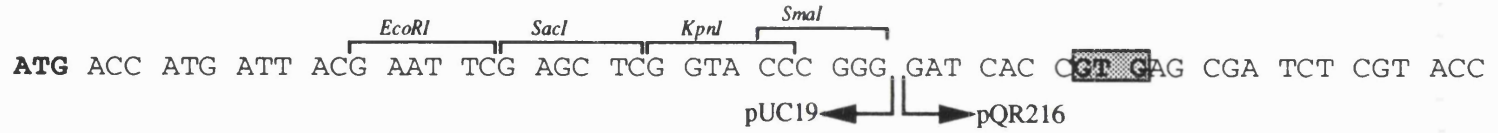
Recombinant colonies were screened as described (2.6.3) in order to find a clone that contained the *BclI* - *Sph*I fragment, this clone was named pQR517 (Figure 5.1). However the *traB* gene was out of frame with respect to *lacZ* as shown in Figure 5.2. In order to bring the *traB* ORF into the correct reading frame, pQR517 was digested to completion with *Sac*I and the resulting 5' overhanging ends made blunt using T_4 DNA polymerase (2.7.5). This DNA was then religated (2.7.2) and used to transform competent *E.coli* JM107 (2.5.2).

When plated on NAXI plates all the resulting colonies were white as expected. A number of these white colonies were picked and screened (2.6.3.), in order to find a construct that contained the ~1.5kb fragment from pQR216 but did not cut with *Sac*I, since this restriction site had been removed by the action of T_4 DNA polymerase. Four prospective constructs were then sequenced (2.9) to see if the *Sac*I site had been removed and to check whether the *traB* ORF was in the correct reading frame. The positive clone was named pQR520 (Figure 5.2), however this clone was in pUC19 which meant that the *traB* ORF was in the wrong orientation with respect to lac, therefore pQR520 was digested with *Eco*RI and *Hind*III and the resulting fragment (~1.5kb) cloned into pUC18 resulting in pQR521, a clone that contained the *traB* ORF downstream of the lac promoter and in frame with respect to the *lacZ* start codon.

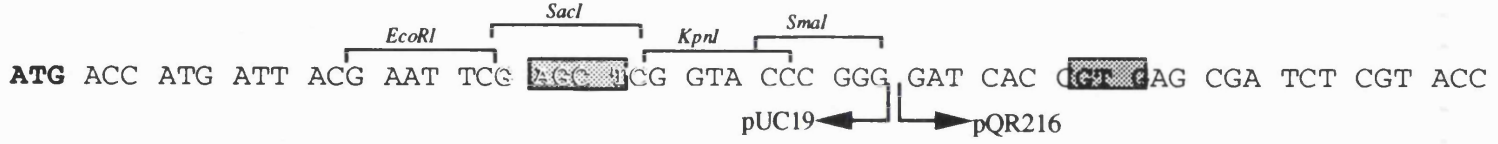
FIGURE 5.1 Construction of pQR517



pQR517

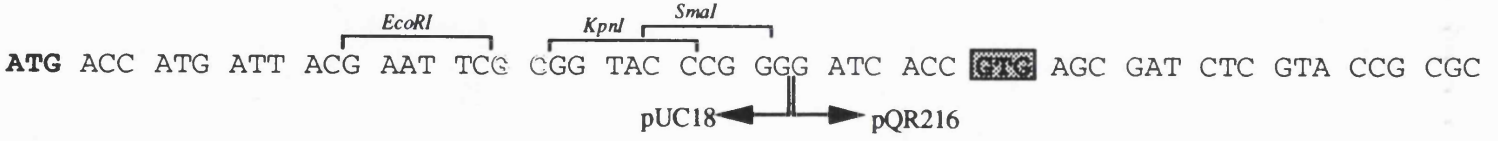


traB ORF is out of frame



DNA digested with *Sac*I and the overhanging ends made blunt with T4 polymerase and religated to form pQR520
pQR520 digested with *Eco*RI/*Hind*III & the resulting 1.5kb fragment cloned into pUC18 creating pQR521.

pQR521



5.2 DNA sequence analysis of pQR521.

pQR521 was sequenced (2.9.4) using the appropriate sequencing primers (Table 2.4) to check the sequence across the junction between *lacZ* and *traB*, the results are shown in Figure 5.3. As can be seen the digestion with T_4 polymerase was successful since 4bp (AGCT) have been removed from the *Sac*I site, thus bringing ORF B into the same reading frame as *LacZ*. This would therefore allow transcription and translation to start at the *lacZ* sequences and read across the junction and through *traB* terminating at the *traB* stop codon.

The complete amino acid sequence of the fusion protein is shown in Figure 5.4. The predicted molecular weight (MW) of the fusion protein being 16.2kDa as compared to the native protein weight of 14.7kDa. The fusion protein is made up of 151 amino acids compared to the 138 amino acids of the native *traB* gene, the additional 13 amino acids originating from the pUC vector.

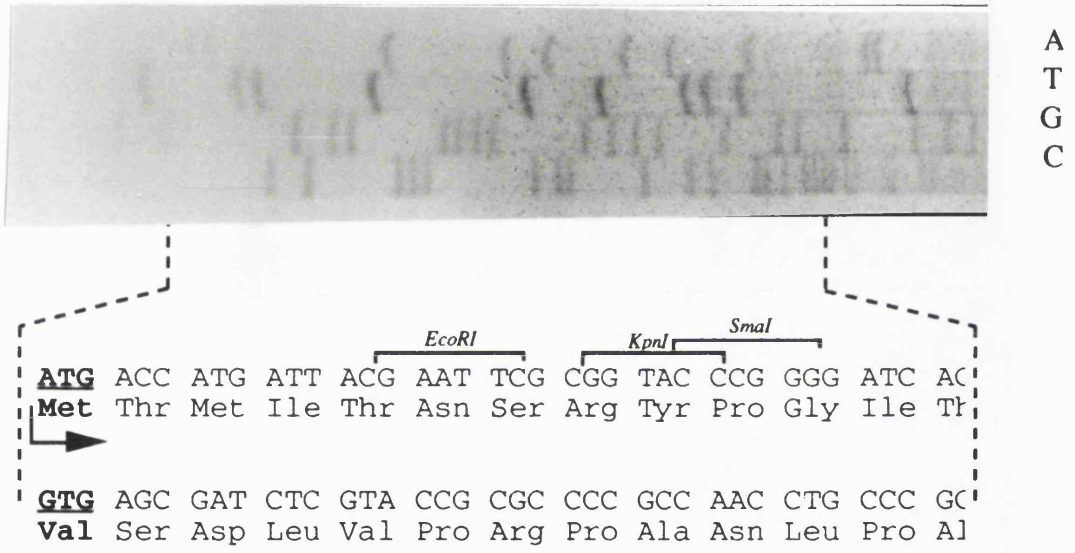
5.3 *In vivo* expression of *traB* using pQR521

Cultures of pQR521 were induced (2.11.1) overnight by the addition of 1mM IPTG and protein samples prepared as described (2.11.3.). 40µg of crude protein extract was then run on a 15% SDS polyacrylamide gel together with the relevant controls (JM107, pUC18 & pQR521 uninduced). Figure 5.5 shows samples of induced pQR521 separated out by PAGE, as can be seen no new band corresponding to the LacZ/TraB fusion protein (~16kDa) is observed. This was repeated using samples induced with IPTG to 2mM, and also taking samples at various time intervals (5,10,24 hours), again the samples were resolved on a 15% SDS polyacrylamide gel but no new proteins bands in the 16kDa region were observed (data not shown).

5.4 *In vitro* expression of *traB* using pQR521

Since no additional protein was observed on induction of pQR521 with IPTG, it was decided to analyse this construct further by carrying out some studies

FIGURE 5.3 DNA sequence across the junction of pQR521



The DNA sequence across the junction of *lacZα* and *traB* in the construct pQR521 is shown above. The sequence was obtained using the method described (2.9). Selected restriction sites, the translational start codon (ATG) of the fusion (underlined - arrow) and the start codon (GTG) of the native (underlined) *traB* gene are shown.

FIGURE 5.4 Nucleotide & amino acid sequence of the *traB* gene in pQR521.

```
CCC AGG CTT TAC ACT TTA TGC TTC CGC CTC GTA TGT TGT GTG GAA TTG TGA GCG GAT AAC

            RBS             start                EcoRI      KpnI    SmaI   BamHI/BclI
AAT TTC ACA CAG GAA ACA GCT ATG ACC ATG ATT ACG AAT TCG CGG TAC CCG GGG ATC ACC
                            Met Thr Met Ile Thr Asn Ser Arg Tyr Pro Gly Ile Thr
traB start
GTG AGC GAT CTC GTA CCG CGC CCC GCC AAC CTG CCC GCC GCC ACC GGC AAG CCG AGC TTT
Val Ser Asp Leu Val Pro Arg Pro Ala Asn Leu Pro Ala Ala Thr Gly Lys Pro Ser Phe

ACC CGC CTG GCC GCC CGC ATG GCG ACG CTC ACC GCG TCC GCC CTC GCG CTG AAA GAG GGC
Thr Arg Leu Ala Ala Arg Met Ala Thr Leu Thr Ala Ser Ala Leu Ala Leu Lys Glu Gly

CTG TGG GCG CTG AAG CGC CGC ATG GAG ACC GAC GCC GAC CAC GCC GAC ATG CTC GCC GAC
Leu Trp Ala Leu Lys Arg Arg Met Glu Thr Asp Ala Asp His Ala Asp Met Leu Ala Asp

CTG TGC GTC GCG GCG GAG GTG GAG CCC CGG TTC ACC GGC CAG ATC AAC GAG GCG GGT ACC
Leu Cys Val Ala Ala Glu Val Glu Pro Arg Phe Thr Gly Gln Ile Asn Glu Ala Gly Thr

GCG CTG CGC AAG GTC GCG GAG GCG TCC GCC GAA CTG GCC CGC GCG GCC GAC CAG GTA CAG
Ala Leu Arg Lys Val Ala Glu Ala Ser Ala Glu Leu Ala Arg Ala Ala Asp Gln Val Gln

CAC GAC TCG CAG GGC CTC CAC GAC GCG CAC CAG GGC GAA TAC CGG GGC GTG TAC GAG GCC
His Asp Ser Gln Gly Leu His Asp Ala His Gln Gly Glu Tyr Arg Gly Val Tyr Glu Ala

GTG AAC GCC TCG GGC GTC CGC CAG GCC AAG CCC GGG TTC TAC CGC ACC CGC TGA
Val Asn Ala Ser Gly Val Arg Gln Ala Lys Pro Gly Phe Tyr Arg Thr Arg END
```

Total number of amino acids = 151 MW = 16.2kDa

using the *in vitro* Coupled Transcription/Translation (CTT) kit (Promega) which allows expression of genes from DNA provided that the necessary control signals are present. This *in vitro* system has previously been used in Section Four in order to examine the products encoded by the *traX* and *traA* genes.

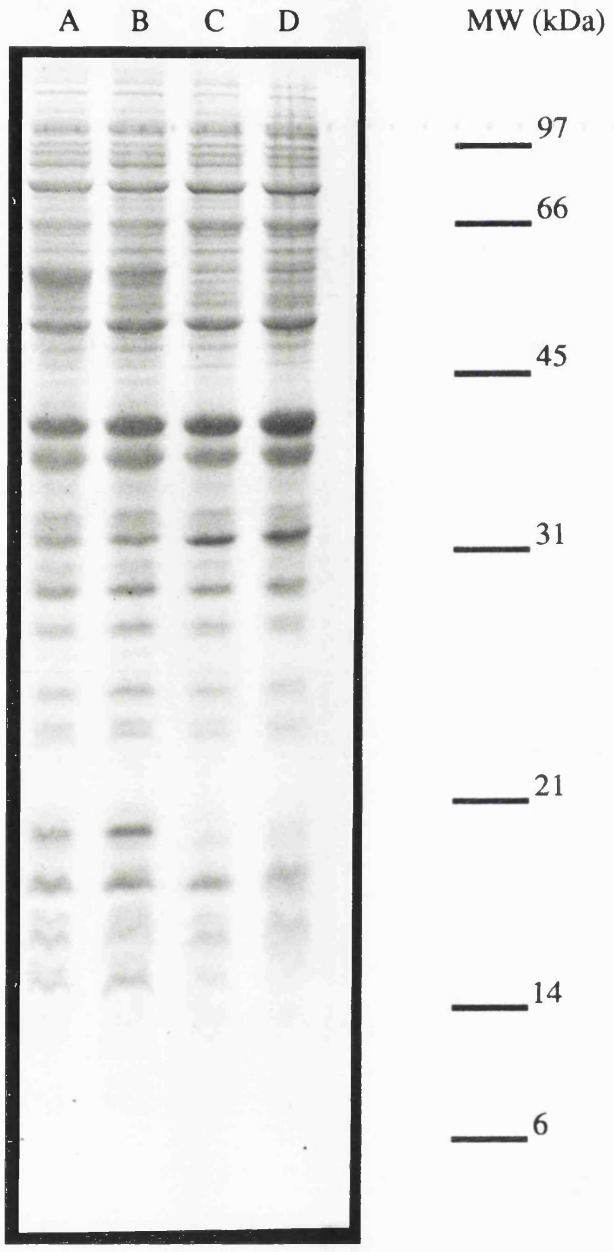
pQR521 and pUC18 DNA was prepared as described (2.12.1) and about 2μg was used to carry out the *in vitro* CTT reactions. The samples were then resolved on a 15% SDS polyacrylamide gel (2.12.2) and the gel dried and exposed to autoradiographic film. As can be seen from Figure 5.6 an additional band corresponding to about 16kDa is observed in pQR521 but not in the pUC18 control. This suggests that the *lacZ/traB* fusion protein may actually be expressed from pQR521.

In order to show that the protein seen in the *in vitro* experiments was actually from the cloned *traB* region, it was decided to digest the template between the pUC18 vector and the *traB* ORF at the unique *Eco*RI site. In theory this would separate the transcription initiation sequences (promoter, RBS etc) from the remainder of the gene preventing the fusion protein from being expressed. pQR521 DNA was digested with *Eco*RI and purified as described (2.7.3), and 2μg of this DNA was used to perform the *in vitro* CTT reactions. When pQR521 DNA digested with *Eco*RI was used the band previously observed with pQR521 disappeared, suggesting that the band was in fact from the *lacZ/traB* region. pUC18 was also used as a control and when this was digested with *Sca*I (which cuts within the gene coding for β- lactamase) the band observed at ~31kb disappeared (data not shown).

This suggests that the *lacZ/traB* fusion protein may well be expressed *in vivo* from pQR521 but is not observed on an SDS polyacrylamide gel because it is expressed at such low levels as not to be detected by coomassie stain and/or is masked by one of the *E.coli* host proteins or in fact is unstable in the host cell.

Since pUC18 with its inducible Lac promoter proved to be insufficient to overexpress the fusion protein, an alternative strategy was necessary. It was decided

FIGURE 5.5 ***In vivo*** **protein expression from pQR521**

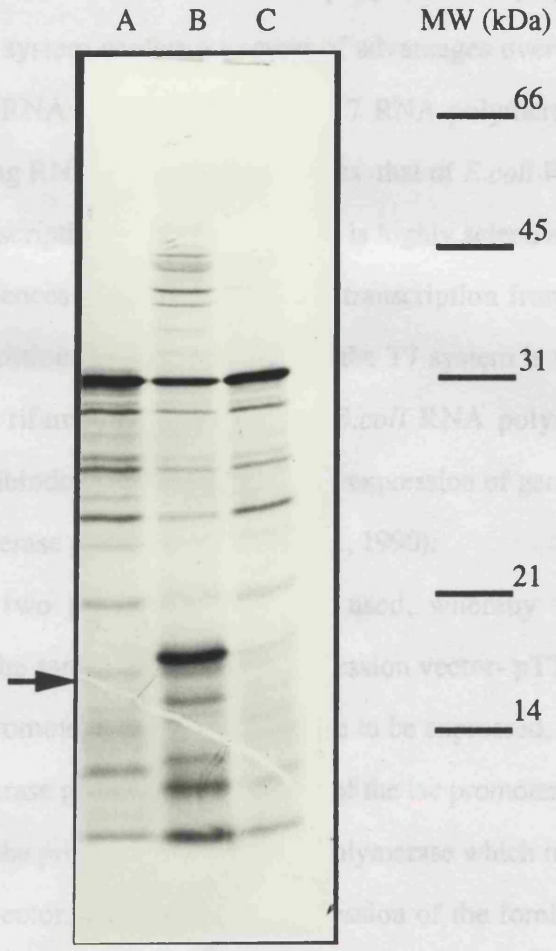


Lane A: *E.coli* JM107 Lane B: pUC18

Lane C: pQR521 (uninduced) Lane D: pQR521 (induced)

IPTG induced overnight cultures of *E.coli* JM107 containing pQR521 and pUC18 (as a control) were used to prepare protein samples as described (2.11). Approximately 40µg of total protein extracts were resolved by SDS PAGE (15%) & stained with coomassie blue. As can be seen no additional protein bands corresponding to the LacZ/TraB fusion were observed.

FIGURE 5.6 *In vitro* protein expression from pQR521



Lane A: pUC18 Lane B: pQR521 Lane C: pQR521 (linear)

In vitro protein extracts synthesised using the CTT kit (2.12) were resolved by 15% SDS PAGE and exposed to X ray film. A protein band corresponding in size (~16kDa) to that expected for the LacZ/TraB fusion is clearly visible as indicated in Lane B, which is absent from pUC18 (Lane A) and pQR521 (Lane C- digested with *Eco*RI).

to opt for a specific expression vector rather than a general cloning vector as in the case of pUC18. The chosen system was the T7 RNA polymerase/promoter system (Tabor & Richardson, 1985) which can be used to express prokaryotic and eukaryotic genes in *E.coli,* by placing them under the control of the phage T7 RNA polymerase promoter.

This system confers a number of advantages over pUC based vectors that rely on *E.coli* RNA polymerase, in that T7 RNA polymerase is a very active enzyme (synthesising RNA at a rate several times that of *E.coli* RNA polymerase). It also terminates transcription less frequently and is highly selective for initiation at its own promoter sequences and does not initiate transcription from any sequences on *E.coli* DNA. An additional advantage of using the T7 system is that it is resistant to antibiotics such as rifampicin which inhibit *E.coli* RNA polymerases, hence the addition of such antibiotics causes the exclusive expression of genes under the control of a T7 RNA polymerase promoter (Studier *et al.*, 1990).

The two plasmid system was used, whereby two plasmids are maintained within the same cell. One (the expression vector- pT7-7) contains the T7 RNA polymerase promoter upstream of the gene to be expressed, the second contains the T7 RNA polymerase gene under the control of the lac promoter, whereby induction with IPTG leads to the production of T7 RNA polymerase which initiates transcription on the expression vector, resulting in the expression of the foreign gene product. A map of the pT7-7 expression vector used in this study is shown in Appendix B.

5.5 Construction of pQR524

5μg of pQR517 (Figure 5.1) was digested to completion with *Eco*RI and *Hind*III releasing two fragments ~1.5kb (ORF B) & ~2.6kb (pUC vector). The digest was run on a low melting point agarose gel and the 1.5kb fragment containing the *traB* ORF was extracted as described (2.8.1). The isolated fragment was then checked on an 1.5% agarose gel and a small fraction of the cleaned DNA used to set up a ligation with pT7-7 vector DNA digested with *Eco*RI and *Hind*III.

The resulting ligation mix was then used to transform competent *E.coli* JM107 cells and the sample plated out onto an LB agar plate supplemented with ampicillin. Since there is no form of selection other than ampicillin for pT7 vectors (such as the blue/white screening of pUC vectors), a number of the resulting colonies were randomly picked and screened (2.6.3) for the presence of a 1.5kb insert upon digestion with *Eco*RI and *Hind*III. The result was a clone, pQR523 which contained the *traB* ORF on a ~1.5kb fragment cloned between the *Eco*RI & *Hind*III sites of pT7-7 in *E.coli* JM107 (Figure 5.7).

5.6 DNA sequence analysis of pQR523

It was predicted that this clone (pQR523) would allow the TraB fusion protein to be expressed since the *traB* ORF was in the correct reading frame with respect to the T7 start site (Figure 5.8). However in order to confirm this, pQR523 was sequenced using the T7 sequencing primer (Table 2.4) in order to check the actual sequence across the pT7-7/*traB* junction (Figure 5.9). As can be seen the *traB* ORF is in fact in the correct reading frame with respect to the T7 promoter and the ATG start codon. Thus transcription and translation should continue uninterrupted through the ATG start codon (of pT7-7) and terminate at the stop codon of the *traB* ORF.

The complete nucleotide and amino acid sequence of this fusion protein is shown in Figure 5.10, together with the predicted molecular weight and total number of amino acids.

5.7 *In vivo* expression of *traB* from pQR524

pQR523 DNA was extracted from an overnight culture and used to transform some competent *E.coli* BL21 (DE3) cells. All the colonies found on an LB agar plate (supplemented with ampicillin & chloramphenicol) were thought to be BL21 (DE3) cells harbouring the pT7-7 recombinant clone. For ease of handling this

FIGURE 5.7 Construction of pQR524

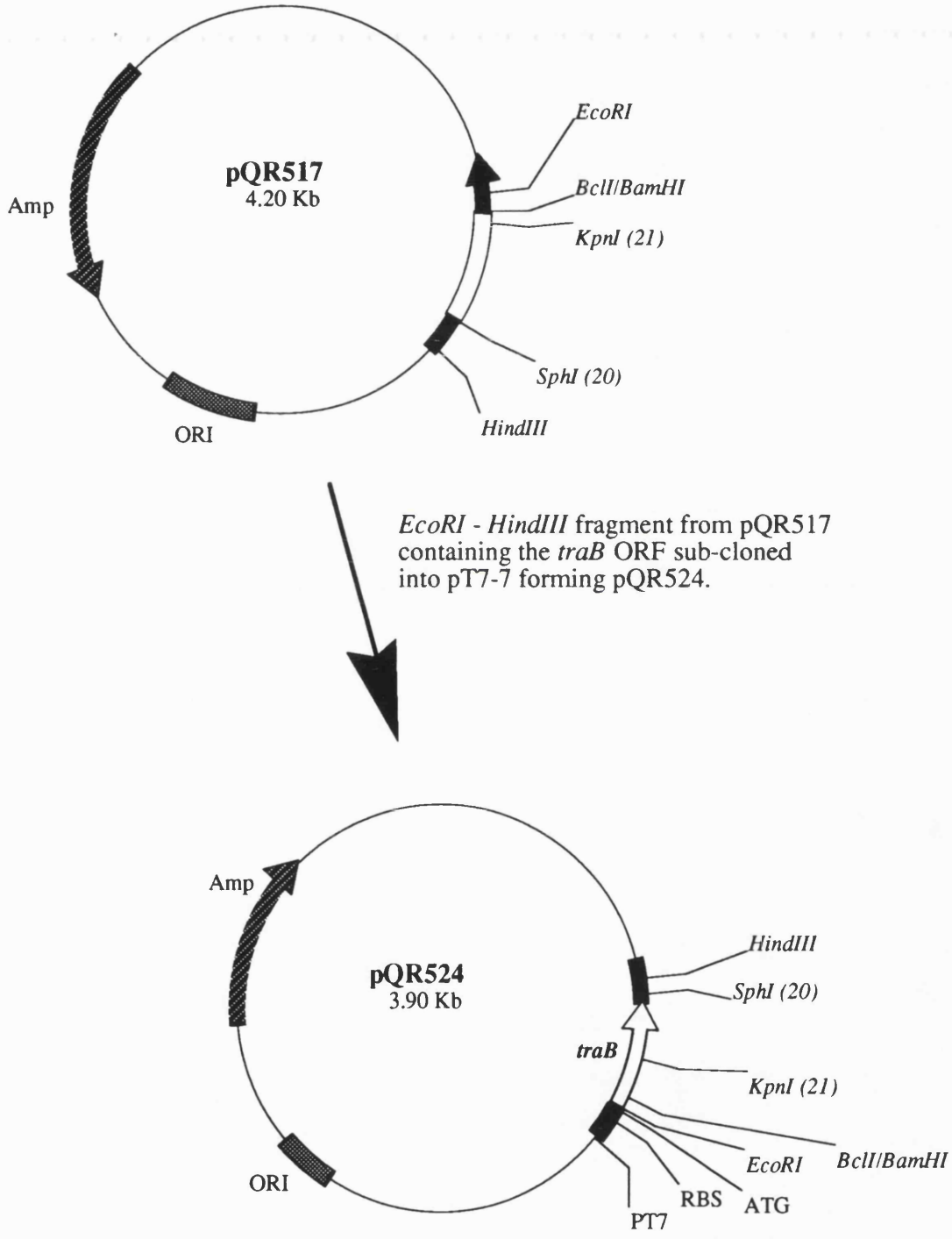


FIGURE 5.8 Construction of pQR524

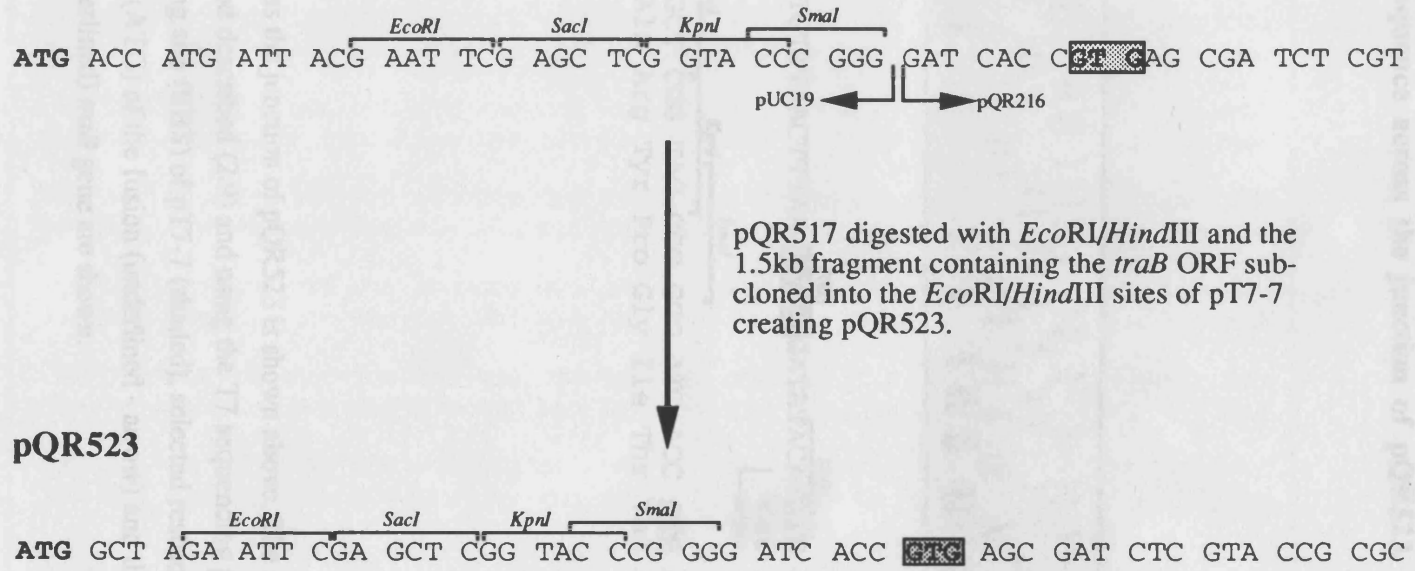
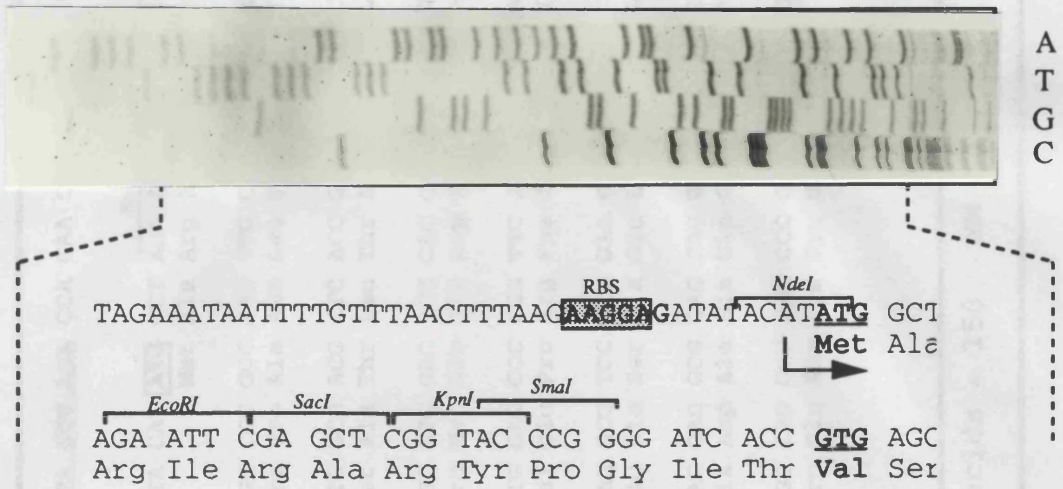


FIGURE 5.9 DNA sequence across the junction of pQR523.



The DNA sequence across the junction of pQR523 is shown above. The sequence was obtained using the method described (2.9) and using the T7 sequencing primer (Table 2.4).The ribosome binding site (RBS) of pT7-7 (shaded), selected restriction sites, the translational start codon (ATG) of the fusion (underlined - arrow) and the start codon (GTG) of the native (underlined) *traB* gene are shown.

FIGURE 5.10 Nucleotide & amino acid sequence of the *traB* gene in pQR523

TAG ACT TCG AAA T**TA ATA CGA CTC ACT ATA GGG AGA** CCA CAA CGG TTT CCC TCT AGA AAT
T7 promoter

RBS; *EcoRI*; *SacI*; *KpnI*; *SmaI*

AAT TTT GTT TAA CTT TAA G**AA GGA GA**T ATA CAT **ATG** GCT AGA ATT CGA GCT CGG TAC CCG
Met Ala Arg Ile Arg Ala Arg Tyr Pro

BamHI/BclI

GGG ATC ACC [illegible] AGC GAT CTC GTA CCG CGC CCC GCC AAC CTG CCC GCC GCC ACC GGC AAG
Gly Ile Thr **Val** Ser Asp Leu Val Pro Arg Pro Ala Asn Leu Pro Ala Ala Thr Gly Lys

CCG AGC TTT ACC CGC CTG GCC GCC CGC ATG GCG ACG CTC ACC GCG TCC GCC CTC GCG CTG
Pro Ser Phe Thr Arg Leu Ala Ala Arg Met Ala Thr Leu Thr Ala Ser Ala Leu Ala Leu

AAA GAG GGC CTG TGG GCG CTG AAG CGC CGC ATG GAG ACC GAC GCC GAC CAC GCC GAC ATG
Lys Glu Gly Leu Trp Ala Leu Lys Arg Arg Met Glu Thr Asp Ala Asp His Ala Asp Met

CTC GCC GAC CTG TGC GTC GCG GCG GAG GTG GAG CCC CGG TTC ACC GGC CAG ATC AAC GAG
Leu Ala Asp Leu Cys Val Ala Ala Glu Val Glu Pro Arg Phe Thr Gly Gln Ile Asn Glu

GCG GGT ACC GCG CTG CGC AAG GTC GCG GAG GCG TCC GCC GAA CTG GCC CGC GCG GCC GAC
Ala Gly Thr Ala Leu Arg Lys Val Ala Glu Ala Ser Ala Glu Leu Ala Arg Ala Ala Asp

CAG GTA CAG CAC GAC TCG CAG GGC CTC CAC GAC GCG CAC CAG GGC GAA TAC CGG GGC GTG
Gln Val Gln His Asp Ser Gln Gly Leu His Asp Ala His Gln Gly Glu Tyr Arg Gly Val

TAC GAG GCC GTG AAC GCC TCG GGC GTC CGC CAG GCC AAG CCC GGG TTC TAC CGC ACC CGC
Tyr Glu Ala Val Asn Ala Ser Gly Val Arg Gln Ala Lys Pro Gly Phe Tyr Arg Thr Arg

TGA
End

Total number of amino acids = 150 MW = 16.1kDa

construct in BL21 (DE3) was named pQR524 as opposed to pQR523 (in JM107).

A single colony of pQR524 was picked and grown overnight in 5ml of L broth (supplemented with ampicillin & chloramphenicol), 150µl of this was then diluted in to fresh L broth (5ml) and grown for several hours until an $O.D._{590}$ of ~ 0.4 was reached at which point the samples were induced by the addition of IPTG to 1mM, and incubation continued for a further 4 hours.

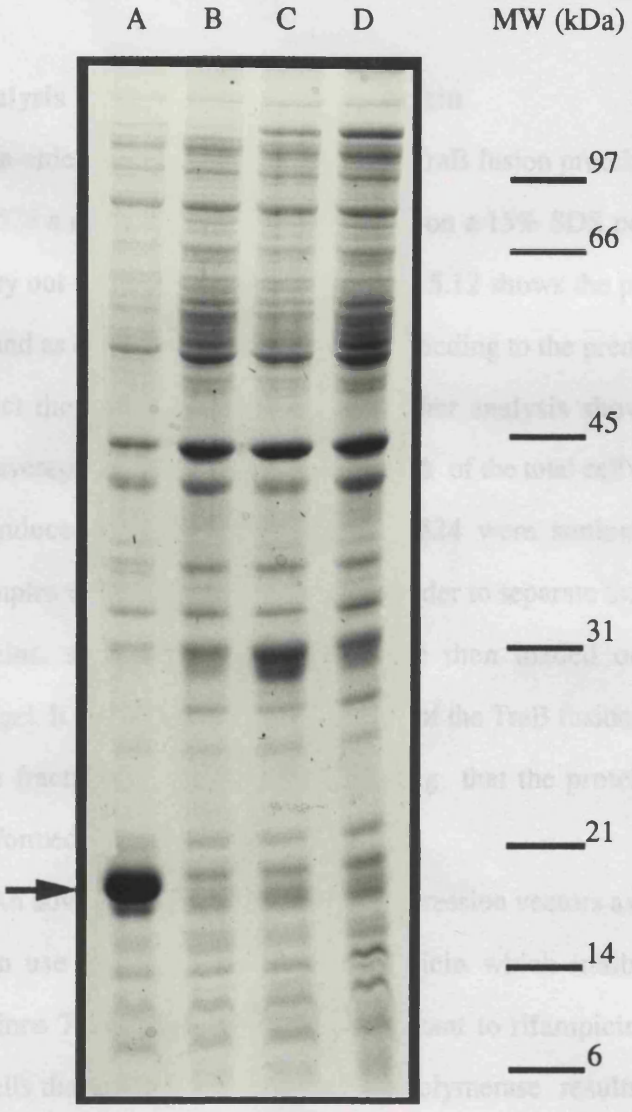
Protein samples were prepared as described (2.11.3) and approximately 40µg of the crude extract resolved on a 15% SDS polyacrylamide gel. Figure 5.11 shows samples of induced pQR524 and the controls separated out by PAGE, as can be observed a distinct band of about 16kDa is seen in the pQR524 induced sample which is absent from all the other tracks.

5.8 Maintenance of pQR524 stocks

pQR524 samples were stored as 30% glycerol stocks (2.4.1), however when these glycerol stocks were used to express TraB fusion protein no new bands were observed. Plasmid instability was not a problem here since the culture remained resistant to ampicillin and chloramphenicol. It was found that if pQR523 DNA was transformed into competent cells the resulting transformants did produce the desired protein product, if on the other hand the transformation plates were stored for more than a few days, or gylcerol stocks were prepared from growing cultures the ability to express the protein of interest was lost.

In order to overcome this problem the clones were stored as DNA in T.E buffer at -20°C and transformed into competent cells prior to protein expression. Tabor & Richardson (1985) reported a similar phenomenon when using the two plasmid system (plasmid containing the T7 RNA polymerase gene & plasmid containing the T7 RNA polymerase promoter), whereby cells that had been stored for more than a few days failed when induced to produce the expected gene product, even

FIGURE 5.11 ***In vivo*** **protein expression from pQR524**



Lane A: pQR524 (induced) Lane B: pQR524 (uninduced)

Lane C: pT7-7 Lane D: *E.coli* BL21(DE3)

E.coli BL21(DE3) cultures containing pQR524 and pT7-7 (as a control) were induced and protein samples prepared as described (2.11). Approximately 40µg of total protein extracts were resolved by 15% SDS PAGE & stained with coomassie blue.

A band corresponding to the TraB fusion protein of about 16kDa is observed, which is absent from the control lanes as expected.

though they retained the resistance to their respective antibiotics. However this phenomenon was attributed to the foreign gene product being mildly toxic to the host cell.

5.9 Analysis of the TraB fusion protein

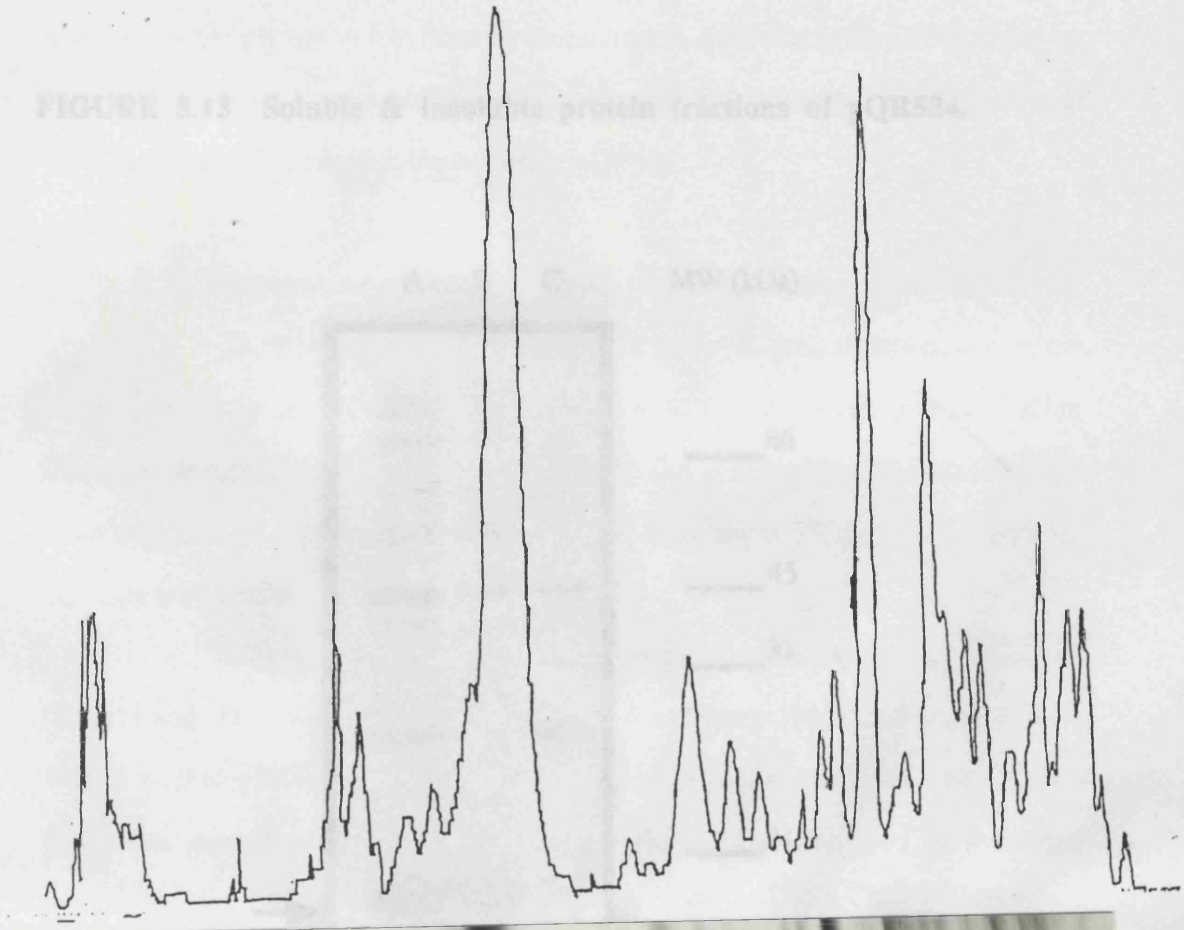
In order to estimate the amount of TraB fusion protein being expressed *in vivo* by pQR524 a number of aliquots were run on a 15% SDS polyacrylamide gel and used to carry out a densitometer scan. Figure 5.12 shows the plot obtained from one such scan and as can be seen the peak corresponding to the predicted TraB fusion protein is in fact the major protein by far. Further analysis shows that this band accounts for an average value of approximately 30% of the total cell protein.

Induced crude samples of pQR524 were sonicated as described (2.11.2), the samples were then spun (2.11.3.) in order to separate the soluble from the insoluble proteins, and aliquots of these were then loaded onto a 15% SDS polyacrylamide gel. It was found that the majority of the TraB fusion protein appeared in the insoluble fraction (Figure 5.13), suggesting that the protein was insoluble, probably being formed as an inclusion body.

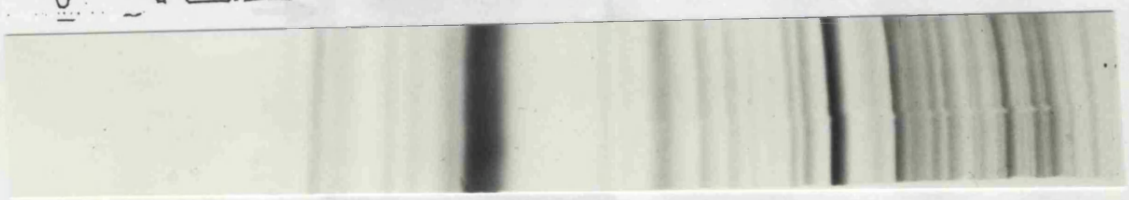
An advantage of using the pT7 expression vectors as described earlier is the ability to use antibiotics such as rifampicin which inhibit *E.coli* RNA polymerases. Since T7 RNA polymerase is resistant to rifampicin, the addition of rifampicin to cells that are producing T7 RNA polymerase results in the exclusive expression of genes under the control of the T7 RNA polymerase promoter. Figure 5.14 shows a protein gel in which one sample has been subjected to rifampicin treatment after inducing with IPTG. As can be seen when rifampicin is added many of the background proteins observed are reduced quite considerably, however the amount of TraB protein being expressed is also reduced.

FIGURE 5.12 Densitometer scan of pQR524 total protein

A



B



C

No	Component name	Locn mm	Height AU	Area AU*mm	Rel ar %					
1		15.32	0.110	0.04460	1.5	11	105.28	0.046	0.12944	4.3
2		15.84	0.068	0.11008	3.7	12	108.80	0.168	0.30972	10.4
3		19.44	0.017	0.04540	1.5	13	113.90	0.029	0.06576	2.2
4		44.92	0.063	0.08840	3.0	14	116.52	0.105	0.15196	5.1
5		47.56	0.039	0.03660	1.2	15	118.36	0.058	0.11848	4.0
6		56.52	0.024	0.05140	1.7	16	121.00	0.054	0.09156	3.1
7		61.24	0.044	0.09108	3.0	17	122.96	0.051	0.08964	3.0
8		64.68	0.183	0.88020	29.4	18	126.36	0.029	0.06248	2.1
9		87.52	0.049	0.16408	5.5	19	130.16	0.076	0.13244	4.4
10		92.44	0.032	0.11912	4.0	20	133.76	0.058	0.10400	3.5
						21	135.48	0.058	0.10460	3.5

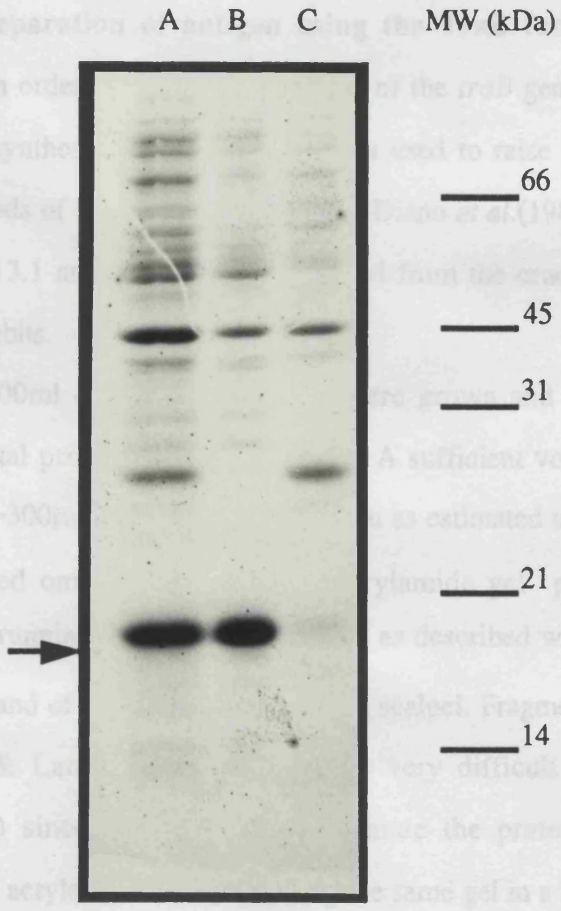
A: Densitometer scan of protein gel (pQR524 total protein extract)

B: 15% SDS gel showing pQR524 protein sample

C: Data showing each band as a % of the total protein.

The band of interest is indicated and the data suggests that it corresponds to approximately 30% of the total protein.

FIGURE 5.13 Soluble & insoluble protein fractions of pQR524.



Lane A: whole cell Lane B: pellet Lane C: supernatant

E.coli BL21(DE3) cultures containing pQR524 were induced and whole cell, pellet and supernatant fractions were prepared as described (2.11.2) and resolved by 15% SDS PAGE and stained with coomassie blue. The majority of the TraB fusion protein appears in the cell pellet fraction (indicated) suggesting that it may be insoluble.

However, the rifampicin treatment did confirm that the protein being observed on the gel was in fact from the cloned region since rifampicin inhibits all other proteins from being expressed. Thus the protein observed on the gel was the predicted TraB fusion protein created in the construct: pQR524.

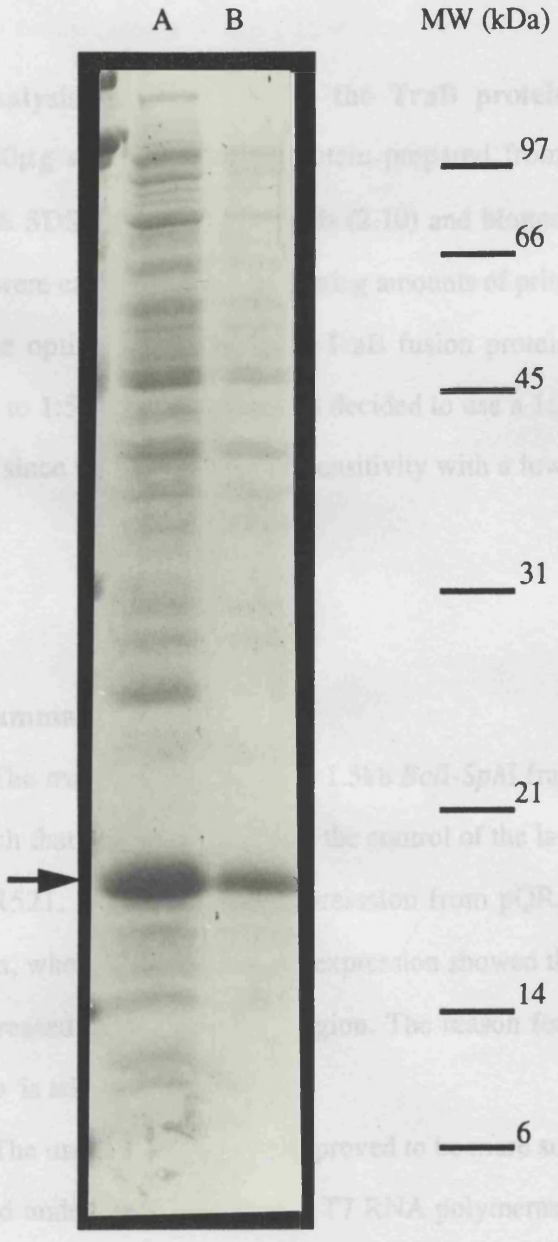
5.10 Preparation of antigen using the TraB fusion protein.

In order to investigate the role of the *traB* gene in *Streptomyces* the fusion protein synthesised from pQR524 was used to raise polyclonal antibodies. Using the methods of Harlow & Lane (1988); Diano *et al.*(1987) & Knudson (1985) described in 2.13.1 antigen could be prepared from the crude protein extracts for injection into rabbits.

500ml cultures of pQR524 were grown and induced as described (2.11.1) and total protein extracted (2.11.2). A sufficient volume of crude protein extract to give ~300μg of TraB fusion protein as estimated using the scan (Figure 5.12) was loaded onto a 15% SDS polyacrylamide gel prepared as described (2.13.1.). After running the gel it was stained as described with $CuCl_2$ (Lee *et al.*, 1987) and the band of interest excised using a scalpel. Fragmentation of the excised band (Harlow & Lane, 1988) proved to be very difficult (T.Kieser - personal communication) since the gel used to separate the proteins contained a high concentration of acrylamide (15%) (running the same gel in a lower concentration of acrylamide proved insufficient to separate the proteins in the region of interest). This method was used to prepare the first sample for injection in rabbits (~100μg per injection - T.Kieser - personal communication), but since preparation of the antigen proved difficult an alternative method was needed for the preparation of subsequent samples for injection.

The alternative approach taken was the use of the method described by Diano *et al.* (1987), which involved the transfer of proteins onto a nitrocellulose

FIGURE 5.14 ***In vivo*** **protein expression from pQR524 in the presence of rifampicin.**



Lane A: pQR524 pQR524 (+rifampicin)

E.coli BL21(DE3) cultures containing pQR524 were induced in the presence and absence of rifampicin.Total protein extracts were resolved by 15% SDS PAGE and stained with coomassie blue. In the presence of rifampicin (which inhibits *E.coli* RNA polymerase) the background proteins are clearly reduced.

membrane, locating the band of interest by staining for total protein, followed by its excision. The excised band was placed in Phosphate Buffered Saline (PBS) and sonicated until the resulting solution could be passed through a 21G hyperdermic needle.

5.11 Analysis of antibodies to the TraB protein.

20μg samples of total protein prepared from pQR524 (2.11) were resolved on 15% SDS polyacrylamide gels (2.10) and blotted as described (2.12.2). Immunoassays were carried out using differing amounts of primary antibodies in order to determine the optimum dilution. The TraB fusion protein could be detected at dilution's of up to 1:5000, however it was decided to use a 1:1000 dilution for future immunoassays since this gave the most sensitivity with a low background (data not shown).

5.12 Summary

The *traB* ORF (within the 1.5kb *BclI-SphI* fragment) was subcloned into pUC18, such that it was placed under the control of the lac promoter this plasmid was called pQR521. *In vitro* protein expresssion from pQR521 failed to show the expected protein, whereas *in vivo* protein expression showed the presence of a protein of ~16kDa expressed from the cloned region. The reason for any protein not being observed *in vivo* is still unclear.

The use of pT7-7 however proved to be more successful, when the *traB* ORF was placed under the control of the T7 RNA polymerase promoter, an obvious new protein was observed. This protein band was shown to be from the region downstream of the T7 RNA polymerase promoter by the use of rifampicin. The fusion protein was predicted to have a molecular weight of 16.1kDa as compared to TraB 14.7kDa, the extra coming from the additional 12 amino acids at the N-terminal end. Further analysis showed that the protein observed accounted for approximately 30% of

the total protein, and was being formed as insoluble protein probably in inclusion bodies.

The above expression system was then used to prepare enough TraB fusion protein for use as antigen which was susequently used to raise polyclonal rabbit antibodies.The antibodies were then shown to bind TraB fusion protein. The identification and analysis of TraB protein in *Streptomyces* is described in the next chapter.

CHAPTER SIX: Analysis of the TraB protein

6.0 Introduction

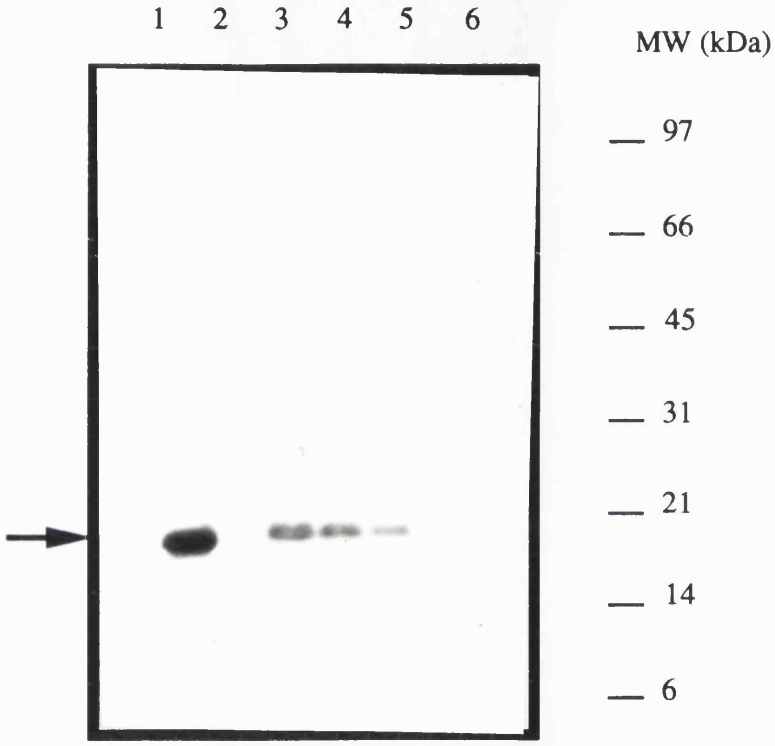
Previous work (Chapter Three) identified the presence of the 14.7kDa (138 amino acid) TraB protein within the transfer region of the *Streptomyces* plasmid SCP2*. Sequence analysis showed no homology to proteins of known structure and function, but analysis for protein motifs identified the presence of a putative Leucine zipper motif (3.7.2), which has been suggested to be involved in nucleotide binding.

Brolle and co-workers (1993) investigated the effect of insertions (using Tn 4560) into various genes of the transfer region of SCP2*. They noted that disruption of the *traB* gene (pIJ8019) resulted in the complete abolition of pocks and a reduced plasmid transfer ability ($2x10^{-5}$). It was postulated that the TraB protein may be involved in a regulatory role, possibly operating via a DNA binding activity. The following section further investigates the role of TraB in the mechanism of plasmid transfer and analyses the DNA binding capabilities of the protein.

6.1 Analysis of the TraB protein in *Streptomyces*.

The polyclonal antiserum raised to the TraB fusion protein, expressed in *E.coli* was analysed by resolving crude pQR524 total protein extracts by SDS polyacrylamide gel electrophoresis. The resolved protein samples were then blotted onto nitro-cellulose as described (2.13.2), and assayed using colorimetric immunodetection (2.13.3). Figure 4.1 shows an example of a western blot, from which it can be observed that the antibodies raised against the TraB protein are quite specific with only a few bands

FIGURE 6.1 Western blot analysis of the TraB protein in *E.coli.*



Lane 1: pT7-7 control total protein extract

Lanes 2-5: decreasing amounts (20 - 2µg) of pQR524 total protein extracts (crude).

Samples were resolved by SDS PAGE (2.10), blotted onto nitrocellulose membranes (2.13.2) and detected using the colorimetric immunodetection system (2.13.3). The above example was assayed using a 1:1000 dilution of antibodies and as can be seen the antibodies are quite specific, bands corresponding to the TraB fusion protein are indicated (arrowed). Molecular weight standards are shown (kDa).

appearing, the band corresponding to the TraB fusion protein being the most significant (as highlighted).

Similarly *Streptomyces* liquid cultures, grown (2.4.2) and total protein extracts (40μg) prepared as described (2.11) were resolved by SDS PAGE and immunodetected. Other than the pQR524 (positive control), no other bands in any of the tracks were observed (data not shown). Samples were prepared from all the *Streptomyces* plasmids listed in Table 2.2 (i.e. pIJ903, pIJ8017, pIJ8019; SCP2*, SCP2^{+}, pIJ101; together with the plasmid free controls).

Since plasmid transfer has not been reported in liquid cultures, it may suggest that growth needs to be on solid media for plasmid transfer to take place, and possibly for the efficient expression of the TraB protein. In fact while isolating RNA from pSN22 Kataoka *et al.* (1991) had reported that the signals were much stronger, when the samples had been harvested from solid media as opposed to those grown in liquid media. Protein samples (as above) from *Streptomyces* grown on solid media (2.11.3) were prepared and assayed using the antibodies raised to the TraB protein. No protein bands were observed other than the positive controls used. Since the colorimetric immunodetection using chloronaphthol as the substrate (2.13.3) proved unsuccessful, it was decided to use the more sensitive technique of enhanced chemiluminescence (ECL). However this detection system also failed to show the presence of any bands corresponding to the putative TraB protein in any of the samples analysed other than pQR524.

It may be that TraB is only present and/or expressed when plasmid transfer is actually taking place, which could account for the results observed so far. In order to examine this hypothesis, plasmid transfer was 'induced' by plating a lawn of recipient spores (TK23 - spc^{R}). This was followed by streaking donor spores (pIJ903 (str^{R}, tsr^{R}), pIJ8017, pIJ8019(str^{R}, tsr^{R}, vio^{R}) onto the recipient lawn (2.4.3). In most instances isolated

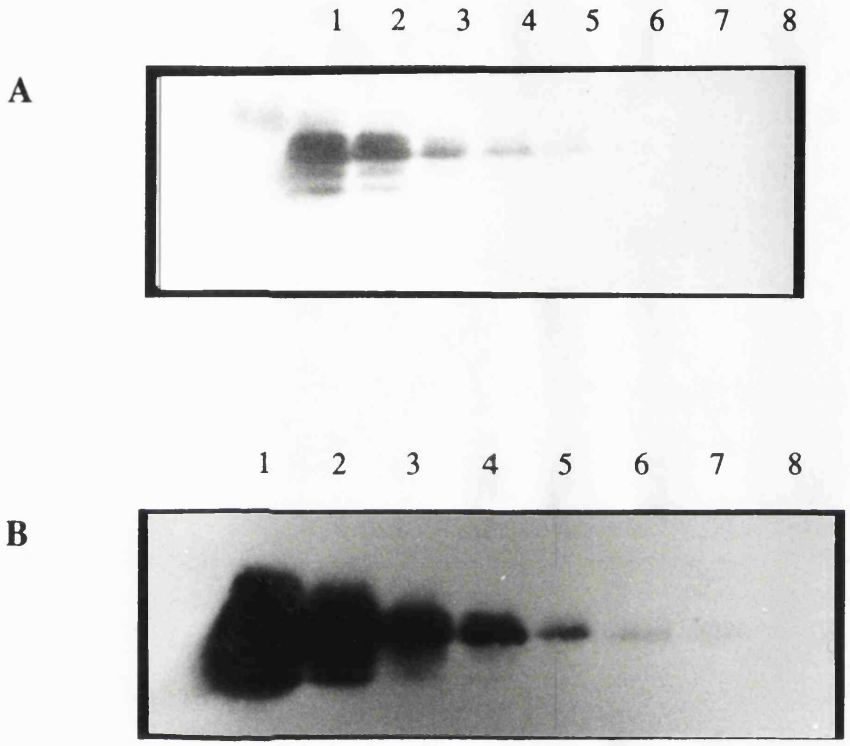
'pocks' were not seen, but a region of differential growth was observed, following the exact contours of the streaked donor strain (pIJ903). No such effect was seen with the transfer deficient plasmids pIJ8017 and pIJ8019. Protein samples of pIJ903, pIJ8017, pIJ8019, TK64, TK23 and the transconjugants were prepared as described (2.11.2 & 2.11.3). Plasmid transfer was checked (2.4.4) by taking a loopful of spores from the prospective transconjugants and plating onto agar plates containing various antibiotics. Growth on plates containing thiostrepton, viomycin and spectinomycin indicated that plasmid transfer had taken place, since the donor (pIJ903) was spc^S, tsr^R, vio^R and the recipient (TK23) was spc^R, tsr^S , vio^S and the growth selected for indicated spc^R, vio^R and tsr^R colonies.

Approximately 40μg of the above samples were resolved by SDS PAGE (2.10), electroblotted (2.13.2) onto nitrocellulose membranes and immunodetected (2.13.3) using both the colorimetric assay using chloronaphthol and the enhanced chemiluminescence kit. Once again neither immunodetection method was able to detect the presence of any TraB protein, other than the pQR524 control (data not shown). Repeats using larger quantities of total protein (~80μg) also failed to show the presence of the TraB protein, as did westerns where soluble and insoluble fractions were used.

6.2 Comparative sensitivities of the immunodetection methods.

Analysis of the TraB protein using the antibodies raised to TraB, in *Streptomyces* proved to be unsuccessful. One of the limiting factors may have been the sensitivity of the assay used for the detection of the proteins. In order to estimate the levels of protein that could be detected, a series of pQR524 protein extracts each containing differing amounts of protein were resolved by SDS PAGE and the gels blotted onto nitrocellulose membranes.

FIGURE 6.2 Comparative sensitivities of the immunodetection methods.



Differing amounts of TraB protein (estimated using the method described 2.10.2 & 5.9) were resolved by SDS PAGE (Lanes 1- 2µg, 2- 1µg, 3- 500ng, 4- 250ng, 5-150ng, 6- 100ng, 7- 50ng, 8- 20ng). Filter A was assayed using the colorimetric system (chloronaphthol as substrate), filter B using the ECL kit (30 sec exposure). In the above example TraB fusion protein was detected down to ~150ng (colorimetric assay) and ~50ng (ECL detection).

The nitrocellulose filters were then assayed using both the colorimetric assay and enhanced chemiluminescence. The results (Figure 6.2) indicated that the colorimetric assay could detect a minimum value of approximately 150ng of TraB fusion protein, whereas the more sensitive ECL kit allowed for a minimum value of approximately 50ng of TraB fusion protein to be detected. Suggesting that the ECL kit was about 3 times more sensitive than the colorimetric based assay. However it must be emphasised here that the methods used to quantify the amounts of protein detectable were extremely crude and were only intended as a comparison of sensitivities.

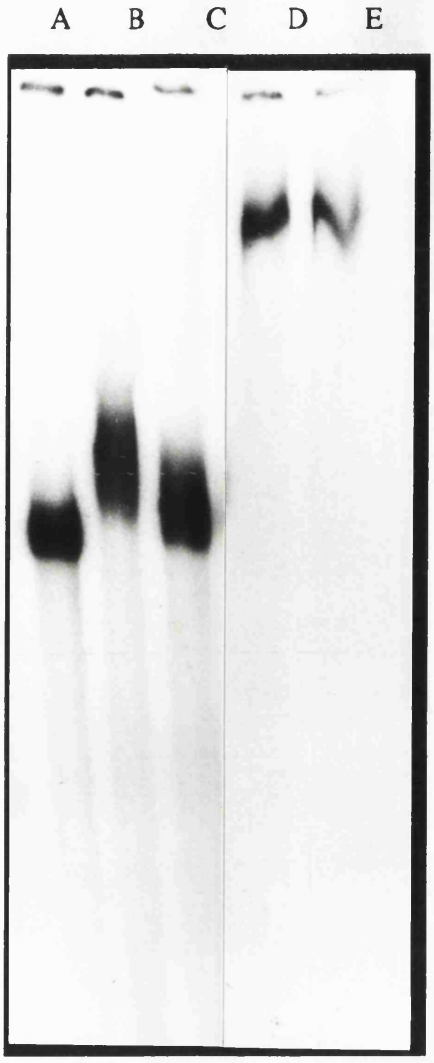
6.3 TraB: a putative DNA binding protein?

Sequence analysis of the *traB* gene identified a probable leucine zipper motif (shown in Figure 3.7), which together with the N-terminal basic region has been associated with DNA binding activity. In order to investigate whether TraB was also a DNA binding protein, gel retardation experiments were carried out (Hennighausen & Lubon, 1987).

Gel retardation relies on the principle that the mobility of DNA in non denaturing polyacrylamide gels is retarded when proteins are bound as compared to free DNA. In most cases pure or semi-purified proteins samples are used and the DNA fragments are small (500 - 100bp). However here we attempted to use crude protein extract and a DNA fragment of about 3kb in length.

The TraB fusion protein over expressed *in vivo* by the construct pQR524 had earlier been reported as being insoluble. Thus in order to obtain a soluble form of the protein, pQR524 cultures were induced using IPTG but grown at 25°C instead of 37°C. This led to a reduced amount of protein being expressed but the majority of the protein was found to be soluble (data not shown).

FIGURE 6.3 Gel retardation assay using crude TraB protein extracts.



Approximately 40µg of *E.coli* crude protein extract containing pQR524 (TraB) and pT7-7 (control) were incubated with labelled (~10 000cpm) DNA (*Sac*I (23)- *Pst*I (18) fragment) as described (2.13.4). After resolving by non denaturing 3% polyacrylamide gel electrophoresis the gel was dried and exposed to X ray film.

Lane A: free DNA Lane B: DNA+pQR524 crude protein extract

Lane C: DNA+pUC18 crude protein extract

Lane D: as lane B+antiserum to the TraB protein

Lane E: as lane B+control antiserum (pre-immune).

The major transfer proteins of pIJ101, pSAM2 and maybe pSN22 have been suggested as being under the regulatory control of a protein found within their respective transfer regions. In order to determine whether a similar situation was found in SCP2*, where TraB was the regulatory protein, pQR210 DNA was digested to completion with *Sac*I & *Pst*I, and the resulting 3.23kb fragment (containing the regions coding TraA and TraB), end labelled using ^{32}P and T4 Polynu cleotide kinase as described (2.7.6). The labelled DNA fragment was then incubated with crude TraB protein extract and resolved by non denaturing polyacrylamide gel electrophoresis (2.13.4). Analysis of the autoradiograms showed some retardation compared to the free DNA but was also accompanied by non-specific binding in the control sample. The shift observed was only minimal, but in order to exaggerate the amount of band shift observed the samples were incubated with the antiserum (raised to the TraB protein). The amount of retardation observed was greatly enhanced (Figure 6.3) but as before the results were inconclusive due to the non specific binding in the control extracts. Further analysis of the samples with non specific competitors also proved to be unsuccessful. Due to time restrictions this avenue could not be further investigated.

6.4 Summary

We have analysed *traB* expression in *Streptomyces* but have been unable to detect any proteins using western blot analysis and the polyclonal sera raised against the TraB fusion protein expressed in *E.coli* using pQR524. However the TraB fusion protein expressed in *E.coli* can be detected using the same antibodies. Plasmid transfer has not been reported, and is thought not to take place in liquid, samples harvested from solid media as well as samples where plasmid transfer had been 'induced' also failed to show the presence of the TraB protein. The reason for this observation may be

due to the small amounts of the TraB protein present in the *Streptomyces* samples used. The sensitivity of the assay system used here, has also been demonstrated as being able to detect levels as low as ~50ng of the TraB fusion protein.

CHAPTER SEVEN: General Discussion

7.0 Introduction

The low copy number *Streptomyces coelicolor* plasmid SCP2* shows great potential as a *Streptomyces* cloning vector, Particularly for cloning large fragments and genes required in unit copy number such as those of the antibiotic biosynthetic pathways.

Currently little is known about the general transfer functions of *Streptomyces* plasmids. At the onset of this research, very little information was available about the plasmid transfer characteristics of SCP2*. This plasmid had previously been shown to be conjugative and induce pock formation. It was suggested that plasmid transfer was accomplished in two stages: intermycelial plasmid transfer from the donor to the recipient, followed by intramycelial spread of the plasmid throughout the recipient mycelia. Lydiate *et al.* (1985) had managed to localise the transfer and spread genes to a *circa* 9kb fragment during the construction of cloning vectors based on SCP2*.

Thus, the main aim of this research was to further investigate the plasmid transfer characteristics of SCP2*. It was decided that the best strategy would be to clone and sequence a large part of the region previously suggested as harbouring the transfer and spread related genes of SCP2*, using a bifunctional derivative pIJ903. It was hoped that the further examination of the genes located from the DNA sequence, by analysis of their predicted amino acid sequences and the subsequent comparison of the putative proteins encoded by them would indicate their probable roles during plasmid transfer and spread of SCP2*. The DNA sequence would also provide the basis for conducting directed genetic experiments and thus bring us closer to understanding the mechanisms of plasmid transfer of *Streptomyces* plasmids.

7.1 Localisation of transfer characteristics by mutagenesis.

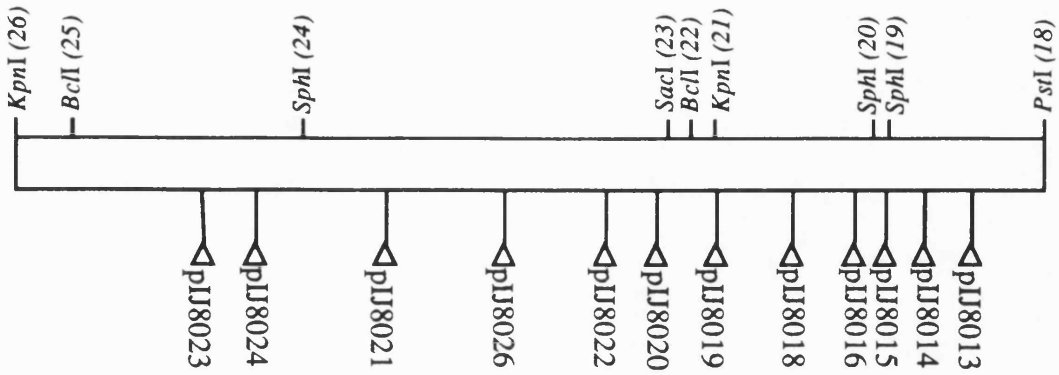
Brolle *et al.* (1993) investigated the transfer region of SCP2* by mutagenesis using the transposon Tn4560. Their investigations revealed that there were potentially five transfer related genes clustered within the previously defined (c 9kb) transfer region as indicated by Figure 1.3.

The transfer genes were identified due to their effects on plasmid transfer frequency, pock formation and chromosome mobilisation ability [CMA]. The five putative genes located were named *traA*, *traB*, *traC*, *traD* and *spd* and their functional locations are indicated by Figure 7.1.

The disruptions to *traA* shown by plasmids pIJ8013 - pIJ8016, generally resulted in the complete cessation of plasmid transfer and the no pock phenotype. Many transposon insertions were found in the genetically defined *traA* loci, and no other mutations showed such dramatic effects on plasmid transfer and pock formation. This observation leads one to speculate that *traA* may be an absolute requirement for intermycelial plasmid transfer from the donor to the recipient. Genetically *traA* shows similar characteristics to the *tra*, *traSA* and *traB* genes of pIJ101, pSAM2 and pSN22 plasmids respectively. These were also shown to encode proteins essential for plasmid transfer.

The single insertions in *traB* (pIJ8019), *traD* (pIJ8021) and the two transposon insertions found in *traC* (pIJ8020, pIJ8022) showed similar characteristics. They all failed to exhibit pocks suggesting that these loci were also involved with the phenomenon of intermycelial transfer. However transfer did not completely cease since the plasmid transfer frequencies were reduced from 1(control plasmid - pIJ903) to values between $2x10^{-5}$ (pIJ8019, pIJ8020 & pIJ8021) to $1x10^{-4}$ (pIJ8022). However *traC* mutant plasmids showed no chromosome mobilisation, whereas *traB* and *traD* mutant plasmids showed a reduced CMA (100 fold and 1000 fold respectively) compared to the pIJ903 control plasmid.

FIGURE 7.1 Localisation of transfer genes of SCP2* by mutagenesis.



Plasmid	Pock formation	Plasmid transfer	CMA	
pIJ903 (control)	+	1	$3x10^{-4}$	
pIJ8013	-	$5x10^{-3}$	$9x10^{-7}$	traA
pIJ8014	-	0	0	
pIJ8015	-	0	0	
pIJ8016	-	0	0	
pIJ8018	+	1	$3x10^{-4}$	traX
pIJ8019	-	$2x10^{-5}$	$5x10^{-6}$	traB
pIJ8020	-	$2x10^{-5}$	0	traC
pIJ8022	-	$1x10^{-4}$	0	
pIJ8026	+	1	$3x10^{-4}$	
pIJ8021	-	$2x10^{-5}$	$1x10^{-7}$	traJ
pIJ8024	+/-	0.5	$3x10^{-4}$	spd
pIJ8023	+/-	0.7	$7x10^{-4}$	

Locations of the mutant plasmids (pIJ8013 - pIJ8026) are indicated by the top figure, which also depicts the restriction enzyme sites with reference to the transposon insertions. Pock formation is indicated by: +, wild-type pock size; +/-, small pocks; -, no pocks. Plasmid transfer frequencies are indicated, the control plasmid pIJ903 was considered to show a 100% transfer frequency. CMA frequencies were quantified as the number of recombinants over the sum of colonies of both parental types.

[Figure adapted from Brolle *et al.*, 1993]

The fifth genetic loci proposed by Brolle *et al.* (1993) to be present in the transfer region of SCP2* was termed *spd*. Insertions into this region lead to no significant change in plasmid transfer but the small pock phenotype was observed suggesting that this region was necessary for plasmid transfer during intramycelial spread. Small pocks were also observed when *spd* functions of pIJ101, pSAM2 and pSN22 were disrupted, suggesting a functional similarity with the *spd* loci of SCP2*. No CMA was observed with the *spd* mutant plasmids (pIJ8023 & pIJ8024).

Two additional plasmids of interest detected by Brolle *et al.* (1993) were pIJ8018 and pIJ8026 showing transposon insertions located between *traB* and *traA*, and *traD* and *traC* respectively. Both these mutant plasmids showed transfer characteristics similar to those observed with the pIJ903 control plasmid. Sequence analysis performed during this research and by Brolle *et al.* (1993) shows that there is an ORF spanning the region between *traB* and *traA*, with the expected 1st, 2nd and 3rd base position GC% and high overall GC% expected from *Streptomyces* protein coding regions. This suggests that this region could in fact be a potential protein coding region, and is thus termed *traX* (Brolle *et al.*, 1993).

7.2 TraA: The major transfer protein of SCP2*.

Derivatives carrying insertions in the *traA* ORF (pIJ8013 - pIJ8016), were unable to form pocks and no plasmid transfer was detected (Brolle *et al.*, 1993). No other mutations had such a dramatic effect on plasmid transfer and pock formation, suggesting that this ORF encoded the major protein required for SCP2* intermycelial transfer from the donor to the recipient. The effects of disruptions to *traA*, were similar to those observed when *tra* of pIJ101 and *traSA* of pSAM2 were disrupted. Similarly plasmid transfer was completely abolished, and these two genes were also proposed to encode the major transfer proteins of pIJ101 and pSAM2.

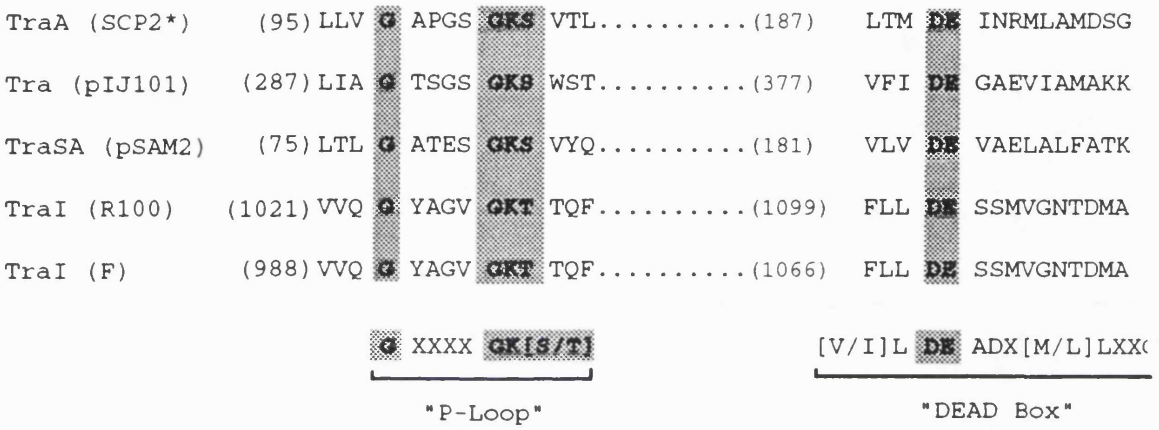
Computer assisted amino acid sequence comparisons of TraA showed no significant homology to any of the proteins in any of the data bases analysed, except for very weak homologies with *tra* of pIJ101 and an *E.coli* DNA helicase (see Chapter 3.0). Yet no obvious sequence homology was noticed between TraA and TraSA. However comparisons between the amino acid sequences of Tra (pIJ101) and TraSA (pSAM2) showed some homology.

Chapter three had also shown the presence of a putative P-loop, ATP/GTP binding domain in the predicted amino acid sequence of TraA, and in the expected position downstream (92 amino acids) of the P-loop a faint similarity to the DEAD box motif common to DNA helicases (Linder *et al.*, 1989). A similar search interestingly revealed the presence of similar P-loop and DEAD box motifs for both Tra (pIJ101) and TraSA (pSAM2). The faint similarity to the DEAD box motif being located 90 and 106 amino acids downstream of the P-loop as indicated by Figure 7.2, This is interesting since no obvious similarity had been observed between the predicted amino acid sequence of the three proteins. Yet they may have similar biological functions and have been shown to possess conserved active proteins domains.

The N-terminus of TraA has previously been shown to have weak homology to an *E.coli* DNA helicase, yet when Tra and TraSA were compared no such homology to the DNA helicase was observed. It may however be possible that all three proteins share a similar biological activity, Possibly in the unwinding of DNA prior to transfer, using the energy from ATP hydrolysis similar to the TraI proteins found in the plasmids (F, R100) of Gram negative bacteria. However sequence comparisons between the TraI proteins and Tra, TraA and TraSA revealed no significant homology.

This work also proposed that the *traA* gene was in fact longer than that previously suggested by Brolle *et al.* (1993). The additional 36 aminoacids at the N-terminus lead to the *traA* ORF being translationally

FIGURE 7.2 P-Loop and DEAD box motifs of potential DNA helicase proteins.



The consensus sequences for the P-Loop and DEAD box motifs are indicated Conserved residues are shaded, X refers to any amino acid and the numbers in brackets refer to the amino acid positions. The distances between the P-Loop and DEAD box motifs are also as expected; 92, 90, 106, 77 and 78 residues TraA, Tra, TraSA, TraI (R100) and TraI (F) respectively. Amino acid sequences were obtained from Kendall & Cohen (1988), Hagege *et al.* (1993), Yoshioka *et al.* (1990), Bradshaw *et al.* (1990) and this work.

coupled with the upstream *traX* ORF, such that the stop codon of *traX* overlapped the start codon of *traA* (see below). This stop/start motif seems to be quite common to operons of Gram negative bacteria.

7.3 The genes encoding TraX and TraA are translationally coupled.

Translational coupling refers to situations in which translation of a downstream coding sequence requires translation and translational termination of an upstream sequence at or near the initiation codon of the downstream sequence (Oppenheim & Yanofsky, 1980). This phenomenon was first described for the tryptophan operon in *E.coli*: translation of the proximal *trpE* gene appeared to be necessary for efficient translation of the distal *trpD* gene (Oppenheim & Yanofsky, 1980). It has been suggested that the most likely explanation for the sequential translation by the same ribosome is that it results from either a high ribosome density near the second start or, an actual sliding of the ribosome to the next start site (Schumperli *et al.*, 1982; Askoy *et al.*, 1984).

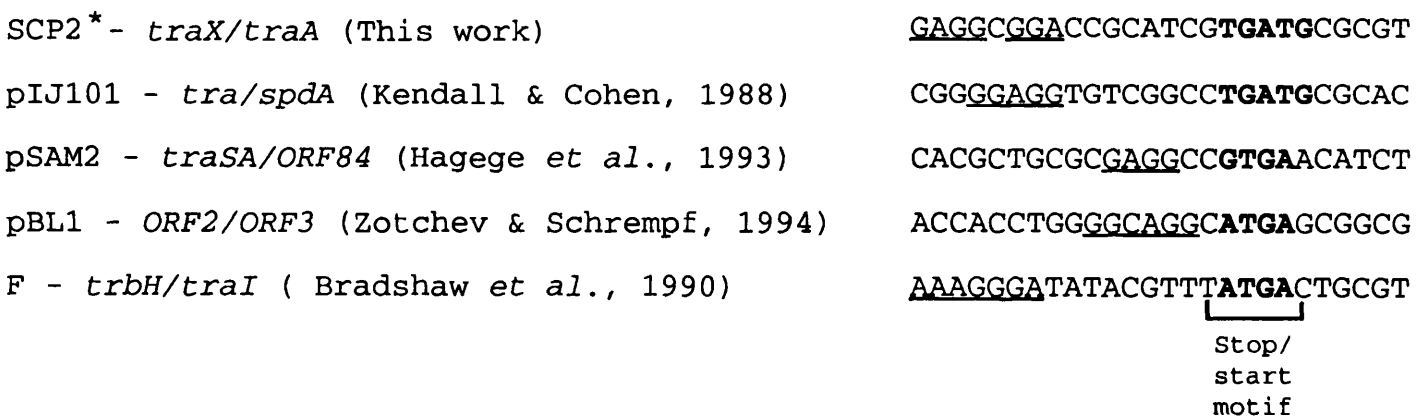
The GAGGCGGA sequence, presumably part of the *traA* ribosome binding signal, lies entirely within the *traX* coding sequence. This overlap of regulatory signals suggests the possibility of a functional coupling of *traX* and *traA* translation. However when *traX* was disrupted no polar effects were observed, as would be expected if *traX* and *traA* are functionally coupled, in fact the *traX* mutant plasmid (pIJ8018) had no effect on plasmid transfer, pock formation or CMA. This phenomenon needs to be further investigated, since there was only a single transposon insertion located within, but very close to the N-terminus of *traX*. Note that there is in fact 78bp of untranslated sequence between the predicted termination codon of *traB* and the initiation codon of *traX*.

Analysis of translational coupling amongst genes involved in plasmid transfer in *Streptomyces* revealed an interesting observation. Many of the genes proposed as encoding the major transfer proteins, like *traA* of SCP2*, seemed to show the stop/start motif common to translationally coupled ORFs. In some cases the ORFs may be translationally coupled to ORFs with unknown functions similar to *traX*. The proteins encoded by these genes, like TraA contained the ATP binding P-loop domain and some weak similarity to the DEAD box motif common to DNA helicases, where the highly conserved D (Aspartic acid) and E (Glutamic acid) residues were present (Hodgman, 1988; Linder *et al.*, 1989).

Both *tra* of pIJ101 and *traSA* of pSAM2 are translationally coupled to an ORF downstream, at their C-termini. An opposite situation is seen for *traA* of SCP2* where *traA* is translationally coupled to an ORF upstream at the N-terminus. As is shown in Figure 7.3 *tra* (pIJ101) is proposed as being translationally coupled to *spdA*, and *traSA* (pSAM2) to *ORF84*. A fourth *Streptomyces* plasmid pBL1 (Zotchev & Schrempf, 1994) may also show a structural overlap of translational regulatory signals at the intercistronic boundary between *ORF2* and *ORF3*, suggesting functional coupling between these two genes. *ORF2* in line with the other *Streptomyces* transfer genes seems to encode a protein essential for intermycelial plasmid transfer. In this work it is postulated that *traA* may be translationally coupled to *traX*. No biological function could be attributed to *traX*, which showed no adverse effect when disrupted, similar to the situation found with *ORF84* of pSAM2. A comparison between the amino acid sequences of TraX and ORF84 showed no homology whatsoever. Similarly no significant homology was observed between the two proteins and any of the proteins in the data bases, except for a very faint similarity between TraX and the Rep protein found in pIJ101.

Earlier it was shown that the Tra, TraSA and TraA proteins all showed faint similarity to the DEAD box motif and all possessed a potential P-

FIGURE 7.3 Intercistronic regions of some transfer genes



The DNA sequence across the termination and initiation regions of some *Streptomyces* plasmid transfer genes is shown, indicating the stop/start motif (bold face) and the putative ribosome binding sites (underlined). In most of the above cases reinitiation of protein synthesis occurs at ATG codons, which are also the most frequently used initiation codons in normal translation. An exception is seen for pSAM2 where a GTG codon is used. The structural overlap of regulatory signals at the intercistronic boundary suggests possible translational coupling between the genes.

loop motif. These three proteins were compared to the TraI protein of the F plasmid which also showed homology to the DEAD box motif and contained a putative ATP binding domain. It is thus interesting to note that the *traI* gene shows a further similar characteristic in that it too is translationally coupled to an ORF (*trbH*) upstream, and therefore resembles the situation found between *tra/spdA, TraSA/ORF84* and *traX/traA*. This may suggest that the genes encoding all putative DNA helicase proteins in plasmid transfer regions show translational coupling with other genes. This however is not the case since *traI* of R100 does not show translational coupling with any other ORF, but like *traI* of F is located within an operon. The presence of the overlapping stop/start translational regulatory signals at the intercistronic boundaries of all the essential transfer genes of *Streptomyces* plasmids suggests that this phenomenon could be a general feature of their genetic regulation.

The results obtained using the *in vitro* coupled transcription/translation system, described in Chapter four are rather difficult to interpret. The single protein of approximately 80kDa observed on the gels, seems to have been the product of the *traX* and *traA* genes, possibly a translational fusion product resulting from the single polycistronic mRNA. Note that in translationally coupled genes it has been suggested that the ribosomes do not actually dissociate from the mRNA but in fact the same ribosomes could be involved in the translation of sequential genes, The ribosomes changing positions from one reading frame to another by sliding to the next translation initiation codon. It is important to take into consideration that the transcription and translation studies have been carried out *in vitro* using a kit based on a heterologous organism. It would thus be interesting to study the translation of the genes in *Streptomyces*.

It could however be possible that the TraX/TraA fusion protein being observed is the result of ribosomal frameshifting. Ribosomal frameshifting is the shifting of ribosomes from one reading frame to another at

a specific sequence on the mRNA during translation which allows the synthesis of a single protein from two (or more) overlapping genes (Atkins *et al.*, 1983; Garcia *et al.*, 1993).

So far, most examples of this type of control have come from the retroviruses, where frameshifting appears to be a mechanism for the regulation of expression of the viral RNA dependant DNA polymerase (Brierley *et al.*, 1989). One termination codon in Rous sarcoma virus (RSV) and human immunodeficiency virus type 1 (HIV-1), and two termination codons in mouse mammary tumour virus (MMTV) are suppressed by (-1) ribosomal frameshifts into alternative overlapping reading frames (Varmus, 1988; Brierley *et al.*, 1989). Ribosomal frameshifting is not however confined to the viruses, for example it has been proposed to be involved in the synthesis of the *E.coli* DNA polymerase III holoenzyme subunits (Tsuchihasi & Kornberg, 1990).

Two different signals in the mRNA have been suggested to induce frameshifting. First, there is a heptanucleotide consensus sequence referred to as the 'slippery' or 'shifty' sequence where the frameshifting actually takes place (Garcia *et al.*, 1993; Tu *et al.*, 1992). This sequence has been generally shown to be composed of a A and/or U rich region (Garcia *et al.*, 1993). The second signal is a structural element just downstream of the heptanucleotide sequence, this region of RNA secondary structure is referred to as a pseudoknot, almost all of which identified to date are hairpin-loop type structures (Garcia *et al.*, 1993; Brierley *et al.*, 1991). The mechanism by which the pseudoknot promotes frameshifting is not yet clear, but it has been proposed (Brierley *et al.*, 1989 & 1991) that ribosomes may slow or stall at the hairpin-loop, increasing the likelihood of a tRNA slippage event.

However, if the large 80kDa protein was the result of the suppression of the termination codon of *traX* by ribosomal frameshifting, the control signals would have to be significantly different to those previously

associated with the process of frameshifting, since there is no A and/or U rich region in the vicinity of the UGA termination codon, or any obvious RNA secondary structure constituting a pseudoknot.

It is noteworthy that the genes encoding the proteins proposed as being the major proteins involved in intermycelial transfer from donor to recipient in *Streptomyces*, all share the properties of containing a potential P-loop ATP binding domain, show faint similarity to the DEAD box motif common to DNA helicases and all show possible translational coupling to other ORFs.

7.4 TraB: A regulatory protein of SCP2* transfer ?

Disruption of *traB* completely abolished pock formation, whereas plasmid transfer was reduced (Brolle *et al.*, 1993). Bibb *et al.* (1981) had previously suggested that the transfer genes of SCP2* were transiently derepressed when the plasmid entered a plasmid free *S.coelicolor* strain. This lead Brolle *et al.* (1993) to speculate that the *traB* mutant plasmid (pIJ8019) may have lost the ability to derepress the transfer genes after entering a plasmid free compartment. This together with the presence of a perfect match to a leucine zipper motif (identified from the predicted amino acid sequence of *traB*), provided the basis for tentatively proposing a regulatory role for the TraB protein.

The structure referred to as the 'leucine zipper', has been proposed to explain how some eukaryotic gene regulatory proteins function (Landschulz *et al.*, 1988; Busch & Sassone-Corsi, 1990). Essentially the leucine zipper consists of an in - phase periodic repetition of leucine residues at every seventh position over a distance of four leucines in the primary sequence (Landschulz *et al.*, 1988; Busch & Sassone-Corsi, 1990). An adjacent region of positively charged or cluster of basic amino acids at the N-terminus, has also been suggested to be present in most leucine zipper proteins (Busch &

Sassone-Corsi, 1990; Giraldo *et al.*, 1989). The leucine zipper appears to be a protein domain that promotes dimer formation (Busch & Sassone-Corsi, 1990) via hydrophobic interactions (Johnson & McKnight, 1989), and the accompanying basic region is suggested as being involved with DNA binding (Busch & Sassone-Corsi, 1990; Vinson *et al.*, 1989).

As previously mentioned, the leucine zipper can be found in many eukaryotic regulatory proteins but there are few examples of such a domain in prokaryotic proteins. The only reported perfect matches to this motif are found in the MetR activator protein of *E.coli* methionine biosynthesis (Maxon *et al.*, 1990), where the leucines in the heptad repeat regions were shown to be involved in the formation of a DNA binding homodimer, and in the *klbC* gene of the *kil-kor* regulon of the RK2 plasmid (Thomson *et al.*, 1993). Thomson *et al.* (1993) suggested that the putative leucine zipper in *klbC* may be involved in the interaction of KlbC monomers with each other or with other proteins. There is no evidence for the biochemical functions of the KlbC protein, but Thomson *et al.* (1993) have postulated a DNA binding role for KlbC.

The heptad repeats of the leucine zipper have also been shown to be important in the formation of tetramers of the Lac repressor (Thomson *et al.*, 1993; Alberti *et al.*, 1991) as well as being proposed to be involved in interactions of plasmid replication initiators (Thomson *et al.*, 1993; Giraldo *et al.*, 1989). Comparisons of the amino acid sequences of proteins involved in the initiation of replication in plasmids of Gram negative bacteria, identified in some of them, a leucine zipper type domain at their amino termini (Giraldo *et al.*, 1989). They also noted that the initiation proteins shared the properties of binding to DNA and of being transcriptional regulators of their own synthesis. However the presence of helix-turn-helix type DNA binding motifs at the C-termini were also shown (Giraldo *et al.*, 1989).

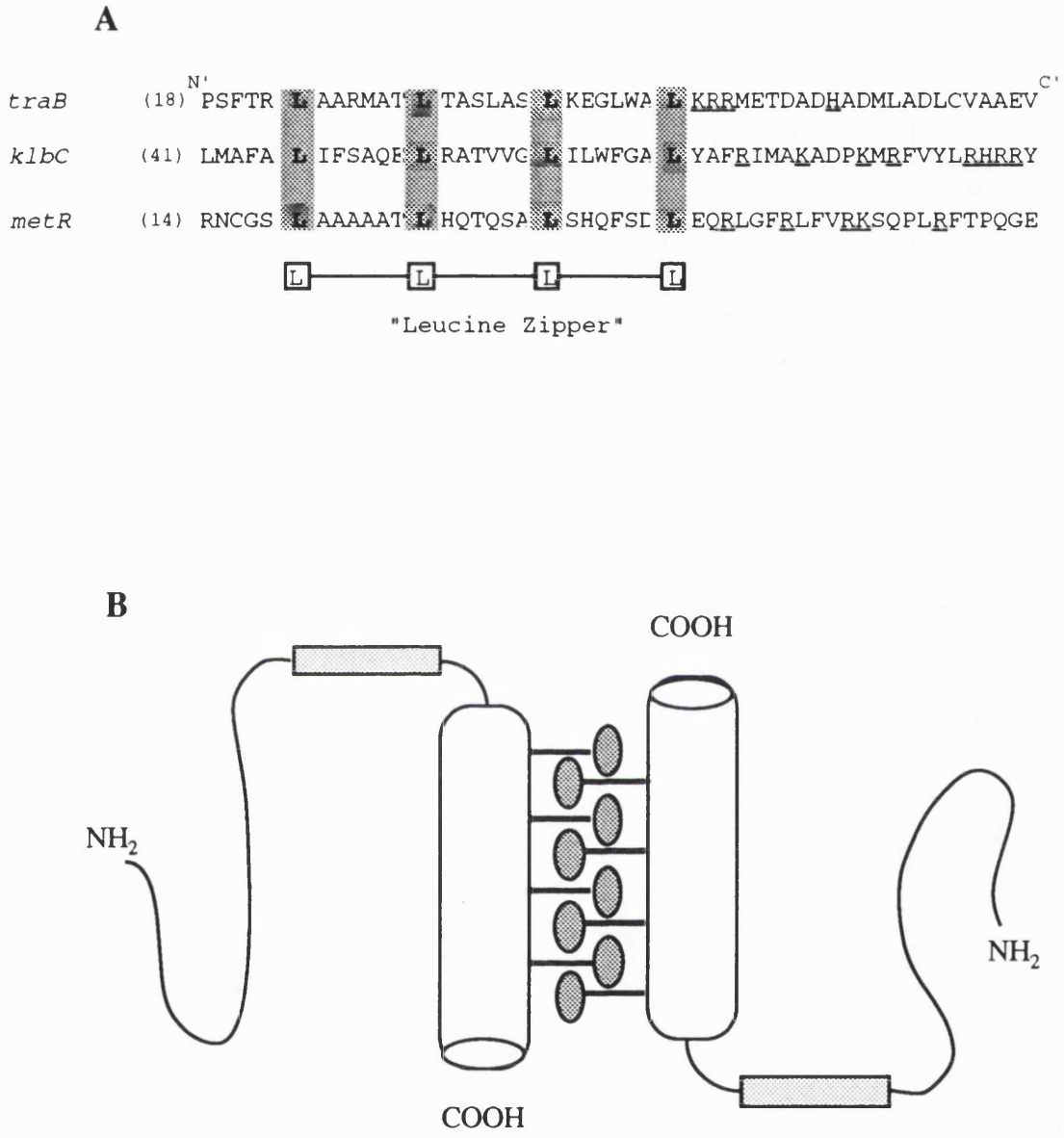
Analysis of the predicted amino acid sequence of *traB* (3.7.2) revealed the presence of a perfect match to the leucine zipper, but no accompanying basic region at the N-terminus was identified, similar to the situation with *klbC*, though a small cluster of basic residues was found at the N-terminus of *metR*. However a cluster of three positively charged amino acids was found at the C-terminus of the heptad repeat (Figure 7.4), similar though smaller than the cluster of basic residues in the *klbC* gene of RK2 (Thomson *et al.*, 1993). Sequence comparisons between the predicted amino acid sequences of *traB* and *klbC* showed no overall homology except for the leucine residues comprising the leucine zipper.

It can be postulated tentatively, that the leucine zipper in *traB* may well be involved in the interaction of TraB monomers with each other forming homodimers or with other proteins (with an analogous leucine zipper) in the formation of heterodimers. Note that there is in fact a small region (LxxxxxxL) at the C-terminus of the *traC* ORF which could possibly be part of a leucine zipper, but since the remaining sequence of *traC* has not been determined no hypothesis regarding this can be made. Given that most known leucine zipper proteins bind DNA (Vinson *et al.*, 1989; Thomson *et al.*, 1993) there is reason to consider the possibility that TraB is also a DNA binding protein, even though the presence of established DNA binding domains, such as the helix-turn-helix or Zinc finger domains were not identified.

7.5 Characterisation of the TraB protein.

Chapter five described the cloning and expression of the TraB fusion protein, under the control of the T7 RNA polymerase promoter, and its subsequent use for the production of polyclonal sera. Chapter six further described the characterisation of the TraB protein using immunoassays and gel retardation experiments. Western blots failed to reveal the presence of protein bands corresponding to TraB in *Streptomyces*. The use of extracts prepared

FIGURE 7.4 The leucine zipper of *traB*.



A- The leucine zipper motifs of *traB* (this work), *klbC* (Thomson *et al.*, 1993) and *metR* (Maxon *et al.*, 1990) are shown. The leucine residues (shaded) are found at intervals of six amino acids. The basic amino acid residues at the C - terminus are underlined.

B- Schematic diagram showing two polypeptide chains of a leucine zipper protein dimerized via hydrophobic interactions between two α helixes. The dimers can be disposed in either an antiparallel (shown) or parallel array with respect to the leucine repeat. Interdigitating protrusions symbolise leucine side chains. Stippled rectangles represent basic regions believed to establish direct contact with DNA.

[Figure adapted from Landschulz *et al.*, 1988; Johnson & McKnight, 1989]

from samples, where plasmid transfer had been shown to have taken place, also proved unsuccessful in identifying the TraB protein. The sensitivity of the assay was investigated, and it was shown that TraB extracts (from *E.coli*) could be detected down to levels of approximately 50ng. This value corresponded to a total percentage of ~0.1 - ~0.2% of total protein loaded, thus suggesting that had TraB been present in the *Streptomyces* samples, the assay used would have been sufficient to detect its presence had it been greater than about 0.1% of the total protein.

The fact that no TraB was detected using the antibodies raised against it, may itself strengthen the earlier hypothesis, that the TraB protein is involved in a regulatory role. Since, one would presume that as a plasmid DNA binding protein, only very small amounts of the protein may be required, this assumption is further strengthened by the very low copy number of the host plasmid SCP2*. Additionally the presence of the leucine zipper, proposed to be involved in dimer formation would further strengthen this argument, since Vinson *et al.* (1989) suggested that tightly associated dimers are able to bind selectively to their DNA targets at far more dilute concentrations than monomers. This is presuming that its own transcription is also regulated, perhaps by itself ?

It would be interesting to determine whether the leucine zipper is in fact involved in dimer formation, either for the purpose of homodimer formation or heterodimer formation. An obvious experiment to carry out would be to determine the size of the TraB protein using gel filtration, and thus provide the answer to the question of whether the leucine zipper of TraB is involved in dimerization. Partial purification (note the size and pI of the TraB protein have been calculated from the DNA sequence) of protein extracts prior to resolution by SDS PAGE and assaying using the antibodies, may allow the detection of TraB. Time spent improving the immunoassay technique or

searching for more sensitive alternatives would itself not be wasteful, since the amount of TraB to be detected could well be the limiting factor.

The investigation into whether the TraB protein is a DNA binding protein and the identification of the binding site, was unsuccessful largely due to time restraints. An obvious way forward here, would be the use of smaller labelled fragments in the gel retardation reactions. It is worth noting here that, as suggested previously the leucine zipper in TraB may be involved in dimerization. This poses the question of whether TraB forms heterodimers via the leucine zipper, which may suggest that another protein is required for TraB to function efficiently ?

The regulatory proteins of other *Streptomyces* plasmid transfer systems (i.e. Kor proteins of pIJ101 and pSAM2) have all been shown to contain helix-turn-helix DNA binding domains (Kendall & Cohen, 1988; Hagege *et al.*, 1993). In the case of pIJ101, the KorA and KorB proteins have been shown to be transcriptional regulators. The KorSA of pSAM2 has yet to be characterised, but has been shown to have similar functional properties. It is likely that it may also be a transcriptional regulator of TraSA, since it contains a helix-turn-helix DNA binding domain and its amino acid sequence shows good homology to that of KorA (Hagege *et al.*, 1993).

However amino acid sequence comparisons of the products of the *kor* genes of pIJ101 and pSAM2 with that of the *traB* gene of SCP2*, failed to show any significant homology other than a weak similarity between KorA and TraB.

Analysis of the amino acid sequence of the *kor* functions of pIJ101 (*korA* and *korB*), pSAM2 (*korSA*) and pSN22 (*traR*) (Hagege *et al.*, 1993), all seem to show some degree of similarity to each other. Their functional characteristics seem to be similar. It is interesting to note that they all also show some degree of homology to other known repressor proteins, all belonging to the GntR family of transcriptional regulators the archetype of

which is found in the gluconate operon of *Bacillus subtilis* (Hagege *et al.*, 1993; Fujita *et al.*, 1986). As previously suggested, the *kor* genes identified in *Streptomyces* (*korA*, *korB*, *korSA*, *traB*) seem to show the presence of one or more helix-turn-helix DNA binding motifs. *Kor* functions are named as such due to their ability to override the lethal effects of a *kil* loci (*tra*, *kilA*, *traSA*, *traR* respectively). It is thus interesting to note that TraB, identified by this work, which could be functionally similar to the Kor proteins, and act as a regulatory protein does not show any significant amino acid sequence homology to the other Kor proteins or to the bacterial repressor proteins of the GntR family.

7.6 Future work

Transposon mutagenesis results obtained by Brolle *et al.* (1993) can only be used as a rough indication to the location and/or effects of the insertions to the genes in the transfer region of the *Streptomyces* plasmid SCP2*. The insertion of the transposon into the plasmid was totally random, and as can be seen from the results many inserts were found in some regions whereas some regions were completely devoid of insertions. Further biological studies are thus necessary (by site directed mutagenesis and the insertion/deletion of specific regions) in order to investigate the transfer genes in more detail. The transposon mutagenesis results can not be used in order to determine the presence of a *kil-kor* type system which has been proposed for all the conjugative *Streptomyces* plasmids studied in detail (pIJ101, pSAM2, pSN22). Therefore it is not known whether such a regulatory system exists in SCP2*.

This work has identified the presence of three proteins all proposed as being involved in the process of plasmid transfer. All the proteins require further investigation. Only 3.3kb of DNA sequence is presented in this

report. This is approximately 1/3 of the region proposed as being involved with inter/intramycelial plasmid transfer. It would thus be necessary to obtain the complete DNA sequence of the entire region in order to be able to investigate the plasmid transfer mechanisms in full.

7.7 Concluding remarks.

The main objectives set out at the beginning of this investigation have been largely fulfilled. A critical part (3.3kb) of the region involved in plasmid transfer in SCP2* has been sequenced and the primary sequence analysed for open reading frames. Sequencing discrepancies have also been identified and studied, suggesting the presence of a different protein to that published. Three putative genes have been identified, the encoded products of which are likely to play a major part in plasmid transfer. Some possible functions have also been hypothesised for some of the encoded protein products after analysis of the predicted amino acid sequences.

The mutagenesis studies indicated that the transfer properties of SCP2* were similar to the other *Streptomyces* plasmids studied, however this work suggests that the transfer genes themselves are very different to those of other *Streptomyces* plasmids.

REFERENCES

Akagawa, H. (1987) Characterization and localization of plasmid functions involved in pock formation and pock resistance of plasmid pSK3* of *Streptomyces kasugaensis* MB273. J. Gen. Microbiol. **133**:1951-1958.

Aksoy, S., Squires, C.L., and Squires, C. (1984) Translational coupling of the *trpB* and *trpA* genes in the *Escherichia coli* tryptophan operon. J. Bact. **157**:363-367.

Alberti, S., Oehler, S., Wilcken-Bergmann, B.V., Krämer, H., and Müller-Hill, B. (1991) Dimmer-to-tetramer assembly of lac repressor involves a leucine heptad repeat. New Biol. **3**:57-62.

Atkins, J.F., Nichols, B.P., and Thompson, S. (1983) The nucleotide sequence of the first externally suppressible - 1 frameshift mutant, and of some nearby leaky frameshift mutants. EMBO **2**(8):1345-1350.

Bairoch, A. (1992) Prosite: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **20**:2013-2018.

Bankier, A.T., Weston, K.M., and Barrell, B.G. (1987) Random cloning and sequencing by the M13/Dideoxynucleotide chain termination method. Methods in Enzymology **155**:51-93.

Bérdy, J. (1980) Recent advances in and prospects of antibiotic research. Process Biochem. **Oct/Nov**:28-35.

Bernan, V., Filpula, D., Herber, W., Bibb, M., and Katz, E. (1985) The nucleotide sequence of the tyrosinase gene from *Streptomyces antibioticus* and characterization of the gene product. Gene **37**:101-110.

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

Bibb, M.J., Findlay, P.R., and Johnson, M.W. (1984) The relationship between base composition and codon usage in bacterial genes and its use for

the simple and reliable identification of protein-coding sequences. Gene **30**:157-166.

Bibb, M.J., Freeman, R.F., and Hopwood, D.A. (1977) Physical and genetical characterisation of a second sex factor, SCP2 from *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. **154**:155-166.

Bibb, M.J., and Hopwood, D.A. (1981) Genetic studies of the fertility plasmid SCP2 and its SCP2* variants in *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. **126**:427-442.

Bibb, M.J., Schottel, J.L., and Cohen, S.N. (1980) A DNA cloning system for interspecies gene transfer in antibiotic-producing *Streptomyces*. Nature **284**:526-531.

Bibb, M.J., Ward, J.M., and Hopwood, D.A. (1978) Transformation of plasmid DNA into *Streptomyces* at high frequency. Nature **274**:398-400.

Bibb, M.J., Ward, J.M., Kieser, T., Cohen, S.N., and Hopwood, D.A. (1981) Excision of chromosomal DNA sequences from *Streptomyces coelicolor* forms a novel family of plasmids detectable in *Streptomyces lividans*. Mol. Gen. Genet. **184**:230-240.

Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513-1523.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.

Bradshaw, H.D., Traxler, B.A., Minkley, E.G., Nester, E.W., and Gordon, M.P. (1990) Nucleotide sequence of the *traI* (Helicase I) gene from the sex factor F. J. Bact. **172**:4127-4131.

Brierley, I., Digard, P., and Inglis S.C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. Cell **57**:537-547.

Brierley, I., Rolley, N.J., Jenner, A.J., and Inglis, S.C. (1991) Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. J. Mol.Biol. **220**:889-902.

Brolle, D.F., Pape, H., Hopwood, D.A., and Kieser, T. (1993) Analysis of the transfer region of the *Streptomyces* plasmid SCP2*. Mol. Microbiol. **10**:157-170.

Busch, S.J., and Sassone-Corsi, P. (1990) Dimers, leucine zippers and DNA-binding domains. Trends in Genetics. **6**:36-40.

Buttner, M.J., and Brown, N.L. (1987) Two promoters from the *Streptomyces* plasmid pIJ101 and their expression in *Escherichia coli*. Gene **51**:179-186.

Buttner, M.J., Fearnley, I.M., and Bibb, M.J. (1987) The agarase gene *(dagA)* of *Streptomyces coelicolor* A3(2):nucleotide sequence and transcriptional analysis. Mol. Gen. Genet. **209**:101-109.

Chater, K.F. (1984) Morphological and physiological differentiation in *Streptomyces* in Microbial Development. Losick, R. and Shapiro, L. (ed.).Coldspring Harbour, New York:Cold Spring Harbour Laboratory Press. pp. 89-115.

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

Chater, K.F., and Hopwood, D.A. (1983) *Streptomyces* genetics in the Biology of Actinomycetes Goodfellow, Mordanski, Williams. (ed) Academic Press London pp. 229-286.

Chater, K.F., Hopwood, D.A., Kieser, T., and Thompson, C.J. (1982) Gene cloning in *Streptomyces*.. Current Topics in Microbiol. and Immunol. **96**:69-95.

Chater, K.F., and Merrick, M.J. (1979) Streptomycetes in Developmental Biology of Prokayotes. Parish J H (eds). Blackwell.pp93-114.

Chen, E.Y., and Seeburg, P.H. (1985) Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA **4** :165-170.

Chowdhury, K. (1991) One step 'miniprep' method for the isolation of plasmid DNA. Nucleic Acids Res. **19** :2792.

Clewell, D.B. (1993) Sex pheromones and the plasmid-encoded mating response in *Enterococcus faecalis* In Bacterial Conjugation. Clewell, D.B. (ed.). New York: Plenum Press, pp. 23-52.

Clewell, D.B. (1981) Plasmids, Drug resistance, and gene transfer in the genus Streptococcus. Microbiol. Rev. **45**:409-429.

Das, A., and Yanofsky, C. (1984) A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. Nucleic Acids Res. **12**:4757-4768.

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

Diano, M., Le Bivic, A., and Hirn, M. (1987) A method for the production of highly specific polyclonal antibodies. Anal. Biochem. **166**:224-229.

Dunny, G.M. (1990) Genetic functions and cell-cell interactions in the pheromone-inducible plasmid transfer system of *Enterococcus faecalis*. Mol. Microbiol. **4**:689-696.

Firth, N., Ridgway, K.P., Byrne, M.E., Fink, P.D., Johnson, L., Paulsen, I.T., and Skurray, R.A. (1993) Analysis of a transfer region from the staphylococcal conjugative plasmid pSK41. Gene **136**:13-25.

Fujita, Y., Fujita, T., Miwa Y., Nihashi, J., and Aratani, Y. (1986) Organization and transcription of the gluconate operon, *gnt*, of *Bacillus subtilis*. J. Biol. Chem. **261**:13744-13753.

Garcia, A., Duin, J.V., Pleij, C.W.A. (1993) Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. Nucleic Acids Res. **21**(3):401-406.

Giraldo, R., Nieto, C., Fernandez-Tresguerres, M., and Diaz, R. (1989) Bacterial zipper. Nature **342**:866.

Hagège, J., Pernodet, J., Sezonov, G., Gerbaud, C., Friedmann, A., and Guérineau, M. (1993) Transfer functions of the conjugative integrating element of pSAM2 from *Streptomyces ambofaciens*: Characterization a *kil-kor* system associated with transfer. J. Bact. **175**:5529-5538.

Harlow, E., and Lane, D. (1988) Antibodies: A laboratory manual. Coldspring Harbour, New York:Cold Spring Harbour Laboratory Press.

Harrison, S.C. (1991) A structural taxonomy of DNA-binding domains. Nature **353**:715-719.

Hattori, M., and Sakaki, Y. (1986) Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**:232-238.

Henninghausen, L., and Lubon, H. (1987) Interaction of protein with DNA *in vitro*. Methds in Enzymology **152**:721-735.

Hodgman, T.C. (1988) A new superfamily of replicative proteins. Nature **333**:22-23.

Holloway, B.W. (1979) Plasmids that mobilize the bacterial chromosome. Plasmid **2**:1-19.

Hopwood, D.A., Bibb, M.J., Bruton, C.J., Chater, K.F., Feitelson, J.S., and Gil, J.A. (1983) Cloning *Streptomyces* genes for antibiotic production. Trends in Biotech. **1**:42-48.

Hopwood, D.A., Lydiate, D.J., Malpartida, F., and Wright, H.M. (1985a) Conjugative sex plasmids of *Streptomyces* in Plasmids and Bacteria. Helinski, D.R., Cohen, S.N., Clewell, D.B., Jackson, D.A., and Hollaender, A. (eds.) Plnum Press, New York and London pp. 615-634.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M., and Schrempf, H. (1985b) Genetic manipulation of *Streptomyces*: A Laboratory Manual, Norwich: The John Innes Foundation.

Hopwood, D.A., Bibb, M.J., Chater, K.F., and Kieser, T. (1987) Plasmid and phage vectors for gene cloning and analysis in *Streptomyces*. Methods in Enzymology **153**:116-166.

Hopwood, D.A., Kieser, T., Wright, H.M., and Bibb, M.J. (1983) Plasmids, recombination and chromosome mapping in *Streptomyces lividans* 66. J. Gen. Microbiol. **129**:2257-2269.

Hopwood, D.A., Kieser, T., Lydiate, D.J., and Bibb, M.J. (1986) *Streptomyces* plasmids: Their biology and use as cloning vectors. The Bacteria. A Treatise on Structure and Function. Vol. IX. Queener, S.W., and Day, L.E. (eds). Orlando: Academic Press, pp. 159-229.

Hopwood, D.A., and Kieser, T. (1993) Conjugative plasmids in *Streptomyces*. In Bacterial Conjugation. Clewell, D.B. (ed.). New York: Plenum Press, pp. 293-311.

Ippen-Ihler, K., and Skurray, R.A. (1993) Genetic organization of transfer-related determinants of the sex factor F and related plasmids In Bacterial Conjugation. Clewell, D.B. (ed.). New York: Plenum Press, pp. 23-52.

Johnson, P.F., and McKnight, S.L. (1989) Eukaryotic transcriptional regulatoty proteins. Annu. Rev. Biochem. **58**:799-839.

Kataoka, M., Seki, T., and Yoshida, T. (1991a) Five genes involved in self-transmission of pSN22, a *Streptomyces* plasmid. J. Bact. **173**:4220-4228.

Kataoka, M., Seki, T., and Yoshida, T. (1991b) Regulation and function of the *Streptomyces* plasmid pSN22 genes involved in pock formation and inviability. J. Bact. **173**:7975-7981.

Kendall, K.J., and Cohen, S.N. (1987) Plasmid transfer in *Streptomyces lividans*: Identification of a *kil-kor* system associated with the transfer region of pIJ101. J. Bact. **169**:4177-4183.

Kendall, K., and Cohen, S.N. (1988) Complete nucleotide sequence of the Streptomyces lividans plasmid pIJ101 and correlation of the sequence with genetic properties. J. Bact. **170**:4634-4651.

Kieser, T. (1984) Factors affecting theisolation of CCC DNA from *Streptomyces lividans* and *Escherichia coli*. Plasmid **12**:19-36.

Kieser, T., and Hopwood, D.A. (1991) Genetic manipulation of *Streptomyces*: Integrating vectors and gene replacement. Methods in Enzymology **20**:430-458.

Kieser, T., Hopwood, D.A., Wright, H.M., and Thompson, C.J. (1982) pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: Functional analysis and development of DNA cloning vectors. Mol. Gen. Genet. **185**:223-238.

Kinashi, K., Shimaji, M., and Sakai, A. (1987) Giant linear plamids in *Streptomyces* which code for antibiotic biosyntesis genes. Nature **328**:454-456.

Kinashi, H., and Shimaji-Murayama, M. (1991) Physical characterization of SCP1, a giant linear plasmid from*Streptomyces coelicolor*. J. Bact. **173**:1523-1529.

Klug, A., and Rhodes, D. (1987) 'Zinc fingers': a novel protein motif for nucleic acid recognition. TIBS **12**:464-469.

Knudsen, K.A. (1985) Proteins transferred to nitrocellulose for use as immunogens. Anal. Biochem. **147**:285-288.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteiophage T4. Nature **227**:680-68.

Landschulz, W.H., Johnson, P.F., McKnight, S.L. (1988) The leucine zipper : A hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759-1764.

Larson, J.L., and Hershberger, C.L. (1986) The minimal replicon of a *Streptomyces* plasmid produces an ultra high level of plasmid DNA. Plasmid **15**:199-209.

Lee, F., Levin, A., and Bronton, D. (1987) Copper staining: A five minute protein stain for STS polyacrylamide gels. Anal. Biochem. **166**:308-312.

Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989) Birth of the D-E-A-D box. Nature **337**:121-122.

Lydiate, D.J., Malpartida, F., and Hopwood, D.A. (1985) The *Streptomyces* plasmid SCP2*: its functional analysis and development into useful cloning vectors. Gene **35**:223-235.

Marck, C. (1988) 'DNA strider': a 'C' program for the fast analysis of DNA protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829-1836.

Maxon, M.E., Wigboldus, J., Brot, N., and Weissbach, H. (1990) Structure-function studies on *Escherichia coli* MetR protein, a putative prokaryotic leucine zipper protein. Proc. Natl. Acad. Sci. USA **87**:7076-7079.

Morrison, D.A. (1979) [21]Transformation and preservation of competent bacterial cells by frezing. Methods in Enzymology **68**:326-331.

Morton, T.M., Eaton, D.M., Johnston, J.L., and Archer, G.L. (1993) DNA sequence and units of transcription of the conjugative transfer gene complex *(trs)* of *Staphylococcus aureus* plasmid pGO1. J. Bact. **175**:4436-4447.

Motallebi-Veshareh, M., Balzer, D., Lanka, E., Jagura-Burdzy, G., and Thomas, C.M. (1992) Conjugative transfer functions of broad-host-range plasmid RK2 are coregulated with vegetative replication. Molecular Microbiology **6**:907-920.

Omer, C.A., and Cohen, S. (1984) Plasmid formation in *Streptomyces*: Excision and integration of theSLP1 replicon at a specific chromosomal site. Mol. Gen. Genet. **196**:429-438.

Oppenheim, D.S., and Yanofsky, C. (1980) Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics **95**:785-795.

Pabo, C., and Sauer, R.T. (1984) Protein-DNA recognition. Ann. Rev. Biochem. **53**:293-321.

Pernodet, J., Simonet, J., and Guérineau, M. (1984) Plasmids in different strains of *Streptomyces ambofaciens:* free and integrated form of plasmid pSAM2. Mol. Gen. Genet. **198**:35-41.

Queen, C., and Korn, L.J. (1984) A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581-599.

Sambrook, J.E., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: A Laboratory Manual. 2nd edn. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463-5467.

Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990) The P-loop - a common motif in ATP- and GTP-binding proteins. Trends in Biochem. Sci. **15**:430-434.

Schrempf, H., and Goebel, W. (1977) Characterization of a plasmid from *Streptomyces coelicolor* A3(2). J. Bact. **131**(1):251-258.

Schümperli, D., McKenney, K., Sobieski, D.A., and Rosenberg, M. (1982) Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. Cell **30**:865-871.

Skurray, R.A., and Reeves, P. (1973) Characterization of lethal zygosis associated with conjugation in *Escherichia coli* K-12. J. Bact. **113**:58-70.

Smokvina, T., Boccard, F., Pernodet, J., Friedmann, A., and Guérineau, M. (1991) Functional analysis of the *Streptomyces ambofaciens* element pSAM2. Plamid **25**:40-52.

Stein, D.S., and Cohen, S.N. (1990) Mutational and functional analysis of the *korA* and *korB* gene products of *Streptomyces* plasmid pIJ101. Mol. Gen. Genet. **222**:337-344.

Stein, D.S., Kendall, K.J., and Cohen, S.N. (1989) Identification and analysis of transcriptional regulatory signals for the *kil* and *kor* loci of *Streptomyces* plasmid pIJ101. J. Bact. **171**:5768-5775.

Strohl, W. (1992) Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res. **20**:961-974.

Studier, F.W., and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**:113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) [6]Use of T7 RNA polymerase to direct expression of cloned genes. Methods in Enzymology **185**:60-89.

Tabor, S. (1990) Expression using the T7 RNA polymerase/promoter system in Current Protoccols in Molecular Biology. Ausubel, F.A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (eds.) Greene Publishing and Wiley-Iterscience, New York, pp. 16.2.1-16.2.11.

Tabor, S., and Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expresion of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074-1078.

Tai, J.T., and Cohen, S.N. (1993) The active form of the *korB* protein encoded by the *Streptomyces* plasmid pIJ101 is a processed product that binds differentially to the two promoters it regulates. J. Bact. **175**:6996-7005.

Thompson, C.J., Kieser, T., Ward, J.M., and Hopwood, D.A. (1982) Physical analysis of antibiotic-resistance genes from *Streptomyces* and their use in vector construction. Gene **20**:51-62.

Thomson, V.J., Jovanovic, O.S., Pohlman, R.F., Chang, C., and Figurski, D.H. (1993) Structure, function, and regulation of the *kilB* locus of promiscuous plasmid RK2. J. Bact. **175**:2423-2435.

Tomich, P.K. (1988) *Streptomyces* cloning: Useful recombinant DNA systems and a summation of cloned genes. Antimirobial Agents and Chemo. **32**(10):1465-1471, and*Streptomyces* cloning: Possible construcion of novel

compounds and regulation of antibiotic biosynthetic genes. Antimicrobial Agents and Chemo. **32**(10):1472-1476.

Tsuchihashi, Z., and Kornberg, A. (1990) Translational frameshifting generates the γ subunit of DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. USA **87**:2516-2520.

Tu, C., Tzebg, T., and Bruenn, J.A. (1992) Ribosomal movement impeded at a pseudoknot required for frameshifting. Proc. Natl. Acad. Sci. USA **89**:8636-8640.

Varmus, H. (1988) Retroviruses. Science **240**:1427-1435.

Vinson, C.R., Sigler, P.B. and McKnight, S.L. (1989) Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science **246**: 911-916.

Vivian, A. (1971) Genetic control of Fertility in *Streptomyces coelicolor* A3(2): Plasmid involvement in the interconversion of UF and IF strains. J. Gen. Microbiol. **69**:353-364.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO **1**:945-951.

Willetts, N. (1988) Conjugation. Methods in Microbiol. **21**:49-77.

Willetts, N., and Skurray, R. (1987) Structure and function of the F factor and mechanism of conjugation. In: Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., Umbarger, H.E., (eds) *Escherichia coli* and *Slamonellla typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C., pp.1110-1133.

Yanisch-Peron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUCI9 vectors. Gene **33**:103-119.

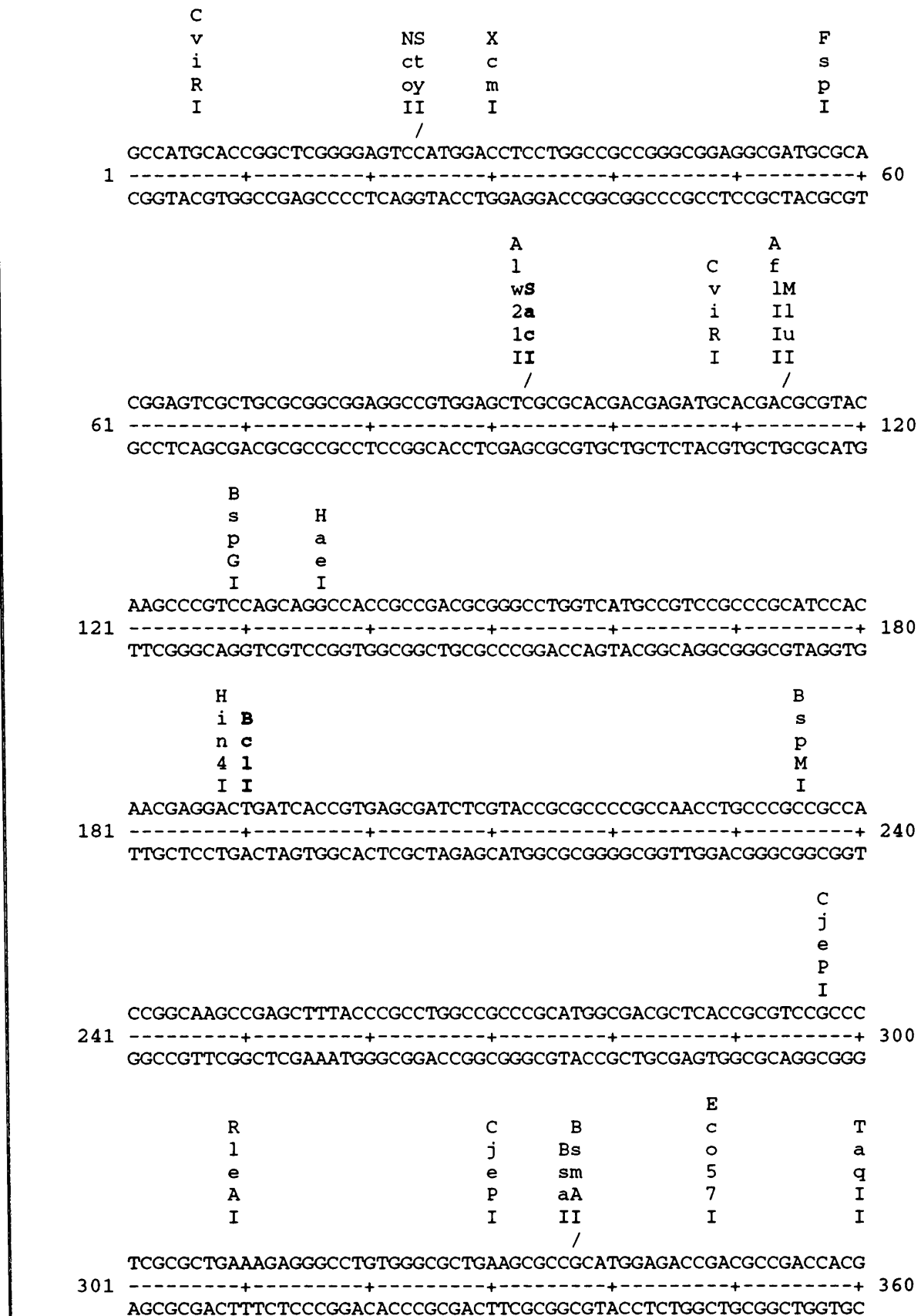
Yoshioka, Y., Fujita, Y., and Ohtsubo, E. (1990) Nucleotide sequence of the promoter-distal region of the *tra* operon plasmid R100, including *traI* (DNA Helicase I) and *traD* genes. J. Mol. Biol. **214**:39-53.

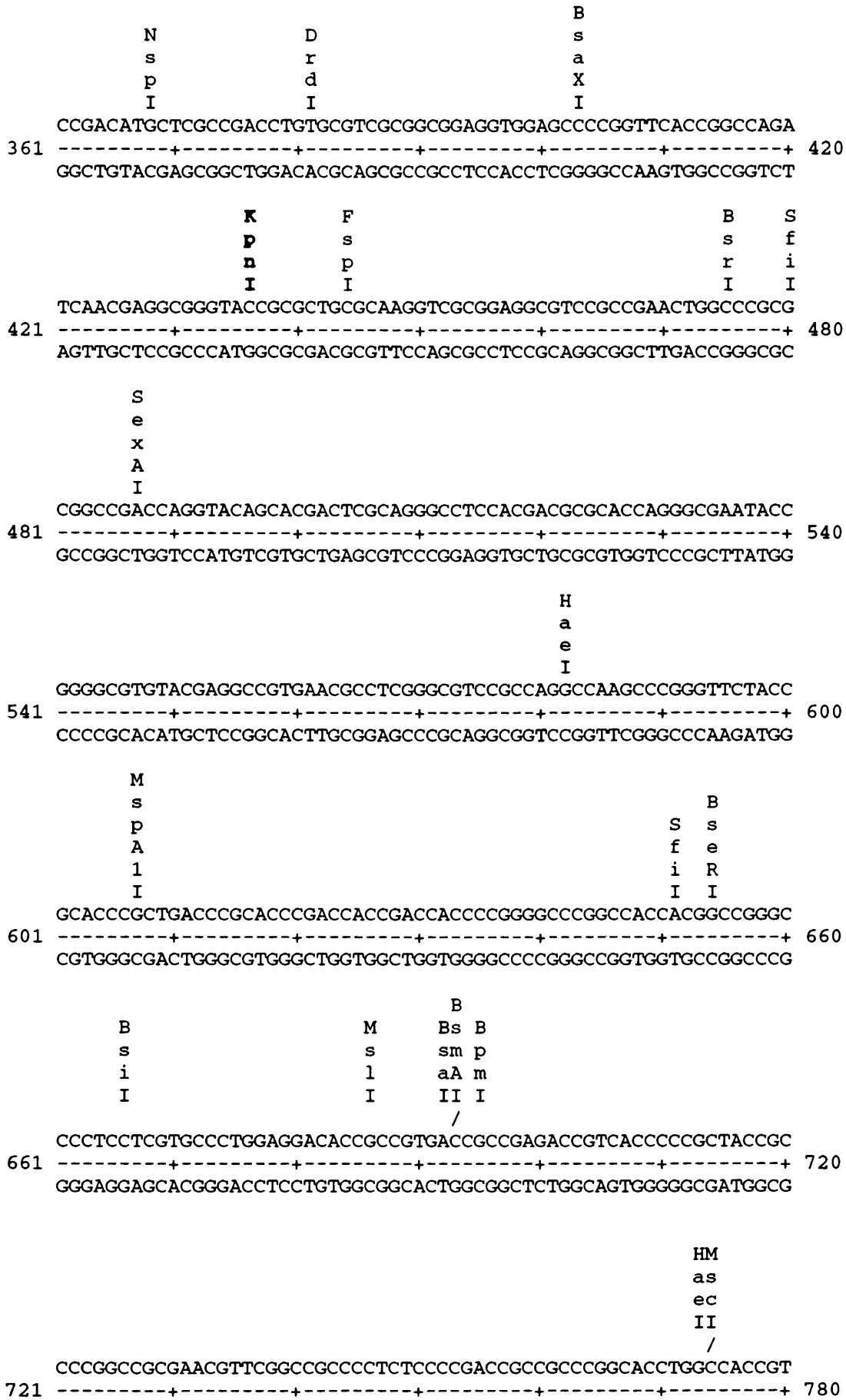
Zaman, S., Richards, H., and Ward, J. (1992) Expression and characterisation of the *korB* gene product from the *Streptomyces lividans* plasmid pIJ101 in *Escherichia coli* and determination of its binding site on the *korB* and *kilB* promoters. Nucleic Acids Res. **20** :3693-3700.

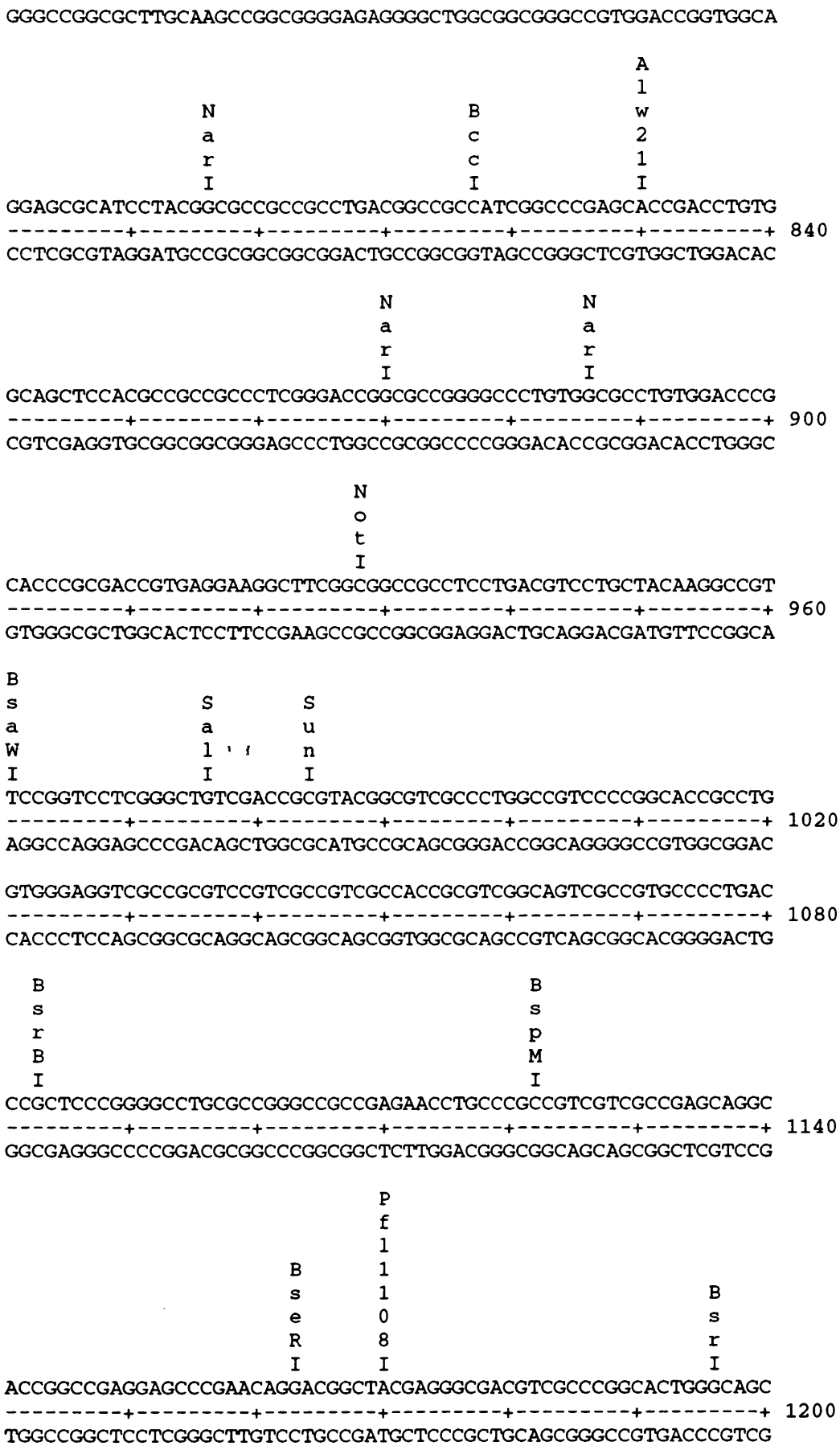
Zotchev, S.B., and Schrempf, H. (1994) The linear *Streptomyces* plasmid pBL1: analyses of transfer functions. Mol. Gen. Genet. **242**:374-382.

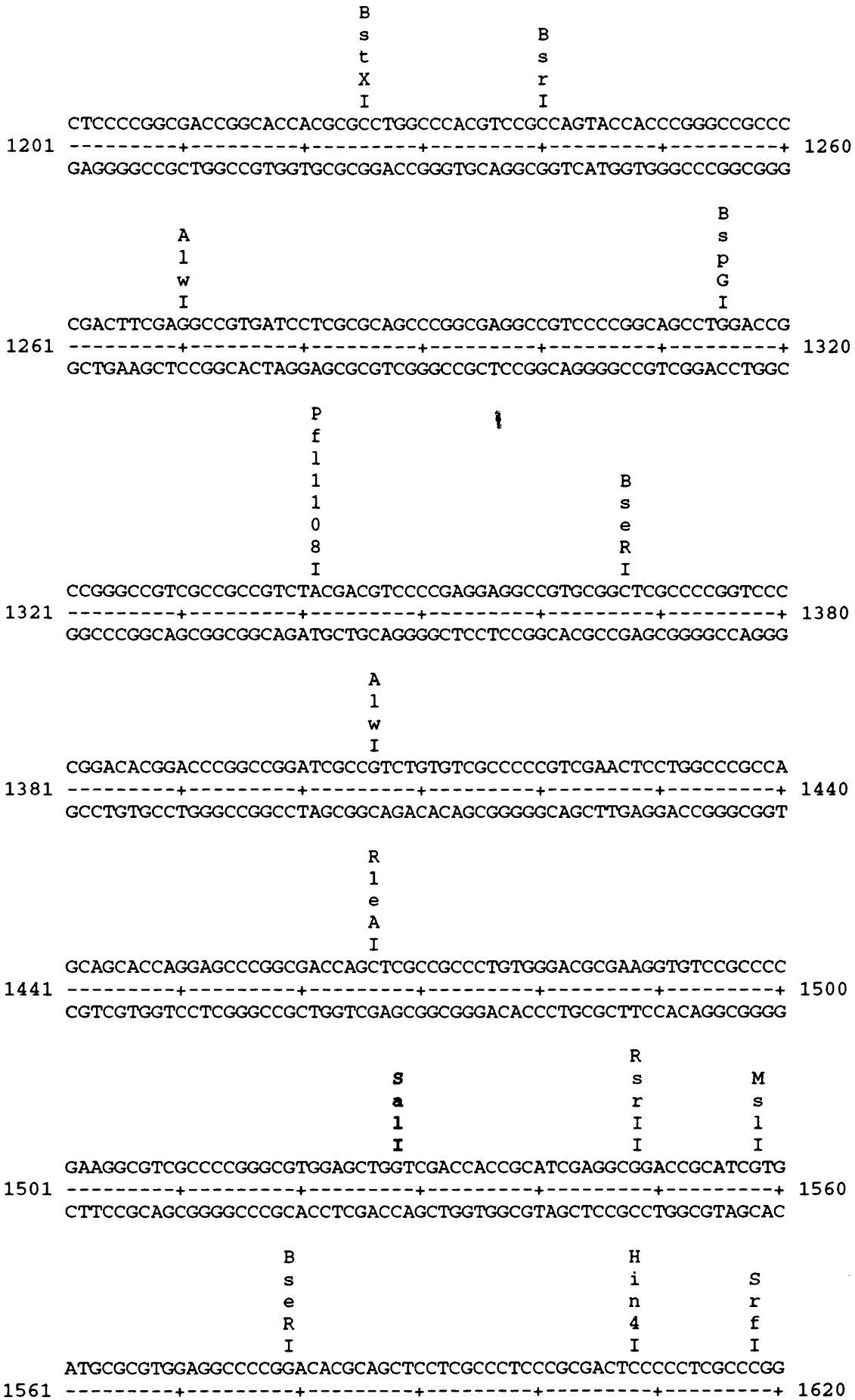
Zubay, G. (1973) *In vitro* synthesis of protein in microbial systems. Ann. Rev. Genet. **7**:267-287.

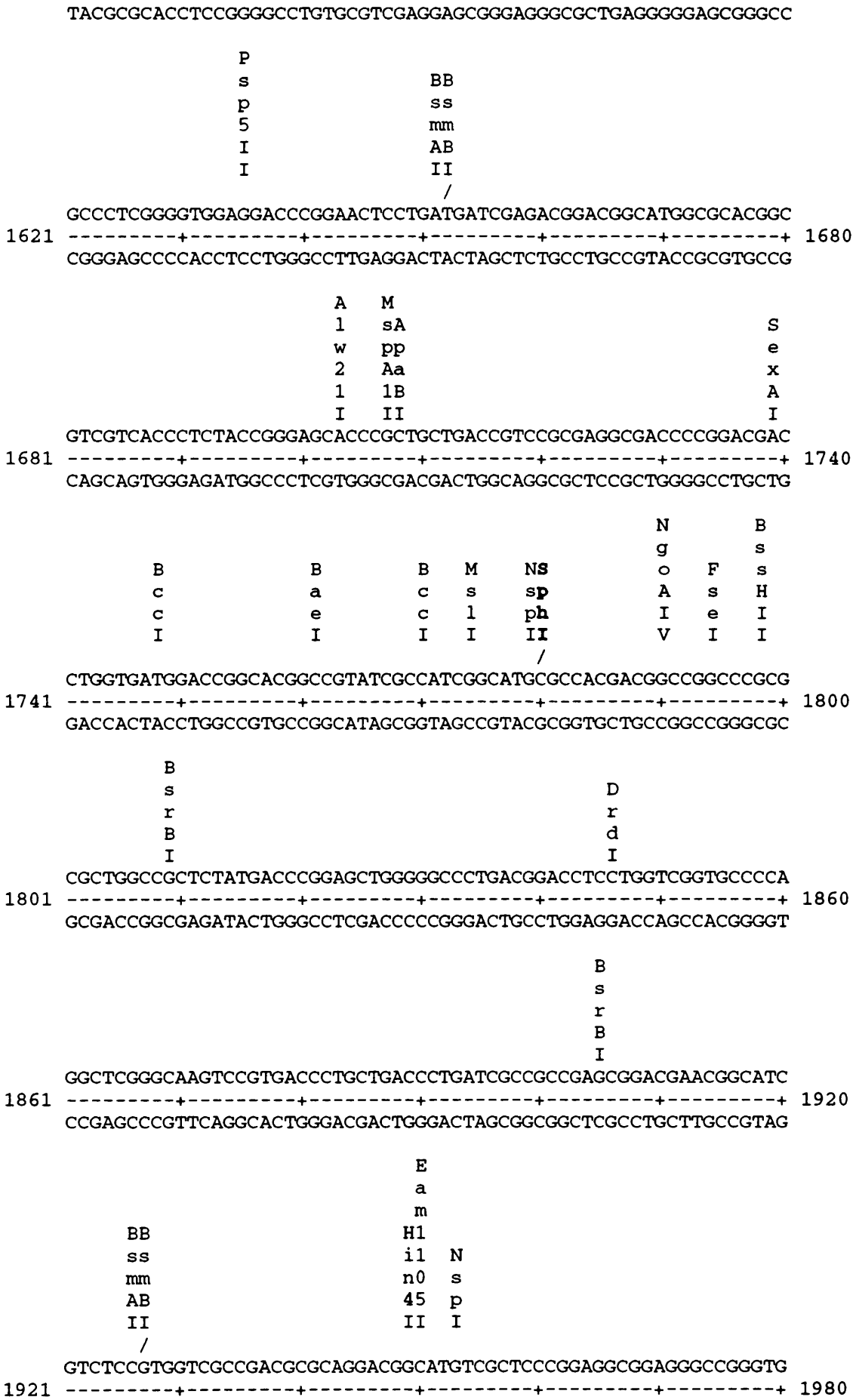
APPENDIX A Restriction Enzyme map of the SCP2* transfer region.

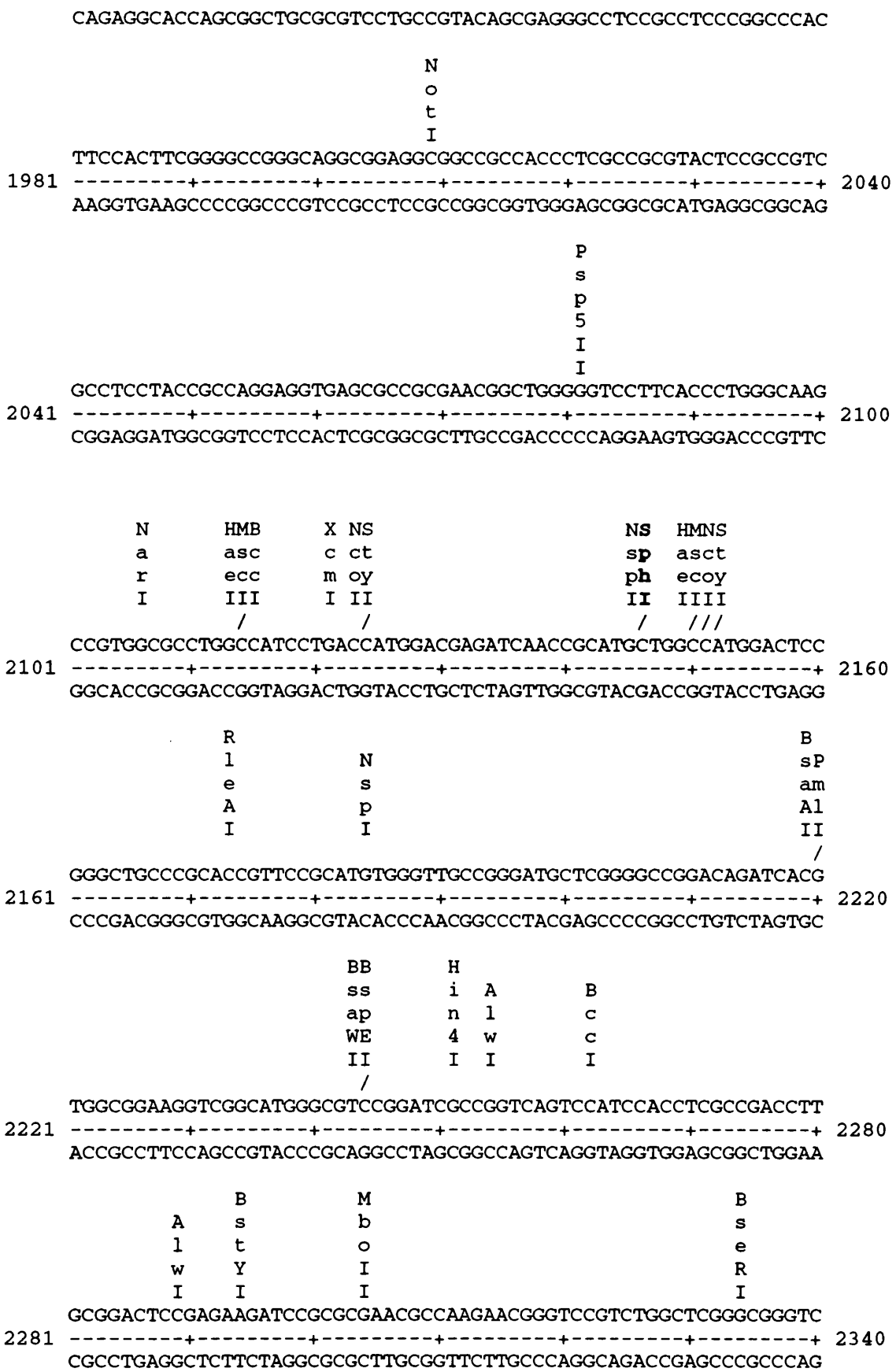


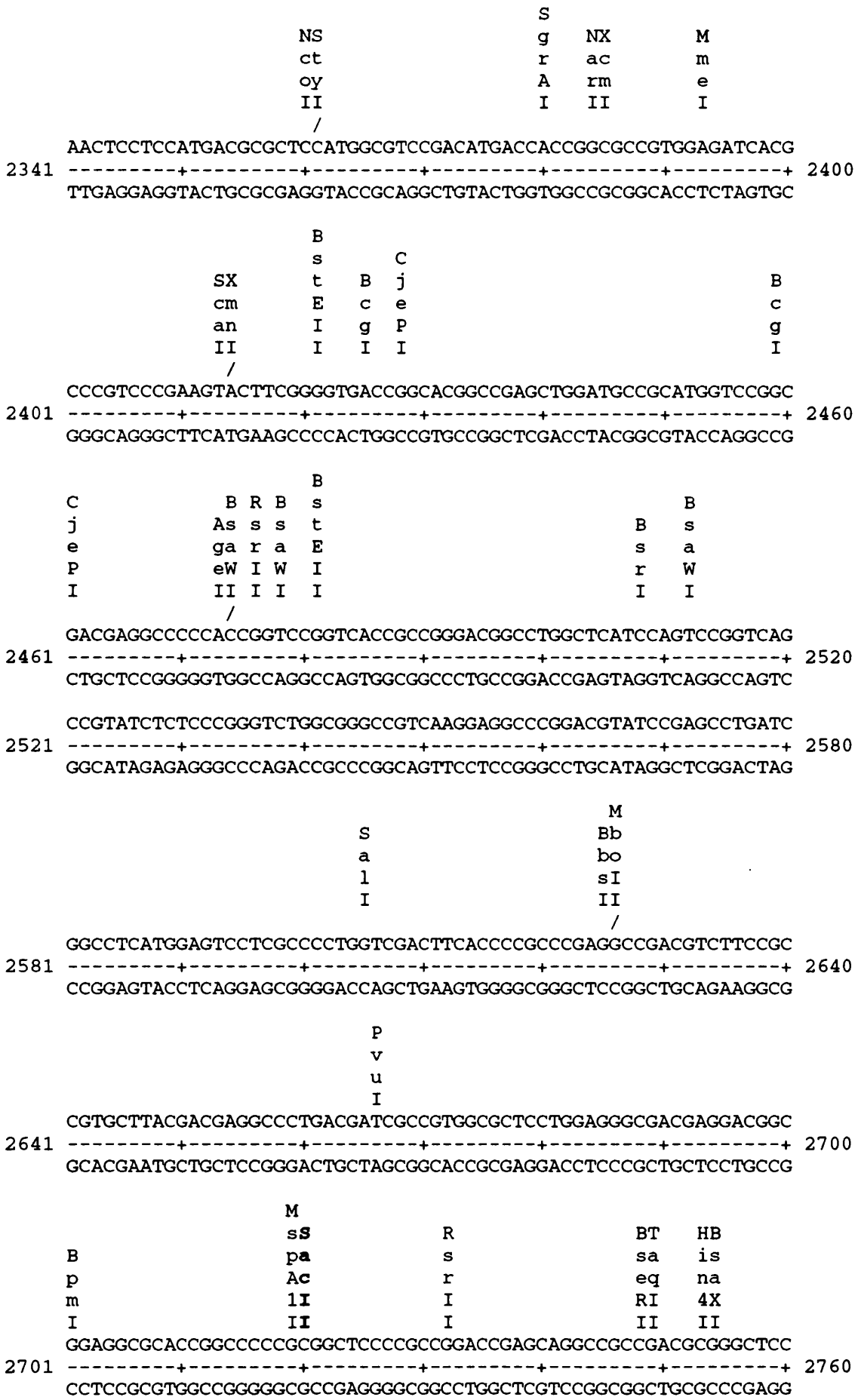

B
N s
D r
s a
p d
X
I I I
CCGACATGCTCGCCGACCTGTGCGTCGCGGCGGAGGTGGAGCCCCGGTTCACCGGCCAGA
361 ---------+---------+---------+---------+---------+---------+ 420
GGCTGTACGAGCGGCTGGACACGCAGCGCCGCCTCCACCTCGGGGCCAAGTGGCCGGTCT
K F B S
p s s f
n p r i
I I I I
TCAACGAGGCGGGTACCGCGCTGCGCAAGGTCGCGGAGGCGTCCGCCGAACTGGCCCGCG
421 ---------+---------+---------+---------+---------+---------+ 480
AGTTGCTCCGCCCATGGCGCGACGCGTTCCAGCGCCTCCGCAGGCGGCTTGACCGGGCGC
S
e
x
A
I
CGGCCGACCAGGTACAGCACGACTCGCAGGGCCTCCACGACGCGCACCAGGGCGAATACC
481 ---------+---------+---------+---------+---------+---------+ 540
GCCGGCTGGTCCATGTCGTGCTGAGCGTCCCGGAGGTGCTGCGCGTGGTCCCGCTTATGG
H
a
e
I
GGGGCGTGTACGAGGCCGTGAACGCCTCGGGCGTCCGCCAGGCCAAGCCCGGGTTCTACC
541 ---------+---------+---------+---------+---------+---------+ 600
CCCCGCACATGCTCCGGCACTTGCGGAGCCCGCAGGCGGTCCGGTTCGGGCCCAAGATGG
M
s
p
A
1
I
B
S s
f e
i R
I I
GCACCCGCTGACCCGCACCCGACCACCGACCACCCCGGGGCCCGGCCACCACGGCCGGGC
601 ---------+---------+---------+---------+---------+---------+ 660
CGTGGGCGACTGGGCGTGGGCTGGTGGCTGGTGGGGCCCCGGGCCGGTGGTGCCGGCCCG
B
B M Bs B
s s sm p
i l aA m
I I II I
/
CCCTCCTCGTGCCCTGGAGGACACCGCCGTGACCGCCGAGACCGTCACCCCCGCTACCGC
661 ---------+---------+---------+---------+---------+---------+ 720
GGGAGGAGCACGGGACCTCCTGTGGCGGCACTGGCGGCTCTGGCAGTGGGGGCGATGGCG
HM
as
ec
II
/
CCCGGCCGCGAACGTTCGGCCGCCCCTCTCCCCGACCGCCGCCCGGCACCTGGCCACCGT
721 ---------+---------+---------+---------+---------+---------+ 780

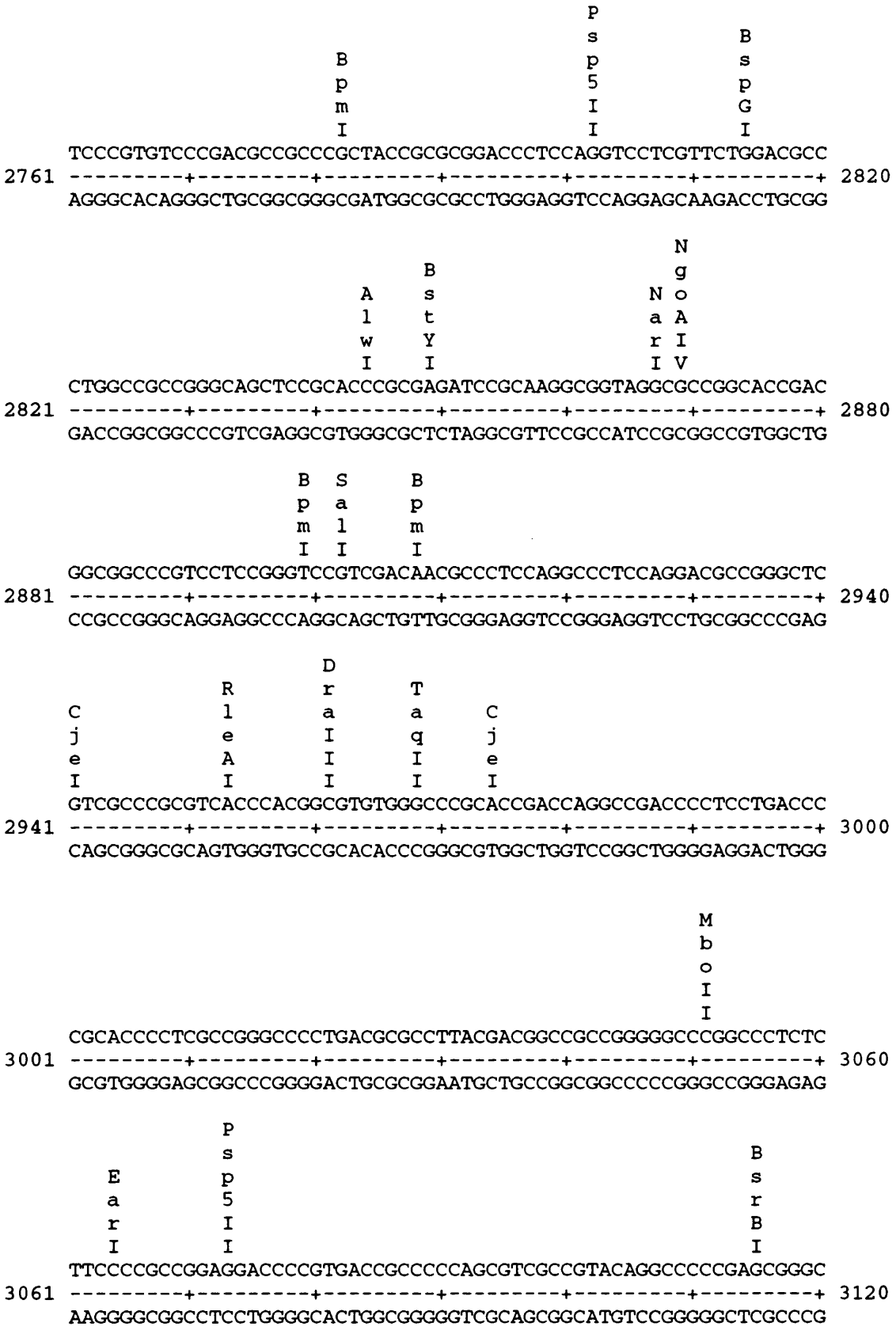

GGGCCGGCGCTTGCAAGCCGGCGGGGAGAGGGGCTGGCGGCGGGCCGTGGACCGGTGGCA
A
1
N B w
a c 2
r c 1
I I I
GGAGCGCATCCTACGGCGCCGCCGCCTGACGGCCGCCATCGGCCCGAGCACCGACCTGTG
781 840
CCTCGCGTAGGATGCCGCGGCGGCGGACTGCCGGCGGTAGCCGGGCTCGTGGCTGGACAC
N N
a a
r r
I I
GCAGCTCCACGCCGCCGCCCTCGGGACCGGCGCCGGGGCCCTGTGGCGCCTGTGGACCCG
841 900
CGTCGAGGTGCGGCGGCGGGAGCCCTGGCCGCGGCCCCGGGACACCGCGGACACCTGGGC
N
o
t
I
CACCCGCGACCGTGAGGAAGGCTTCGGCGGCCGCCTCCTGACGTCCTGCTACAAGGCCGT
901 960
GTGGGCGCTGGCACTCCTTCCGAAGCCGCCGGCGGAGGACTGCAGGACGATGTTCCGGCA
B
s S S
a u
W l n
I I I
TCCGGTCCTCGGGCTGTCGACCGCGTACGGCGTCGCCCTGGCCGTCCCCGGCACCGCCTG
961 1020
AGGCCAGGAGCCCGACAGCTGGCGCATGCCGCAGCGGGACCGGCAGGGGCCGTGGCGGAC
GTGGGAGGTCGCCGCGTCCGTCGCCGTCGCCACCGCGTCGGCAGTCGCCGTGCCCCTGAC
1021 1080
CACCCTCCAGCGGCGCAGGCAGCGGCAGCGGTGGCGCAGCCGTCAGCGGCACGGGGACTG
B B
s s
r p
B M
I I
CCGCTCCCGGGGCCTGCGCCGGGCCGCCGAGAACCTGCCCGCCGTCGTCGCCGAGCAGGC
1081 1140
GGCGAGGGCCCCGGACGCGGCCCGGCGGCTCTTGGACGGGCGGCAGCAGCGGCTCGTCCG
P
f
l
B 1
s 1 B
e 0 s
R 8 r
I I I
ACCGGCCGAGGAGCCCGAACAGGACGGCTACGAGGGCGACGTCGCCCGGCACTGGGCAGC
1141 1200
TGGCCGGCTCCTCGGGCTTGTCCTGCCGATGCTCCCGCTGCAGCGGGCCGTGACCCGTCG


BstXI
BsrI
CTCCCCGGCGACCGGCACCACGCGCCTGGCCCACGTCCGCCAGTACCACCCGGGCCGCCC
1201 ---------+---------+---------+---------+---------+---------+ 1260
GAGGGGCCGCTGGCCGTGGTGCGCGGACCGGGTGCAGGCGGTCATGGTGGGCCCGGCGGG
AlwI
BspGI
CGACTTCGAGGCCGTGATCCTCGCGCAGCCCGGCGAGGCCGTCCCCGGCAGCCTGGACCG
1261 ---------+---------+---------+---------+---------+---------+ 1320
GCTGAAGCTCCGGCACTAGGAGCGCGTCGGGCCGCTCCGGCAGGGGCCGTCGGACCTGGC
Pfl1108I
BseRI
CCGGGCCGTCGCCGCCGTCTACGACGTCCCCGAGGAGGCCGTGCGGCTCGCCCCGGTCCC
1321 ---------+---------+---------+---------+---------+---------+ 1380
GGCCCGGCAGCGGCGGCAGATGCTGCAGGGGCTCCTCCGGCACGCCGAGCGGGGCCAGGG
AlwI
CGGACACGGACCCGGCCGGATCGCCGTCTGTGTCGCCCCCGTCGAACTCCTGGCCCGCCA
1381 ---------+---------+---------+---------+---------+---------+ 1440
GCCTGTGCCTGGGCCGGCCTAGCGGCAGACACAGCGGGGGCAGCTTGAGGACCGGGCGGT
RleAI
GCAGCACCAGGAGCCCGGCGACCAGCTCGCCGCCCTGTGGGACGCGAAGGTGTCCGCCCC
1441 ---------+---------+---------+---------+---------+---------+ 1500
CGTCGTGGTCCTCGGGCCGCTGGTCGAGCGGCGGGACACCCTGCGCTTCCACAGGCGGGG
SalI
RsrII
MslI
GAAGGCGTCGCCCCGGGCGTGGAGCTGGTCGACCACCGCATCGAGGCGGACCGCATCGTG
1501 ---------+---------+---------+---------+---------+---------+ 1560
CTTCCGCAGCGGGGCCCGCACCTCGACCAGCTGGTGGCGTAGCTCCGCCTGGCGTAGCAC
BseRI
Hin4I
SrfI
ATGCGCGTGGAGGCCCCGGACACGCAGCTCCTCGCCCTCCCGCGACTCCCCCTCGCCCGG
1561 ---------+---------+---------+---------+---------+---------+ 1620


TACGCGCACCTCCGGGGCCTGTGCGTCGAGGAGCGGGAGGGCGCTGAGGGGGAGCGGGCC
PspI
BBssmmABII
GCCCTCGGGGTGGAGGACCCGGAACTCCTGATGATCGAGACGGACGGCATGGCGCACGGC
1621 ---------+---------+---------+---------+---------+---------+ 1680
CGGGAGCCCCACCTCCTGGGCCTTGAGGACTACTAGCTCTGCCTGCCGTACCGCGTGCCG
Alw21I
AMsApp Aa1B II
SexAI
GTCGTCACCCTCTACCGGGAGCACCCGCTGCTGACCGTCCGCGAGGCGACCCCGGACGAC
1681 ---------+---------+---------+---------+---------+---------+ 1740
CAGCAGTGGGAGATGGCCCTCGTGGGCGACGACTGGCAGGCGCTCCGCTGGGGCCTGCTG
BccI
BaeI
BccI
MslI
NSspphII
NgoAIV
FseI
BsaHI
CTGGTGATGGACCGGCACGGCCGTATCGCCATCGGCATGCGCCACGACGGCCGGCCCGCG
1741 ---------+---------+---------+---------+---------+---------+ 1800
GACCACTACCTGGCCGTGCCGGCATAGCGGTAGCCGTACGCGGTGCTGCCGGCCGGGCGC
BsrBI
DrdI
CGCTGGCCGCTCTATGACCCGGAGCTGGGGGCCCTGACGGACCTCCTGGTCGGTGCCCCA
1801 ---------+---------+---------+---------+---------+---------+ 1860
GCGACCGGCGAGATACTGGGCCTCGACCCCCGGGACTGCCTGGAGGACCAGCCACGGGGT
BsrBI
GGCTCGGGCAAGTCCGTGACCCTGCTGACCCTGATCGCCGCCGAGCGGACGAACGGCATC
1861 ---------+---------+---------+---------+---------+---------+ 1920
CCGAGCCCGTTCAGGCACTGGGACGACTGGGACTAGCGGCGGCTCGCCTGCTTGCCGTAG
Eam
BBssmmABII
H1i1n045II
NspI
GTCTCCGTGGTCGCCGACGCGCAGGACGGCATGTCGCTCCCGGAGGCGGAGGGCCGGGTG
1921 ---------+---------+---------+---------+---------+---------+ 1980


CAGAGGCACCAGCGGCTGCGCGTCCTGCCGTACAGCGAGGGCCTCCGCCTCCCGGCCCAC
N
o
t
I
TTCCACTTCGGGGCCGGGCAGGCGGAGGCGGCCGCCACCCTCGCCGCGTACTCCGCCGTC
1981 ---------+---------+---------+---------+---------+---------+ 2040
AAGGTGAAGCCCCGGCCCGTCCGCCTCCGCCGGCGGTGGGAGCGGCGCATGAGGCGGCAG
p
s
p
5
I
I
GCCTCCTACCGCCAGGAGGTGAGCGCCGCGAACGGCTGGGGGTCCTTCACCCTGGGCAAG
2041 ---------+---------+---------+---------+---------+---------+ 2100
CGGAGGATGGCGGTCCTCCACTCGCGGCGCTTGCCGACCCCCAGGAAGTGGGACCCGTTC
N HMB X NS NS HMNS
a asc c ct sp asct
r ecc m oy ph ecoy
I III I II II IIII
/ / / ///
CCGTGGCGCCTGGCCATCCTGACCATGGACGAGATCAACCGCATGCTGGCCATGGACTCC
2101 ---------+---------+---------+---------+---------+---------+ 2160
GGCACCGCGGACCGGTAGGACTGGTACCTGCTCTAGTTGGCGTACGACCGGTACCTGAGG
R B
l N sP
e s am
A p Al
I I II
/
GGGCTGCCCGCACCGTTCCGCATGTGGGTTGCCGGGATGCTCGGGGCCGGACAGATCACG
2161 ---------+---------+---------+---------+---------+---------+ 2220
CCCGACGGGCGTGGCAAGGCGTACACCCAACGGCCCTACGAGCCCCGGCCTGTCTAGTGC
BB H
ss i A B
ap n l c
WE 4 w c
II I I I
/
TGGCGGAAGGTCGGCATGGGCGTCCGGATCGCCGGTCAGTCCATCCACCTCGCCGACCTT
2221 ---------+---------+---------+---------+---------+---------+ 2280
ACCGCCTTCCAGCCGTACCCGCAGGCCTAGCGGCCAGTCAGGTAGGTGGAGCGGCTGGAA
B M B
A s b s
l t o e
w Y I R
I I I I
GCGGACTCCGAGAAGATCCGCGCGAACGCCAAGAACGGGTCCGTCTGGCTCGGGCGGGTC
2281 ---------+---------+---------+---------+---------+---------+ 2340
CGCCTGAGGCTCTTCTAGGCGCGCTTGCGGTTCTTGCCCAGGCAGACCGAGCCCGCCCAG


S
NS g NX M
ct r ac m
oy A rm e
II I II I
/
AACTCCTCCATGACGCGCTCCATGGCGTCCGACATGACCACCGGCGCCGTGGAGATCACG
2341 ---------+---------+---------+---------+---------+---------+ 2400
TTGAGGAGGTACTGCGCGAGGTACCGCAGGCTGTACTGGTGGCCGCGGCACCTCTAGTGC
B
s C
SX t B j B
cm E c e c
an I g P g
II I I I I
/
CCCGTCCCGAAGTACTTCGGGGTGACCGGCACGGCCGAGCTGGATGCCGCATGGTCCGGC
2401 ---------+---------+---------+---------+---------+---------+ 2460
GGGCAGGGCTTCATGAAGCCCCACTGGCCGTGCCGGCTCGACCTACGGCGTACCAGGCCG
B
C B R B s B
j As s s t B s
e ga r a E s a
P eW I W I r W
I II I I I I I
/
GACGAGGCCCCCACCGGTCCGGTCACCGCCGGGACGGCCTGGCTCATCCAGTCCGGTCAG
2461 ---------+---------+---------+---------+---------+---------+ 2520
CTGCTCCGGGGGTGGCCAGGCCAGTGGCGGCCCTGCCGGACCGAGTAGGTCAGGCCAGTC
CCGTATCTCTCCCGGGTCTGGCGGGCCGTCAAGGAGGCCCGGACGTATCCGAGCCTGATC
2521 ---------+---------+---------+---------+---------+---------+ 2580
GGCATAGAGAGGGCCCAGACCGCCCGGCAGTTCCTCCGGGCCTGCATAGGCTCGGACTAG
M
S Bb
a bo
l sI
I II
/
GGCCTCATGGAGTCCTCGCCCCTGGTCGACTTCACCCCGCCCGAGGCCGACGTCTTCCGC
2581 ---------+---------+---------+---------+---------+---------+ 2640
CCGGAGTACCTCAGGAGCGGGGACCAGCTGAAGTGGGGCGGGCTCCGGCTGCAGAAGGCG
P
v
u
I
CGTGCTTACGACGAGGCCCTGACGATCGCCGTGGCGCTCCTGGAGGGCGACGAGGACGGC
2641 ---------+---------+---------+---------+---------+---------+ 2700
GCACGAATGCTGCTCCGGGACTGCTAGCGGCACCGCGAGGACCTCCCGCTGCTCCTGCCG
M
sS R BT HB
B pa s sa is
p Ac r eq na
m 1I I RI 4X
I II I II II
GGAGGCGCACCGGCCCCCGCGGCTCCCCGCCGGACCGAGCAGGCCGCCGACGCGGGCTCC
2701 ---------+---------+---------+---------+---------+---------+ 2760
CCTCCGCGTGGCCGGGGGCGCCGAGGGGCGGCCTGGCTCGTCCGGCGGCTGCGCCCGAGG


BpmI
PspPI
BspGI
TCCCGTGTCCCGACGCCGCCCGCTACCGCGCGGACCCTCCAGGTCCTCGTTCTGGACGCC
2761 ---------+---------+---------+---------+---------+---------+ 2820
AGGGCACAGGGCTGCGGCGGGCGATGGCGCGCCTGGGAGGTCCAGGAGCAAGACCTGCGG
AlwI
BstYI
NarI
NgoAIV
CTGGCCGCCGGGCAGCTCCGCACCCGCGAGATCCGCAAGGCGGTAGGCGCCGGCACCGAC
2821 ---------+---------+---------+---------+---------+---------+ 2880
GACCGGCGGCCCGTCGAGGCGTGGGCGCTCTAGGCGTTCCGCCATCCGCGGCCGTGGCTG
BpmI
SalI
BpmI
GGCGGCCCGTCCTCCGGGTCCGTCGACAACGCCCTCCAGGCCCTCCAGGACGCCGGGCTC
2881 ---------+---------+---------+---------+---------+---------+ 2940
CCGCCGGGCAGGAGGCCCAGGCAGCTGTTGCGGGAGGTCCGGGAGGTCCTGCGGCCCGAG
CjeI
RleAI
DraIII
TaqII
CjeI
GTCGCCCGCGTCACCCACGGCGTGTGGGCCCGCACCGACCAGGCCGACCCCTCCTGACCC
2941 ---------+---------+---------+---------+---------+---------+ 3000
CAGCGGGCGCAGTGGGTGCCGCACACCCGGGCGTGGCTGGTCCGGCTGGGGAGGACTGGG
MboII
CGCACCCCTCGCCGGGCCCCTGACGCGCCTTACGACGGCCGCCGGGGGCCCGGCCCTCTC
3001 ---------+---------+---------+---------+---------+---------+ 3060
GCGTGGGGAGCGGCCCGGGGACTGCGCGGAATGCTGCCGGCGGCCCCCGGGCCGGGAGAG
EarI
PspPI
BsrBI
TTCCCCGCCGGAGGACCCCGTGACCGCCCCCAGCGTCGCCGTACAGGCCCCCGAGCGGGC
3061 ---------+---------+---------+---------+---------+---------+ 3120
AAGGGGCGGCCTCCTGGGGCACTGGCGGGGGTCGCAGCGGCATGTCCGGGGGCTCGCCCG

```
                                                      B
                                                      p
                                                      m
                                                      I
     CGACACCAGACCAGCACCGCCCGCGTCCCTGGAGGACCGCCTGACGGCCGTGAGCGCCCT
3121 ---------+---------+---------+---------+---------+---------+ 3180
     GCTGTGGTCTGGTCGTGGCGGGCGCAGGGACCTCCTGGCGGACTGCCGGCACTCGCGGGA

     CATGGACGACCGGCTCACGGTCGCCGCCCTGGCCGTCGCCGTGAACGCCGCCGTGAGCGT
3181 ---------+---------+---------+---------+---------+---------+ 3240
     GTACCTGCTGGCCGAGTGCCAGCGGCGGGACCGGCAGCGGCACTTGCGGCGGCACTCGCA

                                T
                                t
                                h                     B
                    S           1                     s
                    a           1                     p
                    l           1                     G
                    I           I                     I
     CCCCGACGTCGGGCCCGTCGACCTGGCCGACGTCGTCCGCGTCCCCGCCGACGTCCAGCC
3241 ---------+---------+---------+---------+---------+---------+ 3300
     GGGGCTGCAGCCCGGGCAGCTGGACCGGCTGCAGCAGGCGCAGGGGCGGCTGCAGGTCGG

                                P                C
                    S           s              S v P
                    a           h              f i s
                    l           A              c R t
                    I           I              I I I
     CTACCGGCCGCCCGTCGACCTCTACCCGACCCCGGTCGCCGCCGTTCTGCAG
3301 ---------+---------+---------+---------+---------+-- 3352
     GATGGCCGGCGGGCAGCTGGAGATGGGCTGGGGCCAGCGGCGGCAAGACGTC
```

Enzymes that do cut and were not excluded:

AflIII	AgeI	AlwI	Alw21I	ApaBI	BaeI	BbsI	BccI
BcgI	BclI	BpmI	BsaI	BsaAI	BsaWI	BsaXI	BseRI
BsiI	BsmAI	BsmBI	BspEI	BspGI	BspMI	BsrI	BsrBI
BssHII	BstEII	BstXI	BstYI	CjeI	CjePI	CviRI	DraIII
DrdI	Eam1105I	EarI	Eco57I	FseI	FspI	HaeI	Hin4I
KpnI	MboII	MluI	MmeI	MscI	MslI	MspA1I	NarI
NcoI	NgoAIV	NotI	NspI	Pfl1108I	PmlI	PshAI	Psp5II
Psp1406I	PstI	PvuI	RleAI	RsrII	SacI	SacII	SalI
ScaI	SexAI	SfcI	SfiI	SgrAI	SphI	SrfI	StyI
SunI	TaqII	Tth111I	XcmI	XmnI			

Enzymes that do not cut:

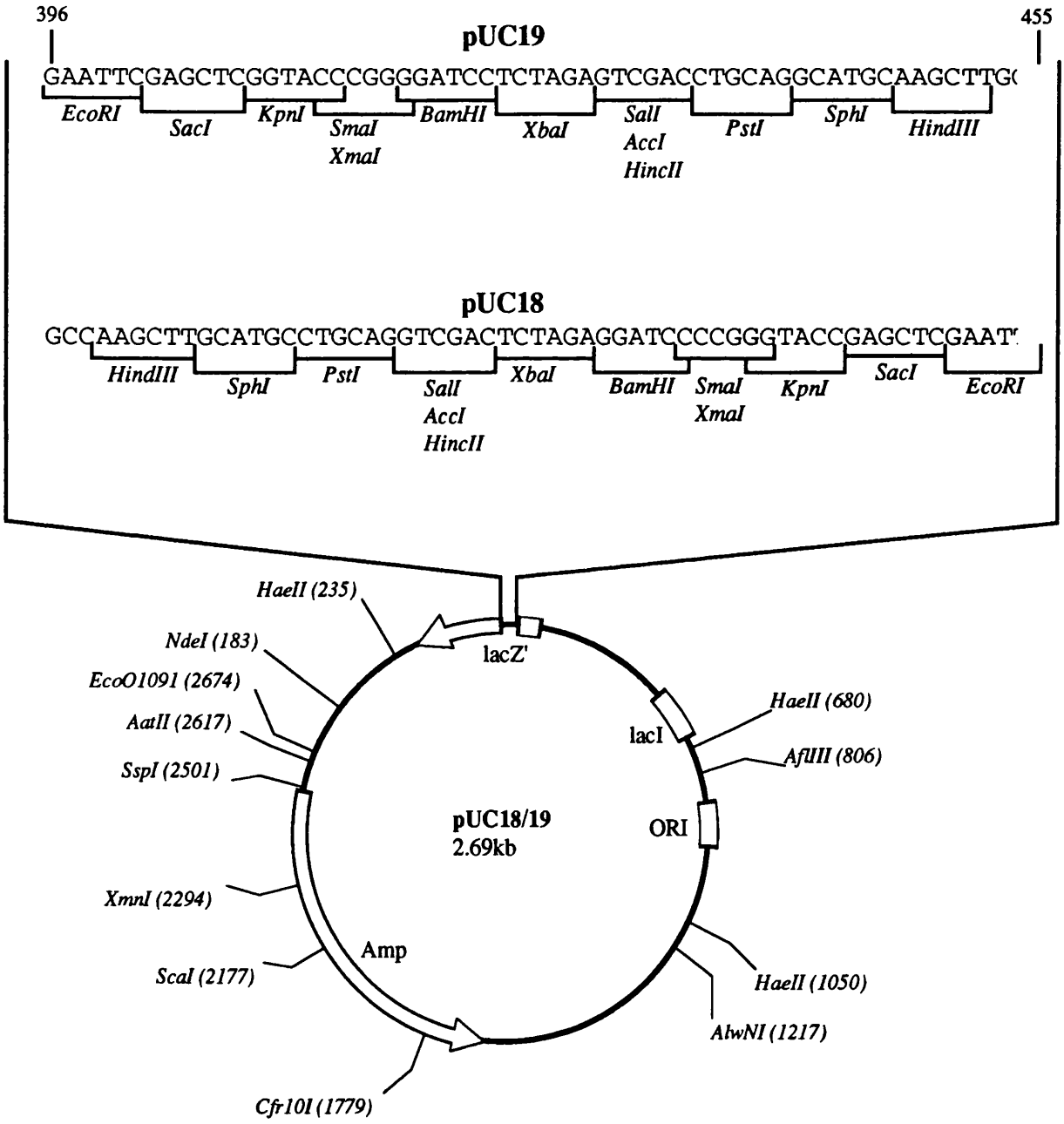
AflII	Alw44I	AlwNI	ApoI	AscI	AvrII	BamHI	Bce83I
BfaI	BglII	Bpu10I	Bpu1102I	BsaBI	BsbI	BsgI	BsmI
Bsp24I	Bsp24I	BspLU11I	BsrDI	BsrGI	Bst1107I	Bsu36I	ClaI
DdeI	DraI	DrdII	Eco47III	EcoNI	EcoRI	EcoRV	HgiEII
HindIII	HpaI	MseI	MunI	NdeI	NheI	NruI	NsiI
NspV	PacI	PflMI	PmeI	PvuII	RcaI	SapI	SgfI
SnaBI	SpeI	Sse8387I	SspI	StuI	SwaI	TfiI	Tsp509I
Tth111II	VspI	XbaI	XhoI				

```
Enzymes excluded; MinCuts: 1  MaxCuts: 6

   AatII     AccI   AceIII     AciI     AluI     ApaI     AvaI    AvaII
    BanI    BanII     BbvI    BcefI     BglI     BmgI    BsaHI    BsaJI
   BscGI    BsiEI     BslI    BsmFI    BsoFI Bsp1286I    BsrFI    Cac8I
   CviJI     DpnI     DsaI     EaeI     EagI     EciI Eco0109I   EcoRII
    FauI     FokI    GdiII    HaeII   HaeIII     HgaI     HhaI   HincII
   HinfI     HphI    MaeII   MaeIII     MnlI     MspI     MwoI     NciI
  NlaIII    NlaIV     PleI     RsaI   Sau96I   Sau3AI    ScrFI    SfaNI
    SmaI     TaqI     ThaI     TseI   Tsp45I    UbaJI
```

A comprehensive restriction enzyme map of the sequenced region (obtained using MAPPLOT Devereux *et al.* 1984) is shown. Only restriction enzymes cutting between 1 and 6 times are indicated, enzymes mentioned in the text are shown in bold face.

APPENDIX B1: Map of pUC18/19



APPENDIX B2: Map of pT7-7