7

AN INVESTIGATION OF THE REPLICATION OF THE BROAD-HOST-RANGE PLASMID pTF-FC2

Rosemary Ann Dorrington

A dissertation submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy in the Faculty of Science, University of Cape Town.

Cape Town, September, 1990

CERTIFICATION OF SUPERVISOR

In terms of paragraph GP 8 in "General rules for Degree of Doctor of Philosophy (PhD)", I as supervisor of the candidate, Rosemary Ann Dorrington, certify that I approve of the incorporation in this dissertation of material that has already been submitted or accepted for publication.

Signed by candidate

Professor Douglas E. Rawlings
Department of Microbiology
University of Cape Town
Cape Town

Opgedra aan my moeder

CONTENTS

Frontispiece: Electron micrograph showing the 12.4 kb plasmid pTF-FC2

ACKNOWLEDGEMENTS

My most sincere gratitude goes to my supervisor, Professor Doug Rawlings, who guided by suggestion and example. He was a constant source of support and encouragement, never too busy to listen and advise.

I am deeply indebted to Di James, whose beautiful sequnecing gels lead to "The Discovery of the Missing Nucleotide"; to Anne-Marie Clennell for her invaluable help with the preparation of diagrams and figures and also to Dr Claire Davis, who helped with the proof-reading of the manuscript. Special thanks are due to my colleagues and friends in the *Thiobacillus* Unit for their patience and encouragement, and especially to Jurg Rohrer for his friendship and support. Thanks also to William Bourn whose honest and critical evaluations I value highly. I would also like to thank all the other members of the Microbiology department for their interest and advice.

My deepest gratitude goes to Sybil and Archie Sieff, who provided me with a home during the past months, to my brother Peter who generously put his computer at my disposal; to Reagan, BB, Barbara, Mona and, last, but not least, Edgar J. for their moral support and sympathy. I owe much to my father, Dr John Dorrington, without whose understanding, support and encouragement this would not have been possible.

I acknowlege financial support from the Foundation for Research and Development of the Council for Scientific and Industrial research. This research was funded by grants from the General Mining Corporation and the Foundation for Research and Development.

ABBREVIATIONS

A	adenine
$Ap^{(R)}$	ampicillin (resistance)
ATP	adenosine triphosphate
bp	base pair(s)
C	cytosine
$Cb^{(R)}$	carbenicillin (resistance)
CIR(s)	complementary inverted repeat(s)
$Cm^{(R)}$	chloramphenicol (resistance)
dA	deoxyadenosine
DNA	deoxyribonucleic acid
dNTP(s)	deoxyribonucleotide triphosphates
ds	double stranded
G	guanine
Glu	Glutamine
IHF	integration host factor
kb	kilobase pair(s)
kD	kiloDalton
$Km^{(R)}$	kanamycin (resistance)
LA	Luria Agar
LB	Luria Broth
Leu	leucine
M9 MM	M9 minimal medium
M_R	relative molecular mass
nt(s)	nucleotide(s)
ONPG	o-nitrophenol-a-o-galactopyranose
ORF(s)	open reading frame(s)
oriC	chromosomal origin of replication
oriV	vegetative origin of replication

Pol I	DNA polymerase I
Pol II	DNA polymerase II
Pol III	DNA polymerase III
RNA	ribonucleic acid
RNA-Pol	RNA polymerase I
RNase	ribonuclease
SDS	Sodim Dodecyl Sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
$Sm^{(R)}$	streptomycin (resistance)
spp	species
ss	single stranded
$Su^{(R)}$	sulfonamide (resistance)
T	thymine
$Tc^{(R)}$	tetracycline (resistance)
ts	temperature sensitive
Val	valine
X-Gal	5-bromo-4-chloro-3-indolyl-b-galactoside

LIST OF TABLES

ABSTRACT

Plasmid pTF-FC2, a 12.4 kilobase pairs (kb) cryptic plasmid, was originally isolated from the acidophilic chemoautotroph *Thiobacillus ferrooxidans* and subsequently cloned into the pMB1-based vector, pBR325. Deletion of the pBR325 origin of replication revealed that pTF-FC2 was able to replicate autonomously in a number of Gram-negative bacteria besides *T. ferrooxidans*.

Constructs carrying the pTF-FC2 origin were able to replicate independently of DNA polymerase I (Pol I) in *Escherichia coli* and this feature was used to distinguish between replication from the *T. ferrooxidans* origin and the pMB1-derived origins of the vectors. A 3.2 kb *Sau*3A partial fragment was obtained which had retained the ability to replicate in a *E. coli polA*$^-$ mutant and also in *Pseudomonas aeruginosa*. A series of deletions of this fragment was used to identify the minimal replicon, the vegetative origin of replication (*ori*V) and the areas determining plasmid incompatibility. The copy number of pTF-FC2 in *E. coli* was estimated at 12 - 15 copies per chromosome and a deletion plasmid was identified which replicated at a reduced copy number. An area which affected the ability of the replicon to replicate in *P. aeruginosa*, was identified.

The nucleotide sequence of the 3.2 kb minimal replicon of pTF-FC2 was determined from overlapping DNA sequence obtained from a series of sequential deletions from each end of the fragment. Analysis of the *ori*V sequence revealed three, tandemly repeated 22 base pairs (bp) DNA sequences and two sets of complementary inverted repeats. The 22 bp direct repeats appeared to be essential for replication and also for incompatibility. The role of one of the two sets of complementary inverted repeats was unclear. The deletion plasmid previously shown to replicate at reduced copy number was found to have half of the second set of repeats deleted.

Analysis of the sequence also revealed at least 5 potential open reading frames (ORFs) preceded by consensus ribosome binding sites (designated ORFs 1 - 5). Four of these produced polypeptides *in vitro*. The products of ORF1 (32 kilodalton [kD]) and ORF2 (33.5 kD) were required for replication in *E. coli*. A deletion plasmid

which was unable to replicate in *P. aeruginosa*, had lost an 8 and a 10 kD protein and these may be involved in replication of pTF-FC2 in certain hosts. At least two promoters exist on the 3.2 kb fragment and a putative site for one of these, *P1* was identified by comparison with the *E. coli* consensus promoter sequence.

An alignment and comparison of the DNA sequence of the minimal replicon of pTF-FC2 with that of the IncQ plasmid RSF1010 showed considerable DNA homology between the *ori*V regions of the two plasmids. In addition, the sequence coding for ORF1 and ORF2 of pTF-FC2 was very similar to the region coding for the RepA and RepC proteins of RSF1010. There was 60% similarity between the 22 bp direct repeats of pTF-FC2 and the sequence immediately adjacent to the repeats with the equivalent area on the RSF1010 *ori*V. The predicted amino acid sequence of ORF1 and ORF2 showed good homology with the amino acid sequences of the RepA and RepC of the IncQ replicon. No homology was detected between ORF3, ORF4 and ORF5 and other proteins encoded by the RSF1010 replicon. Despite these similarities, no incompatibility was detected between derivatives of pTF-FC2 and the IncQ plasmid, R300B. R300B was unable to complement *rep*$^-$ mutations in the pTF-FC2 replicon.

During the course of experiments aimed at the construction of broad-host-range cloning vectors based on the pTF-FC2 replicon, the 3.2 kb pTF-FC2 fragment was found to have an absolute requirement in *cis* for the pUC19 replicon into which it had been cloned. The presence of pDER412 in *trans* enabled the fragment to replicate in the absence of pUC19. An additional 1239 bp, located upstream of the start of the 3.2 kb fragment was identified which restored the ability of this fragment to replicate autonomously. Sequence analysis of the region revealed a single ORF encoding a polypeptide with a predicted momolecular mass (M_R) of 40 kD which was identified *in vitro*. A comparison of the amino acid sequence of this protein with sequence data banks revealed limited homology with only the RepB' primase of the IncQ plasmid, RSF1010. The nature of the complementation by pUC19 of the *repB*$^-$ phenotype of the 3.2 kb fragment, was investigated and a model was proposed for the mechanism by which this fragment was able to replicate.

CHAPTER 1

GENERAL INTRODUCTION

CHAPTER 1

GENERAL INTRODUCTION

Plasmids are generally circular, double stranded (ds) deoxyribonucleic acid (DNA) molecules which replicate autonomously in a host cell. Linear plasmids have been identified in the Streptomycetes and also in *Thiobacillus versutus* (Hirochika *et al.*, 1984; Wlodarczyk and Nowicka, 1988). Plasmids may vary in length from a few to several hundred kilobase pairs (kb). Most plasmids are cryptic, but many provide their hosts with useful, new phenotypic characteristics. These factors include those which confer drug resistance, degradation of organic compounds and virulence factors such as the production of toxins. These types of genes are frequently found within transposable elements which generate flexibility and variability within such plasmids (Couturier *et al.*, 1988).

Plasmids represent an important factor in the evolution of prokaryotes because they allow rapid, short-term adaptability of the host organisms. They provide a means of gene amplification and many plasmids move easily within and between bacterial species. Plasmids have developed a number of mechanisms to ensure that they are stably maintained within a bacterial population. Among these are active partitioning (Austin, 1988), killing of plasmid-free cells (Gerdes *et al.*, 1986) and infectious conjugal transfer (Lundquist and Levin, 1986).

Plasmid copy number may vary from one copy to more than fifty per chromosome and copy number of especially the low copy number plasmids is tightly controlled (Novick, 1987). The existence of related plasmids in the same host cell is prevented by incompatibility determinants which are plasmid-encoded and

may be based on replication or segregation systems (Novick, 1987). Copy number control and incompatibility based on replication invariably involve the control of initiation of replication.

The replication of plasmids is inextricably linked with the chromosomal replication apparatus. Most plasmids, particularly those with a narrow host-range, are almost entirely dependent upon host-encoded proteins for DNA replication. Plasmids such as F, Rts1 and pSC101, carry a single initiator protein which is required for plasmid-specific initiation of replication (Tokino *et al.*, 1986; Kamio *et al.*, 1984; Linder *et al.*, 1985). Other plasmids such as the broad-host-range IncQ plasmids, supply several replication proteins which reduce their dependence upon the host cell (Haring and Scherzinger, 1989).

1.1 DNA REPLICATION

Intense study over the past 20 years has revealed that DNA replication is a complicated, multistep process involving many more enzymes than was anticipated following the discovery of the first DNA polymerase by Arthur Kornberg and his colleagues in 1957. The enzymes involved in DNA replication may be divided into five classes:

(i) DNA polymerases, of which there are three types, namely DNA polymerase I, DNA polymerase II and DNA polymerase III (Pol I, Pol II and Pol III respectively). Pol I fills the gaps between lagging strand fragments and removes ribonucleotide primer sequences. Pol III is the main DNA polymerising agent and is associated with a number of other proteins. The function of Pol II has not been clearly defined yet.

(ii) DNA binding proteins present the DNA template to the Pol III holoenzyme in an acceptable form.

(iii) DNA polymerase accesory proteins convert Pol III into a highly processive enzyme which is capable of replicating a whole genome in one binding event. These proteins are also responsible for reducing the error factor during DNA synthesis.

(iv) DNA "unwinding" proteins (helicases) achieve and maintain the unwinding of the double helix to allow priming of the single stranded (ss) DNA.

(v) Priming proteins synthesise *de novo* small ribo- or deoxyribooligonucleotides which prime the DNA for the polymerase.

Table 1.1 shows some of the proteins involved in the replication of the *E. coli* chromosome and the genes which encode these proteins on the *E. coli* chromosome are shown in Table 1.2.

Table 1.1 Proteins involved in the replication of the *E. coli* chromosome (From Watson *et al.*, 1987)

Protein	Native Mass, kdal	Subunits		Function
SSB	74	4		Single-strand binding
Protein i	66	3		
Protein n	28	2		Primosome assembly
Protein n'	76	1		and function
Protein n''	17	1	c	
DnaC	29	1		
DnaB	300	6		
Primase	60	1		Primer synthesis
Pol III Holoenzyme	(760)	(2)		
α	140	1		
ε	25	1		Processive chain
θ	10	1		elongation
β	37	1	x 2	
γ	52	1		
δ	32	1		
τ	83	1		
Pol I	102	1		Gap filling, primer excision
Ligase	74	1		Ligation
Topoisomerase II (gyrase)	400	4		Supercoiling
GyrA	210	2		
GyrB	190	2		
Rep	65	1		Helicase
Helicase II	75	1		Helicase
DnaA	48			Origin of replication
Topoisomerase I	100	4		Relaxing negative supercoils

Table 1.2 Genes which encode proteins involved in DNA replication in *E. coli* (From Watson *et al.*, 1987)

Gene	Map Location, minutes	In Vivo Phenotype of Mutant	Protein and in Vitro Function
dnaA	82	Slow stop; defective origin initiation	Unknown
dnaB	91	Quick stop	DnaB protein; prepriming
dnaC	99	Slow or quick stop (depending on the mutation)	DnaC protein complexes with DnaB protein
dnaE (polC)	4	Quick stop	Subunit α of Pol III holoenzyme
dnaG	66	Quick stop, defective initiation of fragments	Primase
dnaI	39	Slow stop	Unknown
dnaJ	0.5	Slow stop	Unknown
dnaK	0.5	Slow stop	Unknown
dnaL	28	Quick stop	Unknown
dnaP	84	Phenethyl alcohol resistance; slow stop	Unknown
dnaT	95-99	Regulates termination	Unknown
dnaZ	10	Quick stop	Subunit γ of Pol III holoenzyme; chain growth
cou	82	Coumermycin (and novobiocin) resistance	DNA topoisomerase II subunit β; nicking/ closing
lig	51	Accumulation of replication fragments	DNA ligase; covalently seals DNA nicks
nalA	48	Nalidixate (and oxolinate) resistance	DNA topoisomerase II subunit α; ATPase
polA	85	Defective in DNA repair	DNA polymerase I; gap filling, RNA excision
rep	83	Slowed fork movement	ATP-using helicase
ssb	91	Quick stop	SS DNA binding protein

DNA replication occurs in three basic steps, namely initiation, elongation and termination.

1.1.1 Initiation

This process provides the Pol III holoenzyme with a template carrying a 3'-OH terminus which it requires to begin polymerisation. Initiation occurs on ss DNA either by the action of RNA polymerase (RNA-Pol) or a specialised priming enzyme (primase). Helicases acting in combination with specific initiator proteins recognise sites in the vegetative origin of replication (*ori*V) from where replication of the molecule commences. The duplex DNA is unwound and RNA-Pol and/or the primase uses the exposed ss DNA as a template. Priming may also occur on DNA which has been nicked by a sequence-specific endonuclease. Several proteins are involved in the priming process and the complex of enzymes is known as a primosome. The initation process is specific for each individual origin (Marians *et al.*, 1986).

1.1.2 Elongation

Relatively little is known about the mechanism by which DNA elongation occurs, but there is an absolute requirement for the Pol III holoenzyme, and many systems such as that of the *E. coli* chromosomal origin of replication (*ori*C) and plasmid ColE1 require the topoisomerase DNA gyrase (Kaguni and Kornberg, 1984; Minden and Marians, 1985). In the case of unidirectional replication, the Pol III holoenzyme continues polymerisation of the leading DNA strand while primosomes initiate polymerisation of the lagging strand at intervals along the leading strand to form short (Okazaki) fragments (Okazaki and Okazaki, 1969; Wolfson and Dressler, 1972; Ogawa *et al.*, 1985). The RNA primers are removed by either ribonuclease (RNase) H or Pol I and the gaps between the Okazaki fragments are subsequently filled and ligated to form a continuous strand by Pol I and DNA ligase respectively (Watson *et al.*, 1987). In the case of bidirectional replication, the lagging strand may be synthesised from a single initiation event at a specific site in the *ori*V (Haring and Scherzinger, 1989).

Some plasmids found in Gram-positive bacteria replicate by the asymmetrical rolling circle mechanism (te Riele *et al.*, 1986; del Solar *et al.*, 1987) whereby synthesis of the leading strand results in the displacement of the 5' end from the duplex. The ss DNA is quickly primed by RNA-Pol and the resulting Okazaki fragments filled in and ligated to form continuous duplex DNA.

1.1.3 Termination

The termination process is not well understood and by necessity differs in circular and linear DNA molecules. In circular molecules, termination may occur at the origin as is the case during unidirectional replication. Bidirectional replication can result in termination halfway around the molecules where the replicating forks collide. Alternatively, termination may occur at a specific sequence. A 40 kiloDalton (kD) protein encoded by *E. coli*, the Ter protein, has been shown to induce termination of replication of plasmid R6K *in vitro* by binding specifically to a set of 14 base pair (bp) inverted repeats in the *ori*V of the plasmid (Sista *et al.*, 1989).

This terminiator sequence shows homology with the *ter* sequences in the *E. coli oriC* (Hill et al., 1988; Hidaka et al., 1988). It is possible that the Ter protein may also be involved in termination of the *E. coli* chromosomal replication (Sista *et al.*, 1989). Unwinding of the duplex DNA during replication is countered by the action of topoisomerases which act either by producing single or double stranded nicks in the replicated molecules.

1.2 PLASMID REPLICATION

Unlike the elongation and termination steps, initiation of replication is a molecule-specific step and it is this process which has been the subject of intensive study. The specificity of replication initiation is very important in determining the propagation of a plasmid in a particular host. Initiation of plasmid replication is dependent on various host-encoded enzymes and it is this dependence which determines to some degree, the ability of plasmids to replicate only in a few closely hosts related while others are able to be maintained in a wide range of distanctly related bacteria.

A REPLICATION OF NARROW-HOST-RANGE PLASMIDS

The vast majority of narrow-host-range plasmids which have been isolated to date originate from the *Enterobacteriaceae* (Scott, 1984). These plasmids can be divided into two groups, namely those which require a plasmid-encoded initiator protein and those which do not. The replication of plasmids pSC101 and ColE1 respectively, will be used as examples of each of these groups.

1.3 pSC101

Plasmid pSC101 is a 9.26 kb plasmid (Bernardi and Bernardi, 1984) which was originally isolated from *Salmonella panama* (Cohen and Chang, 1977) and used as a cloning vector in the first recombinant DNA experiment (Cohen *et al.*, 1973). The

plasmid replicates in *E. coli* with a copy number of 5 plasmids per chromosome (Hasunuma and Sekiguchi, 1977) and can be maintained in the absence of Pol I (Cabello *et al.*, 1976).

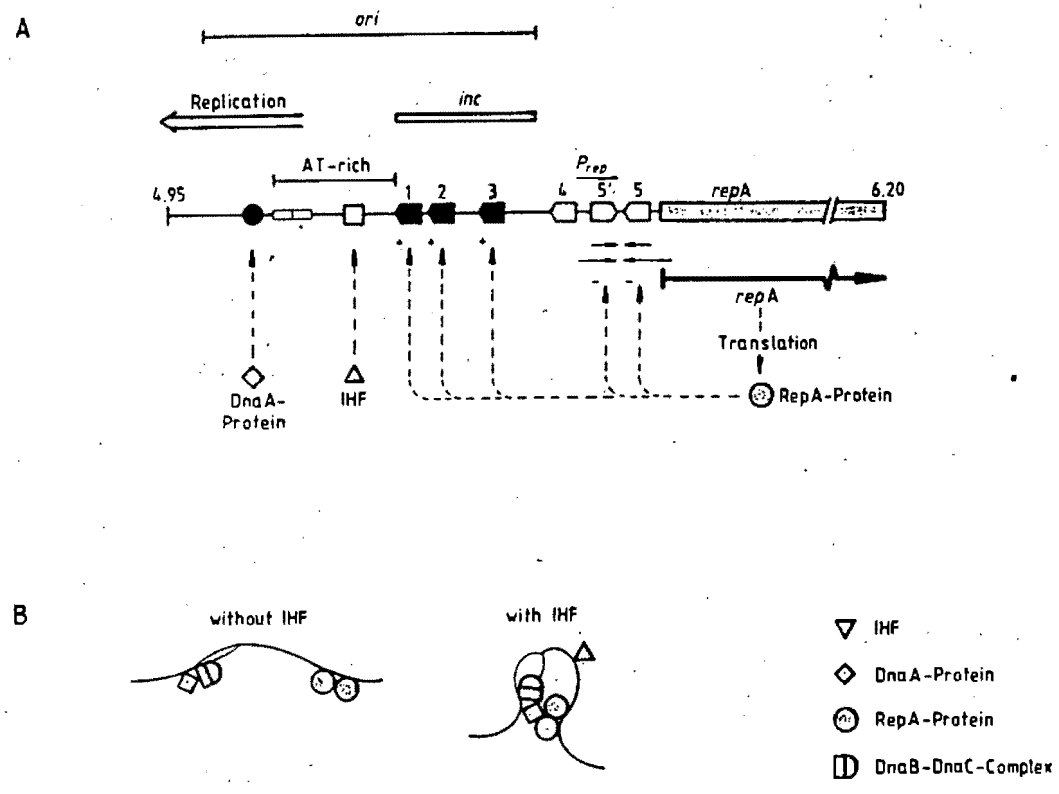

Fig. 1.1 (A) Diagram representing the 1.3 kb pSC101 basic replicon. Symbols shown are a DnaA box as a bold circle, two 13-mer repeats shown as empty rectangles, an IHF-binding site indicated by an empty square, three direct repeats whithin the *ori*V shown as bold arrows and three similar repeats within the promoter region of the *rep*A gene which are shown as empty arrows. The broken lines indicate sites of action of the DnaA, IHF and RepA proteins. **(B)** Schematic representation of the proposed role of the IHF, DnaA, RepA and DnaB-DnaC-complex during initiation of replication of pSC101. Thick and thin lines represent duplex and ss DNA respectively. In the absence of IHF, DNA curvature is not pronounced enough to bring about interaction between the DnaA and RepA molecules bound to the *ori*V DNA as shown in the lefthand side of the diagram. To the right is shown how IHF induces curvature of the duplex DNA which allows interaction between DnaA and RepA and the subsequent priming of replication at the 13-mer repeats by the DnaB-DnaC complex. (From Kues and Stahl, 1989).

1.3.1 The basic replicon

Plasmid pSC101 replicates unidirectionally from a unique origin (Yamaguchi and Yamaguchi, 1984a). The minimal replicon, which is 1.3 kb in size, consists of a 250 bp fragment which is required in *cis*, and a region of approximately 1 kb which encodes a 37.5 kD initiator protein, RepA (Fig. 1.1A). RepA acts in *trans* on the *ori*V (Churchward *et al.*, 1983; Armstrong *et al.*, 1984; Yamaguchi and Yamaguchi, 1984a; 1984b).

1.3.2 Structure of the *ori*V

The *ori*V of pSC101 shows homology with the *ori*C of the *E. coli* chromosome and related members of the *Enterobacteriaceae* (Yamaguchi and Yamaguchi, 1984b). Included in this homologous region are two 13 bp repeated sequences (repeated four times in the *E. coli oriC* [Zyskind and Smith, 1987]), which are found together with a DnaA binding site (*dnaA* box) (Churchward *et al.*, 1983; Fuller *et al.*, 1984). A consensus binding site for the integration host factor (IHF) is located just downstream of the *dnaA* box, within an 82 bp A+T-rich (84% A+T) region (Churchward *et al.*, 1983; Gamas *et al.*, 1986). Three 18 bp direct repeats are situated downstream of the A+T-rich region within the *inc* locus (Fig. 1.1A, repeats 1 - 3) which act as binding sites for the RepA protein (Vocke and Bastia, 1983). Three additional 18 bp repeats (repeats 4. 5' and 5) are located outside the *oriV* region within the putative promoter sequence of *repA* (Linder *et al.*, 1985; Vocke and Bastia, 1985).

1.3.3 Initiation of replication

Replication of pSC101 in *E. coli* is dependent upon the presence of host-encoded DnaA (Hasunuma and Sekiguchi, 1977), DnaB helicase, DnaC and DnaG primase as well as Pol III (Hasanuma and Sekiguchi, 1979; Ely and Wright, 1985).

The first step in the initiation of replication at the pSC101 *ori*V is the binding of RepA to the 18 bp repeats (repeats 1 - 3) within the *ori*V (Vocke and Bastia, 1983). Two host-encoded proteins have also been shown to bind specifically to sequences located in the pSC101 *ori*V.

DnaA, which binds to the DnaA box just upstream of the A+T-rich region, has been found to be essential for replication of not only pSC101, but also the *E. coli* chromosome (Fuller *et al.*, 1984). The precise role of DnaA in the replication of pSC101 is unclear, but it may be similar to that played in the replication of the *E. coli* chromosome. The binding of DnaA to *ori*C is thought to facilitate correct primer synthesis by RNA-Pol and/or DnaG primase (Messer, 1987; Seufert and Messer, 1987). DnaA binding is followed by the formation of a series of complexes which eventuallly leads to the opening of the four 13 bp repeats (similar to the two found in pSC101) to allow entry of the DnaB/DnaC helicase complex (Funnell *et al.*, 1986; Bramhill and Kornberg, 1988). Further duplex opening by DnaB helicase allows priming of DNA synthesis and the forks of bidirectional replication from the *ori*C (Baker *et al.*, 1987).

A second protein, IHF, binds specifically to the IHF consensus binding site within the A+T-rich region (Stenzel *et al.*, 1987). Plasmid pSC101 cannot replicate in the absence of this protein (Gamas *et al.*, 1986). IHF is not essential for replication of the *E. coli* chromosome, but is involved in the regulation of a number of other systems such as the expression of the CIII proteins of phage Lambda (Hoyt *et al.*, 1982). IHF influences expression of a wide variety of genes such as the *ilv* and *inf* genes of *E. coli* (Frieden *et al.*, 1984), has been shown to bind specifically to the ends of IS50 and this binding appears to play a role in the frequency of transposition (Makris *et al.*, 1990). IHF is also necessary for efficient conjugative transfer of the F plasmid (Gamas *et al.*, 1986). The IHF protein is very similar to the HU protein, which is implicated in the formation of nucleosome-like structures (Miller, 1984).

The IHF binding site of pSC101 is found on a region of bent DNA (A+T-rich region) and binding of IHF further enhances bending of this fragment (Stenzel *et al.*, 1987). It is proposed that the natural bending of the A+T-rich region may not be sufficient

to span the distance between DnaA and RepA, which are bound to their respective sites on the *ori*V (Fig. 1.1B). The binding of IHF increases DNA bending, bringing DnaA and RepA into contact. Together, DnaA and RepA facilitate the opening of the 13-mer repeats which enables the DnaB-DnaC complex to prime ss DNA for DNA synthesis (Stenzel *et al.*, 1987).

Mutations in the third codon of RepA which result in the replacement of a glutamine residue by a lysine (residue resulting in altered protein activity) enable pSC101 to replicate in the absence of IHF (Biek and Cohen, 1989a). A mutation in the *topA* gene, coding for topoisomerase I, also results in maintenance of pSC101 in the absence of IHF, but a concomitant mutation in the *gyrB* gene negates this effect (Biek and Cohen, 1989b). Using these and other results, Biek and Cohen (1989b) have proposed the following model to explain the role of IHF in the initation of replication of pSC101. Initation of replication at the *ori*V is directly dependent on the degree of DNA supercoiling in certain loci. Binding of the IHF protein reduces supercoiling in these regions and enables RepA to initiate replication. The mutant RepA proteins could have an altered ability to tolerate supercoiling and are therefore able to initiate replication in the absence of IHF, while the mutations in *topA* and *gyrB* would also directly affect supercoiling of these loci.

1.3.4 Regulation of initiation by RepA

The RepA protein plays both a positive and a negative role in the initiation of replication of pSC101. The binding of RepA to repeats 1 - 3 in the *ori*V stimulates replication (Vocke and Bastia, 1985; Yamaguchi and Yamaguchi, 1984a). The RepA protein negatively regulates replication by controlling its intracellular concentration (Linder *et al.*, 1985; Vocke and Bastia, 1985; Yamaguchi and Masamune, 1985). The putative promoter region of *repA* overlaps with a palindromic arrangement of three RepA binding sites (Fig. 1.1A, repeats 4, 5' and 5) (Vocke and Bastia, 1985; Yamaguchi and Masamune, 1985). RepA binds to the inverted repeats, competes with RNA-Pol for the promoter sequence and inhibits transcription of *repA* (Linder *et al.*, 1985; Vocke and Bastia, 1985). RepA has a higher affinity for inverted repeats

5' and 5 than repeats 1 - 3 in the *ori*V (Vocke and Bastia, 1983). This results in tight control of the intracellular concentration of RepA and consequently the initiation of replication of pSC101.

1.4 THE ColE1-TYPE PLASMIDS

ColE1 is a 6.6 kb *E. coli* plasmid which is maintained at a copy number of approximately 20 plasmids per chromosome (Chan *et al.*, 1985). ColE1 is related to other plasmids of the *Enterobacteriaceae* such as p15A, pMB1, CloDF13 (Selzer *et al.*, 1983) and other colicinogenic plasmids such as ColA (Morlon *et al.*, 1988). Derivatives of p15A such as pACYC184 (Cheng and Cohen, 1978) and of pMB1 such as pBR322 and the pUC plasmids (Balbas *et al.*, 1986; Yanisch-Perron *et al.*, 1985), have been extensively used as cloning vectors in recombinant DNA experiments.

1.4.1 The minimal replicon of ColE1

Replication of ColE1 is initiated within a 0.6 kb *oriV* and proceeds unidirectionally in the o-shaped manner of Cairns-type replication (Inselburg, 1974; Tomizawa *et al.*, 1974; Tomizawa *et al.*, 1975). ColE1 requires only host-encoded proteins for replication (Donoghue and Sharp, 1978). These include RNA-Pol, RNase H, Pol I, DNA gyrase and topoisomerase I (Itoh and Tomizawa, 1978; Orr and Staudenbauer, 1981; Minden and Marians, 1985). DNA gyrase is responsible for unwinding the duplex DNA ahead of the replication forks while topoisomerase I is thought to favour the recognition of the primer promoter by RNA-Pol through modulation of the topology of the *ori*V DNA (Minden and Marians, 1985).

1.4.2 Initiation of replication

There are three mechanisms by which primer synthesis at the ColE1 *oriV* may take place, namely the Type I, Type II and Type III mechanisms. Of these, Type I is the main mechanism by which ColE1 primer synthesis occurs and this system requires

RNase H and Pol I. Type II primer synthesis can occur in the absence of RNase H and Pol I, while Type III primer synthesis relies on the activity of Pol I only (Dasgupta *et al.*, 1987; Masukata and Tomizawa,1987).

1.4.2a Type I mechanism of primer synthesis

Beginning 555 bp upstream from the ColE1 *ori* (Fig. 1.2), a primer precursor of approximately 700 bp, called RNAII is transcribed by RNA-Pol (Itoh and Tomizawa, 1980) and terminated heterogeneously downstream from the *ori* (Tomizawa and Masukata, 1987). The 3' ends of these transcripts persist in forming hybrids with their DNA template near the *oriV* (Itoh and Tomizawa, 1980) and this process (known as coupling) depends on the secondary structure of the 5' end of RNAII (Wong and Polisky, 1985; Masukata and Tomizawa, 1986). If RNAII does not adopt the specific secondary structure required for hybrid formation, coupling does not occur and priming aborts (Masukata and Tomizawa, 1986).

RNase H recognises the *oriV*-proximal stem-and-loop of the DNA-hybridised precursor and cleaves it within a sequence of five A residues to generate the mature primer molecule (Masukata and Tomizawa, 1986). Pol I subsequently uses the primer as a template for the addition of deoxynucleotides (dNTPs) at its 3'-OH end. RNase H then cleaves the primer at a secondary site and the remaining ribonucleotides are removed by Pol I while further cleavage by RNase H removes the primer from the newly synthesised DNA strand (Selzer and Tomizawa, 1982). Termination of replication occurs at *terH*, 17 bp upstream from the start of replication (Dasgupta *et al.*, 1987).

1.4.2b Type II mechanism of primer synthesis

Transcription of RNAII and its hybridisation to the DNA is also required in both Type II and Type III mechanisms of primer synthesis (Dasgupta *et al.*, 1987). The RNA-DNA hybrid results in the formation of a displacement loop (Itoh and Tomizawa, 1980; Selzer *et al.*, 1982; Masukata *et al.*, 1987) and lagging strand synthesis is initiated by DnaG primase on the single non-translated DNA strand (Fig. 1.2). A single stranded DNA loop of at least 40 bp is required and this may be extended by the action of DnaB helicase (Masukata *et al.*, 1987).

RNase H and Pol I. Type II primer synthesis can occur in the absence of RNase H and Pol I, while Type III primer synthesis relies on the activity of Pol I only (Dasgupta *et al.*, 1987; Masukata and Tomizawa,1987).

1.4.2a Type I mechanism of primer synthesis

Beginning 555 bp upstream from the ColE1 *ori* (Fig. 1.2), a primer precursor of approximately 700 bp, called RNAII is transcribed by RNA-Pol (Itoh and Tomizawa, 1980) and terminated heterogeneously downstream from the *ori* (Tomizawa and Masukata, 1987). The 3' ends of these transcripts persist in forming hybrids with their DNA template near the *oriV* (Itoh and Tomizawa, 1980) and this process (known as coupling) depends on the secondary structure of the 5' end of RNAII (Wong and Polisky, 1985; Masukata and Tomizawa, 1986). If RNAII does not adopt the specific secondary structure required for hybrid formation, coupling does not occur and priming aborts (Masukata and Tomizawa, 1986).

RNase H recognises the *oriV*-proximal stem-and-loop of the DNA-hybridised precursor and cleaves it within a sequence of five A residues to generate the mature primer molecule (Masukata and Tomizawa, 1986). Pol I subsequently uses the primer as a template for the addition of deoxynucleotides (dNTPs) at its 3'-OH end. RNase H then cleaves the primer at a secondary site and the remaining ribonucleotides are removed by Pol I while further cleavage by RNase H removes the primer from the newly synthesised DNA strand (Selzer and Tomizawa, 1982). Termination of replication occurs at *terH*, 17 bp upstream from the start of replication (Dasgupta *et al.*, 1987).

1.4.2b Type II mechanism of primer synthesis

Transcription of RNAII and its hybridisation to the DNA is also required in both Type II and Type III mechanisms of primer synthesis (Dasgupta *et al.*, 1987). The RNA-DNA hybrid results in the formation of a displacement loop (Itoh and Tomizawa, 1980; Selzer *et al.*, 1982; Masukata *et al.*, 1987) and lagging strand synthesis is initiated by DnaG primase on the single non-translated DNA strand (Fig. 1.2). A single stranded DNA loop of at least 40 bp is required and this may be extended by the action of DnaB helicase (Masukata *et al.*, 1987).

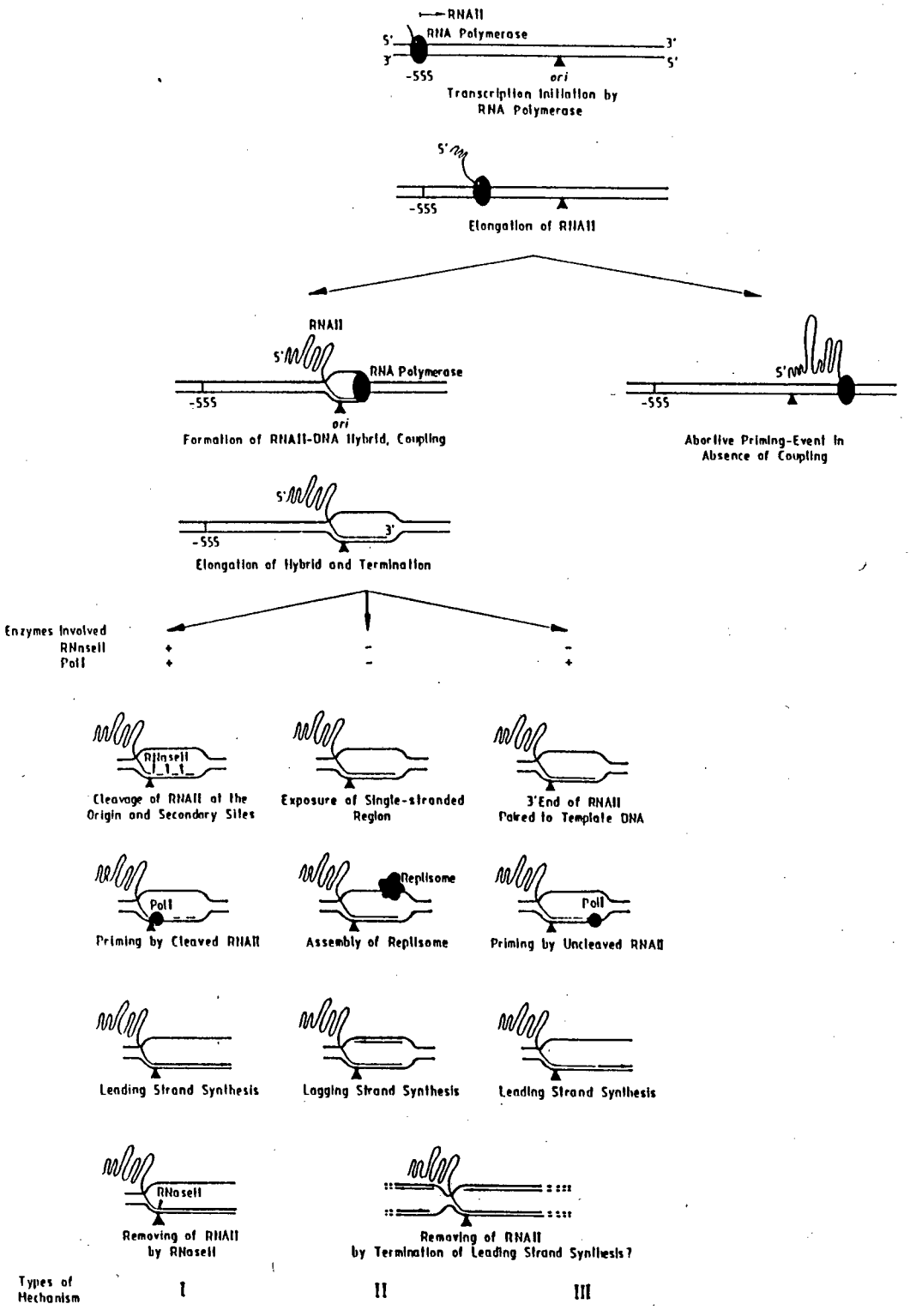

Fig. 1.2 Diagramatic representation of the three mechanisms by which the initiation of ColE1-type plasmid replication occurs. The mechamisms are described in detail in the text. Position -555 corresponds to the start of the RNAII transcript. The triangle marks the position in the *ori* where RNAII processing by RNaseH occurs at the target site for Pol I. DNA and RNA strands are indicated by thick and thin lines respectively and folds in the different RNAs represent secondary structure. The direction of replication is indicated with arrows (from Kues and Stahl, 1989).

1.4.2c Type III mechanism of primer synthesis

In cells which lack a functional RNase H but contain Pol I, replication may start by the Type III or by both Type II and Type III mechanisms. Pol I recognises the RNA-DNA hybrid as a primer without previous cleavage by RNase H (Dasgupta *et al.*, 1987), but this system is normally inefficient *in vitro* (Itoh and Tomizawa, 1980) because only short transcripts of RNAII with a hybridised length of less than 50 bp can be efficiently recognised by Pol I for initiation of replication (Dasgupta *et al.*, 1987; Tomizawa and Masukata, 1987). The RNA primer in the Type II and Type III mechanisms is probably removed by termination of lagging strand synthesis (Dasgupta *et al.*, 1987).

1.4.3 Regulation of initation of replication

The frequency of replication initiation at the *oriV* of ColE1 is in part dependent upon the frequency of RNAII formation by RNA-Pol (Castagnoli, 1987). However, the major regulation of initiation is exerted upon RNAII primer formation by an antisense RNA transcript of 108 bp, RNAI, which is transcribed from a site 445 bp upstream of the *oriV* in the opposite direction to transcription of RNAII (Fig. 1.3) (Tomizawa, 1986). RNAI interacts with part of RNAII to form an RNA-RNA hybrid which prevents RNAII from adopting the conformation required for hybridisation with its DNA template (Tomizawa and Itoh, 1981; Tomizawa *et al.*, 1981; Lacatena and Cesareni, 1983; Masukata and Tomizawa, 1986). RNAI and RNAII adopt specific secondary and tertiary structures as a result of inverted repeats within their sequence which lead to the formation of an RNA-RNA duplex. This process has been shown to occur *in vitro* in two steps (Tomizawa, 1985). The first, which is fast and reversible, involves the interaction of loop sequences in the cloverleaf structure of RNAI with the complementary sequences in the primer precursor ("kissing"). Subsequently, however, by a zipping mechanism starting at the 5' end of RNAI, a complete hybrid is formed. Whether "kissing" alone is sufficient to inhibit primer formation or whether the completion of the hybridisation reaction is required, is not clear (Helmer-Citterich *et al.*, 1988).

Transcription of RNAI occurs from a strong promoter and its half-life is very short, so that the intracellular concentration of the transcript closely follows fluctuations in the copy number of ColE1 (Tomizawa, 1984). The activity of RNAI has been found to be affected by a number of factors. The 5' end, which is involved in duplex formation with RNAII may be cleaved by RNase E which prevents hybrid formation (Tomcsanyi and Apirion, 1985). Furthermore, when *E. coli* (*relA*) strains are grown in media which are limiting for certain amino acids, plasmid copy number increases. This may be due to interaction between the cloverleaf tRNA-like structure of RNAI and tRNA molecules which are overabundant due to the low concentration of their corresponding amino acids (Lin-Chao and Bremer, 1986a; 1986b).

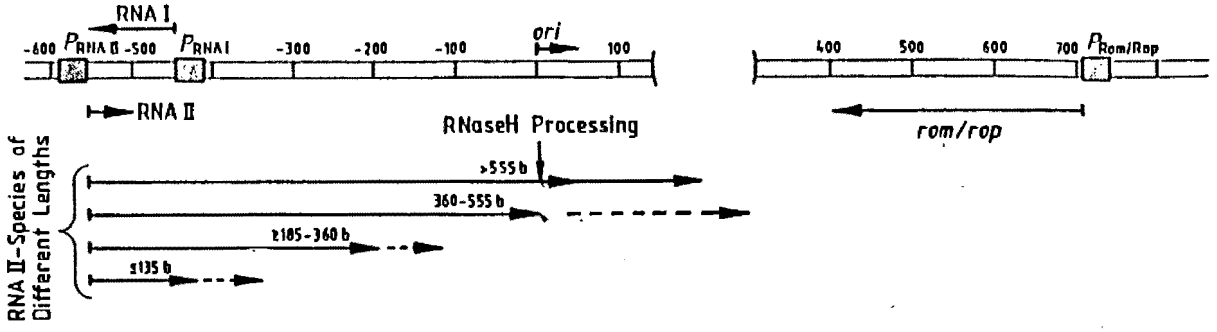

Fig. 1.3 Genetic map of the ColE1 origin of replication. The regions from which RNAI, RNAII and Rop/Rom are transcribed are indicated by arrows below and above the map (from Kues and Stahl, 1989).

A second element, Rop (or Rom) also acts as a negative control of plasmid replication (Twigg and Sherratt, 1980; Som and Tomizawa, 1983; Cesareni *et al.*, 1984; Tomizawa and Som, 1984). The element is coded for by sequences downstream from the *oriV* (Fig. 1.3) in some ColE1-type plasmids (Twigg and Sherratt, 1980; Som and Tomizawa, 1983) and is missing in others (Nijkamp *et al.*, 1986; Morlon *et al.*, 1988). Rop is a small dimeric protein whose monomer consists of two a-helices connected by a sharp bend (Banner *et al.*, 1987). Rop increases the affinity of RNAI for RNAII, speeds up the first reversible "kissing" step of the hybridisation process (Tomizawa and Som, 1984; Tomizawa, 1985). Rop-defective plasmids have an elevated copy number (Helmer-Citterich *et al.*, 1988). Helmer-Citterich *et al.* (1988) have shown that Rop binds to the duplex regions of the RNAI and RNAII molecules. Due to its symmetrical structure, the dimeric Rop protein

may act as an adaptor between RNAI and RNAII in order to direct the orientation of the molecules for the "kissing" reaction (Helmer-Cittirich *et al.*, 1988). Recent studies by Tomizawa (1990a; 1990b) reveal that the process by which RNAI and RNAII hybridise consists of a series of reactions resulting in a very unstable product which produces a more stable intermediate and finally results in the stable hybrid. The interaction with the Rop protein results in a rapid conversion of earlier unstable intermediates to more stable products.

The *E. coli* chromosomal locus *pcnB* is required for replication of ColE1-derived plasmids (Lopilato *et al.*, 1986) and the nucleotide sequence of the gene has been determined (Lui and Parkinson, 1989). The PcnB protein shows considerable homology with the tRNA nucleotidyltransferase which is responsible for binding ATP, CTP and tRNA during the translation process (Masters *et al.*, 1990). Since it has been shown that the PcnB protein is required for the replication of plasmid R1, of which the copy number is also regulated by RNA transcripts, Masters *et al.* (1990) proposed that PcnB is involved along with Rop in the hybridisation of RNAI with RNAII.

B BROAD-HOST-RANGE PLASMIDS

Plasmids belonging to the IncC IncJ, IncN, IncP, IncQ and IncW groups are capable of replication in a wide range of Gram-negative bacteria (Pinkney and Thomas 1985). Plasmid pLS1 has been shown to replicate in some Gram-positive and Gram-negative bacteria (Lacks *et. a.*, 1986).

Plasmids belonging to the IncP and IncQ groups (also known as the IncP-1 and IncP-4 groups in *Pseudomonas*) have by far the greatest flexibility and are able to replicate in virtually all Gram-negative bacteria (Kues and Stahl, 1989). These broad-host-range plasmids differ from their narrow-host-range counterparts in that many more plasmid-encoded factors are involved in replication and the genes and sequences encoding these factors are dispersed throughout their genomes.

1.5 THE INCP PLASMIDS

The IncP plasmids may be divided into two major groups according to the similarity between their replication and transfer systems, namely IncP$^-$ and IncPß, represented by RK2 and R751, respectively (Chikami *et al.*, 1985). The best studied of these, RK2, is indistinguishable from other members of the IncPa group such as RP4, RP1, R18 and R68 (Smith and Thomas, 1989).

1.5.1 The basic replicon of RK2

RK2 is 56 kb in size (Meyer and Helinski, 1977) and is maintained at a copy number of 4 - 7 and 3 copies per chromosome in *E. coli* and *Pseudomonas aeruginosa* respectively (Grinter, 1984; Itoh *et al.*, 1984). The three major regions involved in replication are dispersed accross 20 kb on the RK2 genome (Thomas *et al.*, 1980) and are separated by a single copy of Tn*1* and genes conferring resistance to tetracycline (Pansegrau and Lanka, 1987) (Fig. 1.4).

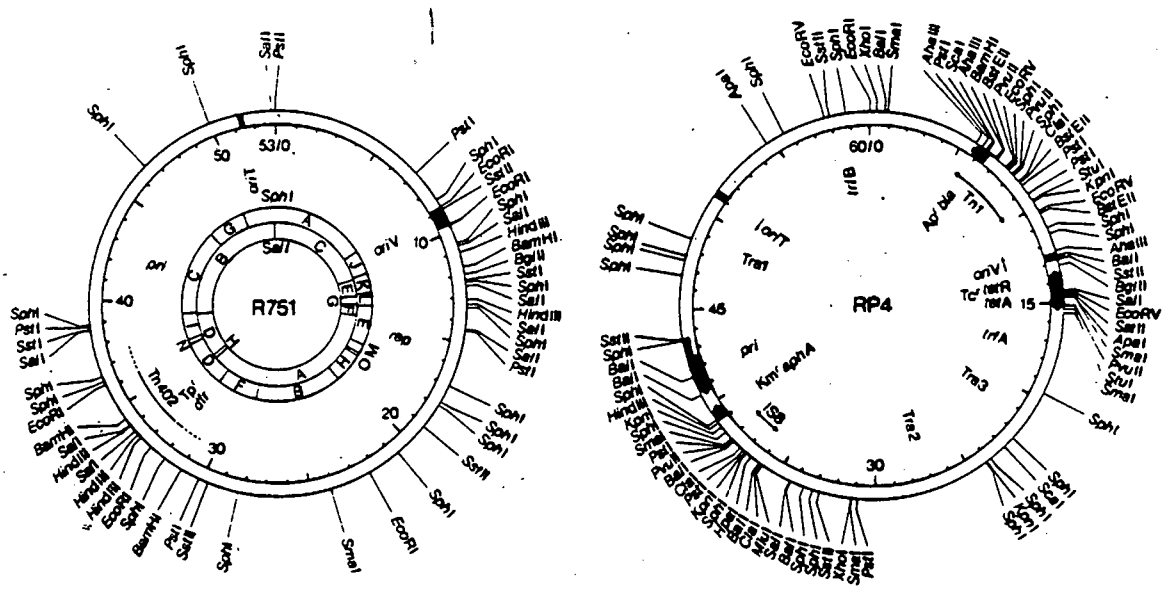

Fig. 1.4 Comparison of the restriction and functional maps of IncP plasmids RK2 and R751 (from Pansegrau and Lanka, 1987).

The first of these regions comprises the *ori*V while the others, which contain the *trfA* and *trfB* operons, act in *trans* on initiation (TrfA) and regulation (TrfA and TrfB) of replication at the *ori*V (Figurski and Helinski, 1979; Schreiner *et al.*, 1985; Theophillus *et al.*, 1985).

Only the *ori*V and the *trfA** operon (derived from the *trfA* operon by deletion of a regulatory determininant) are required for replication in *E. coli, P. aeruginosa, Pseudomonas putida, Azotobacter vinelandii* and *Alcaligenes eutrophus*, while additional regulatory determininants are required for replication in *Rhizobium meliloti, Agrobacterium tumefaciens, Caulobacter crescentus, Acinetobacter calcoaceticus* and *Rhodopseudomonas sphaeroides* (Schmidhauser and Helinski, 1985).

1.5.2 Structure of the *ori*V

The *ori*V is located on a 393 bp fragment (Thomas *et al.*, 1981), but sequences upstream and downstream of this region appear to modulate *ori*V activity so that the *ori*V actually occupies approximately 1000 bp between coordinates 12 and 13 kb on the standard map of RK2 as shown in Fig. 1.5 (Thomas *et al.*, 1984; Smith and Thomas, 1989).

The *ori*V region consists of a cluster of five 17 bp direct repeats separated by 22 - 23 bp spacer sequences (Stalker *et al.*, 1981) (direct repeats 5 - 9. Located upstream relative to the direction of replication is a single copy and a further cluster of three unevenly spaced repeats (repeats 1 - 4) which are all highly conserved and orientated in the same direction relative to each other. Two less well conserved repeats are located downstream of repeats 5 - 9 in both direct and inverted orientations (Smith and Thomas, 1985). Repeats 1 - 4 are involved in the determination of incompatibility and the control of copy number (Thomas *et al.*, 1981). When repeats 1 - 9 are present at high copy number, coresident RK2 replicons are rapidly lost from the population. The presence of repeats 1 - 4 results in a similar loss, while repeats 5 - 9 exert a weaker effect which is only significant on plasmids producing limiting amounts of the initiator protein, TrfA (Thomas *et al.*, 1981). Only repeats 5 - 9 are required for *ori*V activity while the role of the less conserved repeats downstream has yet to be investigated (Smith and Thomas, 1989).

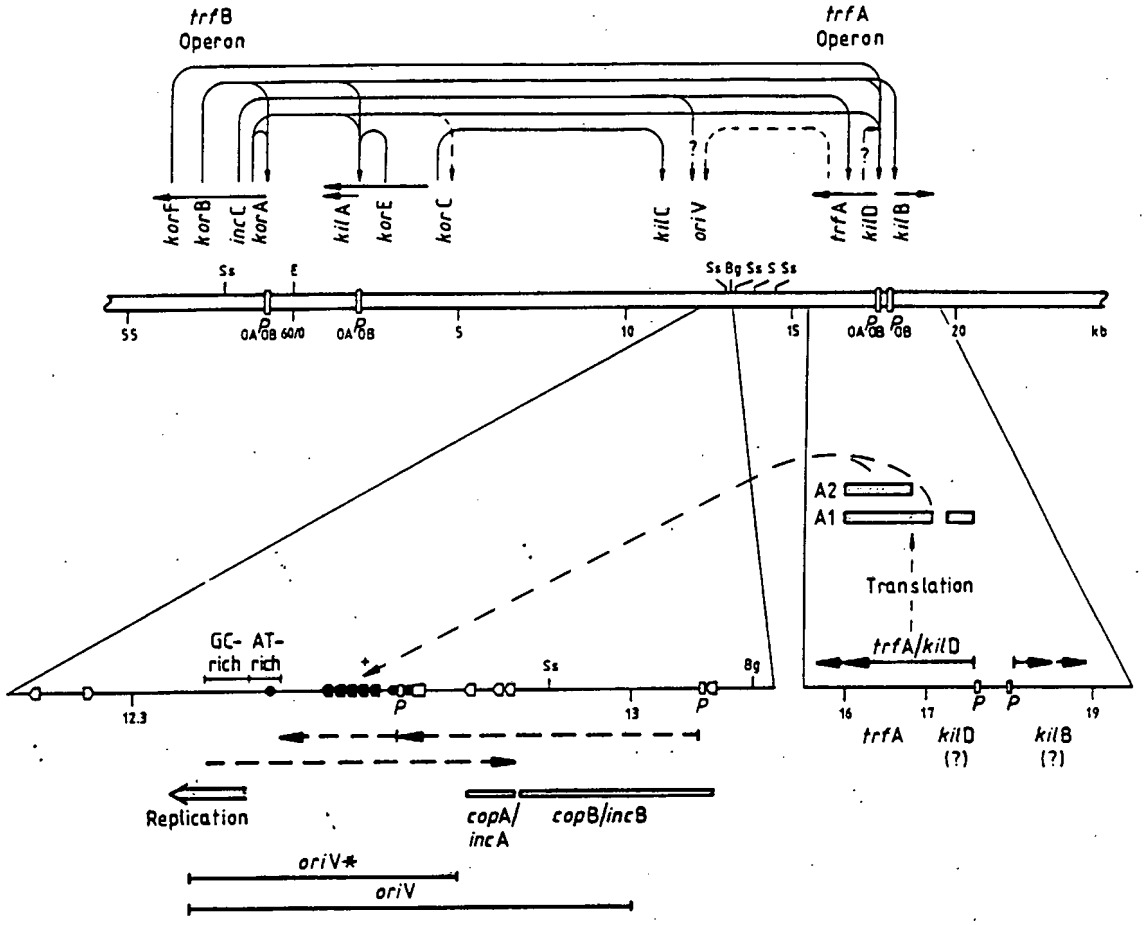

Fig. 1.5 Regions involved in initiation and regulation replication of RK2 . The loci involved in replication are shown in the upper half of the diagram (*trfA, trfB, incC,* incompatibility determinants and the *kil* and *kor* regulatory lelements. P_{OA}, P_{OB} and $_{OA}P_{OB}$ show promoters which contain operators for KorA, KorB or both KorA and KorB, respectively. Dashed and solid arrows indicate promoters which are negatively and positively controlled by *kil* and *Kor* elements, while question marks indicate proposed interactions. Restriction enzyme sites for *Bgl*II, *Eco*RI *Sal*I and *Sst*I are indicated by Bg, E,S and Ss respectively while the plasmid coordinates are shown in kb. Solid and empty arrows denote direct and inverse 17 bp repeats within the origin and the putitive promoter for the *trfA* products, A1 (P_{382}) and A2 (P_{285}).. A solid circle and empty square show DnaA and IHF binding sites while *P* together with an empty square denotes putitive promoters. The extent of the *ori*V required for replication in different hosts is indicated below the map (from Kues and Stahl, 1989).

A 49 bp A+T-rich sequence (74% A+T) followed by a 67 bp G+C-rich region (79% G+C) is located between repeats 9 and 10. These regions are conserved between the IncPa and IncPb plasmids. The G+C-rich region has much potential secondary structure and one inverted repeat which is highly conserved between RK2 and R751 has been thought to represent a primase recognition site involved in origin function. However, use of mutant M13 phages to detect such a site has proved unsuccessful (Smith and Thomas, 1989).

A pair of inverted DnaA boxes are located between repeats 2 - 4 and 5 - 9 and a third DnaA box is found within the A+T-rich region. All three sequences have been shown to bind DnaA protein *in vitro* (Gaylo *et al.*, 1987). The inverted DnaA boxes are essential for replication and deletion of the third, which is not conserved between the IncPa and IncPb plasmids does not appear to affect replication (Gaylo *et al.*, 1987). An additional feature of the *ori*V is the presence of an IHF binding site between repeats 2 - 4 and the inverted DnaA boxes which is flanked by A+T-stretches (Stalker *et al.*, 1981; Kues and Stahl, 1989).

Two putative promoter sites are present in the *ori*V region; the first overlaps with direct repeat 1 and the second is surrounded by the inverted DnaA boxes (Kues and Stahl, 1989). There are three open reading frames (ORFs) in the *ori*V, of which two (*copA/incA* and *copB/incB*) are involved in determination of incompatibility and control of plasmid copy number (Thomas *et al.*, 1984).

Different structural features of the *ori*V are required for replication of RK2 in different hosts. The 393 bp fragment which includes the pair of inverted DnaA boxes, direct repeats 5 - 9 and the A+T- and G+C-rich regions, are required for replication in *E. coli* and *P. putida*. Deletion of the DnaA boxes from this fragment does not affect *ori*V activity in *P. aeruginosa* (Schmidhauser *et al.*, 1983). Deletion of repeat 9 does not affect replication in *E. coli* but inactivates the *ori*V in *P. aeruginosa* and *P. putida* (Krishnapillai *et al.*, 1984) and the converse is true when repeat 5 is either removed or the adjacent sequences interfered with (Cross *et al.*, 1986).

1.5.3 Host proteins involved in the replication of RK2

Replication of RK2 in an *E. coli*-derived *in vitro* system is dependent on the presence of Mg^{++} ions, ATP, dNTPs, DnaA, DnaB helicase, and DnaG primase (Pinkney *et al.*, 1988). These workers also found a requirement of the system for DNA gyrase and DNA Pol III but not for RNA-Pol. Previously, Firshein *et al.* (1982) used a membrane-bound *in vitro* system and showed that replication of RK2 was dependent on RNA-Pol activity. However, it is now thought that this requirement was due to the need for expression of the *trfA* and *trfB* operons during initiation of replication and not the synthesis of ribonucleotide primers at the *ori*V (Smith and Thomas, 1989).

RK2 appears to have different requirements for DnaA in different hosts. Replication in *E. coli* and *P. putida* but not *P. aeruginosa* requires the inverted pair of DnaA boxes in the *ori*V (Cross *et al.*, 1986). This implies that the replicon relies on DnaA in some hosts but not in others. The presence of an IHF site in the *ori*V suggests that this protein may be involved in replication, but this remains to be confirmed.

1.5.4 The *trfA* operon

There are two cistrons in the *trfA* operon. The first is known as *kilD* and the second, *trfA*, codes for two proteins of 382 amino acids (43 kD) and 285 amino acids (32 kD). The TrfA proteins are translated from alternative translation starts within the same ORF (Smith and Thomas, 1986). At high concentration, both proteins exhibit non-specific DNA binding properties, but at low concentrations, they bind specifically to the *ori*V (Pinkney *et al.*, 1988). At least the larger protein, P_{382} and possibly the smaller protein, P_{285}, is required for replication in *P. aeruginosa*, while the smaller protein, P_{285} is sufficient to support replication in other bacteria (Durland and Helinski, 1987).

Transposon mutagenisis of the *trfA* operon has produced mutants which were unable to replicate in *E. coli* but could be maintained in *Pseudomonas* spp (Krishnapillai *et al.*, 1987). Several *trfA* mutants show species-specific temperature sensitivity which is possibly dependent upon divalent cation concentrations in *E. coli* (Smith and Thomas, 1989).

The intracellular concentration of the TrfA protein is not directly responsible for control of plasmid copy number. A two to three-fold increase in intracellular TrfA has been found to result in a 30% increase in plasmid copy number, but further increases in TrfA concentration do not elevate copy number further (Durland and Helinski, 1990).

Six independent mutations within the *trfA* operon which cause RK2 to replicate at an elevated copy number, are the result of base transitions within a 24 amino acid region of TrfA (Durland *et al.*, 1990). The effect of the mutations could be suppressed by the presence of wild-type TrfA, but this was directly dependent on the amount of mutant protein present. These results and those of Durland and Helinski, (1990) have lead Durland *et al.* (1990) to suggest that TrfA plays a regulatory role in replication which is not directly linked to the intracellular concentration of the protein

1.5.5 Regulation of replication

Expression of the TrfA proteins is regulated and modulated by the products of the *trfB* operon. The *korA* locus codes for two proteins of 37 kD and 30 kD (TrfB), which are expressed from two translational starts within the same ORF (Meyer and Hinds, 1982; Thomas, 1986). TrfB may either regulate initiation of replication at the *ori*V or alternatively modulate the expression of TrfA at the level of translation (Thomas, 1986).

The *trfA* operon is negatively regulated by the products of the *korA*, *korB* and *kor*F genes which are located along with the *incC* determinant in the *trfB* operon (Thomas, 1986). KorA and KorB contain αhelix-turn-αhelix motifs within their protein sequences which are typical of DNA binding proteins (Kornacki *et al.*, 1987) and have been shown to bind to operator sequences upstream of the -35 and -10 regions of the promoters which they regulate. KorF binds to an operator located just upstream from the transcriptional start of the *trfA* operon (Smith and Thomas, 1989). The product of *korE* also regulates expression of the *trfA* operon (Young *et al.*, 1987). The combined effect of these repressors is to reduce *trfA* expression to a level which is limiting for replication (Thomas and Hussain, 1984).

KorA, KorB, KorE and KorF form part of a regulatory system which enables RK2 to adapt to conditions in different hosts (Fig. 1.5) (Schreiner *et al.*, 1985). The regulation network consists of various *kil* determinants (which are potentially lethal to the host cells) and *kor* determinants which prevent the lethal effect of the *kil* gene products (Bechhofer *et al.*, 1986). There are *six kor* genes (*korA, korB, korC, korE* and *korF*) and four *kil* genes (*kilA, kilB, kilC* and *kilD*) (Young *et al.*, 1987).

KorA, a 101 amino acid protein, inhibits *kilA* expression and also expression of the *trfB* operon by autorepression (Theophillus *et al.*, 1985; Young *et al.*, 1985). KorA is a positive regulator of the *korC* gene (Young *et al.*, 1984). KorB, a 39 kD protein (Kornacki *et al.*, 1987), represses the *trfB* operon and also the *kilB* locus (Smith and Thomas, 1989). KorB also acts as a repressor in the presence of KorA on the expression of *kilA* (Young *et al.*, 1987). KorC is an 88 amino acid (9.15 kD) protein with an ahelix-turn-ahelix motif which in combination with KorA, acts as a repressor of the *kilC* and *kilE* genes (Kornacki *et al.*, 1990).

KorE is able to substitute for KorB as a corepressor of *kilA*, while it has a similar effect as KorB on the expression of *trfA* and *trfB* (Donoghue *et al.*, 1987; Thomas, 1988). KorA is able to counteract the activity of KorA/KorB on the *trfA* operon (Thomas, 1988) but the mechanism by which this occurs is not known. The molecular action of the *kil* genes is also still unclear.

The regulation system including multiple repression of the *trfA* operon is central to the ability of RK2 to replicate in a wide range of Gram-negative bacteria. Rearrangements within or deletions of regulator genes result in loss of the ability to replicate in specific hosts (Thomas *et al.*, 1982; Thomas, 1983; Barth *et al.*, 1984; Schmidhauser *et al.*, 1984; Schreiner *et al.*, 1985; Theophillus *et al.*, 1985).

1.6 THE INCQ PLASMIDS

The IncQ incompatibility group of plasmids consists of a number of small, low copy number (8 - 12 copies per chromosome in *E. coli*) plasmids which usually specify resistance to streptomycin (Sm^R) and sulfonamide (Su^R) (Barth and Grinter, 1974; Grinter and Barth, 1976; Frey and Bagdasarian, 1989). These plasmids are easily mobilised between a remarkably wide range of Gram-negative bacteria by conjugative plasmids belonging to the IncP group (Bagdasarian *et al.*, 1981; David *et al.*, 1981; Meyer *et al.*, 1982a). The wide host-range of the IncQ plasmids has resulted in their extensive use in the construction of broad-host-range cloning vectors (Bagdasarian *et al.*, 1982; Bagdasarian and Timmis, 1982; Frey *et al.*, 1983; Priefer *et. al.*, 1995).

The most prominent members of of the IncQ group are the probably identical plasmids RSF1010 and R300B, which were independently isolated from *E. coli* (Guerry *et al.*, 1974) and *Salmonella typhinurium* (Barth and Grinter, 1974), respectively. The very similar R1162 was isolated from *P. aeruginosa* (Bryan *et al.*, 1972) The replication of RSF1010 and R1162 has been the subject of intense study and results from experiments with these plasmids will be discussed in detail.

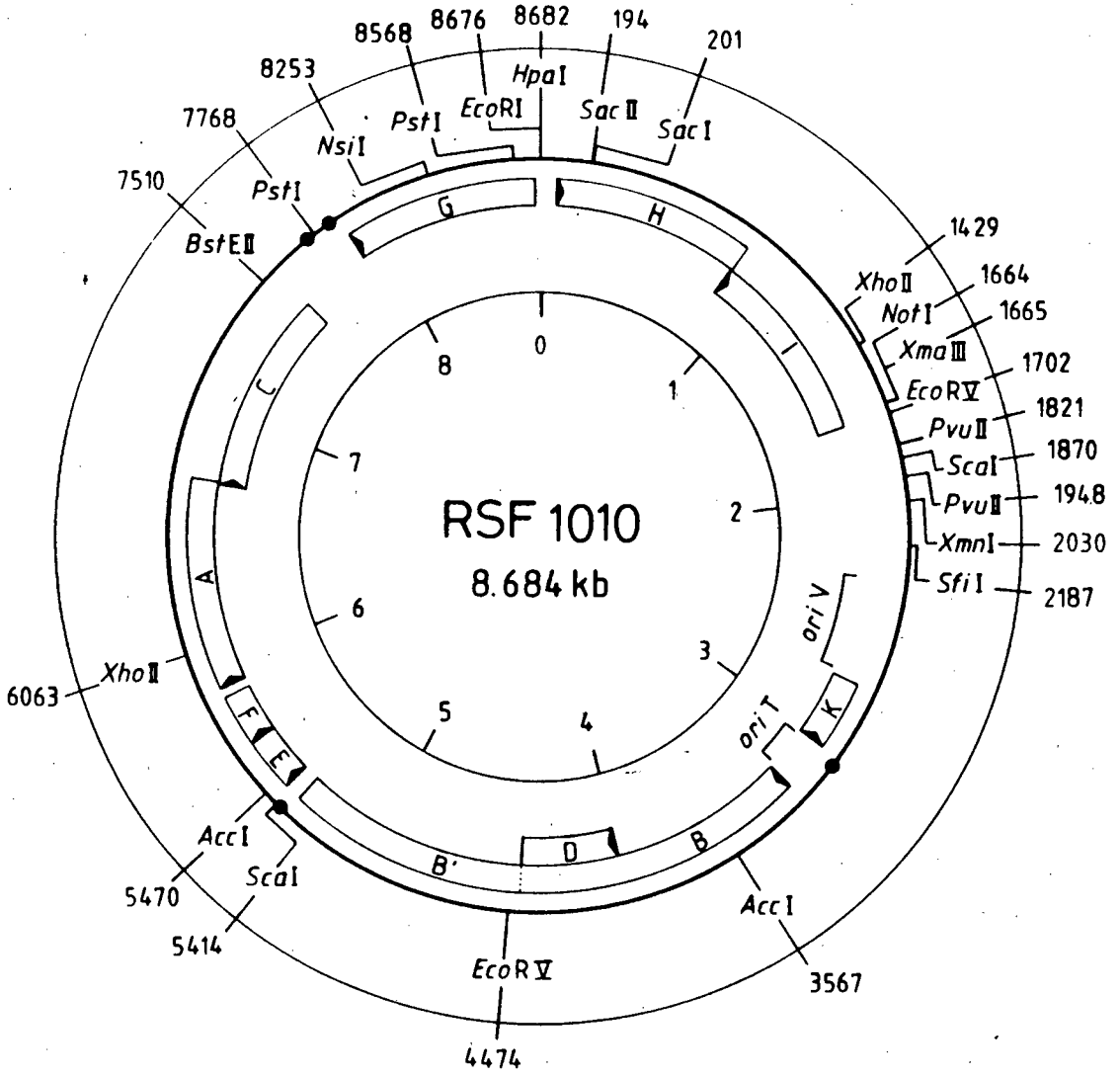

Fig. 1.6 Physical and genetic map of RSF1010. Sites for restriction endonucleases which cut the plasmid once or twice are indicated on the map while A, B/B' and C correspond to *repA, repB/B'*Physical and genetic map of RSF1010. Sites for restriction endonucleases which cut the plasmid once or twice are indicated on the map while A, B/B' and C correspond to *repA, repB/B'* and *repC* respectively. The positions of 7 other ORFs are indicated, with black circles showing putative promoters. Also shown are the *ori*V and origin of transfer (*ori*T) (from Haring and Scherzinger, 1989).

1.6.1 Structure of the IncQ replicon

RSF1010 and R1162 have been reported to be 8684 bp and 8.7 kb in size respectively (Scholz *et al.*, 1989; Meyer *et al.*, 1982a). The entire nucleotide sequence of RSF1010 (Scholz *et al.*, 1989) and part of R1162 (Meyer *et al.*, 1985) has been determined and this, along with *in vivo* and *in vitro* complementation (deGraaf *et al.*, 1978; Scherzinger *et al.*, 1984; Haring *et al.*, 1985; Kim and Meyer,1985) has resulted in a

detailed genetic map of the plasmids. A comparision of these sequences indicates that the plasmids are virtually identical. The maps and nucleotide sequence for RSF1010 and R1162 have been reported in opposite orientation and for the purpose of simplicity, all map coordinates will specifically refer to the map of RSF1010 (Fig. 1.6). The genes conferring Sm^R and Su^R resistance are located between coordinates 7.8 kb and 1 kb relative to the unique *Eco*RI restriction site and are transcribed from a promoter at 0.9 kb (Bagdasarian *et al.*, 1981; Meyer *et al.*, 1985). The mobilisation functions are found between 3.0 kb and 4.3 kb with the origin of conjugative transfer (*ori*T) located at 3.1 kb (Nordheim, *et al.*, 1980; Meyer *et al.*, 1982a; Derbyshire *et al.*, 1987). The *ori*V is located at coordinate 2.5 kb and the region required for replication lies between 4.4 kb and 7.6 kb (Scherzinger *et al.*, 1984; Meyer *et al.*, 1985).

1.6.2 Structure of the *ori*V

The *ori*V of RSF1010 and R1162 (Fig. 1.7) is located on contiguous *Hpa*II fragments of 370 bp and 210 bp (Meyer *et al.*, 1985). The determinants for incompatibility are found on the 370 bp fragment of which 166 bp distal to the junction between the two fragments may be deleted without any effect on replication (Meyer *et al.*, 1985; Persson and Nordström, 1986). The two fragments appear to encode two domains, of which the relative orientation can be altered without affecting replication *in vivo*. An increase of greater than 2 kb in the distance between the two domains reduces the rate of replication *in vitro* (Kim *et al.*, 1987).

Analysis of the DNA sequence of the *ori*V (Meyer *et al.*, 1985; Persson and Nordström, 1986) has revealed several striking features. The most prominant are three perfectly conserved 20 bp direct repeats which are separated by nonconserved 2 bp spacer sequences all located on the 370 bp fragment. These repeats have been implicated in the determination of plasmid incompatibility and the control of copy number (Persson and Nordström, 1986; Lin *et al.*, 1987). There is convincing evidence that the 20 bp repeats act as binding sites for an initiator protein (Haring *et al.*, 1985) which is believed to be responsible for incompatibility and the control of plasmid copy number (Persson and Nordström, 1986; Lin *et al.*, 1987). Directly

adjacent to the repeats lies a 28 bp G+C-rich region followed by a 31 bp A+T-rich region. In RSF1010, deletion of the region between the repeats and the 210 bp *Hpa*II fragment does not affect replication *in vitro* (Haring and Scherzinger, 1989) implying that the G+C- and A+T-rich regions are not required for replication in *E. coli*.

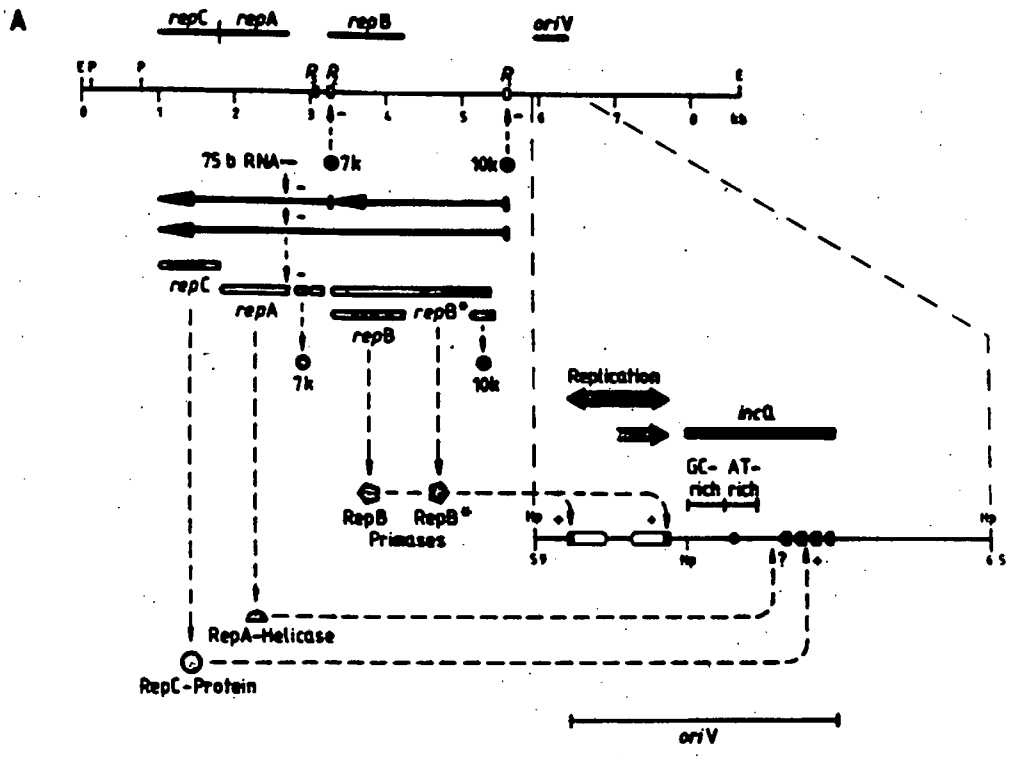

Fig. 1.7 The basic replicon of RSF1010. The extent of essential *rep* genes and of the *ori*V is shown by black lines above the map. *P1*, *P2* and *P3* show the positions of putitive promoters, while horizontal arrows show the extent and direction of the *rep* transcripts. In the expanded section, which shows the *ori*V, thick arrows represent three and one half 20 bp repeated sequences while two empty arrows denote the palindromic sequence. A circle indicates a nonfunctional DnaA box and A+T-rich and G+C-rich areas are shown above the map. Symbols E, Hp and P denote recognition sites for restriction enzymes *Eco*RI, *Hpa*II and *Pst*I (from Kues and Stahl, 1989).

The 210 bp fragment has a large 152 bp complementary inverted repeat (Meyer *et al.*, 1985; Scholz *et al.*, 1989) which has the potential to form a stem-and-loop structure with a stem of 40 bp and an intervening loop of 37 bp (Meyer *et al.*, 1985). The presence of dA clusters interrupted by G+C stretches in turn with the helix as well

as the presence of purine-pyrimidine stretches adjacent to the stem-and-loop have been proposed by Kues and Stahl (1989) to support the cruciform structure which these sequences have the potential to form. Deletion of one-half of the palindromic sequence results in loss of replication *in vitro* (Kim *et al.*, 1987) implying that the secondary structure may be important in the replication process.

Replication is initiated bidirectionally or unidirectionally (deGraaf *et al.*, 1978) from within the 210 bp fragment (Kim *et al.*, 1987) at two conserved 10 bp sequences at the base of the stem-and-loop structure formed by the 152 bp palindromic sequence (Lin and Meyer, 1987). These repeats encode two plasmid-specific ss DNA initiation signals designated *ori*L or *Ssi*A and *ori*R or *Ssi*B repectively (Honda *et al.*, 1988; Haring and Scherzinger, 1989).

1.6.3 Dependence of IncQ replication on host-encoded proteins

The requirement of RSF1010 replication for host-encoded proteins has been tested *in vivo* in *E. coli* (Diaz and Staudenbauer, 1982; Scherzinger *et al.*, 1984; Haring and Scherzinger, 1989). Replication of RSF1010 requires DNA gyrase and the DnaZ protein (gamma subunit of the Pol III holoenzyme). The absence of DnaA, DnaB helicase, DnaC, DnaG primase, DnaT (i protein) and RNA-Pol have no effect on the replication of RSF1010. The plasmid can also replicate independently of certain mutations in the *polA* gene (Grindley and Kelly, 1976).

1.6.4 Plasmid-specific proteins

Three plasmid-encoded proteins are essential for replication of the IncQ plasmids *in vitro* and are coded for by genes located at a distance from the *ori*V (Fig. 1.7) (Scherzinger *et al.*, 1984; Meyer *et al.*, 1985).

The *repB* gene product, RepB, is found in two molecular forms. The larger form, RepB (75.9 kD), cross-reacts with antibodies raised against the smaller product, RepB' (35.9 kD) (Haring and Scherzinger, 1989). DNA sequence analysis data has

shown that both proteins are translated from the same ORF, with translation of the smaller protein starting from the middle of the ORF (Scholz *et al.*, 1989). There is an absolute requirement for either RepB or RepB' for replication of RSF1010. Deletion of part of *repB* leaving the region encoding RepB' intact has no effect on replication (Scherzinger *et al.*, 1984).

RepB' occurs as a monomer *in vivo* and has been shown to have DNA primase activity. The protein can synthesise a DNA primer *de novo* (Haring and Scherzinger, 1989), independently of the host-encoded primase or RNA-Pol. RepB' is directly responsible for initiation of replication at the *ori*V by the synthesis of primers at both or either *ori*L and *ori*R (Haring and Scherzinger, 1989, Honda *et al.*, 1989).

RepA and RepC are encoded by overlapping genes situated downstream from *repB* (Scherzinger *et al.*, 1984; Meyer *et al.*, 1985; Scholz *et al.*, 1989). The *repA* gene codes for a 30 kD protein (Scherzinger *et al.*, 1984) which functions as a DNA helicase (Haring and Scherzinger, 1988). Activity of this protein is dependent on ATP hydrolysis and is stimulated by the presence of ss DNA. The RepA protein occurs as a hexamer of 164.3 kD *in vivo* and shows structural but little sequencev similarity with the DnaB (helicase) protein of *E. coli* and Phage P22 gene 12 (which has helicase activity) (Haring and Scherzinger, 1989).

RepC is a DNA binding protein of 30.98 kD (Scholz *et al.*, 1989) which occurs *in vivo* as a dimer of 60.6 kD (Haring and Scherzinger, 1989). The copy number of RSF1010 and R1162 has been found to be dependent on the concentration of RepC which binds specifically to the 20 bp repeats of the *ori*V (Haring *et al.*, 1985; Kim and Meyer, 1985). Analysis of the interaction of RepC with the *ori*V by electron microscopy has revealed that the proteins bind specifically and co-operatively to the duplex DNA in the region of the three repeats and then bind non-specifically to a sequence 500 bp downstream to form a loop structure (Haring and Scherzinger, 1989).

The Rep proteins are expressed from three promoters, namely p1, p2 and p3 (Fig. 1.7) (Bagdasarian *et al.*, 1982). Promoter p1 lies upstream from RepB/RepB' and transcription results in a transcript which includes RepB, RepA and RepC (Bagdasarian *et al.*, 1987). Deletion of a 10 kD protein which is transcribed from an ORF situated just downstream of p1 results in an increase in transcription from p1 and also the loss of the ability to replicate in *P. putida* and *P. aeruginosa* (Frey and Bagdasarian, 1989). Transcription from p2 results in a transcript which includes RepA and RepC, and activity of the promoter is regulated by a 7.6 kD protein produced by gene *cac* which is located downstream of p2 and precedes *rep*A (Frey and Bagdasarian, 1989; Scholz *et al.*, 1989). Deletion of this protein results in elevated plasmid copy number in *E. coli* although the changes in copy number in *P. aeruginosa* are not of the same order of magnitude. Deletion of the regulators of p1 and p2 may lead to runaway replication of the RSF1010 replicon which could result in the death of transformants containing these plasmids (Frey and Bagdasarian, 1989).

1.6.5 A model for the initiation of replication of the IncQ replicon

Haring and Scherzinger (1988) have proposed the following model for the initiation of replication from the *ori*V of RSF1010. Firstly, RepC forms a complex with the duplex DNA in the vicinity of the 20 bp repeats in the *ori*V. Subsequent binding of RepA to the repeats is co-operative and results in the interaction of the complex with a second site on the DNA at about 500 bp downstream, forming a duplex loop structure which contains *ori*L and *ori*R.

Secondly, the duplex DNA in the loop is unwound to enable RepB' to synthesise a DNA primer at *ori*L and *ori*R. Loop formation could result in localised unwinding exposing ss DNA where the RepA hexamer binds and by means of its helicase activity, opens the remainder of the ds loop. No interaction has as yet been detected between RepA and RepC, but co-expression of the two proteins suggests that RepC serves to focus RepA on the *ori*V. Alternatively, gyrase could unwind the double stranded DNA within the loop, the structure of which serves to stabilise the single stranded DNA.

In the third step, RepB' synthesises a DNA primer at *ori*L and *ori*R and the Pol III holoenzyme begins the polymerisation of the DNA. RepA functions as a helicase, unwinding the DNA ahead of the replication fork.

1.7 *THIOBACILLUS FERROOXIDANS*

Thiobacillus ferrooxidans is an industrially important acidophilic chemoautotroph used in the bioleaching of certain minerals from their ores. This process involves the solubilisation of metals from low grade ores which would normally be unworkable if conventional chemical methods were used (Brierly, 1982). Bioleaching is considered more energy efficient than the conventional chemical leaching process, but a major drawback is that the bioleaching process takes far longer.

T. ferrooxidans is a Gram-negative, rod-shaped bacterium which fixes atmospheric nitrogen and carbon and uses either the oxidation of ferrous to ferric iron or the reduction of sulphur compounds to sulphuric acid as an energy source. The bacterium generally uses O_2 as an electron acceptor, but under anaerobic conditions, ferric iron may be used for the oxidation of reduced sulphur compounds (Suglio *et al.*, 1985; Suzuki *et al.*, 1990). Sulphuric acid, which is a by-product of these reactions, lowers the enviromental pH, favouring the acidophilic *T. ferrooxidans* (pH 1.5 - 3.5).

This organism is highly adapted to its unusual enviroment and it is this unique physiology which makes it particularly interesting for molecular biologists. In addition, the industrial importance of *T. ferrooxidans* makes it a candidate for genetic manipulation to improve its industrial potential (Rawlings, 1990). The major problem with a genetic investigation into the molecular biology of this organism is the lack of a suitable genetic system by which recombinant DNA may be introduced into the *T. ferrooxidans* cells. A study of the plasmids and plasmid-encoded genes found in *T. ferrooxidans* is the first step into the development of a genetic system for

the organism. Plasmid vectors based on replicons which are isolated from the bacterium are the most obvious candidates for cloning vehicles.

1.7.1 Plasmids in *T. ferrooxidans*

A number of plasmids have been identified and isolated from different *T. ferrooxidans* strains (Martin *et al.*, 1981; Martin *et al.*, 1983; Rawlings *et al.*, 1983). Attempts to assign phenotypic markers such as uranium resistance to these plasmids in *T. ferrooxidans* have proved fruitless (Martin *et al.*, 1983). Four plasmids from *T. ferrooxidans* strains 36, FC and 33020 have been cloned into the pBR325 and pBR322 vectors (Rawlings *et al.*, 1984; Holmes *et al.*, 1984; Rawlings and Woods, 1985).

T. ferrooxidans strain FC, an arsenic resistant strain isolated from the acid leach liquor of a South African mine (Rawlings *et al.*, 1983), was found to contain three plasmids. Of these, the most prevalent was pTF-FC2, a 12.4 kb plasmid, which was cloned into pBR325. The resulting construct pDER401, was cryptic in *E. coli*, but deletion of the pBR325 origin of replication revealed that pTF-FC2 possessed a broad-host-range origin (Rawlings *et al.*, 1984). This construct, pDER412, has been shown to be stably maintained in a number of Gram-negative bacteria such as *P. aeruginosa* (Rawlings *et al.*, 1986a), *Thiobacillus novellus* (Rawlings *et al.*, 1986b), *Salmonella typhimurium*, *Klebsiella pneumoniae* and *Agrobacterium tumefaciens* (Rawlings, unpublished results). The plasmid is easily mobilised between *E. coli* and other Gram-negative bacteria by members of the IncP group (Rawlings *et al.*, 1986b; Rawlings and Woods, 1985).

The small size and promiscuous nature of pTF-FC2 makes it one of relatively few broad-host-range plasmids which have been identified to date. In addition to its potential as a cloning vehicle for genetic manipulation of *T. ferrooxidans*, an investigation of the molecular biology of its replicon may contribute to the understanding of the factors which determine broad-host-range replication of plasmids.

The purpose of this study was to identify and characterise the basic replicon of pTF-FC2 with the view to the eventual construction of a cloning vector for the introduction of genes into *T. ferrooxidans*. Described in this dissertation is the functional and genetic characterisation of the basic replicon of pTF-FC2. An analysis of the nucleotide sequence revealed DNA sequences and proteins which are involved in the replication of the plasmid in its different hosts. A comparison of the basic replicon with those of other broad-host-range plasmids has shown a number of structural similarities between pTF-FC2 and the replicons of the IncQ plasmids. These similarities were used to assign functions to the identified features of the pTF-FC2 basic replicon.

Results included in this dissertation have been accepted for publication in the Journal of Bacteriology (Dorrington and Rawlings, 1989; 1990).

CHAPTER 2

DELINEATION AND FUNCTIONAL CHARACTERISATION OF THE REGIONS REQUIRED FOR REPLICATION OF pTF-FC2

CHAPTER 2

DELINEATION AND FUNCTIONAL CHARACTERISATION OF THE REGIONS REQUIRED FOR REPLICATION OF pTF-FC2

2.1 SUMMARY

Plasmids containing the pTF-FC2 origin were shown to replicate independently of DNA polymerase I (Pol I) in *Escherichia coli*. This feature was used to distinguish between replication from the *Thiobacillus ferrooxidans* origin and the pMB1-derived origins of the vectors. A 3.2 kb *Sau*3A partial fragment which had retained the ability to replicate in the *E. coli polA*$^-$ mutant and also in *Pseudomonas aeruginosa*, was obtained. A series of deletions of this fragment was used to identify the minimal replicon, the *ori*V and the areas determining plasmid incompatibility. The copy number of pTF-FC2 in *E. coli* was estimated at 12-15 copies per chromosome and a deletion was identified which allowed replication at a reduced copy number. An area which affected the ability of the replicon to replicate in *P. aeruginosa*, was identified.

2.2 INTRODUCTION

In order to understand the replication mechanism of a broad-host-range plasmid, it is necessary to functionally characterise the regions required for replication by identifying areas involved in the determination of plasmid copy number, incompatibility and the ability to replicate in various hosts.

The delineation of the basic replicon involves the testing of various deletions for their ability to replicate autonomously in a particular host. The plasmid *ori*V comprises the region required in *cis* for replication, provided that the diffusible products such as proteins are provided in *trans* on a coresident plasmid (Figurski and Helinski, 1979). The major problem in the identification of the *ori*V is that the sequence does not provide a functional origin for plasmid maintenance. If a functional origin from another plasmid is supplied, the problem remains of distinguishing between replication initiated at the vector origin of replication from the test *ori*V.

There are a number of ways in which the *ori*V may be identified. The first is to construct plasmids containing the *ori*V, from which the vector origin may be easily deleted leaving the test *ori*V and an antibiotic marker (Meyer *et al.*, 1985). Transformation of these constructs into strains containing the wild-type replicon allows replication of the satellite plasmids. Purification of these plasmids is, however, difficult, as they only replicate in the presence of the wild-type plasmid.

Other workers such as Shon *et al.* (1982) and Lane *et al.*, (1984), have overcome this problem by exploiting the ability of certain plasmids such as R6K and F to replicate in *E. coli polA*$^-$ mutants while others, such as the pMB1/ColE$_1$-based origins are unable to function in these mutants (Kingsbury and Helinsky, 1970). Deletion fragments carrying the *ori*V of Pol I-independent plasmids are cloned into vectors based on Pol I-dependent origins facilitating plasmid maintenance. Activity of the Pol I-independent *ori*V is then tested by transforming the constructs into a *pol*A$^-$ mutant containing the wild-type replicon.

A number of different *polA*$^-$ mutations of *E. coli* have been characterised, each with different effects on the activity of the enzyme (Grindley and Kelley, 1976) and some of which are temperature sensitive (ts). Grindley and Kelley (1976) found that the ability of replicons to be stably maintained was dependent on the *polA* allele. The advantage of using ts *polA*$^-$ mutants is that they allow for distinction between plasmid establishment and maintenance. This is achieved by allowing the plasmid to be established by the Pol I-dependent vector origin at the permissive temperature

and then shifting the culture to the non-permissive temperature to test for the ability of the Pol I-independent origin to maintain its copy number. The problems of this system are firstly that the temperature does not always inactivate the enzyme completely and a low level of Pol I-dependent replication can occur. Secondly, if the copy number at which the vector origin replicates is much higher than that of the Pol I-independent origin, a drop in the maintenance ability of the latter origin may not be detectable. Thus a combination of *polA*⁻ mutants should be used to analyse the *ori*V.

This chapter details the functional characterisation of the pTF-FC2 replicon. A series of deletions was used to identify the minimal replicon. Regions that affected replication were identified using the ability of the pTF-FC2 replicon to replicate in *polA*⁻ *E. coli* mutants. In this way, sequences involved in the determination of incompatibility and plasmid copy number in *E. coli* were mapped on the basic replicon. In addition, a region which affected the ability of pTF-FC2 to replicate in *P. aeruginosa* was identified.

2.3 MATERIALS AND METHODS

2.3.1 Bacterial strains and plasmids

Bacterial strains used in the course of this work are listed in Appendix 1. They include the *E. coli* strain LK111 which was used for plasmid maintenance (Zabeau *et al.*, 1982). Strains GW125a and AB1157(*polA*12) (gifts from Graham Walker, M.I.T) are *polA*⁻ derivatives of *E. coli* AB1157 (Appendix 1) carrying the *polA*1 and *polA*12 alleles respectively, which were used to test for Pol I-independent plasmid replication. Plasmids were maintained in AB1157(*polA*12) by incubation at 29 °C and then shifted to the non-permissive temperature of 42 °C to test for Pol I-independent replication. *P. aeruginosa* PAO1162 was used to test for broad-host-range replication of plasmids. Cloning vectors and plasmids not originating from this study are described in Appendix 2 while plasmids generated during the course

of this chapter are shown in Figs. 2.1 and 2.2. All plasmids constructed during the course of the work described in this dissertation are listed and described in Appendix 3.

2.3.2 Media and general recombinant DNA techniques

Plasmid DNA was prepared from *E. coli* and *P. aeruginosa* strains using the method of Ish-Horowicz and Burke (1981) (Appendix 4). The DMSO-transformation method of Chung and Miller (1988) (Appendix 4.2) was used to transform *E. coli* strains. Transformation rates for strains GW125a and AB1157(*polA*12) were routinely in excess of 10^4 colonies per μg of plasmid DNA. The method used to transform *P. aeruginosa* PAO1162 is described in Appendix 4. Unless otherwise stated, growth conditions for *E. coli* and *P. aeruginosa* cultures and concentrations of the relevant antibiotics were as described in Appendix 4. General recombinant DNA techniques were essentially as described by Maniatis *et al.* (1989). Sequential deletions were constructed from pTV400 using the Exonuclease III-shortening method of Henikoff (1984). DNA probes were nick-translated using the Amersham Nick-Translation Kit and the hybridisation conditions were as described in Maniatis *et al.* (1989).

2.3.3 Determination of plasmid copy number

All plasmids tested for copy number were grown in the *polA*$^-$ *E. coli* strain, GW125a. Two methods were used to determine plasmid copy number. In the first, total DNA was isolated from cells carrying the relevant plasmids (Appendix 4). Precise quantities of total DNA were digested with *PstI* restriction endonuclease and the fragments separated by electrophoresis in agarose gels. The gels were blotted and hybridized to nick-translated pDER412. The intensity of the hybridization signals was compared with that of signals obtained from known concentrations of plasmid. The quantity of plasmid relative to the total *E. coli* chromosomal plus plasmid DNA was calculated using the relative sizes of the *E. coli* chromosome of 4000 kb (Watson *et al.*, 1987) pDER412 (14.7 kb) and pTV100 (8 kb). In this way it was possible to determine the number of plasmids per chromosome.

In the second method, plasmid copy number relative to pTV100 was determined using the single cell resistance to ampicillin (Ap) method of Nordström *et al.* (1980). Bacteria were grown overnight in Luria broth (LB) (Appendix 4) containing Ap at 100 μg/ml and 10^5, 10^6 and 10^7-fold dilutions were spread onto Luria agar (LA) (Appendix 4) plates containing 400, 600, 800, 1000 and 1200 μg/ml Ap. Colonies were counted after incubation overnight and the amount of Ap required to give a 50% survival rate was used as a relative measure of plasmid copy number.

2.3.4 Test for incompatibility

The ability of fragments of pTF-FC2 cloned in pUC19 or Bluescript SK which is ampicillin resistant (Ap^R) to displace pDER412 which is resistant to chloramphenicol (Cm^R), was used as the test for incompatibility. Plasmids to be tested were transformed into *E. coli* LK111 containing pDER412 and transformants plated on LA containing Ap. Six colonies from each transformation were streaked onto separate Ap plates to give single colonies. Ten colonies from each plate were toothpicked onto LA plates containing Cm. Assuming that there are approximately 10^9 cells in a single colony after incubation overnight and that 10^3 cells represent 10 generations from a single cell, then each cell in a colony will be the result of 30 generations of cell division. The cells tested on Cm would therefore have undergone 60 generations since transformation without selection pressure for pDER412.

2.4 RESULTS

2.4.1 General functional map of pTF-FC2

A restriction map of pTF-FC2 is shown in Fig. 2.1. Plasmid pTF-FC2 is 12.4 kb in size and contains single recognition sites for *Pst*I, *Mlu*I, *Xho*I, *Kpn*I, *Eco*R1, *Apa*I, *Cla*I, *Nde*I, *Pvu*II and *Stu*I restriction endonucleases. The area determining the

mobilisation function is located in the region between the *Nde*I and *Cla*I siltes (Rawlings and Woods, 1985; DE Rawlings, unpublished results; J Rohrer, unpublished results) while Rawlings *et al.* (1986b) located the origin of replication between the *Cla*I and *EcoR*1 sites by construction of various deletion plasmids from pDER412 (Fig. 2.1).

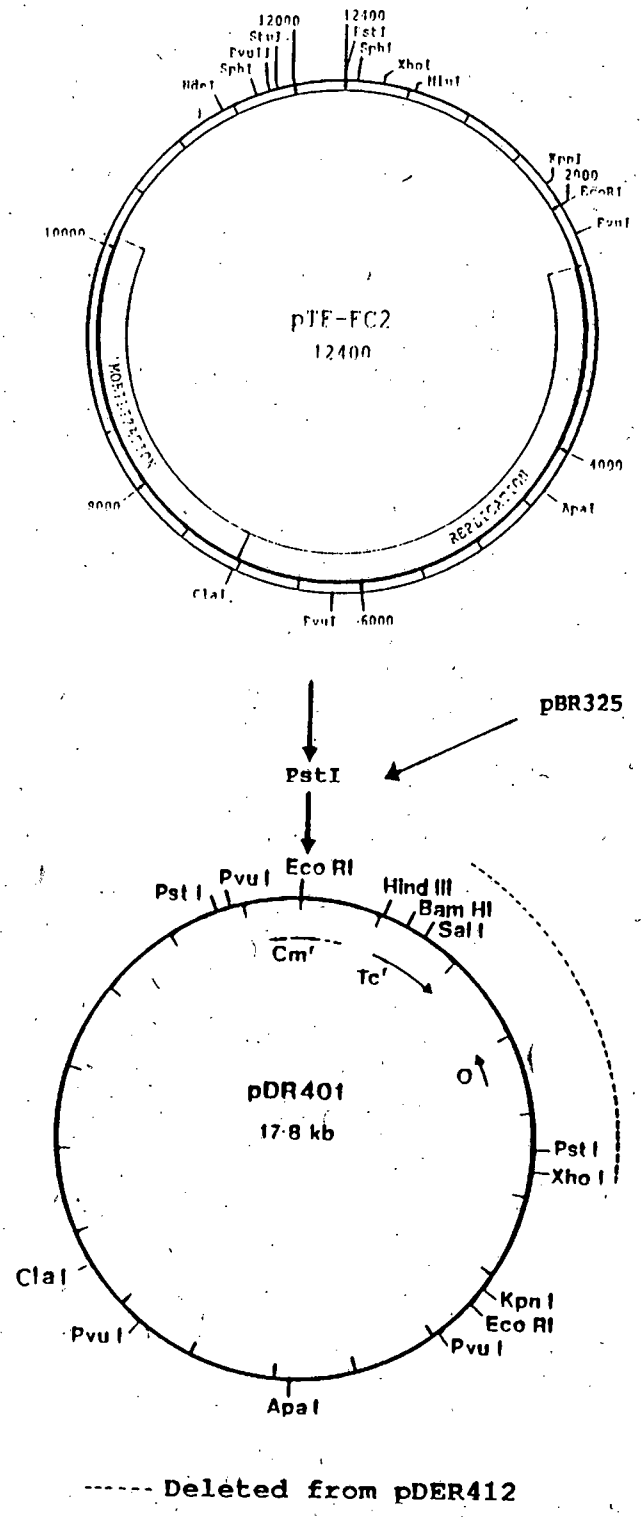

Fig. 2.1 Restriction and functional map of pTF-FC2 and its derivatives pDER401 and pDER412. The regions involved in mobilisation and replication are indicated on the maps.

2.4.2 Replication of pTF-FC2 in *E. coli*

2.4.2a DNA polymerase I-independence

To investigate whether replication of pTF-FC2 was independent of *E. coli* DNA Pol I, plasmids pBR325, pDER401 and pDER412 were transformed into the *polA1* mutant GW125a. Transformants were selected by growth on LA containing Cm and the presence of the relevant plasmid was confirmed by restriction analysis. Only pDER401 and pDER412, which contain pTF-FC2 plasmid DNA were able to replicate in GW125a. Plasmids based on the pMB1/ColE1 origin such as pBR325 and related plasmids, are unable to replicate in *polA*1 mutants (Kingsbury and Helinsky, 1970). Since the replication of the pTF-FC2 plasmids was Pol I-independent, the ability to replicate in GW125a could be used to identify regions of pTF-FC2 involved in replication.

The ability of the pTF-FC2 origin to replicate in the temperature-sensitive *polA*12 mutant was investigated. Plasmids pUC19 (Appendix 2) and pDER412 were transformed into AB1157(*polA*12) and transformants incubated at the non-permissive temperature of 42 °C. Colonies were obtained from cells transformed with pDER412 after 24 h of incubation and cells transformed with pUC19 produced very small colonies after 36 h. Since the distinction between Pol I-independence and -dependence was not as clear in AB1157(*polA*12) as had been shown in GW125a, it was decided to use GW125a to test for Pol I-independent replication.

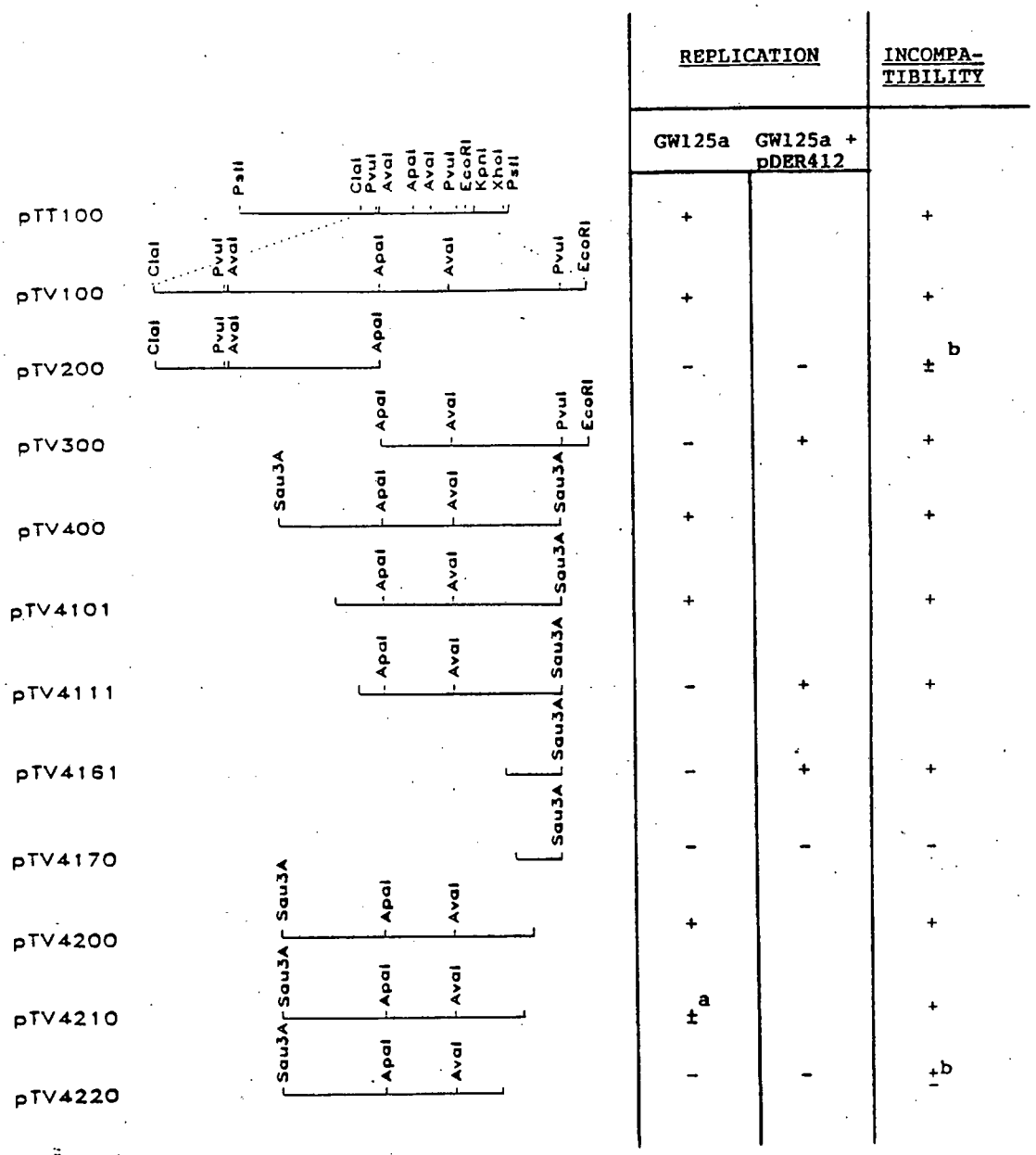

Fig. 2.2 Deletions of pTF-FC2 used in defining areas required for replication, the *ori*V and the incompatibility determinants in *E. coli*. Only pTF-FC2-derived fragments are shown.

[a] indicates colonies of reduced size
[b] indicates partial plasmid incompatibility

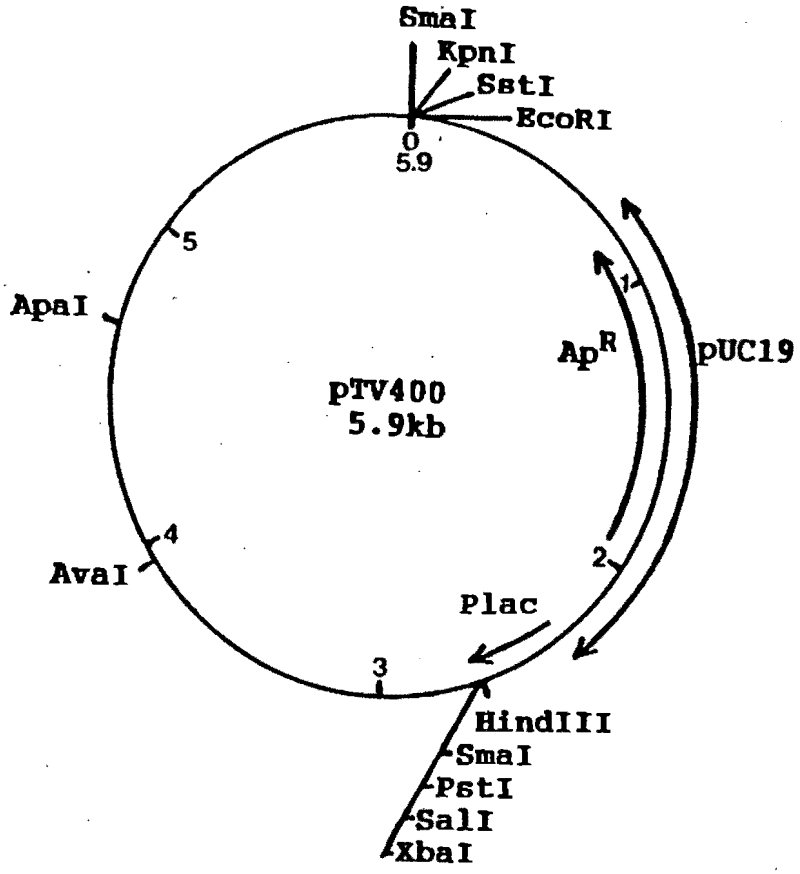

Fig. 2.3 Restriction map of pTV400.

2.4.2b Determination of the minimal replicon of pTF-FC2

Plasmid pTT100 (Fig. 2.2) was constructed by cloning pTF-FC2 into the *Pst*I site of the vector Bluescript SK (Appendix 2). The 5 kb *Cla*I - *EcoR*1 fragment from pTT100 was subcloned into Bluescript SK and the resulting plasmid, pTV100 (Fig. 2.2), was able to replicate in GW125a. As the *Apa*I site cut the pTF-FC2 sequence into two fragments of approximately equal size, it was tested whether the basic replicon lay on either of these two fragments. Deletion of the *Cla*I - *Apa*I fragment (pTV300) or the *Apa*I - *EcoR*1 fragment (pTV200) resulted in the inability of the plasmids to replicate in GW125a (Fig. 2.2). The sequences on both sides of the *Apa*I restriction site are therefore part of the pTF-FC2 minimal replicon. Plasmids pTV200 and pTV300 were transformed into GW125a containing pDER412 to identify the fragment which contained the sequence required in *cis* for replication of pTF-FC2. Only pTV300 was able to replicate, implying that the *ori*V was located on the *Apa*I - *EcoR*I fragment.

Due to the lack of 6 bp restriction enzyme recognition sites within the *Cla*I - *Eco*RI fragment, a *Sau*3A-partial library of pTV100 was constructed using the *Bam*HI site of the vector pUC19 and the recombinants transformed into GW125a. Several plasmids were isolated of which the smallest was pTV400 (Figs 2.2 and 2.3). Plasmid pTV400 contained a 3.2 kb insert which included the *Apa*I site. The insert was orientated in such a way that the *Apa*I site lay within 1.25 kb of the *Sma*I site of the vector on the left hand side and within 2 kb of the *Sal*I site of pUC19 on the right hand side. The *lacZ* promoter of pUC19 runs from beyond the *Hind*III site throught the *Sal*I and *Sma*I sites in the direction of the *Eco*RI site.

Exonuclease III shortening was used to produce a series of plasmids in which the 3.2 kb pTF-FC2 fragment of pTV400 had been progressively deleted from either end. The *Kpn*I and *Pst*I restriction endonucleases, producing restriction fragments with 3' overhanging ends resistant to *Exo*III digestion were used to prevent vector digestion, while enzymes *Sma*I and *Sal*I were used to create susceptable ends for *Exo*III digestion from the left and right ends of pTV400 repectively (Figs 2.2 and 2.3).

Plasmids containing shortened pTV400 fragments were tested for their ability to replicate in GW125a and the results are summarised in Fig. 2.2. Only the deletions which flanked a change in the ability to replicate are shown. With respect to the deletions from the left *Sau*3A site of pTV400, plasmid pTV4101, which had a 600 bp deletion was able to replicate in GW125a, while pTV4111, with an 850 bp deletion, had lost this ability.

Three clearly distiguishable phenotypes were detected in the deletions from the right hand *Sau*3A site of pTV400. A 300 bp deletion (pTV4200) had no effect on replication, while a plasmid with an 850 bp deletion (pTV4220) was unable to replicate. *E. coli* GW125a cells containing pTV4210, which had a 400 bp deletion, produced markedly smaller colonies than those with either pTV400 or pTV4200 which was 100 bp larger. The minimal replicon of pTF-FC2 in *E. coli* is therefore contained within the 2.5 kb common to pTV4101 and pTV4200.

2.4.2c Identification of the *oriV*

In order to identify the DNA sequence required in *cis* for replication, deletions of pTV400 which were unable to replicate in *E. coli* GW125a, were transformed into the same strain containing pDER412 (Fig. 2.2). The region of pTV400 required in *cis* was located as being between pTV4161 and pTV4170 on the one side and pTV4210 and pTV4220 on the other. The 184 bp fragment common to both pTV4161 and pTV4210 was identified as the minimum region required in *cis* for replication of pTF-FC2 *in vivo*.

2.4.2d Location of the incompatibility determinants

The ability of different fragments of pTF-FC2 to displace pDER412 after 60 generations was used to identify the region containing the plasmid incompatibility determinants. From the results in Table 2.1 (summarised in Fig. 2.2), it is clear that a major incompatibility determinants is located on the 184 bp of DNA common to pTV4161 and pTV4210. and all colonies containing vectors pUC19 and Bluescript SK retained pDR412 after 60 generations. All colonies containing pTV4161 and pTV4210 had completely displaced pDER412 after 60 generations. In contrast, pTV4220 and pTV200, which have the *oriV* deleted, showed an intermediate degree of incompatibility (37% and 20% respectively of the colonies tested had lost pDER412). There is therefore a secondary incompatibility region located on pTV200 which is normally masked by the determinants located on the *oriV*.

Table 2.1 Identification of regions carrying the incompatibility determinants of pTF-FC2

Plasmid	No of Cm^R colonies (/60)	% Incompatibility (% Cm^S colonies)
pUC19	60	0
Bluescript SK	60	0
pTT100	0	100
pTV100	0	100
pTV200	12	20
pTV300	0	100
pTV400	0	100
pTV4101	0	100
pTV4111	0	100
pTV4161	0	100
pTV4170	60	0
pTV4200	0	100
pTV4210	0	100
V4220	38	37

2.4.2e Plasmid copy number

The copy number of plasmids pDER412 and pTV100 was estimated by the hybridization method at 12 - 15 plasmids per chromosome. All plasmids capable of Pol I-independent replication, with the exception of pTV4210 conferred an identical level of Ap resistance (800 µg/ml) to *E. coli* GW125a. The presence of pTV4210, resulted in a decrease in the resistance to Ap to 600 µg/ml.

2.4.3 Replication of pTF-FC2 in *P. aeruginosa*

Plasmid pTV400 and the deletions capable of replication in *E. coli* GW125a were transformed into *P. aeruginosa* PAO1162 and transformants selected for their ability to grow on carbenicillin (Cb). The presence of the respective plasmids was confirmed by restriction analyis of plasmid DNA isolated from the transformants. Plasmids with deletions from the right end of pTV400 which replicated in *E. coli* were also capable of replication in *P. aeruginosa* PAO1162 (Fig. 2.4). Plasmids pTV400 and pTV4200 were maintained in PAO1162, while pTV4210-containing colonies were smaller than those containing pTV400 or pTV4200. When deletions from the left-hand side of pTV400 were tested, the situation was different. Neither pTV4101 nor pTV4100 (which is 260 bp larger than pTV4101) were capable of replication in *P. aeruginosa* although both were maintained in *E. coli*.

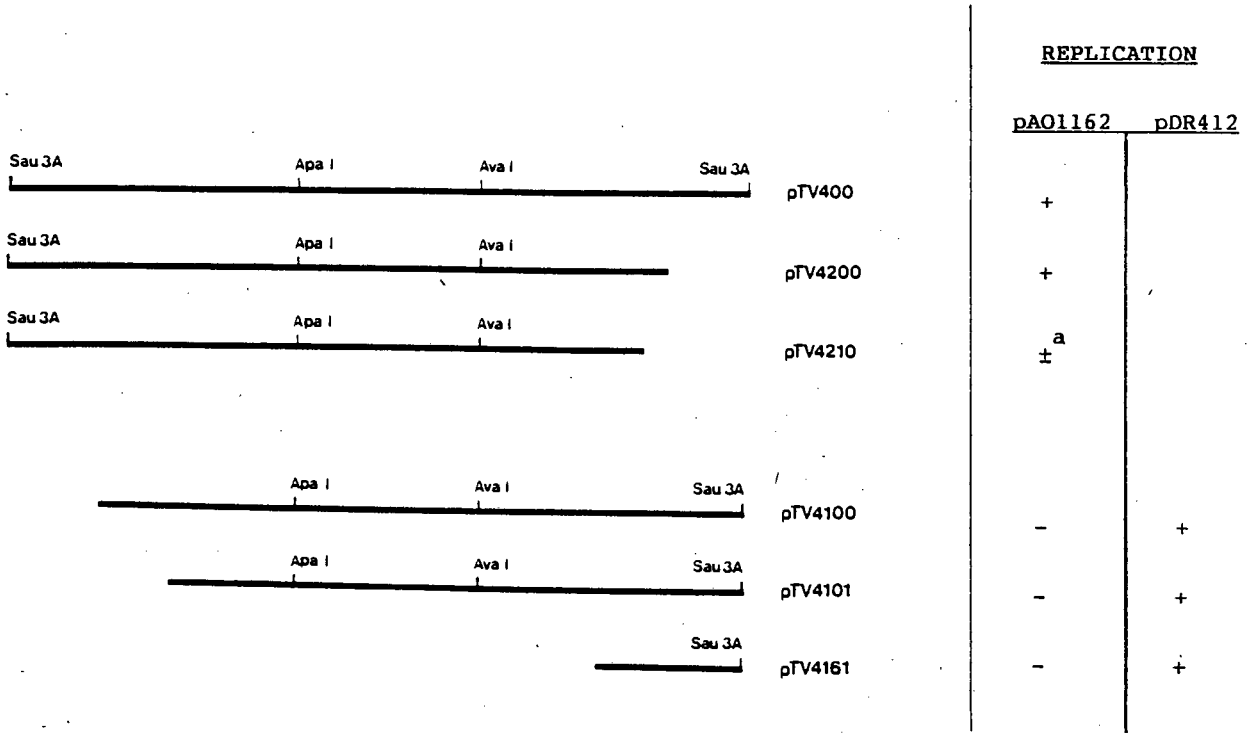

Fig. 2.4 Deletions of pTF-FC2 used to determine minimal replicon *P. aeruginosa*. Only pTF-FC2-derived sequences are shown.

[a] indicates small colonies

The complementation of deletions which were unable to replicate in PAO1162 by pDER412 was tested. Plasmids pTV4100, pTV4101 and pTV44161 all produced Cb^R colonies. Unlike *E. coli* GW125a, *P. aeruginosa* PAO1162 is not *recA*$^-$, so the Cb^R colonies could have been due to an homologous recombination event between pDER412 and the plasmid that was tested. However, since the transformation frequency for the test plasmids was equivalent to that obtained with the positive control (pTV400), and since homologous recombination after transformation would be expected to occur at a much lower frequency, it is most likely that complementation had occured.

2.5 DISCUSSION

A fragment of 3.2 kb has been identified which retained the replication properties of the intact pTF-FC2 replicon. The ability of pTF-FC2 to be replicated in a *polA*$^-$ *E. coli* mutant enabled replication at the pTF-FC2 origin of replication to be distinguished from that which occurred at the pMB1 origin of pBR325, Bluescript SK and pUC19. The background levels of Pol I-dependent replication made the ts mutant, AB1157(*polA*12) less suitable for these experiments and plasmids were only tested for replication in GW125a.

A series of deletions of the 3.2 kb fragment enabled the location of functional regions on the minimal replicon. Based on the deletions described here, only 2.25 kb appear to be required for replication in *E. coli*, while a further 340 bp adjacent to the left hand *Sau*3A site of pTV400 is necessary for plasmid replication to occur in *P. aeruginosa*.

The *ori*V appears to consist of at least a 184 bp fragment common to pTV4161 and pTV4210. As the copy number of pTV4210 was reduced relative to pTV400, it is possible that the *ori*V may extend beyond this fragment in the direction of the right hand *Sau*3A site of pTV400. The area assigned to the *ori*V also encodes the major incompatibility determinant, although there is another weaker incompatibility determinant located elsewhere on the minimal replication. These results are

consistent with reports that incompatibility determinants are associated with the *ori*V of a wide variey of narrow-host-range plasmids such as plasmids F and P1 (Novick, 1987) as well as the broad-host-range IncQ plasmids (Meyer *et al.*, 1985; Persson and Nordström, 1986).

Deletion of the first 360 bp of pTV400 resulted in the inability of the plasmid to replicate in *P. aeruginosa*. Plasmid replication was, however restored in the presence of pDER412, suggesting that a diffusible product produced or regulated by this region is involved in the determination of the host range of pTF-FC2.

CHAPTER 3

THE NUCLEOTIDE SEQUENCE AND IDENTIFICATION OF THE POLYPEPTIDES ENCODED BY THE MINIMAL REPLICON OF pTF-FC2

CHAPTER 3

THE NUCLEOTIDE SEQUENCE AND IDENTIFICATION OF THE POLYPEPTIDES ENCODED BY THE MINIMAL REPLICON OF pTF-FC2

3.1 SUMMARY

The nucleotide sequence of the 3202 bp minimum replicon of pTF-FC2 was determined from a series of overlapping sequences from sequential deletions obtained from each end of the fragment. Analysis of the *ori*V sequence revealed three tandemly repeated 22 bp DNA sequences and two sets of complementary inverted repeats. The 22 bp direct repeats appeared to be essential for replication and for incompatibility. The role of one of the two sets of complementary inverted repeats is unclear while deletion of half of the second set resulted in a drop in plasmid copy number. Analysis of the sequence revealed at least 5 potential ORFs preceded by consensus ribosome binding sites. Four of these ORFs were identified as polypeptides *in vitro*. Proteins of 33.5 and 32 kD which corresponded to the products of ORF2 and ORF1 respectively, were required for replication in *Escherichia coli*. A deletion which was unable to replicate in *Pseudomonas aeruginosa* had lost 8 and 10 kD proteins and these may be involved in replication of pTF-FC2 in certain hosts. At least two promoters exist on the 3202 bp fragment and a putative site for one of these, *P1* has been identified by comparison with the *E. coli* consensus promoter sequence.

3.2 INTRODUCTION

In order to understand the mechanism by which pTF-FC2 replicates, a functional map of the basic replicon is required. The functional regions may then be correlated with the nucleotide sequence to identify the sequences involved in replication, incompatibility, the control of plasmid copy number and determination of host-range. The primary structure of the DNA may reveal features such as repeated sequences and other binding sites for host- and plasmid-encoded proteins such as the DnaA protein (Hansen and Yarmolinsky, 1986; Masai and Arai, 1987; Pinkney *et al.*, 1988) and integration host factor (IHF) (Stenzel *et al.*, 1987). Analysis of the nucleotide sequence allows the prediction of ORFs which may encode proteins involved in the initiation and control of plasmid replication. In addition, computer analysis reveals secondary features such as stem-and-loop structures which may regulate gene expression or act as priming sites for DNA replication (Honda *et al.*, 1988). A comparison of the functional map with the structural analysis of the nucleotide sequence allows the allocation of functions to the structural characteristics an essential step towards understanding the mechanism of replication.

This chapter deals with the analysis of the nucleotide sequence of the 3.2 kb *Sau*3A partial pTF-FC2 fragment and the polypeptides for which it codes. Proteins corresponding to the predicted ORFs were identified and a preliminary investigation was made of their expression.

3.3 MATERIALS AND METHODS

3.3.1 Bacterial and plasmid strains

Plasmid DNA was maintained and prepared in *E. coli* LK111 while Pol I-independence was tested in *E. coli* GW125a and AB1157(*polA*12) (Appendix 1) as described in Chapter 2.3.1. *E. coli* HMS2562 (Appendix 1) containing the kanamycin resistant (Km^R) plasmid pGP1-2 (Appendix 2) which has the T7 RNA

gene under control of the temperature-inducible P_R promoter, was used to produce polypeptides *in vivo* (Tabor and Richardson, 1985). Plasmid vectors used in this chapter are described in Appendix 2 and other plasmids are summarised in Appendix 3.

3.3.2 DNA sequencing

The ordered deletions of pTV400 generated previously (Chapter 2.4.2b) were used as templates for DNA sequencing. Plasmid DNA was prepared by a scaled-up Ish-Horowicz and Burke (1981) method (Appendix 4) and purified by caesium chloride density ultracentrifugation. Sequencing templates were labelled with [α-^{35}S]dATP (specific activity > 1000 Ci/mM, Amersham) by the chain termination method (Sanger *et al.*, 1977) using the Sequenase sequencing Kit (United States Biochemical Corp., Cleveland, Ohio). Templates were run as described in Maniatus *et al.* (1989) on 6% acrylamide gels. DNA sequence was obtained from both strands using overlapping fragments. The sequence was analysed using the UWGCG (version 6.1) software (Devereux *et al.*, 1984) and Genepro (version 4.1) packages.

3.3.3 Analysis of polypeptides

3.3.3a *In vitro* protein synthesis

Translation products were identified *in vitro* using a prokaryotic DNA directed transcription-translation kit (Amersham, code N.380) and the polypeptides were labelled with L-[^{35}S]- methionine (specific activity approximately 1500 Ci/mMol, Amersham) and separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Maniatis *et al.*, 1989).

3.3.3b *In vivo* protein synthesis

Polypeptides were expressed *in vivo* from the T7 promoter on Bluescript SK (Appendix 2) using the system developed by Tabor and Richardson (1985). Plasmids were transformed into *E. coli* HMS262 containing plasmid pGP1-2. Cells were grown overnight at 30 °C with selection for Ap^R and Km^R in M9 minimal medium (M9 MM) supplemented with all amino acids except for cysteine and

methionine (Appendix 4.. The cultures were diluted 1/50 into 25 ml MM with Ap and Km, grown to OD_{600} = 0.2 at 30 °C and then shifted to 42 °C to induce the T7 RNA polymerase. After 20 min, rifampicin was added to a final concentration of 200 µg/ml (to inhibit the *E. coli* RNA polymerase) and the cells incubated for 15 min at 42 °C. The cultures were shifted to 30 °C for 20 min. Three ml of culture was removed, 2 ml [^{35}S]-methionine added and incubated at 30 °C for 5 min before the cells were precipitated and resuspended in 30 ml lysis buffer (50 mM Tris.HCl pH6.8; 1% SDS, 2-mercaptoethanol; 10% glycerol; 0.025% bromophenol blue). The labelled cells were disrupted by heating to 95 °C for 5 min, centrifuged to remove undissolved cellular debris and 10 ml of each sample was used for SDS-PAGE analysis of the labelled polypeptides.

3.3.3c SDS-polyacrylamide gel electrophorectic (SDS-PAGE) analysis of polypeptides

Polypeptides produced *in vitro* and *in vivo* were analysed by SDS-PAGE. Two concentrations of acrylamide were used to resolve the 7 - 16 kD and 16 - 50 kD polypeptides (20% and 12.5% respectively). Using the method of Maniatis *et al.* (1989), 0.5 mm thick gels were prepared and run in a Hoefer protein gel apparatus for approximately 6.5 h at a constant voltage of 150 mV. Gels were stained with Coomassie Blue, dried and autoradiographed for 24 h - 1 week, depending on the amount of radioactive labelling.

3.3.4 β-Galactosidase assays

The method of Miller and Stadtman(1972) was used to assay for b-galactosidase activity. Cells containing the plasmids to be tested were grown overnight at 37 °C in LB containing antibiotics and the OD_{660} recorded for the culture. One ml of cells was centrifuged, the pellet resuspended in 1 ml phosphate buffer (0.1 M K_2HPO_4 pH 7.0) and the cells incubated on ice for 20 min. An equal volume of Z buffer (0.1 M Na_2HPO_4, pH 7.0; 10 mM KCl; 1 mM $MgSO_4$; 50 mM ß-mercaptoethanol), 20 µl chloroform and 10 µl 0.1% (w/v) SDS was added to 200 µl of cells. The mixture was vortexed for 10 s and 100 ml of a freshly prepared solution of o-nitrophenyl-a-o-galactopyrenose (ONPG) (4 µg/ml) added, after which the tubes were incubated at

28 °C for up to 1 h. The reaction was stopped by adding 250 µl of a 1 M Na_2CO_3 solution and the tubes microfuged to remove the cells and chloroform; 800 µl was removed and the OD_{420} read on a spectrophotometer to determine the amount of o-nitrophenol released. The OD_{550} was recorded as a measure of the effect of cell debris on the OD_{420} reading (Miller and Stadman, 1972). The β-galactosidase activity (Miller units) was calculated using the following equation as a factor of time (min) and cell volume (µl).

$$\text{Activity (Miller units)} = \frac{1000 \times [A_{420} - (1.75 \times A_{550})]}{\text{Reaction time (min)} \times A_{660} \text{ of cells} \times \text{volume of cells (µl)}}$$

3.4 RESULTS

3.4.1 Nucleotide sequence of pTV400

The nucleotide sequence of the pTF-FC2 minimal replicon is shown in Fig. 3.1. The *Sau*3A fragment from pTV400 (Fig. 3.2) was 3202 bp in length and contained very few 6 bp recognition restriction enzyme digestion sites relative to its size. Previously mapped *Apa*I and *Ava*I restriction enzyme sites were located at nucleotides (nts) 1246 and 2066 respectively. Sequence analysis revealed unique sites for *Ava*II (nt 161), *Bal*I (nt 2416), *Dra*I (nt 2104), *Hinc*II (nt 1386) and *Xho*II (nt 1079). The 3202 bp fragment had a G+C ratio of 59%. This corresponds closely with an average G:C ratio of 58.7% for the *T. ferrooxidans* genes previously cloned and sequenced in the laboratory (Pretorius *et al.*, 1987; Rawlings *et al.*, 1987; Rawlings, 1988 Ramesar *et al.*, 1989; Berger *et al.*, 1990) and to that calculated for the *T. ferrooxidans* chromosome by Harrison (1984).

```
1     GATCGAGCGCGACAACTCGCCGGGAATGAGTCTCTAGCGTTGCGTGGTGGTTGTGATATA

61    CTTGTATAGCGTTTTCAGAACAGGAGCCGAAACATGCTTGCAATCCGACTGCCCGCCGAA
                              ORF3 →   M  L  A  I  R  L  P  A  E

121   GTGGAAACCCGCCTTGAAGCACTGGCGCAGGCCACAGGGCGGACCAAGACTTTCTATGCC
       V  E  T  R  L  E  A  L  A  Q  A  T  G  R  T  K  T  F  Y  A

181   CGCGAAGCCATCCTTGAGCACTTGGATGACCTCGAAGATTTGTACCTTGCAGAGCAACGC
       R  E  A  I  L  E  H  L  D  D  L  E  D  L  Y  L  A  E  Q  R

241   CTGATCGACATTCGCGCAGGCAAAACCCAAACCGTGCCACTCGAAGAAGTGATGAAACGC
       L  I  D  I  R  A  G  K  T  Q  T  V  P  L  E  E  V  M  K  R
                                                        ORF4 →

301   TATGGCATGGAAGGTTGAACTCGACCCAGCCGCCGAGCGCGAGCTAGGCAAGATCGACCA
       Y  G  M  E  G  *
          M  A  W  K  V  E  L  D  P  A  A  E  R  E  L  G  K  I  D  Q

361   GCAGACCGCCCGCCGCATCCTCGCTTTTTTGCATGGCCGTGTCGCCCAGCTCGACGACCC
       Q  T  A  R  R  I  L  A  F  L  H  G  R  V  A  Q  L  D  D  P

421   GCGCAGCATTGGCGAAGCCCTCAAAGGCTCCAAACTGGGAGCCTTCTGGAAATACCGCGT
       R  S  I  G  E  A  L  K  G  S  K  L  G  A  F  W  K  Y  R  V

481   TGGGGATTGGCGAATCATCGCCAGCATCGAGGACGGTGCTTTGCGCATCCTCGTTATGCG
       G  D  W  R  I  I  A  S  I  E  D  G  A  L  R  I  L  V  M  R

541   CATCGGCAATCGTAAGGAGGTTTACCGCCAATGATCGAATACAGCTACCAGATCGACCCG
       I  G  N  R  K  E  V  Y  R  Q  *
                                 ORF5 →   M  I  E  Y  S  Y  Q  I  D  P

601   CGCCCCTCCGACCTTGGCGGCGGCTGGCGGTTGCGCCTGTTGGAAAGCGGCGAGGAAGTC
       R  P  S  D  L  G  G  G  W  R  L  R  L  L  E  S  G  E  E  V

661   GGCGGCGGAGTGTTCCCGTTGTCCGAGTACGCCACAGCAGAGAACGCAGAAGAAGCGGCC
       G  G  G  V  F  P  L  S  E  Y  A  T  A  E  N  A  E  E  A  A

721   ACGTACGCCTATGAGGACGCCTTGGCCGAGGCTTCGGCGTGGCTGGCATCGAGGGGCGAA
       T  Y  A  Y  E  D  A  L  A  E  A  S  A  W  L  A  S  R  G  E

781   AATTGAGCGGCGCGGCAGGGGATTGCGGCCCCGGCAGCGCCTAACCACAACTGTCTGAAA
       N  *

841   AGGAGACAAGCATGGCTTTAGACATTATGGCGGCCTTCACCAATGAGCCGCCAGAACTTG
         ORF1 →   M  A  L  D  I  M  A  A  F  T  N  E  P  P  E  L

901   ATTTCATCTGGCCCGGATTCTTGGCCGGAACCGTGGGCGCACTTGTCGCACCTGGCGCAA
       D  F  I  W  P  G  F  L  A  G  T  V  G  A  L  V  A  P  G  A

961   CTGGCAAGAGCTTTTTTGCTCTTGAAGCGGCCATGTCAATCGCTTGCAGTGTGGCAGGCG
       T  G  K  S  F  F  A  L  E  A  A  M  S  I  A  C  S  V  A  G

1021  GCGACCTTGTGGGACTAACCCCGGCGCACACCGGGCGCGTGGTTTATCTCGCTGGCGAAG
       G  D  L  V  G  L  T  P  A  H  T  G  R  V  V  Y  L  A  G  E

1081  ATCCACAGCCCGCCCTTGTGCGACGTGTCCACGCCATCGGCCAGCACCTCAACCAGTCGG
       D  P  Q  P  A  L  V  R  R  V  H  A  I  G  Q  H  L  N  Q  S

1141  CCCGCGAAGCCATCGCTGAGAACCTGATGCTTGAGCCGATCATGGGCAAGCGGCTAAACG
       A  R  E  A  I  A  E  N  L  M  L  E  P  I  M  G  K  R  L  N

1201  TGATGGACGACGCGCACTTGCGCCGCGTCATCGACTACAGCGCAGGGGCCCGGCTGATTG
       V  M  D  D  A  H  L  R  R  V  I  D  Y  S  A  G  A  R  L  I

1261  TGCTGGACACCCTGAGCCGGATTCACATCCTCGACGAGAACAGCAATGGCGACATGGCCC
       V  L  D  T  L  S  R  I  H  I  L  D  E  N  S  N  G  D  M  A

1321  ACCTTGTTTCCGTGTTGGAACACATCGCGGCGACCACCGGGGCGGCTGTCCTGTACCTGC
       H  L  V  S  V  L  E  H  I  A  A  T  T  G  A  A  V  L  Y  L

1381  ACCACGTCAACAAGGGCAGCGCCCGCGACGGCCAGACCGACCAGCAGCAGGCAGCGCGGG
       H  H  V  N  K  G  S  A  R  D  G  Q  T  D  Q  Q  Q  A  A  R

1441  GCGCGTCTGCCCTGATCGACAACGCCAGATGGTGCGGCTATGTCGCCAAAATGACGGAGC
       G  A  S  A  L  I  D  N  A  R  W  C  G  Y  V  A  K  M  T  E

1501  AGGAAGCCGAGCGCATGAGTGACCGGGGCTTTGATCGTTCGCCATCCGGCAACGAGCGGC
       Q  E  A  E  R  M  S  D  R  G  F  D  R  S  P  S  G  N  E  R

1561  GCGGCCTTTTTGTCCGCTTTGGCGTGAGCAAGCAGAACTACGACGCGACCCCGCTAGACC
       R  G  L  F  V  R  F  G  V  S  K  Q  N  Y  D  A  T  P  L  D
```

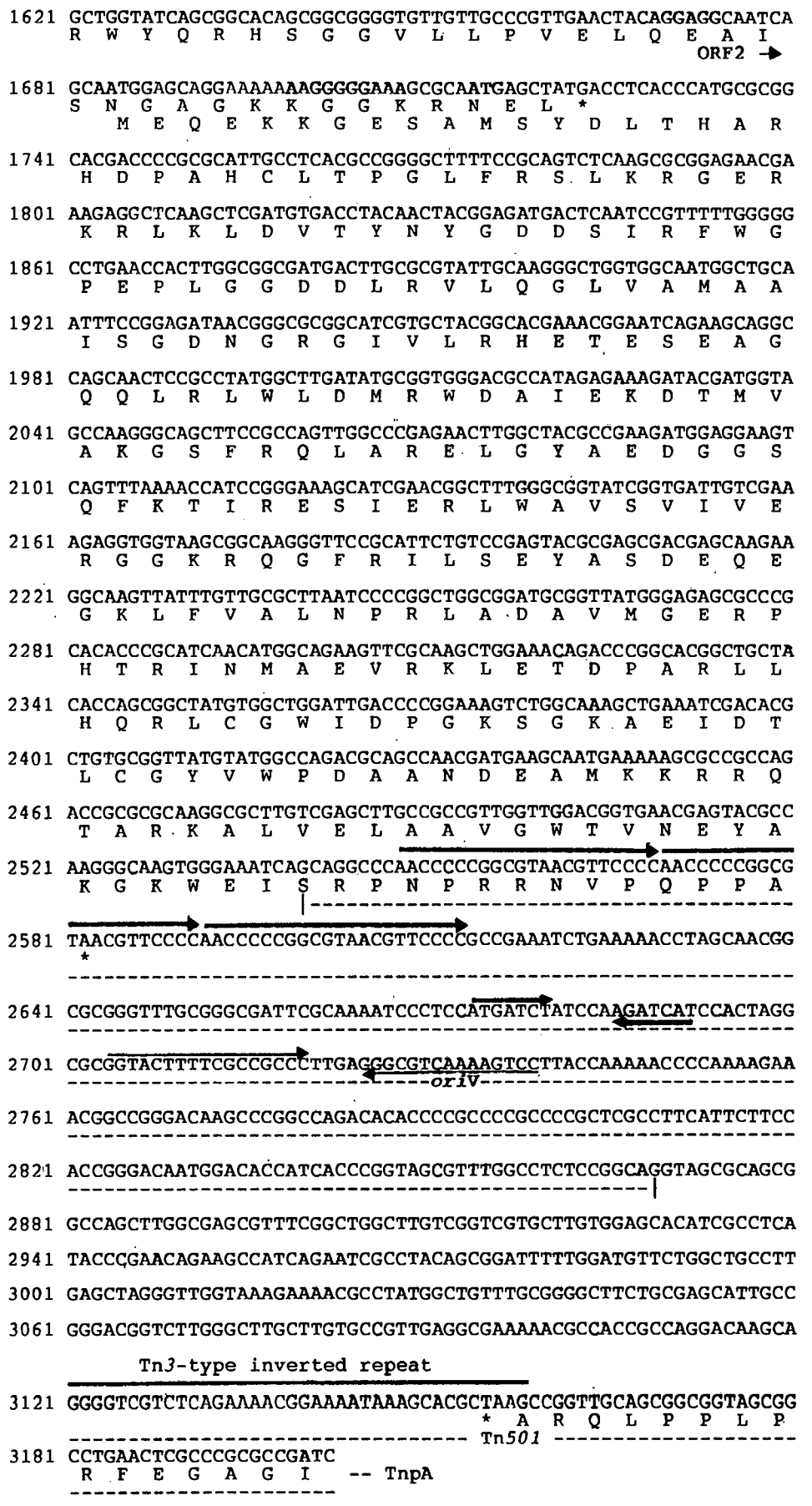

Fig. 3.1 (Legend on following page)

Fig. 3.1 Nucleotide sequence of the 3202 bp *Sau*3A partial DNA fragment of pTF-FC2 contained within pTV400. The predicted amino acid sequences for the five predicted ORFs are shown below the nucleotide sequence. Consensus sequences for ribosome binding sites are indicated in bold type as are the ATG start codons. Stop codons are indicated by asterisks. The extent of the *ori*V is shown by a dotted line below the sequence. The three 22 bp direct repeats within the *ori*V are indicated by thick arrows as is a fourth, less conserved repeat upstream of the start of ORF1. Arrows above and below the sequence show the two sets of complementary inverted repeats (CIRs) A and B.

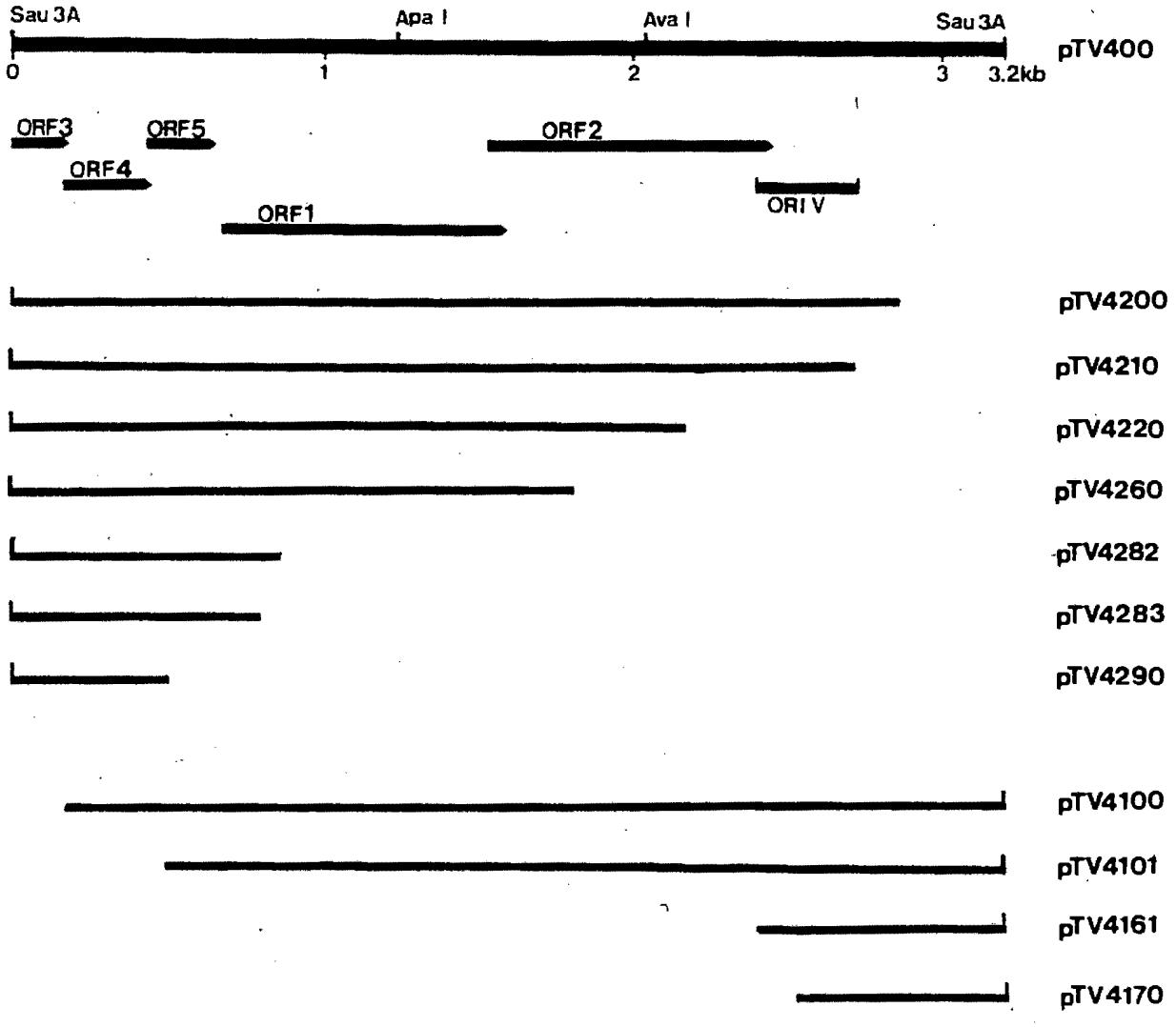

Fig. 3.2 Plasmids constructed and used in this chapter. A restriction endonuclease cleavage map of pTV400 is shown with the positions and orientations of the five potential ORFs. The deletion plasmids used to identify corresponding polypeptides *in vitro* are also shown.

3.4.1a Structural features of the *ori*V

The *ori*V of pTF-FC2 (nts 2540 to 2868), which corresponds to the region common to pTV4161 and pTV4200 (Fig. 3.2) is shown in Fig. 3.3. The most notable feature of this region is the presence of three perfectly conserved, tandemly repeated 22 bp sequences (nts 2548 - 2613). A region between nts 655 and 677 contains a 14 bp match with the 22 bp repeated sequences (Fig. 3.1). Analysis of the base composition of the *ori*V revealed a 28 bp A+T-rich region (65.5%) from nts 2734 to 2761 and a 49 bp G+C-rich region (80%) from nts 2762 to 2810.

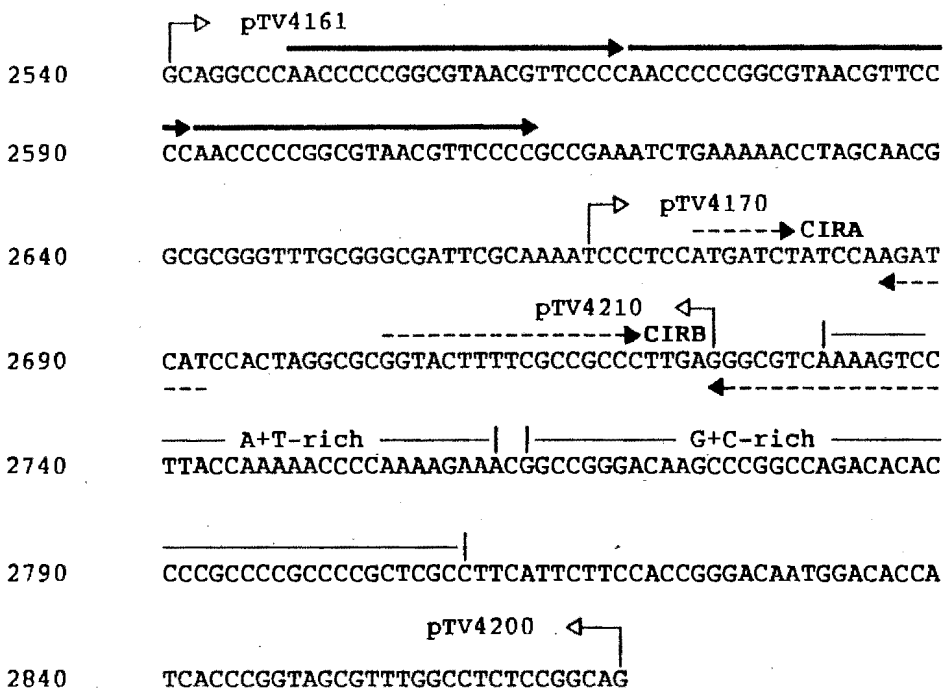

Fig. 3.3 Nucleotide sequence of the pTF-FC2 *ori*V. The position of the 22 bp direct repeats are indicated by thick arrows, while the complementary inverted repeats are shown as dotted lines. Solid lines represent areas of A+T and G+C bias. The start (pTV4161 and pTV4170) and ends (pTV4200 and pTV4210) of deletion plasmids are also shown.

There are two sets of complementary inverted repeated sequences (CIRs), the first (A) from nts 2674 to 2691 and the second (B) from nt 2704 to 2739. These sets of repeats are able to form potential stem-and-loop structures (Fig. 3.4) of ΔG = -5.4 and -14.7 kcal/mol respectively (Salser, 1977).

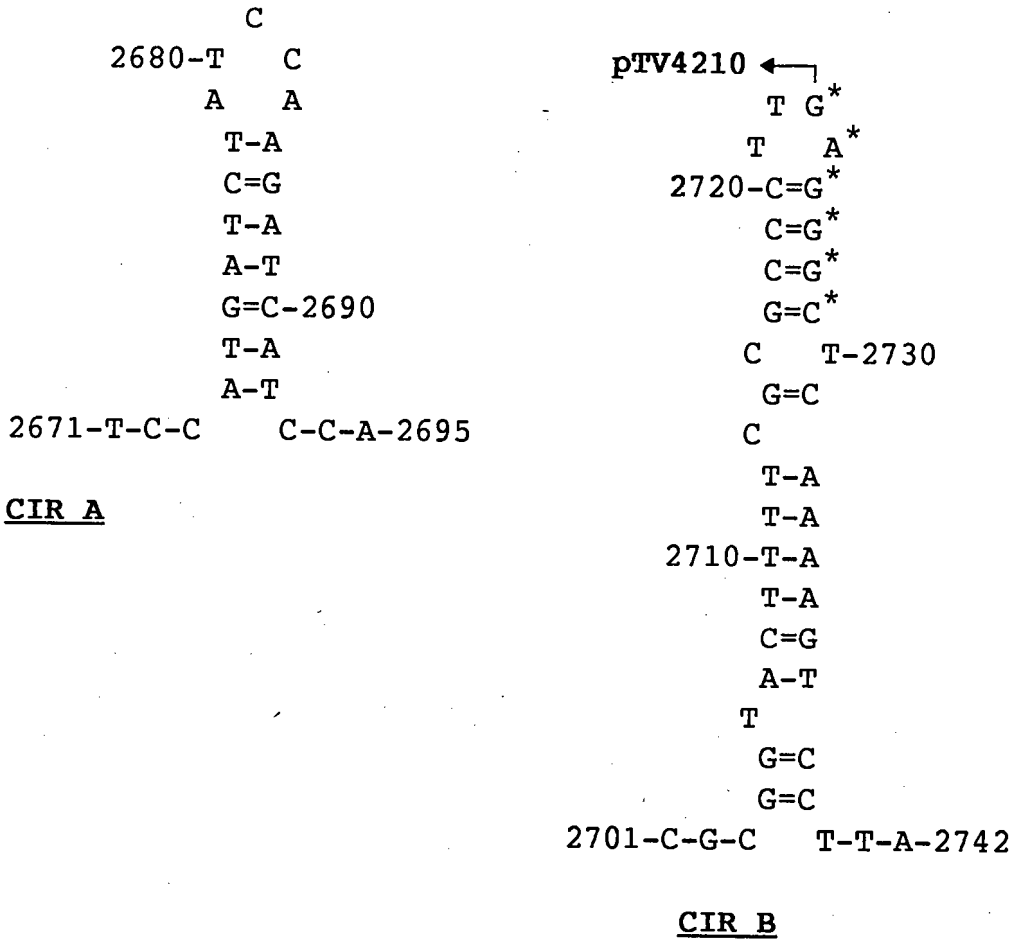

Fig. 3.4 Secondary structure of the two sets of CIRs found in the *ori*V of pTF-FC2. * denotes consensus with n' protein-recognition sequence.

Plasmid pTV4170 (Fig. 3.2), which could not replicate with pDR412 in *trans* (Chapter 2.4.2b), had lost the three 22 bp repeats (Fig. 3.3). This plasmid did not exhibit any incompatibility with pDR412 (Table 2.1). The 22 bp repeats appear to be directly associated with replication and the determination of incompatibility. It is interesting that pTV4210 (Fig. 3.2), which appeared to be maintained at a lower copy number than pTV4200 in *E. coli* (Table 2.1) and produced smaller colonies in *P. aeruginosa* (Chapter 2.4.3), had lost one arm of CIR B (Figs 3.3 and 3.4). The deleted sequence shows some homology to the consensus sequence for the n' protein-recognition site (Fig. 3.4) (van der Ende *et al.*, 1983). No consensus sequences could be identified for the binding sites of the host-encoded DnaA (TAAC/ACACTTATC/ACACA) or IHF (YAA----TTGATW) proteins (Fuller *et al.*, 1984; Gamas *et al.*, 1986).

A search of the available databases for sequences homologous to the 3202 bp fragment revealed a region downstream from the *ori*V with strong homology to the terminal repeats of transposable elements belonging to the Tn3 family (Fig. 3.1). Directly adjacent to this sequence was part of an ORF of which the amino acid sequence was highly homologous to the carboxyterminus of the transposase of Tn*21* (Ward and Grinsted, 1987).

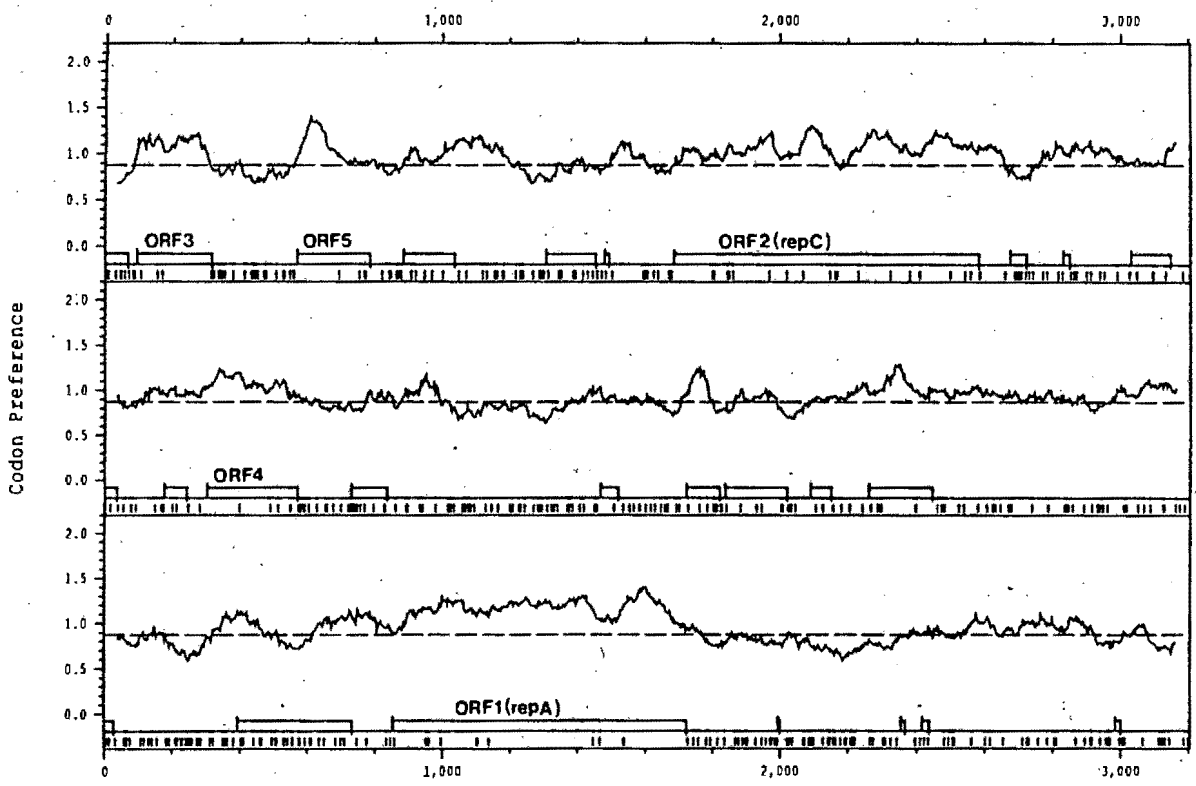

Fig. 3.5 Potential ORFs and their position on pTV400. A codonpreference plot was generated using the UWGCG vax package. The combined codon usage of ORFs 1 and 2 was used as a codon usage standard. The position of rare codons is indicated below open boxes which denote ORFs.

3.4.1b Potential ORFs encoded by the minimal replicon

Computer analysis of the 3202 bp fragment using the Codonpreference and Testcode subroutines of the UWGCG package indicated the presence of at least five potential coding regions, each preceded by a consensus ribosome binding site (Shine and Delgarno, 1975) (Fig. 3.5). The extent and M_R of the polypeptides predicted for these ORFs is shown in Table 3.1. ORF2 has two possible ATG start codons at nts 1684 and 1714, each preceded by a consensus ribosome binding site. A protein with a predicted M_R of 33.778 kD would be produced if translation occurred from the first ATG and one of 32.659 kD if translation commenced from the second. The carboxyterminus of ORF2 overlaps with the first 36 bp of the three 22 bp direct repeats in the *ori*V.

Table 3.1 Potential polypeptides produced by the pTF-FC2 minimal replicon

ORF	Position	Translational initiation sites*	Stop codon	No. Nts	No. aas#	M_R (kD)
ORF1	852-1721	CUGUCUGAAAAGGAGACAAGCAUG	UGA	870	290	31 252
ORF2	1684-2580	AACUACAGGAGGCAAUCAGCAAUG	UAA	897	299	33 778
	1714-2580	GAAAAAAAGGGGGAAAGCGCAAUG	UAA	867	289	32 659
ORF3	94-315	UUUCAGAACAGGAGCCGAAACAUG	UGA	222	74	8 462
ORF4	302-571	UCGAAGAAGUGAUGAAACGCUAUG	UGA	270	90	10 318
ORF5	571-783	UCGUAAGGAGGUUUACCGCCAAUG	UGA	213	71	7 702

* The start codons are underlined as are the nucleotides in the pre-start region that could air with the 3' end of the *E. coli* 16S rRNA (AUUCCUCCACU...5')

3.4.2 Analysis of proteins encoded by the minimal replicon

3.4.2a Polypeptides produced *in vitro*

Translation products from the pTF-FC2 minimum replicon were analysed *in vitro* to ascertain whether polypeptides equivalent to ORFs predicted from the nucleotide

sequence were translated. Polypeptides were expressed from the cloned fragments and deletion plasmids (shown in Fig. 3.2) using an *in vitro* DNA directed transcription-translation kit and the products analysed by SDS-PAGE.

Translation of pUC19 vector resulted in the 30.5 kD ß-lactamase protein (Figs 3.6 lane 5 and 3.7 lane 1) which was present in all the lanes. Plasmid pTV400, with the intact 3202 bp fragment, produced additional polypeptides of M_R 33.5 kD, 32 kD (Fig. 3.6 lane 1), 10 kD and 8 kD (Fig. 3.7 lane 2). These values corresponded closely to the polypeptides predicted for ORFs 2, 1, 4 and either 3 or 5 respectively. Plasmid pTV4220 (nts 1 - 2363), had the *ori*V and 65 amino acid residues deleted from the carboxyterminus of ORF2. The 33.5 kD polypeptide corresponding to ORF2 disappeared. A new polypeptide M_R 28.5 kD which corresponded to the predicted M_R of 25.68 kD for the truncated 33.5 kD protein (Fig 3.6 lane 2) appeared, while the other polypeptides remained unchanged. Plasmid pTV4260 (nts 1 - 1739), which has the oriV and all except the first 17 amino acid residues of ORF2 deleted, showed a complete loss of the 33.5 kD protein (Fig. 3.6 lane 3), while the remaining polypeptides were unchanged (Fig. 3.7 lane 4). Plasmids pTV4283 (nts 1 - 801) and pTV4282 (nts 1 - 870) which have both ORF1 and ORF2 deleted, no longer produced the 33.5 or 32 kD proteins (Figs 3.6 lane 4 and 3.8A lane 1; Fig. 3.7 lane 5) but still produced the 10 and 8 kD proteins.

From the codon preference data (Fig. 3.3) it was predicted that the ORF3 (8 kD) and ORF4 (1 kD) proteins are located within the first 571 bp of pTV400. This prediction correlated well with what was found *in vitro*. Plasmid pTV4290 (nts 1- 660) produced proteins of 10 and 8 kD (Fig. 3.7 lane 6) while pTV4100 (nts 360 - 3202) and pTV4101 (nts 615 - 3202) produced the 33.5 and 32 kD proteins but not the 8 and 10 kD proteins (Figs 3.8A lane 2 and 3.7 lane 3). Additional proteins of 30 kD and 28 kD (Fig. 3.6 lanes 1, 2 and 3) were observed. These proteins disappeared when ORF1 was deleted (Fig. 3.6 lane 4) and may be degradation products of ORF1. A protein of 16 kD was observed (Fig. 3.8A lane 2) which disappeared when ORF 2 was deleted (Fig. 3.8A lane 1) and is probably a degradation product of ORF2. This polypeptide was only observed if the protein samples had been stored for several days.

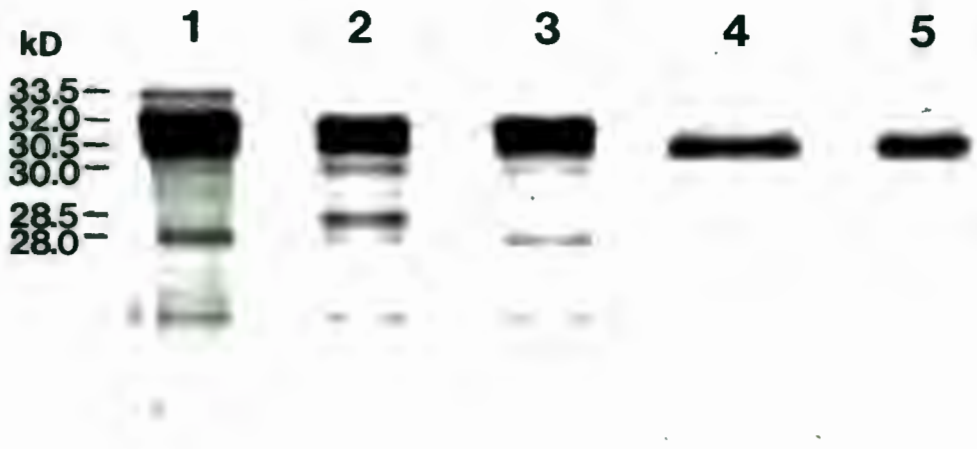

Fig. 3.6 SDS-PAGE analysis of polypeptides translated *in vitro* from pTV400 and its deletions. Lanes 1 to 5 show polypeptides from pTV400, pTV4220, pTV4260, pTV4283 and pUC19 respectively. Proteins were separated in a 12% acrylamide gel.

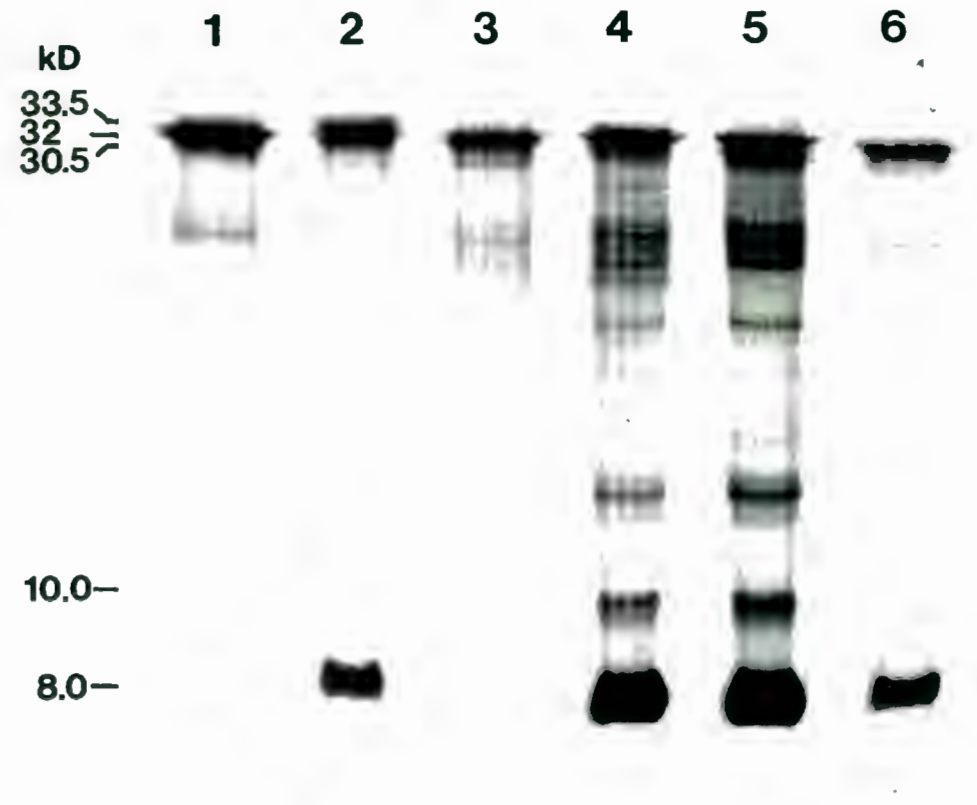

Fig. 3.7 SDS-PAGE analysis of polypeptides from pTV400 and its deletions separated in 20% acrylamide. Lanes 1 to 6 show pUC19, pTV400, pTV4101, pTV4260, pTV4282 and pTV4290 respectively.

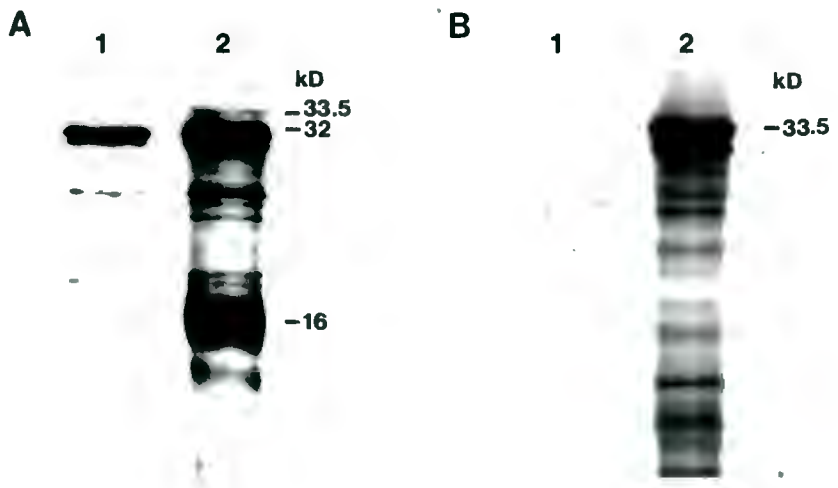

Fig. 3.8 SDS-PAGE analysis of polypeptides produced by *in vitro* transcription and translation and *in vivo* using the T7 transcription system. **A** Lanes 1 and 2 show polypeptides derived from plasmids pTV4283 and pTV4100 respectively. The proteins were separated in 20 % acrylamide **B** Lane 1 shows polypeptides derived from Bluescript SK and Lane 2 shows polypeptides derived from pTV300. Polypeptides were separated in 12% acrylamide gel.

3.4.2b Codon usage of the ORFs

Codon usage of the five ORFs encoded by the pTF-FC2 minimal replicon is shown in Table 3.2. The average G+C ratio for the five ORFs is 59.89%. The ORFs share a similar bias for codons ending in either a G or C residue where there is redundancy, except in the case of Glu where ORFs 2 and 3 show a bias in favour of the GAA codon, Leu, where ORFs 1, 2 and 3 show a bias for the CUU codon and Val where ORFs 2 and 4 show a bias for the GUU codon. This bias is different from the codon bias of the *T. ferrooxidans* chromosomal genes where the bias for Leu and Val is in favour of the G and C residues (Table 3.3).

Table 3.2 Codon usage of the ORFs encoded by the pTF-FC2 minimal replicon.

Amino Acid	Codon	ORF1	ORF2	ORF3	ORF4	ORF5	Total	Fraction
Ala	GCG	10	10	1	0	2	23	0.23
	GCA	8	8	4	1	3	24	0.25
	GCU	6	2	0	2	1	11	0.11
	GCC	14	8	4	7	5	38	0.39
Arg	AGG	0	2	0	0	1	3	0.04
	AGA	1	1	0	0	0	2	0.02
	CGG	7	10	1	0	1	19	0.26
	CGA	1	2	1	1	0	5	0.05
	CGU	2	2	0	2	0	6	0.08
	CGC	10	13	5	8	2	38	0.52
Asn	AAU	4	1	0	1	1	7	0.30
	AAC	8	7	0	0	1	16	0.69
Asp	GAU	3	9	2	1	0	15	0.30
	GAC	14	10	2	5	3	34	0.69
Cys	UGU	0	1	0	0	0	1	0.20
	UGC	2	2	0	0	0	4	0.80
Gln	CAG	11	6	1	3	1	22	0.73
	CAA	0	5	2	1	0	8	0.26
Glu	GAG	9	7	2	4	5	27	0.39
	GAA	7	18	8	2	6	41	0.61
Gly	GGG	5	6	1	1	0	13	0.17
	GGA	6	7	0	1	1	15	0.19
Gly	GGU	0	4	1	1	0	6	0.07
	GGC	18	10	2	5	7	42	0.55
His	CAU	0	2	0	1	0	3	0.16
	CAC	10	4	1	0	0	15	0.83
Ile	AUA	0	1	0	0	0	1	0.2
	AUU	3	4	1	1	0	9	0.23
	AUC	10	7	3	7	2	29	0.74
Leu	UUG	5	3	2	2	4	16	0.19
	UUA	1	1	0	0	0	2	0.02
	CUG	7	6	3	1	2	19	0.22
	CUA	5	4	0	1	0	10	0.11
	CUU	8	8	4	0	1	21	0.25
	CUC	3	6	2	5	0	16	0.19
Lys	AAG	5	13	1	3	0	22	0.62
	AAA	3	5	2	3	0	13	0.37
Met	AUG	9	8	3	2	1	23	1.00
Phe	UUU	5	3	0	1	0	9	0.50
	UUC	3	3	1	1	1	9	0.50
Pro	CCG	4	4	0	1	2	11	0.34
	CCA	3	2	1	1	0	7	0.21
	CCU	1	1	0	0	0	2	0.06
	CCC	3	7	1	0	1	12	0.37
Ser	UCG	2	1	0	0	2	5	0.12
	UCA	1	2	0	0	0	3	0.07
	UCU	1	1	0	0	0	2	0.05
	UCC	2	2	0	1	2	7	0.17
	AGU	2	2	0	0	0	4	0.10
	AGC	8	6	0	2	2	18	0.46
Thr	ACG	1	5	0	0	1	7	0.22
	ACA	0	1	1	0	1	3	0.09
	ACU	1	0	1	0	0	2	0.06
	ACC	9	5	4	1	0	19	0.61
Trp	UGG	3	8	0	3	2	16	1.00
Tyr	UAU	3	2	2	0	1	8	0.34
	UAC	3	5	1	2	4	15	0.65
Val	GUG	10	5	3	0	1	19	0.41
	GUA	0	4	0	0	0	4	0.08
	GUU	3	5	0	4	0	12	0.26
	GUC	7	2	0	1	1	11	0.23
End	UGA	1	0	1	1	1	4	0.80
	UAG	0	0	0	0	0	0	0.00
	UAA	0	1	0	0	0	1	0.20

% G+C content of the three positions in the condons used:

1st position	67.01	66.67	69.33	65.93	63.88	66.34
2nd position	45.36	45.00	38.67	42.86	51.38	44.87
3rd position	72.85	63.67	58.67	74.42	72.78	68.50
Average	61.74	58.45	55.56	61.07	62.63	59.89

Table 3.3 Codon usage of *T. ferrooxidans* chromosomal genes.

Amino Acid	Codon	*glnA*	*recA*	*nifHDK*	*ntrA*	ORF1	Total	Fraction
Ala	GCA	4	13	5	5	2	29	0.12
	GCC	29	51	22	20	6	128	0.51
	GCG	9	32	8	22	7	78	0.31
	GCU	1	6	5	2	1	15	0.06
Arg	AGA	0	2	0	0	0	2	0.01
	AGG	2	2	0	0	0	4	0.03
	CGA	0	2	1	1	0	4	0.03
	CGC	10	29	10	17	11	77	0.49
	CGG	2	12	6	6	8	34	0.22
	CGU	4	12	3	10	6	35	0.22
Asn	AAC	10	31	6	5	3	55	0.62
	AAU	7	11	6	9	0	33	0.38
Asp	GAC	19	56	10	23	2	110	0.61
	GAU	6	26	10	17	11	70	0.39
Cys	UGC	4	14	1	2	1	22	0.76
	UGU	0	6	1	0	0	7	0.24
Gln	CAA	0	4	0	9	4	17	0.16
	CAG	8	34	14	22	11	89	0.84
Glu	GAA	18	46	16	19	9	108	0.56
	GAG	9	44	9	16	7	85	0.44
Gly	GGA	5	6	2	2	2	15	0.06
	GGC	16	92	18	14	6	146	0.60
	GGG	4	11	2	4	4	25	0.10
	GGU	11	16	14	8	7	56	0.23
His	CAC	7	27	5	6	1	45	0.56
	CAU	7	13	3	5	6	35	0.44
Ile	AUA	0	2	2	1	2	7	0.04
	AUC	16	70	19	14	6	125	0.75
	AUU	4	10	6	8	6	34	0.20
Leu	CUA	1	2	0	0	0	3	0.01
	CUC	8	18	8	11	5	50	0.21
	CUG	18	53	20	29	18	138	0.57
	CUU	4	4	3	3	3	17	0.07
	UUA	1	0	1	2	2	6	0.02
	UUG	2	12	1	10	3	28	0.12
Lys	AAA	4	23	8	8	2	45	0.30
	AAG	20	64	12	8	2	106	0.70
Met	AUG	16	44	8	16	8	92	1.00
Phe	UUC	14	50	3	4	3	74	0.70
	UUU	3	12	6	6	4	31	0.30
Pro	CCA	1	5	2	1	1	10	0.08
	CCC	13	31	4	10	6	64	0.49
	CCG	8	18	4	10	2	42	0.32
	CCU	4	4	2	1	3	14	0.11
Ser	AGC	5	16	5	13	5	44	0.26
	AGU	4	10	5	1	2	22	0.13
	UCA	1	4	1	0	0	6	0.04
	UCC	11	23	10	6	3	53	0.31
	UCG	5	16	4	4	3	32	0.19
	UCU	5	5	1	3	0	14	0.08
Thr	ACA	1	3	0	2	0	6	0.05
	ACC	10	43	11	10	5	79	0.61
	ACG	6	16	1	7	3	33	0.26
	ACU	1	5	1	2	2	11	0.09
Trp	UGG	5	16	1	4	0	26	1.00
Tyr	UAC	11	27	8	6	3	55	0.60
	UAU	6	22	3	4	2	37	0.40
Val	GUA	4	4	2	1	1	12	0.06
	GUC	10	25	10	5	5	55	0.30
	GUG	16	43	13	18	12	102	0.55
	GUU	1	7	1	3	4	16	0.09
End	UGA	1	0	1	1	1	4	0.57
	UAG	0	0	0	0	0	0	0.00
	UAA	0	3	0	0	0	3	0.43

G+C ratios for the first, second and third position of the five ORFs are very similar with the exception of ORF3 which has a lower ratio in the second and third positions. Whether or not this is due to a real difference in the codon usage pattern or simply an aberration due to the small sample size is unknown.

3.4.2c Transcription of the ORFs encoded by the pTF-FC2 minimal replicon

All plasmids used for expression of proteins *in vitro* and *in vivo* were constructed in such a way that the β-galactosidase gene of the vectors pUC19 and Bluescript SK was translated in the opposite direction to that of the pTF-FC2 ORFs. Thus the β-galactosidase promoter of the vector should not have had any effect on the expression of the ORFs on the minimal replicon.

Expression of proteins from pTV400 and deletions pTV4100 and pTV4101 indicated that at least two promoters were resident on the 3202 bp fragment. The first, *P1*, lies between nts 1 and 81 upstream of the ribosome binding site of ORF3, while the second, *P2* , must be located between nts 615 (at the start of pTV4101) and the putative ribosome binding site of ORF1 at nucleotide 839. Evidence for the existence of *P2* is that the 32 and 33.5 kD proteins corresponding to ORF1 and ORF2 respectively, were still produced when the first 615 bp of the 3202 bp fragment were removed (pTV4100; Fig. 3.8A, lane 2).

An examination of the nucleotide sequence upstream of the start of ORF3 revealed a putative consensus promoter with homology to a number of broad-host-range promoters from the IncP plasmid RK2 and the IncQ plasmid RSF1010 (Fig. 3.9). Whether this putative promoter is active in *E. coli* and *P. aeruginosa* is not known. No clear consensus promoter sequence could be found in the 237 bp *P2* region.

Plasmid pTV300 has the first 1250 bp up to the *Apa*I site of the 3202 bp fragment deleted (Fig. 2.2). The intact ORF2 coding region together with the last 158 codons of ORF1 are situated downstream of the T7 promoter of Bluescript SK. When this plasmid was expressed *in vivo* without initiation of the T7 promoter , the 33.5 kD ORF2 product was not detected (results not shown). However, when transcription

from the T7 promoter was induced, the 33.5 kD protein was detected (Fig. 3.B lane 2), which implied that the promoter from which ORF2 was transcribed was situated either within the first 400 bp of the coding region of ORF1 or that the two ORFs are transcribed from the same promoter upstream of ORF1. The latter possibility seems the most likely.

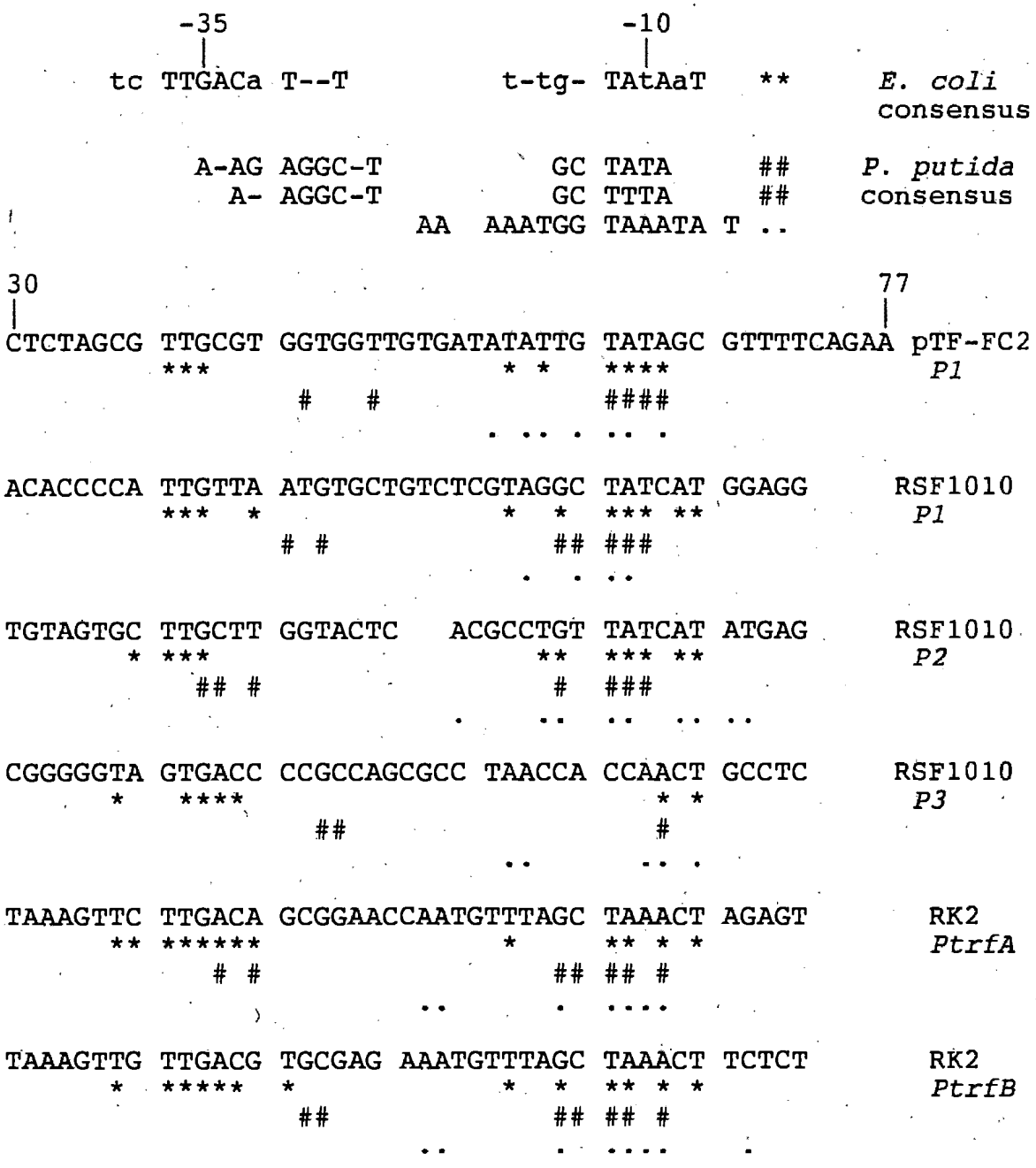

Fig. 3.9 Comparison between the sequence for the putative *P1* promoter of the pTF-FC2 replicon and promoters originating from the broad-host-range plasmids RK2 and RSF1010. Consensus sequences for the -35 and -10 regions of the *E. coli* (Hawley and McClure 1983) and three different *P. putida* promoters are shown (Mermod *et al.*, 1984). Symbols: * denotes homology with the *E. coli* consensus, # denotes homology with two similar *P. putida* consensus sequences and . denotes homology to a different *P. putida* consensus promoter sequence.

The secondary structure of a potential transcript at the start of ORF1 is interesting (Fig. 3.10). CIRs in this region have the potential to form a stable stem-and-loop structure of ΔG = -40.81 kCal. If this structure was able to form *in vivo*, the ribosome binding site and start codon of ORF1 would be obscured by a ds stem which could interfere with translation. The secondary structure around the start of ORF1 is strongly reminiscent of the cloverleaf structure typical of antisense transcripts (Simons, 1988). It is possible that the translation of ORF1 is regulated by an antisense RNA transcribed in this region. This phenomenon is common in the regulation of plasmid *rep* proteins (Thomas, 1988).

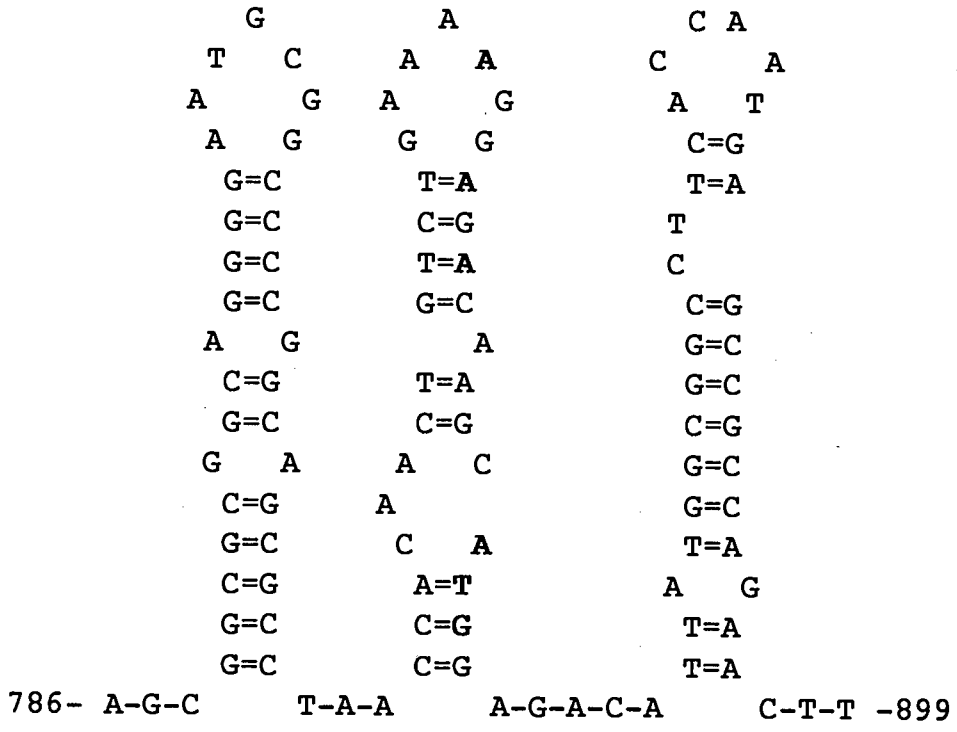

Fig. 3.10 Secondary structure of DNA around the start of ORF1. Bold type denote the ribosome binding site and ATG start codon.

3.4.2d Construction of an ORF1-LacZ fusion

To study the activity of the possible *P2* further, the fragment containing *P2* was cloned into a β-galactosidase promoter probe, pMC1403 (Fig. 3.11). Plasmid pMC1403 consists of the *lac* operon from which the promoter, ribosome binding site and start codon of the β-galactosidase gene have been removed (Cassadaban *et al.*, 1983). Three restriction sites, *EcoRI*, *Sma*I and *BamHI* allow the insertion of DNA fragments which, provided they contain a promoter, ribosome binding site and ATG start codon in frame with the truncated β-galactosidase protein, will result in the expression of β-galactosidase.

Analysis of the nucleotide sequence of pTV4101 revealed a *Sau3A* site at nt 1076 which could generate an in-frame fusion between the start of ORF1 and the *lacZ* gene if the fragment was inserted into the *BamHI* site of pMC1403 (Fig. 3.11). Plasmid pTV4101 was digested with *EcoRI* and *Sau3A*, and the DNA ligated with pMC1403 which had been digested with *EcoR*1 and *BamHI*. The transformants were plated onto LA plates containing Ap and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). Colonies containing plasmids with the promoter fragment turned blue while colonies containing promoterless plasmids remained white. Restriction and sequence analysis of the plasmid DNA isolated from blue colonies confirmed that plasmid p*P21* contained a DNA fragment which extended from nts 615 to 1076.

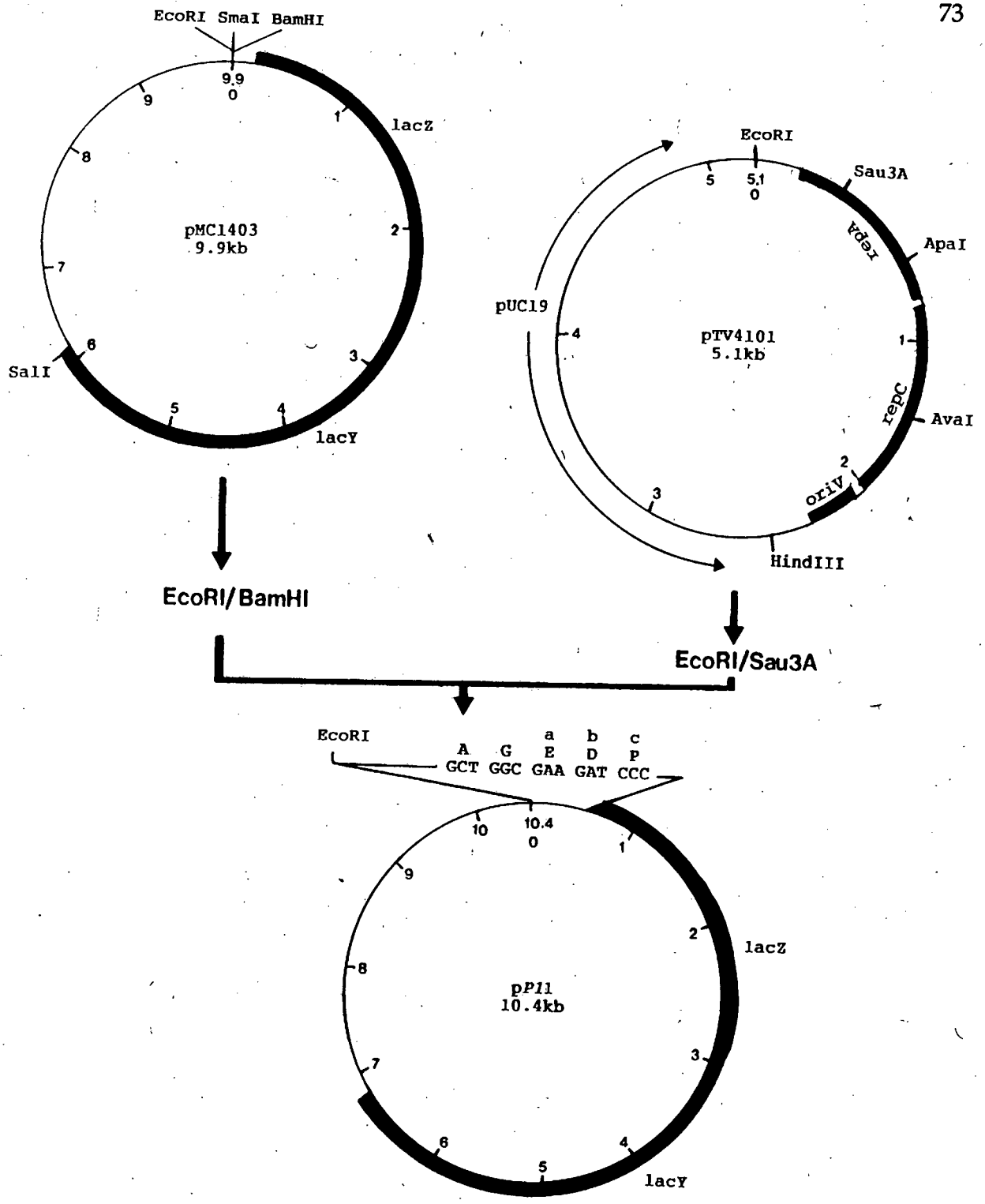

Fig. 3.11 Cloning strategy for the construction of p*P*21. The position of the *repA*, *repC* genes and the *ori*V on pTV4101 is shown as well as the codons around the fusion site between RepA and β-galactosidase.

[a] the kast ReoA codon

[b] fusion codon

[c] firt β-galactosidase codon

3.4.2e Regulation of putative promoter *P2*

β-galactosidase activity which is dependent on the activity of *P2* was assayed in *E. coli* LK111 containing p*P21* and pDR412 in *trans* to determine whether the promoter was regulated by diffusible elements from pDR412. Results are shown in Table 3.4.

Table 3.4 β-galactosidase assay for activity of p*P21*

Plasmids	β-Galactosidase activity (Miller units)
pUC 19	8.96
pMC1403	0.00
p*P21*	10.953
p*P21* + pDR412	11.295

The β-galactosidase activity measured in the assays was very low and the results indicate that *P2* has a very low level of activity, similar to that of the uninduced pUC19 *lacZ* promoter. The presence of pDER412 did not appear to affect activity of *P2*.

3.4.3 Subcloning of the *ori*V to determine the minimum sequence required in *cis* for replication

In Chapter 2.4.2c the *ori*V was located on the sequence common to deletion plasmids pTV4161 and pTV4210. However, as pTV4210 replicated at a reduced copy number, it was possible that this may not be all the sequence required in *cis* for replication. To locate the *ori*V sequence more closely, three subclones containing all or part of the 329 bp common to pTV4161 and pTV4200 were constructed and tested for their ability to replicate in *E. coli* GW125a in the presence of pDER412. Analysis of the sequence revealed two *Sau3*A (nts 2676 and 2687) and an *Hinf*I site which were used to construct the subclones.

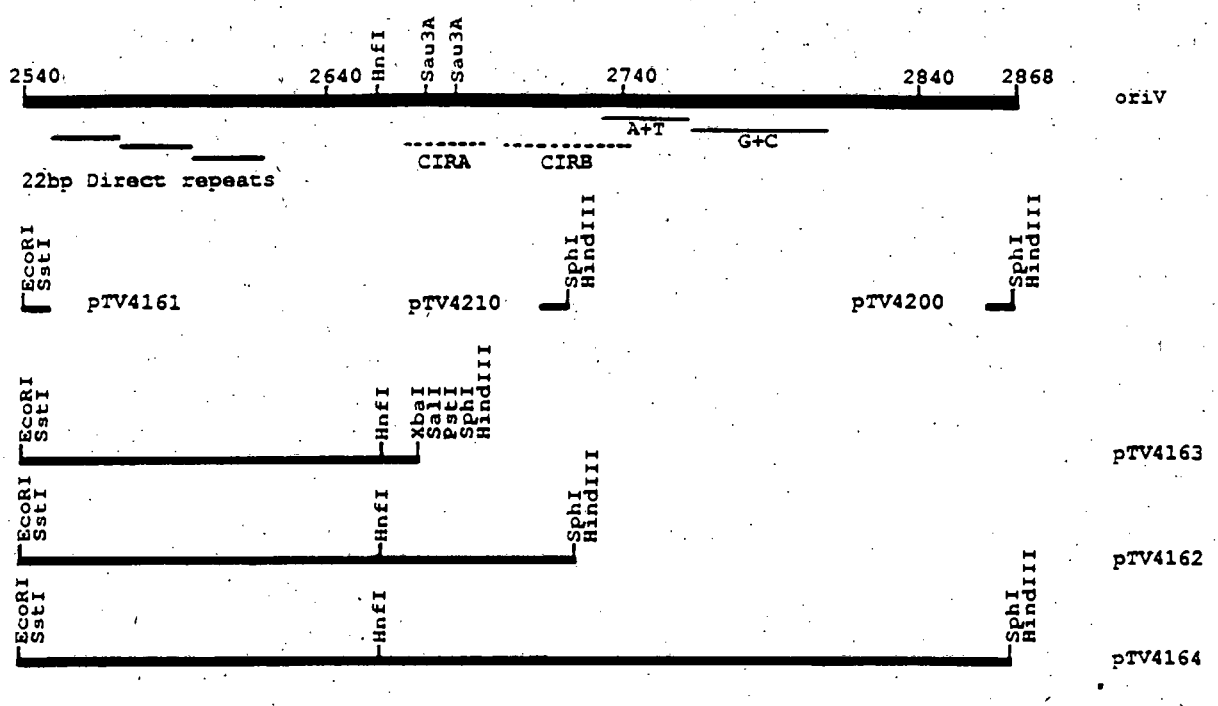

Fig. 3.12 Map of pTF-FC2 *ori*V with subclones for testing extent of the *ori*V. Below the map of the *ori*V, the position of the 22 bp direct repeats are indicated by bold arrows while the two sets of CIRs are shown by dotted lines. The regions of A+T and G+C bias are indicated by thin lines.

Plasmid pTV4163 was constructed by cloning a 148 bp *Eco*RI - *Sau*3A fragment (containing the three 22 bp direct repeats and 61 bp adjacent sequences but excluding the two sets of CIRs) from pTV4161 into pUC19 digested with *Eco*RI and *Bam*HI (Fig. 3.12). Plasmid pTV4162 was constructed by digesting pTV4161 with *Eco*RI and *Hinf*I and pTV4210 with *Hind*III and *Hinf*I and the DNA from both plasmids ligated with pUC19 digested with *Eco*RI and *Hind*III. The recombinants contained a fragment of 184 bp (Fig. 3.12) which included the three direct repeats and CIR A but had one half of CIR B deleted. Plasmid pTV4164 was constructed in the same way from plasmids pTV4161 and pTV4200 and contained the 329 bp common to the two plasmids. Because of the small sizes of the inserts in these three constructs, the DNA was sequenced to confirm that the correct fragments had been cloned.

Plasmids pTV4161, pTV4162, pTV4163, pTV4164 and pUC19 were transformed into *E. coli* GW125a containing pDER412. Ap^R colonies were detected from cells transformed with pTV4161 and pTV4164 but not from those transformed with pTV4162, pTV4163 and pUC19. To determine whether the failure to detect cells transformed with pTV4162 and pTV4163 was due to the inability of these plasmids

to be established upon transformation, plasmids pTV4161, pTV4162, pTV4163, pTV4164 and pUC19 were transformed into *E. coli* AB1157(*polA*12), transformants plated onto LA (Ap and Cm) and incubated at 28 °C. After 24 h, colonies were detected on all the plates, but those transformed with pTV4162, pTV4163 and pUC19 were very small. Colonies were picked from each plate transferred to LA (Ap) and incubated at 45 °C for 16 to 36 h. Colonies containing pTV4161 and pTV4164 were detected after 16 h at the non-permissive temperature while very small colonies containing pTV4162, pTV4163 and pUC19 were detected after 36 h. Due to the similarity in behavior between colonies containing pTV4162 and pTV4163 with those containing pUC19, it was concluded that these plasmids were unable to be maintained in AB1157(polA12) at the non-permissive temperature.

3.5 DISCUSSION

There is a marked lack of 6 bp restriction endonuclease recognition sites in the nucleotide sequence of the 3202 bp minimum replicon of pTF-FC2. This feature has also been noted in other broad-host-range plasmids (Walker, 1985) and it has been suggested that this phenomenon may be an essential feature for the survival of promiscuous plasmids.

A general comparison of the structural organisation of the pTF-FC2 *ori*V with those of the narrow host range F (Lane *et al.*, 1982), the α and β *ori*Vs of R6K (Shon *et al.*, 1982; Shafferman *et al.*, 1987) and pSC101 (Churchward *et al.*, 1983) plasmids, and the broad-host-range plasmids RK2 (Stalker *et al.*, 1981), RSF1010 (Scholz *et al.*, 1989) and R1162 (Meyer *et al.*, 1985) shows a number of common features. Firstly, the origins all contain a series of directly repeated sequences which are central to the control of replication, plasmid copy number and the determination of plasmid incompatibility. Secondly, a prominent feature of these origins are regions of dyad symmetry which in the case of the IncQ plasmids, have been shown to be the site of initiation of DNA synthesis (Lin and Meyer, 1987; Honda *et al.*, 1988). A third common feature is the presence of A+T-rich sequences usually followed by

G+C-rich sequences which play a role in the initiation of replication (Kues and Stahl, 1989).

It has been shown in narrow-host-range plasmids such as F (Rokeach *et al.*, 1985) and pSC101 (Armstrong *et al.*, 1984) and in the broad-host-range plasmids RSF1010 (Haring *et al.*, 1985) and RK2 (Krishnapillai, 1986) that the repeated sequences of the *ori*V function as binding sites for plasmid-encoded proteins which initiate DNA replication. It is probable that the 22 bp repeats serve a similar function in pTF-FC2. In some plasmids such as pSC101, additional copies of these repeats are found in the promoter area of the genes encoding these proteins, allowing autoregulation of these proteins (Linder *et al.*, 1985). This may be the role of the degenerate (14/22 bp) repeat sequence found upstream from the start of ORF1. Plasmid pTV200, which showed partial incompatibility with pDER412 (Chapter 2, Table 2.1), contains this repeat. It is tempting to suggest that the presence of the degenerate repeat is responsible for this incompatibility determinant.

Whether the two CIRs play any role in replication is unclear. The deletion in pTV4210 and pTV4262 of part of CIR B would prevent the formation of a stem-and-loop structure. In addition, the deleted sequence contains a possible n' protein-recognition site. In the IncQ plasmid RSF1010, an n' protein-recognition site is found within a region of dyad symmetry from whence DNA synthesis is initiated (Honda *et al.*, 1988). Deletion of part of one arm of the stem-and-loop reduces the rate of replication *in vitro* and inactivates the *ori*V *in vivo*. A further deletion results in the loss of replication *in vitro* (Haring and Scherzinger, 1989). There may be an analogous situation in the case of pTF-FC2. Replication of a fragment which lacked both sets of CIRs (pTV4163) and a second which had one half of CIR B deleted (pTV4162) could not be detected in either *E. coli* GW125a or AB1157(*polA*12). It is possible that replication activity on pTV4162 may be detectable in a more sensitive *in vitro* system such as that developed by Diaz and Staudenbauer (1982) where plasmid replication is measured by the incorporation of radiolabelled nucleotides in cell-free lysates. These results do however show that pTV4162 lacks an element which is involved in replication in *cis*.

Computer analysis of the nucleotide sequence revealed at least five ORFs on the 3202 bp fragment of pTV400. The arrangement of these ORFs shows considerable economy with respect to space with four ORFs sharing overlapping coding sequences. The aminoterminus of ORF2 overlaps with the carboxyterminus of ORF1, the aminoterminus of ORF4 overlaps with the carboxyterminus of ORF3 and the consensus ribosome binding site of ORF5 is situated within the coding region of ORF4 with its ATG start codon overlapping the TGA stop codon of ORF4. In addition, the carboxyterminus of ORF2 overlaps with the first 36 bp of the direct repeats of the *ori*V.

Although the *P2* promoter activity was detectable in the β-galactosidase fusion plasmid p*P2*1, on LA plates containing X-Gal, direct enzyme assays using ONPG revealed a very low level of activity. Use of ONPG rather than X-Gal is a less sensitive assay for β-galactosidase activity (D. Berger, personal communication). This explains the relatively good levels of expression observed on LA plates containing X-Gal as compared with the relatively low levels of expression when assaying with ONPG. The presence of pDR412 had no effect on the activity of *P2*. Because of the low activity of the RepA-β-galactosidase fusion, it is possible that any effect exerted by pDER412 could be undetectable.

The lack of a clear consensus promoter sequence in the region of *P2* could indicate a low affinity for *E. coli* RNA polymerase to the site and result in low promoter activity in *E. coli*. If an antisense RNA is produced from the region around the start of ORF1, it may also be produced from p*P2*1 and continue to negatively regulate translation of the ORF1-*lac*Z fusion. This would further repress expression from an already weak promoter.

Attempts to identify the precise location of *P2* using primer extension from a point within the aminoterminus of ORF1 were unsuccessful. The failure of the primer extension to detect a transcript from *P2* could be due to fact that *P2* is tightly regulated and that the transcript has a very low copy number.

Of the five ORFs identified from the DNA sequence, four were shown to produce proteins. Deletion of ORFs 3, 4 and 5 did not affect plasmid copy number in *E. coli*. It seems therefore that only the 32 kD and 33.5 kD products of ORFs 1 and 2 respectively are required for replication of pTF-FC2 in *E. coli*.

This, however, is not the case for *P. aeruginosa*, where a deletion of 360 bp from the front end of pTV400 (and the concomitant loss of at least the products of ORF3 and ORF4) resulted in the inability of pTV4100 to replicate in this host. There are two possible reasons for this. Either the proteins produced by ORF3, ORF4 and possibly ORF5 are required for replication, or alternatively, the *P2* is not expressed in *P. aeruginosa* and the loss of *P1* results in loss of expression of the products of ORF1 and 2.

In a review of the replication of plasmids in Gram-negative bacteria, Kues and Stahl (1989) aligned promoters from broad-host-range plasmids RK2 and RSF1010 with promoters from plasmids ColE, pSC101 and R401 which replicate only in *E. coli* and other closely related *Enterobacteriaceae*. They noted that there was no noticable difference between broad- and narrow-host-range promoters.

Kues and Stahl (1989) postulated that the inability of narrow-host-range plasmids to replicate in other hosts may not be due to the failure of their promoters to be expressed in these bacteria. This may be due instead, to the inability of regulators to modulate the activity of these promoters at levels required for replication in the particular host. This hypothesis is corroborated by the results of Lodge *et al.*, (1990) who showed that promoters tested for activity in *E. coli* and in *P. aeruginosa* were equally active in both hosts. Furthermore, dependence of activity upon promoter sequence was the same in both *E. coli* and *P. aeruginosa*. Furthermore, there is evidence that deletion of a protein which negatively regulates expression of the RSF1010 *rep* proteins, results in increased activity of the promoter and also in the loss of the ability of the plasmid to replicate in *P. aeruginosa* (Frey and Bagdasarian, 1989). Based on this evidence, it is more probable that the host range of pTF-FC2 is determined by one or more of the three small proteins encoded by ORF3, ORF4 or ORF5.

CHAPTER 4

A COMPARISON BETWEEN THE REPLICONS OF pTF-FC2 AND THE INCQ PLASMIDS

CHAPTER 4

A COMPARISON BETWEEN THE REPLICONS OF pTF-FC2 AND THE INCQ PLASMIDS

4.1 SUMMARY

The DNA sequence of the minimal replicon of pTF-FC2 was aligned and compared with that of the sequence of the IncQ plasmid RSF1010. There was considerable DNA homology between the *ori*V regions of the two plasmids. There was also similarity between the sequence coding for ORF1 and ORF2 of pTF-FC2 and the region coding for RepA and RepC of RSF1010. The 22 bp direct repeats of pTF-FC2 and the sequence immediately adjacent to the repeats showed 60% homology with the equivalent area on the RSF1010 *ori*V. On the protein level, there was considerable conservation between the predicted amino acid sequence of ORF1 and ORF2 and the RepA and RepC proteins of the IncQ replicons. No homology was detected between ORF3, ORF4 and ORF5 and other proteins encoded by the IncQ replicon. Despite these similarities, no incompatibility was detected between the two types of plasmids and the IncQ replicon, R300B, was unable to complement mutations in the *rep* proteins of pTF-FC2 .

4.2 INTRODUCTION

A powerful tool in an investigation into a new biological system is the use of comparison with other systems which have already been extensively studied and characterised. A comparison of the structure and organisation of the pTF-FC2 replicon with those of other, better studied broad-host-range plasmids such as the

IncP and IncQ plasmids might shed light on the functions of the various elements identified in the functional and sequence characterisation of pTF-FC2. Furthermore a comparison between the different groups of broad-host-range plasmid replicons may help to formulate a general model for broad-host-range replication.

In this chapter, the pTF-FC2 minimal replicon is compared and contrasted with that of the IncQ plasmids at the DNA and amino acid level. As the two plasmids were clearly related, an investigation into the compatibility and the ability of the IncQ plasmids to complement pTF-FC2 mutants was carried out.

4.3 MATERIALS AND METHODS

4.3.1 Bacterial and plasmid strains

E. coli strain LK111 was used for general plasmid maintenance while strains GW125a and AB1157(*polA*12) were used to test for PolA-independent replication. (Appendix 1, Chapter 2.3.1). Plasmid pKE462, a derivative of R300B, (Appendix 2) was used to test for incompatibility of pTF-FC2 with the IncQ replicons and also to test for complementation of *rep* mutations in the pTF-FC2 origin by the IncQ replicon.

4.2.2 Recombinant DNA techniques and plasmid constructions

General techniques were as described by Maniatis *et al.* (1989). Plasmids pTV4102 and pTV4103 were constructed as follows. Plasmid pTV4100 DNA (Fig. 3.2) was digested with restriction endonuclease *Apa*I which has a single site within the coding region of ORF1. The 3' overhanging ends created by *Apa*I were flushed using the 3'-5' Exonuclease activity of the Klenow fragment of DNA polymerase I. One unit of enzyme was added per mg DNA in 10 ml restriction buffer and the reaction

incubated at room temperature for 15 min after which the DNA was ligated and transformed into *E. coli* LK111. Plasmid DNA from Ap^R colonies was analysed by restriction with *Apa*I and plasmid pTV4102, which had lost the *Apa*I site, was identified. Plasmid pTV4103 was constructed in a similar fashion by flushing the 5' overhanging ends created after cleavage of the single *Ava*I site of pTV4100 (Maniatis *et al.*, 1989), religation and selection for a plasmid which had lost the *Ava*I site.

The UWGCG and Genepro packages were used for computer analysis, comparison and alignment of DNA sequences.

4.4 RESULTS

4.4.1 DNA sequence alignment of the replicons of pTF-FC2 and RSF1010

A dot matrix alignment between the nucleotide sequences of the 3202 bp minimal replicon of pTF-FC2 and the basic replicon of the IncQ plasmid, RSF1010 (Scholz *et al.*, 1989) revealed two distinct regions of RSF1010 showing greater than 60% DNA homology with the 3202 bp fragment (Fig. 4.1). The first region corresponded to the *ori*V sequences of the two plasmids and the second area of similarity lay between the regions encoding the RepA and RepC proteins of RSF1010 and ORF1 and ORF2 of pTF-FC2. There was no further significant homology between the first 800 bp and the last 470 bp of the 3202 bp fragment and the RSF1010 replicon.

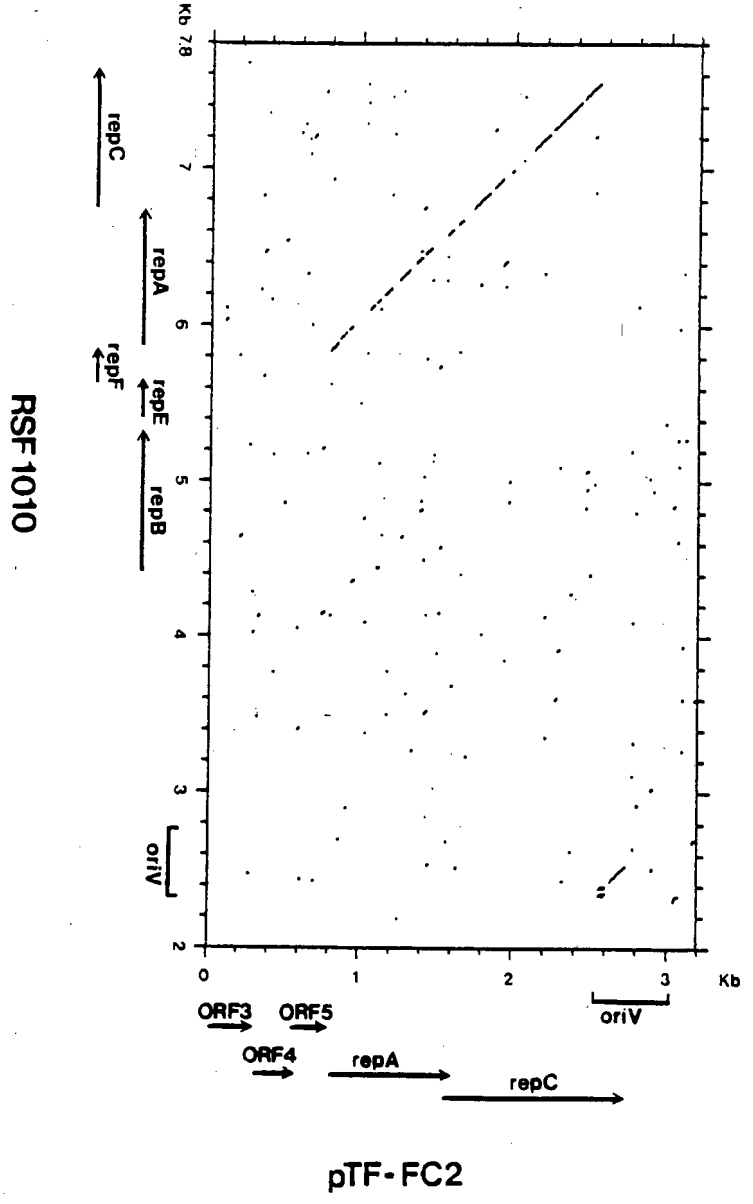

Fig. 4.1 Alignment of the nucleotide sequence of the 3202 bp DNA fragment contained within pTV400 with nts 2000 to 7800 of RSF1010 (Scholz *et al.*, 1989). This region of RSF1010 contains all sequences involved in replication and mobilisation, but excludes the sulfonamide and streptomycin resistance coding regions.

4.4.2 Comparison between the *ori*V regions of pTF-FC2 and RSF1010

The DNA sequence of the 329 bp fragment which contained the *ori*V of pTF-FC2 (Chapter 3.4.1) was aligned and compared with the DNA sequence of the *ori*V of RSF1010 (nts 2347 and 2742, Scholz *et al.*, 1989). Substantial homology was found as shown in Fig. 4.2. The 20 bp repeated sequences (including the 2 bp spacer nucleotides) of the IncQ plasmids had a 40/66 bp homology (60.1%) with the three 22 bp repeated sequences of pTF-FC2 . There was 75% similarity between the DNA sequence found immediately adjacent to the repeats (pTF-FC2 nts 2613 - 2727, RSF1010 nts 2415 - 2520). Beyond this region the two sequences showed only limited DNA homology.

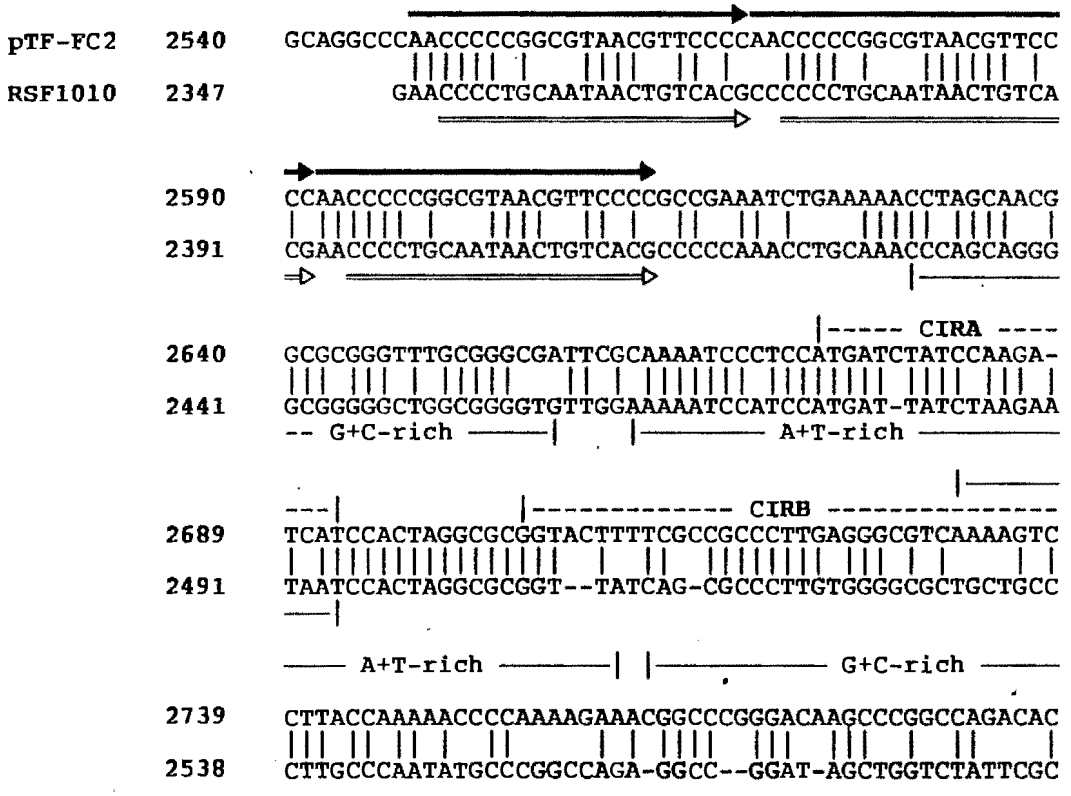

Fig. 4.2 A comparison and alignment of the *ori*V sequences of pTF-FC2 and RSF1010. Bold arrows denote the 22 and 20 bp direct repeats of pTF-FC2 and RSF1010 respectively, while solid lines delineate areas with an A+T and G+C sequence bias. Dotted lines show the position of CIRs A and B on the pTF-FC2 *ori*V.

Meyer *et al.* (1985) aligned the repeats of R1162 with those of the broad-host-range IncP plasmid RK2 and found sequence conservation between the two sets. The sequence of the repeats from the pTF-FC2 *ori*V was added to this alignment (Fig. 4.3) and the areas conserved between the IncQ plasmids and pTF-FC2 do not entirely reflect the consensus found between the IncP and IncQ repeats.

```
- -   C C C - - - - - - - A A C - - T C - - -      Consensus

A A C C C C C G G C G T A A C G T T C C C C      pTF-FC2
| | | | | |   |       | | | |     | |   |
A A C C C C T G C A A T A A C T G T C A C G      IncQ
      | | | |   |       | | | | | | | |
C C G C C C T - C - - - A A C T G T C A          RK2
```

Fig. 4.3 Alignment of the nucleotide sequence of the repeated sequences found in the origins of the broad-host-range plasmids pTF-FC2, R1162 and RK2. Consensus between the repeats is shown above.

The sequence CPyCTCA is, with few exceptions, present in the direct repeats of plasmids F, R6K, RK2 and P1. This motif is thought to be important in the binding of the initiator protein to the repeated sequences (Filutowicz *et al.*, 1985). Persson and Nordström (1986) reported that this motif also occurred in RSF1010 (CTC(G)TCA) where the C in position 4 was not conserved. This sequence was not found in the repeats of pTF-FC2.

4.4.3 Similarity between the ORFs of the pTF-FC2 minimal replicon and the *rep* proteins of RSF1010

From the alignment of the DNA sequences of the pTF-FC2 and RSF1010 replicons (Fig. 4.1) it was clear that there was considerable homology between the area encoding ORF1 and ORF2 of pTF-FC2 and the region coding for *rep*A and *rep*C on RSF1010. The predicted amino acid sequence of the ORF1 and ORF2 polypeptides were aligned and compared with the amino acid sequences of RepA and RepC of RSF1010 (Scholz *et al.*, 1989).

```
TF-FC2    RepA  1    MALDIMAAFTNEPPELDFIWPGFLAGTVGALVAPGATGKSF
RSF1010   RepA  1    MATHKPINILEAFAAAPPPLDYVLPNMVAGTVGALVSPGGAGKSM

                42    FALEAAMSIACSVAGGDLVGLTPAHTGRVVYLAGEDPQPALVRRV
                46    LALQLAAQIA---GGPDLLEVGELPTGPVIYLPAEDPPTAIHHRL

                87    HAIGQHLNQSAREAIAENLMLEPIMGKRLNVMDDAHLRRVIDYSA
                89    HALGAHLSAEERQAVADGLLIQPLIGSLPNIMAPEWFDGLKRAAE

               132    GARLIVLDTLSRIHILDENSNGDMAHLVSVLEHIAATTGAAVLYL
               133    GRRLMVLDTLRRFHIEEENASGPMAQVIGRMEAIAADTGCSIVFL

               177    HHVNKGSARDGQTDQQQAARGASALIDNARWCGYVAKMTEQEAER
               178    HHASKGAAMMGAGDQQQASRGSSVLVDNIRWQSYLSSMTSAEAE-

               222    MSDRGFDRSPSGNERRGLFVRFGVSKQNYDATPLDRWYQRHSGGV
               222    ------EWGVDDDQRR-FFVRFGVSKANYGAPFADRWFRRHDGGV

               267    LLPVELQEAISNGAGKKGGKRNEL
               260    LKPAVLERQRKSKGVPRGEA


TF-FC2    RepC  1    MEQEKKGESAMSYDLTHARHDPAHCLTPGLFRSLKRGERKRLKLDVTYNY
RSF1010   RepC  1          VKPKNKHSLSHVRHDPAHCLAPGLFRALKRGERKRSKLDVTYDY

                51    GDD-SIRFWGPEPLGGDDLRVLQGLVAMAAISGDNGRGIVLRHETESEAG
                43    GDGKRIEFSGPEPLGADDLRILQGLVAMAGPNG-----LVLGPEPKTEGG

               100    QQLRLWLDMRWDAIEKDTMVAKGSFRQLARELGYAEDGGSQFKTIRESIE
                88    RQLRLFLEPKWEAVTAECHVVKGSYRALAKEIGAEVDSGGALKHIQDCIE

               150    RLWAVSVIVERGGKRQGFRILSEYASDEQEGKLFVALNPRLADAVMGERP
               138    RLWKVSIIAQNGRKRQGFRLLSEYASDEADGRLYVALNPLIAQAVMGGGQ

               200    HTRINMAEVRKLETDPARLLHQRLCGWIDPGKSGKAEIDTLCGYVWPDAA
               188    HVRISMDEVRALDSETARLLHQRLCGWIDPGKTGKASIDTLCGYVWPSEA

               250    NDEAMKKRRQTARKALVELAAVGWTVNEYAKGKWEISRPN-----PRRNV
               238    SGSTMRKRRQRVREALPELVALGWTVTEFAAGKYDITRPKAAGLPPPLYC

               295    PQPPA
               288    KQDTF
```

Fig. 4.4 Alignment of the amino acid sequences of the RepA and RepC proteins of pTF-FC2 and RSF1010 using the Genepro computer sequence analysis package. Identical and conservative amino acid changes are indicated by double and single dots respectively between the sequences.

The predicted ORF1 protein is 11 amino acid residues larger than RepA, and on alignment the two proteins have a similarity of 47% based on the perfectly conserved amino acid residues (Fig. 4.4). An alignment between the larger of the two possible ORF2 polypeptides and RepC of RSF1010 is also shown in Fig. 4.4 and the percentage similarity between these proteins was calculated at 60%. Homology between both sets of proteins was evenly distributed across the length of the sequence, with some divergence at the amino- and carboxytermini of the polypeptides. Due to the similarity between the ORF1 and ORF2 proteins of pTF-FC2 and RepA and RepC of RSF1010, the corresponding pTF-FC2 genes were named *rep*A and *rep*C respectively.

Analysis of the amino acid sequence of pTF-FC2 RepA and RepC genes revealed a consensus helix-turn-helix motif in each. This motif is conserved in a variety of DNA binding proteins such as the RK2 KorA protein (Kornacki *et al.*, 1987), the RepA protein from pLS1 (del Solar *et al.*, 1989) and the RepE protein of plasmid F (Masson and Ray 1988). Similar motifs were found in the RepA and RepC proteins of RSF1010, but the position of the consensus sequence on RepC was different from that on the pTF-FC2 proteins. The region of the RSF1010 RepC protein corresponding to the DNA binding motif of the pTF-FC2 RepC contains the conserved amino acid residues, but the spacing between alanine and glycine is not conserved. Similarly, the spacing of the consensus amino acid residues of the pTF-FC2 protein which corresponds with the consensus sequence of the RSF1010 RepC protein, is not conserved. An alignment of the helix-turn-helix sequences of the pTF-FC2 and RSF1010 RepA and RepC proteins with the consensus sequence of Pabo and Sauer (1984) is shown in Fig. 4.5.

Protein		Position	Sequence
			1 5 10 15 20
			Helix Turn Helix
Consensus			. . . * A . . * G * . . . . I V L . . * * .
pTF-FC2	RepA	161	L E H I A A T T G A A V L Y L H H V N K
RSF1010	RepA	162	M E A I A A D T G C S I V F L H H A S K
pTF-FC2	RepC	73	L V A M A A I S G D N G R G I V L R H E
RSF1010	RepC	176	L I A Q A V M G G G Q H V R I S M D E V
pLS1	RepA	17	L E K M A R E M G L S K S A M I S V A L
RK2	KorA	37	Q A T F A T S L G L T R G A V S Q A V H
F	RepE	64	V A K Y A E I F G F T S A E A S K D I R

Fig.4.5 Helix-turn-helix regions in the RepA and RepC proteins of pTF-FC2 and RSF1010. The consensus pattern of αhelix-turn-αhelix regions of known DNA-binding proteins is shown above. The highly conserved amino acid residues in positions 5, 9 and 15 are shown in bold type while asterisks represent non-polar and dots represent hydrophilic amino acid residues.

A comparison of the predicted amino acid sequence of ORFs 3, 4 and 5 with the proteins of RSF1010 using the Fasta subroutine of the UWGCG Vax package, revealed no similarities between the proteins. A similar search of the GenEMBL sequence data bank revealed no homology between the polypeptides produced by these ORFs and sequences in the data bank.

4.4.4 Incomaptibility between pTF-FC2 and the IncQ replicons

Previously it had been shown that the major incompatibility determinants of pTF-FC2 lay in the 184 bp common to pTV4161 and pTV4210 (Chapter 2.4.2d). The 20 bp repeated sequences of the IncQ *ori*V have been found to be responsible for the determination of incompatibility (Persson and Nordström 1986; Lin *et al.*, 1987). Since there was a considerable degree of homology between the repeated sequences and adjacent DNA regions of the origins of pTF-FC2 and corresponding areas in the IncQ replicon, it was dtermined whether any incompatibility could be detected

between pTF-FC2 and R300B. Since R300B confers only Su^R and Sm^R and the *E. coli* strains used for the experiments were Sm^R, a derivative of R300B, pKE462, was used. Plasmid pKE462 has a 7.5 kb *EcoR*I DNA fragment carrying arsenic and tetracycline (Tc) resistance from plasmid R46 inserted into the single *EcoR*I site of R300B. The derivative therefore consists of an intact IncQ replicon with a Tc^R marker inserted into the Sm-Su operon.

Incompatibility between the pTF-FC2 and IncQ replicons was tested in two ways. In the first, pKE462 was transformed into a strain already containing a pTF-FC2 replicon and secondly, a series of pTF-FC2 deletions were transformed into a strain containing pKE462. In this way, it was possible to test for incompatibility exerted by the IncQ replicon on pTF-FC2 and the reverse effect of the pTF-FC2 origin on the replication of the IncQ plasmid.

Plasmid pKE462 was transformed into *E. coli* GW125a (*polA*$^-$) containing pDER412 and the cells tested for Cm^R after 60 generations of selection on Tc. All 60 colonies tested contained pDER412, implying that the presence of pKE462 had no detectable effect on the maintenance of pDER412 in GW125a. Plasmids pDER412, pTV400, pTV4210, pTV44220, pTV4161, pTV300 (Fig 2.2) and pUC19 were transformed into *E. coli* LK111 containing pKE462 and the transformants grown under Cm or Ap selection for 60 generations. None of these plasmids were able to displace pKE462 which showed that the presence of the pTF-FC2 incompatibility determinants had no detectable effect on the replication of pKE462 and that the two replicons were compatible.

4.4.5 Complementation of pTF-FC2 *rep*$^-$ mutations and deletions by the IncQ plasmid R300B

Since there was a high degree of homology between the RepA and RepC proteins of pTF-FC2 and RSF1010, it was important to determine whether the IncQ proteins were able to complement mutations and deletions in the pTF-FC2 replicon. Deletion plasmid pTV4161, containing only the region nts 2540 - 3202 (Fig. 3.2), which includes the intact *oriV* but none of the ORFs, was able to replicate in *E. coli* GW125a

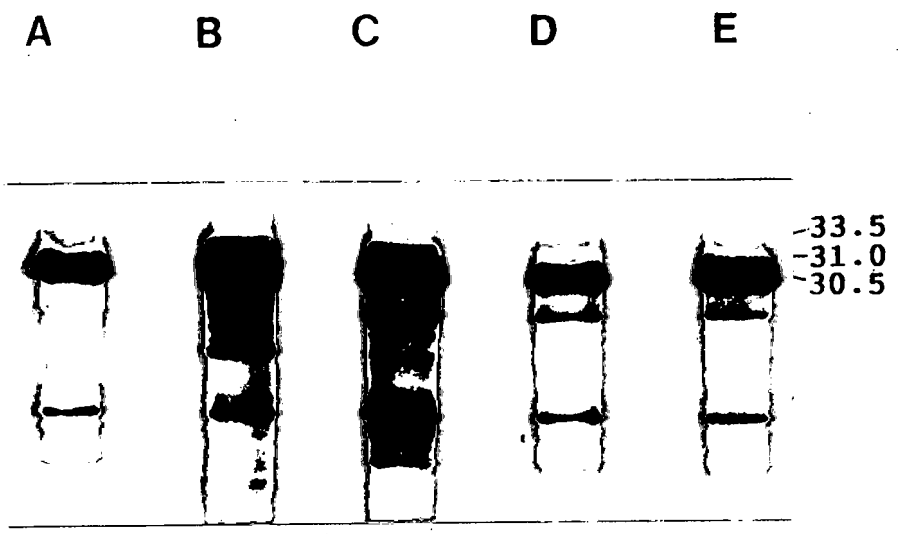

Fig. 4.6 SDS-PAGE analysis of proteins produced *in vitro* showing the inactivation of *rep*A and *rep*C in plasmids pTV4102 and pTV4103 respectively. Lanes A to E represent plasmids pUC19, pTV400, pTV4100, pTV4102 and pTV4103 respectively.

Plasmids pTV4102 and pTV4103 were subsequently used to investigate whether the individual RepA and RepC proteins of R300B could complement mutations in their respective equivalents on the pTF-FC2 origin. Plasmids pTV4102 and pTV4103 were transformed into *E. coli* GW125a containing pKE462. No transformants were obtained from cells transformed with these plasmids while more than 1000 colonies were obtained from cells transformed with an equivalent amount of pTV400. It was concluded that the IncQ replicon of pKE462 was unable to complement either the *rep*A or *rep*C mutants.

It was possible that the mutations in *rep*A and *rep*C could have affected plasmid establishment, but that pKE462 could still complement plasmid maintenance. The temperature sensitive *polA*$^-$ system was used to address this question. Plasmid pKE462 was transformed into AB1157(*polA*12) and transformants selected by growth on on LA containing Tc at 28 °C. When the colonies were transferred to 42 °C, small colonies were detected after at least 24 h.

When broth cultures which were to be used for the preparation of competent cells were grown overnight at 37 °C, the culture took 48 h to grow to stationary phase and competent cells could only be prepared from cultures grown overnight at 28 °C. It was concluded that the R300B replicon was unstable at 42 °C in AB1157(*polA*12) and that this system would not be suitable for testing for complementation of the *rep*⁻ mutants by pKE462. Support for this result was found subsequently in a paper by Grindley and Kelley (1976) who reported instability of an IncQ plasmid in an *E. coli* carrying the *polA*⁻*ts*12 mutation.

4.5 DISCUSSION

There are several similarities between pTF-FC2 and the 8.7 kb IncQ plasmids, RSF1010, R1162 and R300B. Both types of plasmid have a wide host-range, are able to replicate independently in *E. coli polA*1 mutants and they have a comparable copy number in *E. coli* (Chapter 2.4.2e; Barth and Grinter, 1974). There are, however, some substantial differences. In RSF1010 and R1162 the minimum replicon required for broad-host-range replication is spread over approximately 5.5 kb of DNA and is interspersed with regions not required for replication (Bagdasarian *et al.*, 1982; Meyer *et. al.*, 1982b). The pTF-FC2 replicon seems to be more compact and is confined to a 2.8 kb DNA fragment (Chapter 2.4.3).

Plasmid pTF-FC2 and the IncQ plasmids have been shown to be similar in the structure and nucleotide sequence of their replicons. From an alignment of the nucleotide sequences, it is clear that there is considerable similarity between the *ori*Vs and some of the *rep* proteins. Comparisons between the 329 bp *ori*V fragment of pTF-FC2 and its equivalent on the IncQ plasmids, R1162 and RSF1010, indicated 65% similarity between the first 200 bp of both origins, after which the sequences diverge considerably.

An alignment of the two sets of repeated sequences (including the 2 spacer nucleotides of the IncQ plasmids) implies that the unit length of the IncQ repeats

may be 22 bp, with the first two nucleotides imperfectly conserved. The similarity between these two sets of repeats suggests that they share a common function in their respective origins. Alignment of these repeats with IncP iterons did not reflect the similarities which Meyer *et al.* (1985) postulated may indicate a broad-host-range characteristic shared by the IncQ and IncP plasmids. Rather, sequence consensus may reflect a structural conservation which is required for the binding of initiator proteins to the DNA. Support for this possibility is provided by Smith and Thomas (1985), who reported similarities between the repeats of the IncP plasmids and *E. coli* plasmids from six different incompatibility groups, while Kues and Stahl (1989) have found similarities between the broad-host-range IncQ and IncP plasmids, *E. coli* plasmid F and the *Methylomonas clara* plasmid pBE-2.

Scholz *et al.* (1989) identified a G+C-rich area (24/28 bp) followed by an A+T-rich stretch (24/31 bp) of nucleotides in the area directly adjacent to the repeated sequences of RSF1010. This area shows 75% similarity with the equivalent region on the pTF-FC2 *ori*V but the G+C-rich and A+T-rich regions appear to be less pronounced (20/28 and 19/31 bp respectively). In the IncQ plasmids, this region may be deleted without affecting replication (Haring and Scherzinger, 1989). The pTF-FC2 *ori*V contains areas with a high G+C and A+T content further downstream in the region which bears little homology to its IncQ equivalent (Fig. 3.4), but whether these sequences are involved in the replication process is unknown.

It has been shown in R1162 that two distinct domains, contained on adjacent 370 and 210 bp *Hpa*II fragments, are required in *cis* for plasmid replication (Kim *et al.*, 1987). These domains are able to direct plasmid replication even when the distance between them is increased. The *ori*V of pTF-FC2 has extensive homology with sequences contained entirely within the 370 bp HpaII fragment. The 210 bp *Hpa*II fragment of R1162 has been shown to contain two 10 bp inverted complementary repeats separated by a region of extensive dyad symmetry (Meyer *et al.*, 1985; Scholz *et al.*, 1989) and it has been shown *in vitro* that initiation of DNA replication

occurs bidirectionally at these two sequences (Lin and Meyer, 1987; Honda *et al.*, 1988). It is possible that the absence of homology between this region and the pTF-FC2 *ori*V may reflect differences in the replication mechanisms of the two plasmids.

The high degree of homology between some of the *rep* proteins of the two groups of plasmids is maintained at both the DNA and amino acid levels, which along with the observation that the similarity is distributed evenly across both proteins, is an indication that the two sets of proteins have similar functions in the replication of their respective origins. In spite of these similarities, the IncQ RepA and RepC proteins were unable to complement mutations in their equivalent proteins on the pTF-FC2 origin.

The RepA protein of RSF1010 has been reported to be a helicase and there is convincing evidence that RepC is a DNA binding protein (Haring and Scherzinger, 1989). RepC is thought to be involved in initiation of replication by binding specifically to the 20 bp repeats of the IncQ *ori*V (Haring *et al.*, 1985). These repeats have been found to be responsible for plasmid incompatibility and copy number control in RSF1010 and R1162 (Persson and Nordström, 1986; Lin *et al.*, 1987). Despite 60% DNA homology between the repeated sequences of the two plasmids, no incompatibility was detected. Lin *et al.* (1987) postulated that plasmid incompatibility was directly dependent on the ability of RepC to recognise and bind to the 20 bp repeats of the *ori*V. The absence of detectable incompatibility between the pTF-FC2 *oriV* and R300B, suggests that the inability of the IncQ RepC protein to complement a pTF-FC2 *rep*C mutant could be due to its inability to recognise and bind to the pTF-FC2 repeated sequences.

Besides the *rep*A and *rep*C gene products, the IncQ plasmids have an absolute requirement for the product of *rep*B gene for replication to occur *in vitro* in *E. coli* (Scherzinger *et al.*, 1984). The 35.5 kD RepB protein has been shown to be involved in primer synthesis at the *oriV* during initiation of replication (Honda *et al.*, 1989). The carboxy-terminus of the *rep*B gene is situated within 500 bp of the start of *repA*.

No equivalent of this protein was found in the 850 bp of sequence upstream of the start of pTF-FC2 *rep*A start.

The ability of pTF-FC2 derivatives to be maintained in *P. aeruginosa* appeared to be linked to the presence of at least two proteins of 8 kD and 10 kD which corresponded to the products of ORF3 and ORF4 respectively. In the IncQ plasmid, RSF1010, a small 10 kD protein produced by a gene upstream of *rep*B has been implicated in the determination of host range (Scholz *et al.*, 1985; Frey and Bagdasarian, 1989). There was no significant homology between this protein and the products of ORF3, ORF4 or ORF5.

The IncQ RepA and RepC proteins are expressed from a polycistronic mRNA (Scholz *et al.*, 1985) and their expression is controlled at the level of translation by an antisense RNA complementary to the ribosome binding site and start of the *repA* gene (Kim and Meyer, 1986). In addition, expression of IncQ *rep*C appears to be translationally coupled to expression of *rep*A (Scherzinger *et al.*, 1984). A potential stem-and-loop structure which could obscure the ribosome binding site and ATG start codon of RepC in a transcript, is proposed by Haring and Scherzinger (1989) to be responsible for the translational coupling of RepA and RepC. Translation of RepA would make the start of RepC available for ribosome binding and subsequent translation. In contrast, Meyer *et al.* (1985) were able to express the R1162 *rep*C independently of RepA translation from a *tac* promoter and complement *repC*$^-$ derivatives of R1162. Whether these conflicting results indicates a difference between RSF1010 and R1162 or an experimental artefact in either system can only be explained by comparison of the nucleotide sequences of the two plasmids.

Although the pTF-FC2 *rep*A and *rep*C genes are structurally similar to their counterparts on the IncQ origin, there is evidence that expression of these genes may be regulated differently. The expression of RepC does not appear to be coupled to that of RepA. Evidence for this is the production of RepC from the T7 promoter of Bluescript SK in the absence of the *rep*A gene (Chapter 3 Fig. 3.8A), and that RepC is produced even when *rep*A has been inactivated by a frameshift mutation as reported here. The secondary structure of the DNA at the start of *rep*A suggests that

translation of at least RepA, may be regulated by an antisense RNA molecule as is the case in the IncQ replicon, but there is no significant dyad symmetry in the sequence around the end of *rep*A and the start of *rep*C.

A comparison of the structure of the pTF-FC2 replicon with that of the IncQ plasmids shows that in both types of plasmids, the *rep*A genes, *rep*C genes and *ori*V regions have a considerable amount of nucleotide homology. In pTF-FC2 the two regions that show homology are contiguous, whereas in RSF1010 the *oriV* is separated from the *rep*A and *rep*C by a 2484 bp fragment that includes the Su^R and Sm^R genes. These genes have a different codon usage pattern from the rest of the genes found on RSF1010. Scholz et al. (1989) suggest that these genes may have been incorporated at a later stage in the development of the IncQ plasmids. The finding that the two homologous regions are contiguous in pTF-FC2 indicates that the two replicons may be derived from a common ancestor from which they have subsequently diverged. In the case of the IncQ plasmids, the *rep*A and *rep*C genes may have been separated from the *ori*V by insertion of a transposable element such as Tn*4* which carries Su^R and Sm^R markers.

The presence of sequences showing strong homology to an inverted repeat and the carboxyterminus of the transposase of members of the Tn*3* family adjacent to the *ori*V (see Chapter 3 section 3.4.1a) suggests that a smilar event may have occured in pTF-FC2. Preliminary sequence from the region downstream of the righthand *Sau3*A site of pTV400 has revealed the presence of a transposase-like gene which shows strong homology with that of Tn*501* and Tn*21* (D.E. Rawlings, personal communication). A sequence homologous to the terminal repeat has been found approximately 3 kb downstream adjacent to the *mob* region as well as an ORF which is homologous to the *merR* gene of Tn*501* (J. Rohrer, personal communication). These results suggest that a Tn*501*-like transposable element may have been inserted between the *mob* and *rep* regions of pTF-FC2 at some stage during the evolution of pTF-FC2.

CHAPTER 5

CHARACTERISATION OF AN ADDITIONAL 1239 BP FRAGMENT WHICH IS REQUIRED FOR AUTONOMOUS REPLICATION OF THE pTF-FC2 REPLICON

CHAPTER 5

CHARACTERISATION OF AN ADDITIONAL 1239 BP FRAGMENT WHICH IS REQUIRED FOR AUTONOMOUS REPLICATION OF THE pTF-FC2 REPLICON

5.1 SUMMARY

During the course of experiments aimed at the construction of a broad-host-range cloning vector based on the pTF-FC2 replicon, the 3202 bp basic replicon was found to have an absolute requirement for the pUC19 replicon in the absence of coresident pDER412. An additional 1239 bp, located upstream of the start of the 3202 bp fragment, was identified which restored the ability of this fragment to replicate autonomously. Sequence analysis of the region revealed a single ORF encoding a polypeptide with a predicted M_R of 40 kD which was identified by *in vitro* transcription and translation. A comparison of the amino acid sequence of this protein with the data banks revealed limited homology with RepB' primase of the IncQ plasmid, RSF1010. The nature of the complementation by pUC19 of the *repB*$^-$ phenotype of the 3202 bp fragment, was investigated.

5.2 INTRODUCTION

The use of the pTF-FC2 replicon as a cloning vector for the introduction of recombinant genes into *T. ferrooxidans* was one of the aims of this project. In addition to its importance as a genetic tool in *T. ferrooxidans*, the apparently wide host-range of this replicon could result in its use as a cloning vector for many other bacteria. The relatively low copy number of the replicon makes it useful for the isolation of potentially lethal genes in heterologous systems. The small size of the

replicon compares very favourably with the sizes of available broad-host-range vectors currently available (Bagdasarian *et al.*, 1981; Bagdasarian and Timmis, 1982; Priefer *et al.*, 1985).

Comparison of pTF-FC2 with the IncQ plasmids has revealed a very interesting feature of the 3202 bp identified previously as the minimal replicon of pTF-FC2. The IncQ plasmids are unable to replicate in the absence of the RepB' primase protein (Haring and Scherzinger, 1989). No equivalent of this protein could be found on the pTF-FC2 fragment. This result implied that pTF-FC2 used the host priming system (DnaG or RNA-Pol) which is unexpected in view of the similarity between the RepA and RepC proteins and the *ori*V of the replicons.

This chapter details how attempts to construct pTF-FC2-based vectors led to the discovery of a primase-like protein located outside the 3202 bp fragment.

5.3 MATERIALS AND METHODS

5.3.1 Bacterial strains and plasmids

E. coli strains LK111 and a *recA*$^-$ derivative of LK111, LK112, were used for the maintenance and preparation of plasmid DNA. The ability of plasmids to replicate in the absence of Pol I was tested in *E. coli* GW125a (Chapter 2.3.1 and Appendix 1). Vectors Bluescript SK and pUC19 are described in Appendix 2. Plasmid pDER412 (Fig. 2.1) was used to complement *rep*$^-$ mutants of the pTF-FC2 replicon. All other plasmids used during the course of these experiments are shown in Fig. 5.3.

5.3.2 General recombinant DNA methods and DNA sequencing

All cloning methods were essentially as described in Maniatis *et al.* (1989). Ordered deletions were generated from pTV100 and pTV101 using the *Exo*III method of Henikoff (1984). Deletions were generated from the *Eco*RI site of pTV100 by protecting with *Pst*I and digesting with *Exo*III from the 5' *Eco*RI overhang. Deletions were generated from the *Cla*I end of pTV101 by protecting with *Kpn*I and *Exo*III

digestion from a 5'overhang resulting from digestion with *Cla*I. These templates were subsequently used to generate overlapping sequence in both directions, of the region between the *Cla*I site and the start of pTV400. The sequencing techniques used were as described in Chapter 3.3.2, and analysis of proteins by PAGE was as described in Chapter 3.3.3.

5.4 RESULTS

5.4.1 Identification of an autonomously replicating pTF-FC2 fragment

The first step towards the construction of a broad-host-range vector was to link the pTF-FC2 basic replicon with an antibiotic resistance marker and then to delete the vector sequences. Plasmid pTV400, containing the 3202 bp fragment (which retained the ability to replicate in *P. aeruginosa* and in the *polA*$^-$ *E. coli* strain, GW125a) was chosen for vector construction due to its convenient restriction sites. Two antibiotic resistance markers were selected, namely the Cm^R marker from pBR325 (Balbas *et al.*, 1986), and the kanamycin resistance (Km^R) marker originating from Tn*5* (De Bruin and Lupski, 1984).

A 950 bp *Asu*II fragment carrying the Cm^R marker from pBR325 was inserted into the compatible *Cla*I site of Bluescript SK (Fig. 5.1). The resulting plasmid, pSKCm1 was digested with *Kpn*I and *Hind*III and the 950 bp fragment was gel-purified and ligated with a gel-purified 3.2 kb *Kpn*I - *Hind*III fragment containing the pTF-FC2 basic replicon from pTV400 (Fig. 5.2). No Cm^R transformants were obtained, despite numerous repetitions of the experiment.

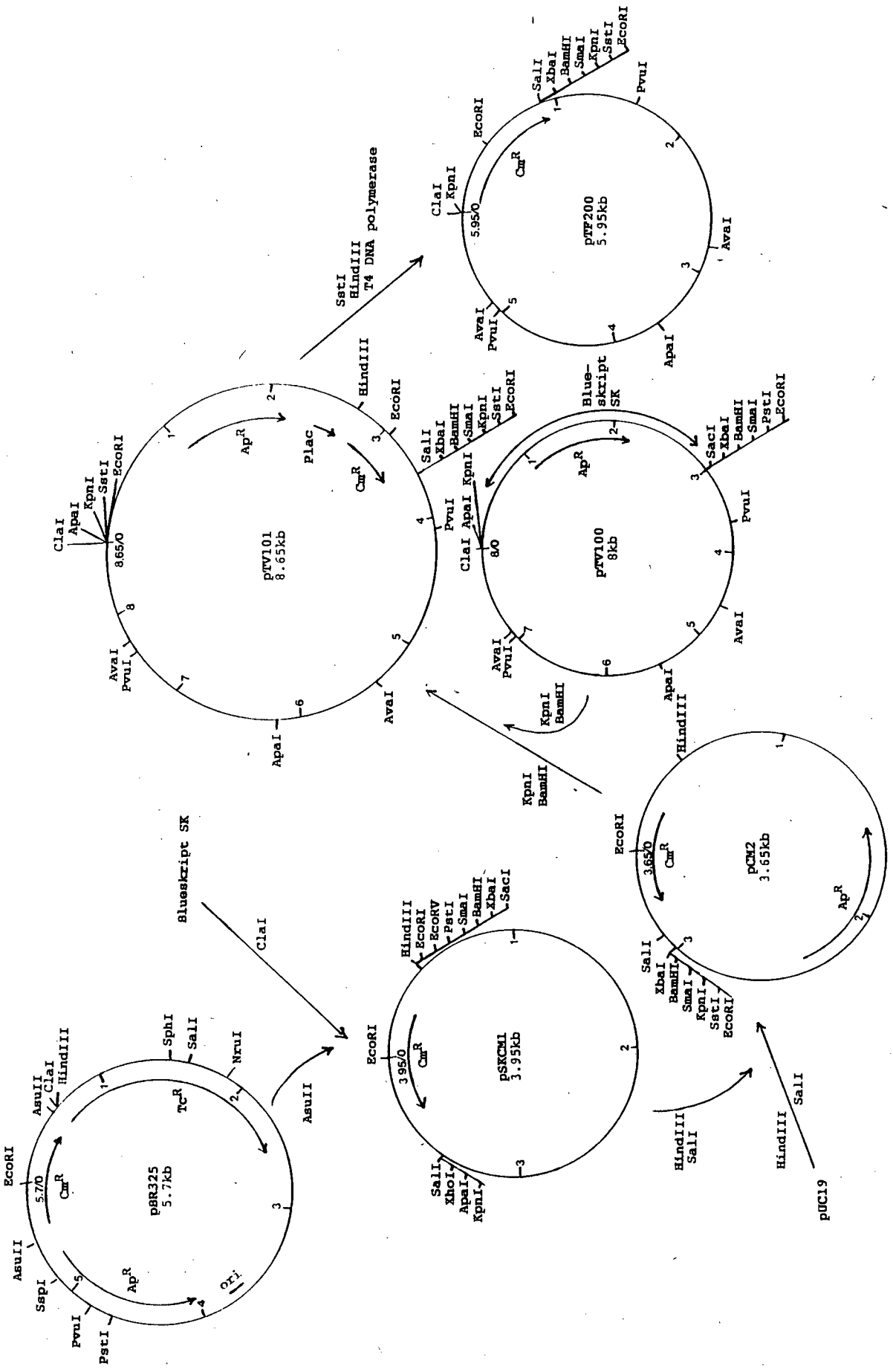

Fig. 5.1 Construction of various Cm^R plasmids;.

into *E. coli* LK111. Km^R transformants were picked onto agar plates containing Ap, where the inability to grow would indicate loss of the pUC19 replicon. Despite several repetitions of the experiment, no Ap^S colonies were obtained.

To determine whether the requirement for pUC19 could be complemented by the providing pDER412 (Chapter 2.4.1), the ligations from the experiment described above were transformed into *E. coli* LK112 containing pDER412. Km^R transformants were transferred to LA containing Ap and a number of Ap^S colonies were identified. These colonies were all resistant to Cm. Due to the *recA*⁻ nature of LK112, it was unlikely that the Km^R Cm^R phenotype was due to rescue of the Km^R marker by homologous recombination with pDER412. Restriction analysis of plasmids isolated from Km^R Cm^R Ap^S colonies confirmed the presence (in addition to pDER412) of a satellite plasmid corresponding to pTF100 from which the pUC19 replicon had been deleted. These results indicated that pDER412 was supplying a diffusible factor which enabled the 3202 bp fragment to replicate independently of the pUC19 replicon. The 3202 bp fragment of pTV400 did not therefore contain the whole of the minimal replicon of pTF-FC2 and pUC19 was thus complementing this deficiency.

Experiments were carried out to investigate whether the larger, 5.0 kb *Cla*I - *Eco*RI fragment (pTV100), which extends beyond the borders of pTV400, was able to replicate autonomously in the absence of a ColE1-type origin.

First, a Cm^R marker was inserted between the pTF-FC2 fragment and the Bluescript SK sequences (Fig 5.1). The 950 bp *Sal*I - *Hind*III fragment which carries the Cm^R marker was subcloned from pSKCm1 into pUC19 to generate convenient restriction sites. This construct, pCM2, was digested with *Kpn*I and *BamH*I and the DNA ligated with pTV100 DNA which had likewise been digested with *Kpn*I and *BamH*I. Transformants were plated on LA containing Ap and Cm and the resulting recombinant, pTV101, was able to confer Ap^R and Cm^R upon *E. coli* GW125a.

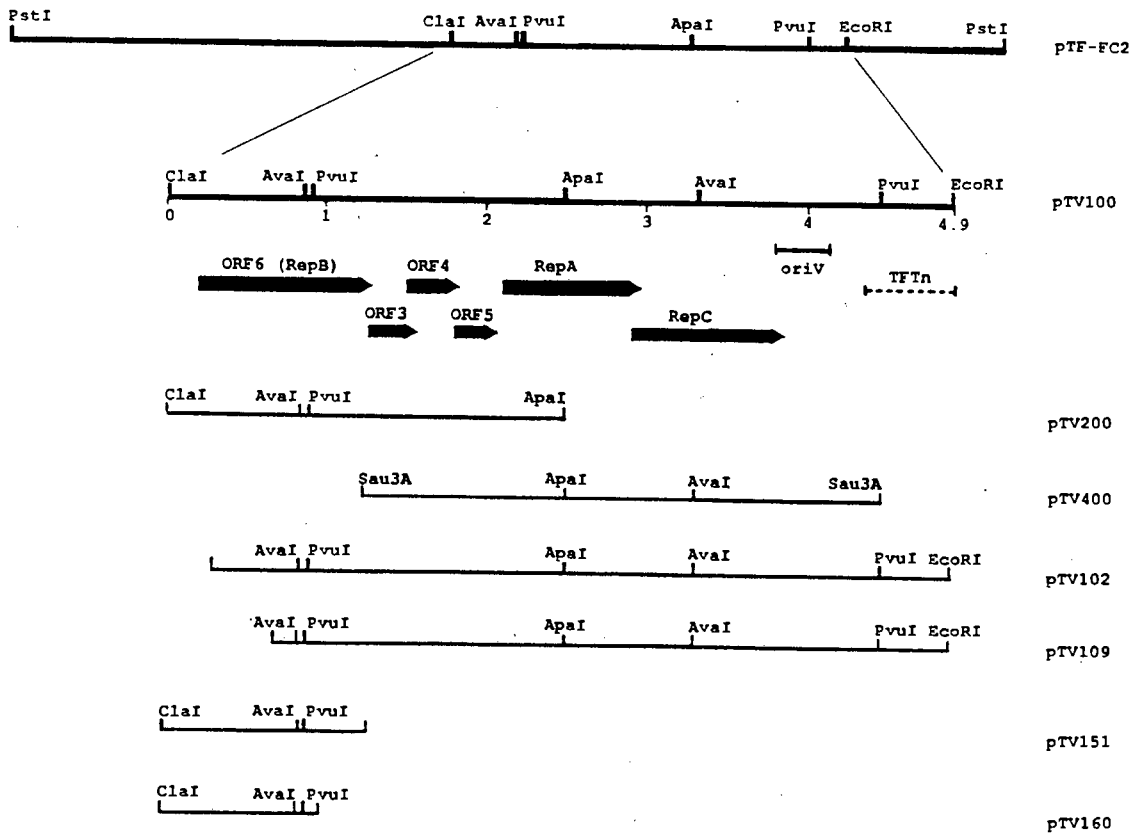

Fig. 5.3 Restriction maps of plasmid used in Chapter 5. ORFs are indicated as solid lines below the map of pTV100.

Plasmid pTV101 was digested with *Sst*I and *Hind*III (to release the pUC19 replicon but retain the linked Cm^R and pTF-FC2 fragment), the overhanging ends were flushed using T4 DNA polymerase and the ligations transformed into *E. coli* LK111. Cm^R transformants were tested for Ap^S and a number of Cm^R Ap^S colonies were obtained. Restriction analysis of plasmid DNA from these colonies confirmed that the recombinant, pTF200, had lost the 2.7 kb pUC19 fragment.

A series of ordered deletions of pTV101 was generated from the *Cla*I site using the *Exo*III shortening method. Two of these deletions, pTV102 and pTV109 (Fig. 5.3) were digested with *Sst*I and *Hind*III, the ends flushed with T4 DNA polymerase and religated. No Ap^S Cm^R transformants were obtained. Deletion of 300 bp from the *Cla*I site (pTV102), therefore appeared to have resulted in the loss of the ability to replicate autonomously.

To discover the nature of the complementation provided by pUC19, the Ap^R recombinants obtained from attempts to delete pUC19 from pTV102 and pTV109 were characterised by restriction analysis. In all cases, the pUC19 fragment had been religated to the pTF-FC2 fragments. However, the disappearance of the *Sst*I and *Hind*III sites indicated that the religation was not due to incomplete flushing of the overhanging ends by T4 DNA polymerase. In addition, the pUC19 fragment was present in both orientations relative to the pTF-FC2 fragment.

5.4.2 Analysis of the nucleotide sequence of the region required in addition to the 3202 bp fragment of pTF-FC2 for replication

The first 1400 bp extending from the *Cla*I site towards and including the start of pTV400 are shown in Fig. 5.4. As was the case with the 3202 bp fragment, there were very few 6 bp recognition restriction endonuclease sites in the sequence, namely an additional *Ava*I site (nt 850) and a *Pvu*I site (nt 905). The start of pTV400 was found to be 1239 bp downstream from the *Cla*I site.

Analysis of the sequence data revealed an ORF (ORF6) of 1056 nucleotides starting with an ATG start codon at nt 217 (preceded by a consensus ribosome binding site) and ending with a TAG stop codon at nt 1273 (Fig. 5.4). The potential protein encoded by ORF6 consists of 352 amino acid residues and has a predicted M_R of 40.155 kD. A codonpreference plot of ORF6 using the codonpreference subroutine of the UWGCG package indicated that the codon usage of the potential ORF6 protein was similar to that of the other proteins identified on the 3202 bp fragment (Fig. 5.5).

```
                pTV100 ─▷
                ClaI
1       ATCGATATTCACGACTGGCGCAACCGGGACAGCACGCTTGCCGCCTTGCAGCTATCGGCG

51      CAGAAGTGGGGCAGCTTCACCGTAACCGGGAACGACGAATACAAGGCCATGTGCGGGAAG

121     TTAGCAGCAGAGCACGGCTTCAAGATCACAAATGCAGAGCTTCAGGAGAGCATCCAGCAG

181     GAGCGGCAGCGGATACAGCAGGAGAGGGCGCAGGCGATGAAATCGGAGCAGCTAAAGCAG
                                        RepB ─▶  M  K  S  E  Q  L  K  Q

241     TTCGAGCGGTACGCGGAAGCGGTAGGCGCGGAGCGCTACCGGGTAACGTCCATCAAGATG
         F  E  R  Y  A  E  A  V  G  A  E  R  Y  R  V  T  S  I  K  M
            pTV102 ─▷
301     CAGGCAGACGGAAGGAAGCAAACCTTCATCCTCGACAAGAAGGACGGCATCACGCGGGGT
         Q  A  D  G  R  K  Q  T  F  I  L  D  K  K  D  G  I  T  R  G

361     TTCACACCGCAGGAGATCGAGCAGCGCACGCCGGAGATGCAGCGCCTACAGCGCCGGGGC
         F  T  P  Q  E  I  E  Q  R  T  P  E  M  Q  R  L  Q  R  R  G

421     GAAAACCTCTACTACACGCCGCTATCGGACAAGAAGCATCACATCCTCATCGACGACATG
         E  N  L  Y  Y  T  P  L  S  D  K  K  H  H  I  L  I  D  D  M

481     AACCGGGAGAAGCTGGAGCGGCTTATCAAAGACGGCTACCAGCCCGCCGCCGTGCTGGAA
         N  R  E  K  L  E  R  L  I  K  D  G  Y  Q  P  A  A  V  L  E

541     TCCAGCCCCGGCAACTATCAGGCCATCATCACCGTGTCGAAGCTGGGGACGGCCCACGAT
         S  S  P  G  N  Y  Q  A  I  I  T  V  S  K  L  G  T  A  H  D

601     AAGGACGTGGGCAACCGCCTGAGCGATGCCCTGAACCGTGAATACGGCGACCCGAAGCTA
         K  D  V  G  N  R  L  S  D  A  L  N  R  E  Y  G  D  P  K  L

661     TCGGGAGCCATCCACCCGCACCGCGCACCCGGCTACGAGAACCGCAAGCCCAAGCACCAG
         S  G  A  I  H  P  H  R  A  P  G  Y  E  N  R  K  P  K  H  Q

721     CGGGAGGACGGCAGCTATCCAGAGGTGCGCTTGCTCAAGGCCGAGCGCCGGGAGTGCGTC
         R  E  D  G  S  Y  P  E  V  R  L  L  K  A  E  R  R  E  C  V

781     AAGGCGCTGGCCTTGTCCAGCCAGATCGACGCCGAGTATCAGCGGCAAGCGGCCTTGAAG
         K  A  L  A  L  S  S  Q  I  D  A  E  Y  Q  R  Q  A  A  L  K
                  AvaI
841     GCGCAGCAGCCCGAGCGCACGAAAGCCAAGCCCGCCTTGGAGCTTGCAGCGGCCAGCGGC
         A  Q  Q  P  E  R  T  K  A  K  P  A  L  E  L  A  A  A  S  G
               PvuI
901     AGCGCGATCGACGCCTACCAGCGGCATTACCGCGACGTGCTCAAGCGGCAGCGTGGCGGC
         S  A  I  D  A  Y  Q  R  H  Y  R  D  V  L  K  R  Q  R  G  G

961     GAGGTGGACTTGTCCCGCGTGGATTCCATGATTGCCGTGCGTATGCGCGTCACCGGCCAC
            ◀─ pTV160
         E  V  D  L  S  R  V  D  S  M  I  A  V  R  M  R  V  T  G  H

1021    GATCAAGCGGCCATCGAGGGCGCTATCCGCCAGTGCGCACCGGCCACCCGGCAGAAAGAC
         D  Q  A  A  I  E  G  A  I  R  Q  C  A  P  A  T  R  Q  K  D

1081    GAGGGCCGCGATTGGAACGACTACGCGCAGCGCACCGCCCGCTATGCCTACAGCGCCGCA
         E  G  R  D  W  N  D  Y  A  Q  R  T  A  R  Y  A  Y  S  A  A

1141    CAGCCGCAAGCCGCCGATCTTGGCAAGTACCGGCAGCAGTGGGAGAAGCTGGAAGGGCGC
         Q  P  Q  A  A  D  L  G  K  Y  R  Q  Q  W  E  K  L  E  G  R
                                          pTV400 ─▷
1201    GAGCCTGTACGACAGCAGGAGCAGGCAAAGGCGCAGAAGATCGAGCGCGACAACTCGCCG
         E  P  V  R  Q  Q  E  Q  A  K  A  Q  K  I  E  R  D  N  S  P
                        -35                 -10
1261    GGAATGAGTCTCTAGCGTTGCGTGGTGGTTGTGATATACTTGTATAGCGTTTTCAGAACA
         G  M  S  L                      ◀─ pTV151

1321    GGAGCCGAAACATGCTTGCAATCCGACTGCCCGCCGAAGTGGAAACCCGCCTTGAAGCAC
           ORF3─▶  M  L  A  I  R  L  P  A  E  V  E  T  R  L  E  A

1381    TGGCGCAGGCCACAGGGCGG
         L  A  Q  A  T  G  R
```

Fig. 5.4 Nucleotide sequence of the first 1400 bp extending from the *Cla*I site of pTV100. The start codons and ribosome binding sites of RepB and ORF3 are indicated in bold type while amino acid sequences are shown below the nucleotide sequence. The extent of deletion plasmids described in Fig. 5.3 is indicated above the sequence.

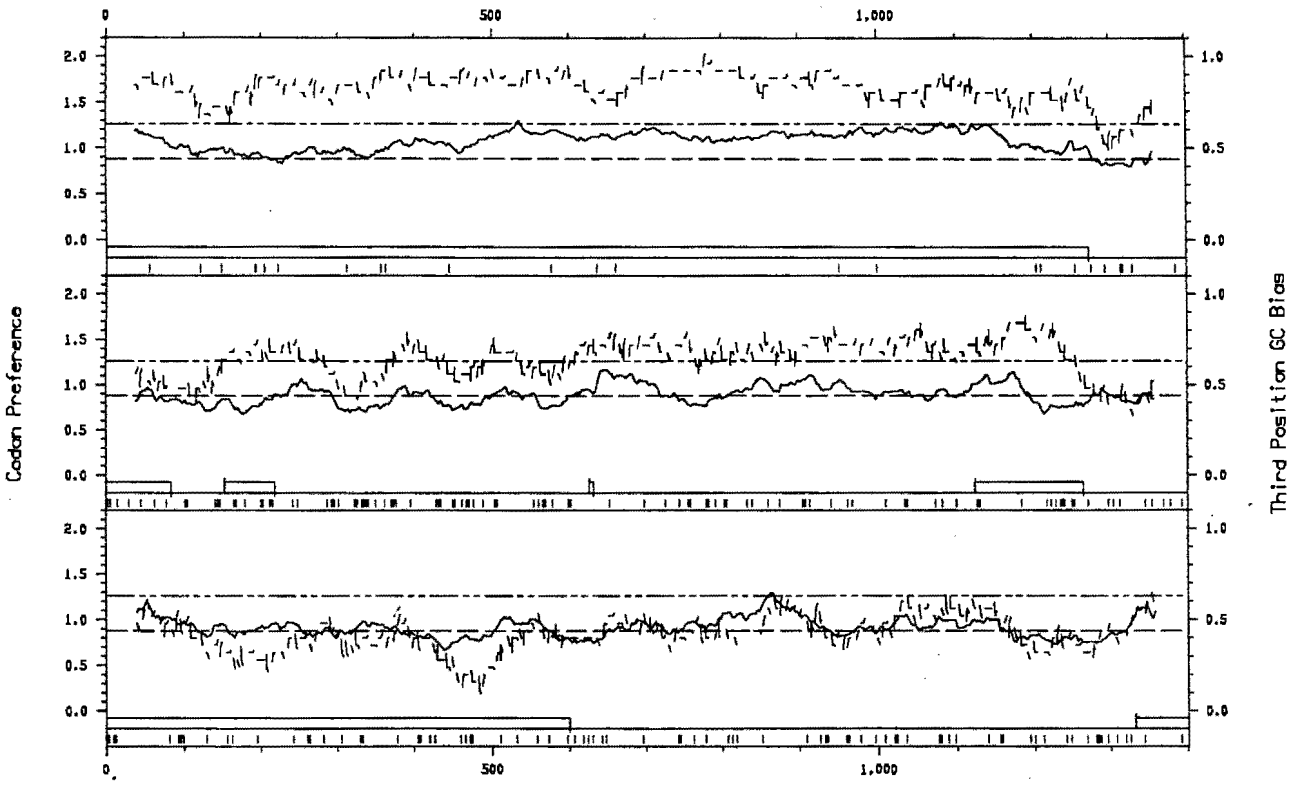

Fig. 5.5 Codon preference and bias plot of the nucleotide sequence of the first 1400 bp extending from the *Cla*I site of pTV100. The codon preference plot was generated using the combined codon usage table for the 5 ORFs encoded by 3.2 kb fragment. The extent of the ORF encoding RepB is indicated by an open box and rare codons are shown as bars beneath.

5.4.3 Translation products of the 1400 bp adjacent to the *Cla*I site

To determine whether ORF6 produced a translational product, plasmids pTV200, pTV151 and pTV160 (Figs. 5.3 and 5.4) were subjected to *in vitro* transcription and translation and the products analysed by PAGE. Plasmid pTV200, which extended from the *Cla*I site to the *Apa*I site within pTV400, contained the whole of ORF 6 and produced a translation product of 40 kD which was not present in the Bluescript SK control (Fig. 5.6 lanes 2 and 1 respectively). Plasmid pTV151, which extended from the *Cla*I site to nt 1296 and also contained the intact ORF6, produced the same 40 kD protein (Fig. 5.6 lane 3). Plasmid pTV160 (nts 1 - 980) which had 98 amino acid residues deleted from the carboxyterminus of ORF6 should have produced a truncated protein with a M_R of 29.025 kD. Translation of this plasmid did not

produce the 40 kD protein, but a new polypeptide which was resolved at 35 kD (Fig. 5.6 lane 4). Ann examination of the nucleotide sequence showed that a fusion between the truncated ORF6 and Bluescript SK could could result in a polypeptide of predicted M_R 35.14kD, which correlates well with the polypeptide produced by plasmid pTV161.

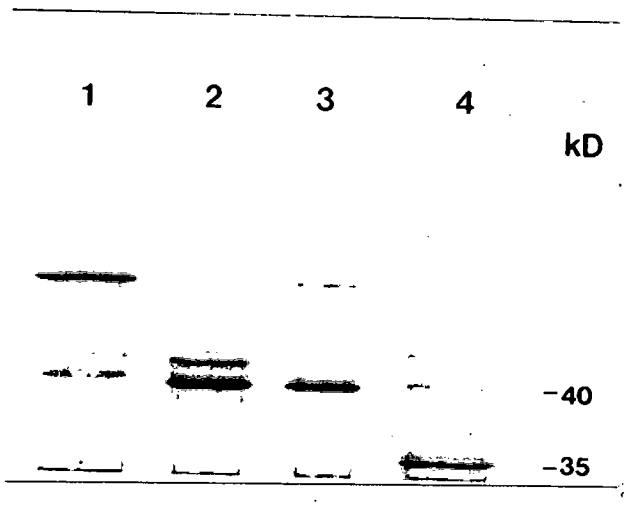

Fig. 5.6 SDS-PAGE analysis of polypeptides produced from the 1400 bp fragment extending from the *ClaI* site of pTV100. Lanes 1 - 4 represent polypeptides derived from pUC19, pTV200, pTV151 and pTV160 respectively.

5.4.4 Expression of ORF6

Due to the fact that the *lacZ* promoter of Bluescript SK was orientated in the opposite direction to ORF6, it is probable that the protein was expressed from a promoter located upstream of the start of the ORF at nt 217. Analysis of this region did not reveal an *E. coli* consensus promoter sequence and confirmation of the presence of such a promoter will have to be obtained from primer extension experiments.

5.4.5 Function of the ORF6 protein

The amino acid sequence of the ORF6 protein was compared with sequences in the GenEMBL data bank using the Tfasta subroutine of the UWGCG package. Only one

sequence, that of the RepB' protein of the IncQ plasmid RSF1010, showed limited, but significant homology (23% identical amino acid residues) (Fig. 5.7). There was a particularly well conserved region with 11/19 identical amino acids (residues 99 - 118 of ORF6 and 91 - 109 of RSF1010. On the strength of this evidence, the ORF6 was named *repB* and its 40 kD protein RepB.

```
pTF-FC2     1   MKSEQLKQFERYAEAVGAERYRVTSIKMQADGRKQTFILDKKDGITRGFTPQEIEQRTPE
                :. .:   :.:          .     : :           :  .  :. : ::
RSF1010     1       MKN-DRTLQAIGRQLKAMGCERFDIGVRDAT----TGQMMNREWSAAEVLQNTPW

           61   MQRLQRRGENLYYTPLSDKKH-HILIDDMNREKLERLIKDGYQPAAVLESSPGNYQAIIT
                : .  :  .:  :    .:  .:.::     :.    .:  :: :.:.:: :::: .
           51   LKRMNAQGNDVYIRPAEQERHGLVLVDDLSEFDLDDMKAEGREPALVVETSPKNYQAWV-

          120   VSKLGTAHDKDVGNRLSDALNREYG-DPKLSGAIHPHRAPGYENRKPKHQREDGSYPEVR
                :.. :   ..    .   :  ::  ::      :  :  :. ::: ::    :  : :
          110   --KVADAAGGELRGQIARTLASEYDADPASADSRHYGRLAGFTNRKDKHTTRAGYQPWVL

          179   LLKAERRECVKALALSSQIDAEYQRQAALKAQQPERTKAK-PALELAAASGSAIDAYQRH
                :     .      . ::  :                   :   :     .:.: :
          168   LRESKGKTATAGPALVQQAGQQIEQAQRQQEKARRLASLELPERQLSRHRRTALDEYRSE

          238   YRDVLKRQRGGEVDLSRVDSMIAVRMRVTGHDQAAIEGAIRQCAPATRQKDEGRDWNDYA
                  ..::  :   :::. :   : .   :     :  :    ::   .  : .  ::
          228   MAGLVKR-FGD--DLSKCDFIAAQKLASRGRSAEEIGKAMAEASPALAERKPGHE-ADYI

          299   QRTARYAYSAAQPQAADLGKYRQQWEKLEGREPVRQQEQAKAQKIERDNSPGMSL
                ::          : :     :   . : .
          284   ERTVSKVMGLPSVQLARAELARAPAPRQRGMDRGGPDFSM
```

Fig. 5.7 Amino acid alignment of the pTF-FC2 RepB and RSF1010 RepB' proteins using the Genepro software package. Identical and conservative amino acid changes are indicated by colons and single dots respectively between the amino acid sequences.

5.5 DISCUSSION

The experiments presented in this chapter showed that the discrimination between replication from the pTF-FC2 and the ColE1-type vector origins based on Pol I-independence was misleading. Deletion of the pUC19 replicon from pTV400 and pTF100 resulted in loss of the ability of the 3202 bp pTF-FC2 fragment to replicate in

the absence of pDER412. An additional 1239 bp upstream of the start of pTV400 was required to restore activity of the pTF-FC2 origin. This region was found to encode a 40 kD protein, RepB, which had some homology with the RepB' protein of the IncQ plasmid RSF1010.

Previously, Rawlings *et al.* (1984) had attempted to delete the *Pvu*I fragment which included the *Cla*I site (Fig. 2.1) from pDER412. They were unable to delete this fragment and found that it was always present in the original orientation. This led Rawlings *et al.* (1984) to propose that sequences around the *Pvu*I site were essential for replication of pTF-FC2. These results were in conflict with those obtained from experiments with the *Sau*3A partial 3202 bp fragment which did not extend to the *Pvu*I site. The *repB* gene is located in the region which includes the *Pvu*I site. The inablity to delete this region was therefore due to the fact that pDER412, which has the pBR325 origin of replication deleted, would have been unable to replicate in the absence of RepB.

In RSF1010, the *repB'* lies within a larger ORF, *repB*, the product of which is essential for mobilisation and can also substitute for RepB' during replication. RepB' is translated from a point within the RepB coding sequence (Haring and Scherzinger, 1989). The absence of any stop codons upstream of the start of the pTF-FC2 RepB ORF and codon bias data (Fig. 5.4) suggests that RepB may also be part of a larger protein. Preliminary results from DNA sequencing upstream of the *Cla*I site supports this possibility (J. Rohrer, personal communication)). If this were indeed the case, then it is very interesting that in pTV100, the RepB protein appears to be expressed from a promoter located within approximately 200 bp of its start. Evidence for this is the production of the RepB protein *in vitro* (Fig. 5.6 lanes 2 and 3) without the influence of external promoters. The expression and regulation of this promoter remains to be investigated.

RepB' is essential for replication of RSF1010 *in vitro* (Scherzinger *et al.*, 1984). The 35 kD protein has been found to have primase activity and is responsible for the *de novo* synthesis of primers at the *ssi* sites in the *ori*V (Haring and Scherzinger, 1989; Honda *et al.*, 1989). The pTF-FC2 RepB also appears to be essential for replication, although

this needs to be confirmed by *in vitro* experiments. In view of its similarity with the IncQ RepB protein, it seems likely that RepB also fulfils the role of a primase activity during the replication of pTF-FC2. A consensus n' binding site was previously identified on a set of complementary inverted repeats downstream of the direct repeats within the *ori*V of pTF-FC2 (Chapter 4.4.2). In RSF1010 a similar sequence is contained within a plindromeic region which serves as a recognition site for the RepB primase (Honda *et al.*, 1988; 1989).

The role of pUC19 in the replication of the 3202 bp pTF-FC2 fragment in the absence of RepB is intriguing. Although pUC19 was unable to replicate in *E. coli* GW125a (Chapter 2.4.2a) and in *P. aeruginosa* (Chapter 2.4.3), it is nevertheless able to complement a pTF-FC2 origin from which the RepB has been deleted. The pUC19 replicon is required in *cis*, although its orientation and distance relative to the pTF-FC2 *ori*V did not seem to affect replication (pTF100 has a 1.5 kb fragment inserted between the pTF-FC2 *ori*V and the pUC19 replicon). The pUC19 replicon in pTF100 has been replaced with pACYC184, a p15A-based vector (Chang and Cohen, 1978) without affecting replication of the construct (S. Bardiem, personal communication). This implies that the requirement of the 3202 bp fragment is not necessarily for a specific sequence of pUC19 only.

Similar observations have been made with the IncQ plasmid R1162. The orientation of the RepB priming site relative to the 20 bp direct repeats does not affect replication and these domains may be separated by up to 2 kb before replication is affected (Meyer *et al.*, 1987). In addition, the presence of Bluescript SK appears to enable a deletion which has lost the *ssi* sites to replicate (personal communication from Richard Meyer via Doug Rawlings)).

The host range of the ColE1-type plasmids is confined to the enteric bacteria (Selzer *et al.*, 1983) and the replicons are not maintained in *E. coli polA1* mutants (Grindley and Kelley, 1978). Dasgupta *et al.* (1987) have shown that little if any DNA synthesis occurs from the ColE1 origin in *E. coli* cells which lack Pol I (*polA1*), but contain RNase H. These workers argue that this is because RNase H cleaves the RNAII precursor, which results in a RNA-DNA hybrid which is too short to provide

sufficient strand displacement for initiation of replication via the Type II mechanism (see Chapter 1.4.2b). Plasmid pACYC184 has been shown to possess an *ssi* site within its origin (Bahk *et al.*, 1988), while plasmids ColE1 and pBR322 carry two *ssi* sites (Nomura *et al.*, 1982; Abarzua *et al.*, 1984). Primosome assembly has been shown to occur at the *ssi* sites of pBR322 (Abarzua *et al.*, 1984).

In view of the structural and genetic similarities between the origins of pTF-FC2 and the IncQ plasmids, it is probable that replication of the wild-type pTF-FC2 replicon occurs in much the same way as is proposed for the IncQ plasmids (Haring and Scherzinger, 1989). Accordingly, RepC would recognise the 22 bp repeats of the *ori*V and bind to the duplex DNA, whereafter RepA would begin the unwinding of duplex DNA in the direction of the priming site. RepB is responsible for synthesis of a primer at this site which is then used by the DNA polymerase holoenzyme as a template for strand elongation.

Based on these observations, a speculative model may be proposed for the events leading to initiation of replication of pTV400 in an *E. coli polA*$^-$ mutant. In the absence of RepB, the RepC-RepA complex may continue to unwind the DNA past the pTF-FC2 priming site and into the pUC19 replicon until the origin is reached. The strand displacement produced by the pTF-FC2 proteins would allow the *E. coli* primosome assembly, resulting in initiation of DNA synthesis from this point. This could be investigated by testing for the ability of wild-type and *repB*$^-$ pTF-FC2 replicons to replicate in *E. coli dnaG* mutants.

A number of observations provide support for the hypothesis. Firstly, pTV400 replicates at 12 - 15 plasmids per *E. coli* chromosome (Chapter 2.4.2e), while pUC19 replicates at a copy number of about 70 plasmids per chromosome. Clearly, the replication rate is controlled by the pTF-FC2 origin and this would be expected to be the case if priming at the pUC19 origin was dependent upon the action of the RepC and RepA proteins of the 3202 bp fragment. Additional evidence is that mutations in or deletion of either *rep*A and *rep*C result in the inability of such plasmids to replicate in the *polA*$^-$ mutant (Chapter 4.4.5).

With the absence of pTF-FC2-specific primase (RepB), the priming site of the pTF-FC2 *ori*V would not be functional. The ability of pACYC184 to substitute for pUC19 in pTV400 suggests that the requirement of the 3202 bp pTF-FC2 fragment may be for an *ssi* site recognisable to the *E. coli* DnaG primase.

If the above hypothesis is correct, it remains difficult to explain why pTV4210 replicates at a reduced copy number. If the palindromic sequence is the site at which priming occurs, then deletion of part of this region should not affect replication in the absence of RepB. A drop in the copy number of pTV4210 may be due to altered expression of RepA and RepC. It is possible that the palindromic sequence may be involved in the transcriptional regulation of RepA and RepC expression, for instance by acting as a *rho*-dependent terminator of transcription. Deletion of part of the palindromic sequence could therefore affect the replication rate by reducing the amount of RepA and RepC. Alternatively, binding of RepC and/or RepA may be dependent upon sequences in this region. Deletions such as that in pTV4210 may therefore decrease the efficiency by which the RepC/RepA complex initiates replication.

CHAPTER 6

GENERAL CONCLUSIONS

Prior to the start of this investigation, very little was known about pTF-FC2 except that it was able to replicate in a number of Gram-negative bacteria and that it could be mobilised between these hosts by IncP plasmids. No phenotypes other than *mob* and *rep* functions could be assigned to the plasmid. The aim of this investigation was therefore to characterise the basic replicon of the plasmid in order to further the understanding of the mechanisms by which different plasmids relicate in a wide range of bacteria and also with the view to its eventual use as a cloning vector for *T. ferrooxidans*.

The basic replicon of pTF-FC2 was found to be contained within a 4.14 kb DNA fragment. The entire nucleotide sequence of this fragment was determined (Appendix 5). Of this, 329 bp was required in *cis* for replication while the remainder encodes at least 5 proteins of which three, namely RepA and RepB and RepC have been shown to be essential for replication *in vivo* in *E. coli*. Furthermore the presence of at least one of the two remaining proteins and possibly the product of a third ORF is associated with the ability of derivatives of pTF-FC2 to replicate in *P. aeruginosa*.

A comparison of the pTF-FC2 replicon with those of other well studied broad-host-range plasmids belonging to the IncP and IncQ incompatibility groups showed considerable similarity between pTF-FC2 and RSF1010, R1162 and R300B, all members of the IncQ group, but little similarity with the IncP plasmids.

The IncQ plasmids are similar to pTF-FC2 in that they are small in size (8.7 kb and 12.4 kb respectively). The IncQ plasmids appear to be easily mobilised and

maintained in the same wide range of Gram-negative bacteria. They replicate (at least in *E. coli*) at the same copy number and the basic replicon is contained within a relatively small region. In contrast, the IncP plasmids are much larger in size (60 kb), are self transmissible and regions involved in replication are dispersed over a large area on the plasmids. The IncQ and IncP plasmids carry a number of antibiotic resistance markers which were probably acquired as a result of insertion of various transposable elements. In contrast, pTF-FC2 confers no antibiotic resistance. There is evidence that the replicon acquired a Tn3-like transposable element at some stage during its evolution. Preliminary sequence data from the region adjacent to the *mob* region has indicated that an ORF whose predicted amino acid sequence shows homology to a protein, MerR, which is involved in the regulation of the mercury resistance operon of TN*501* (J Rohrer, personal communication). The plasmid may thus be able to confer resistance to mercury, but this remains to be confirmed.

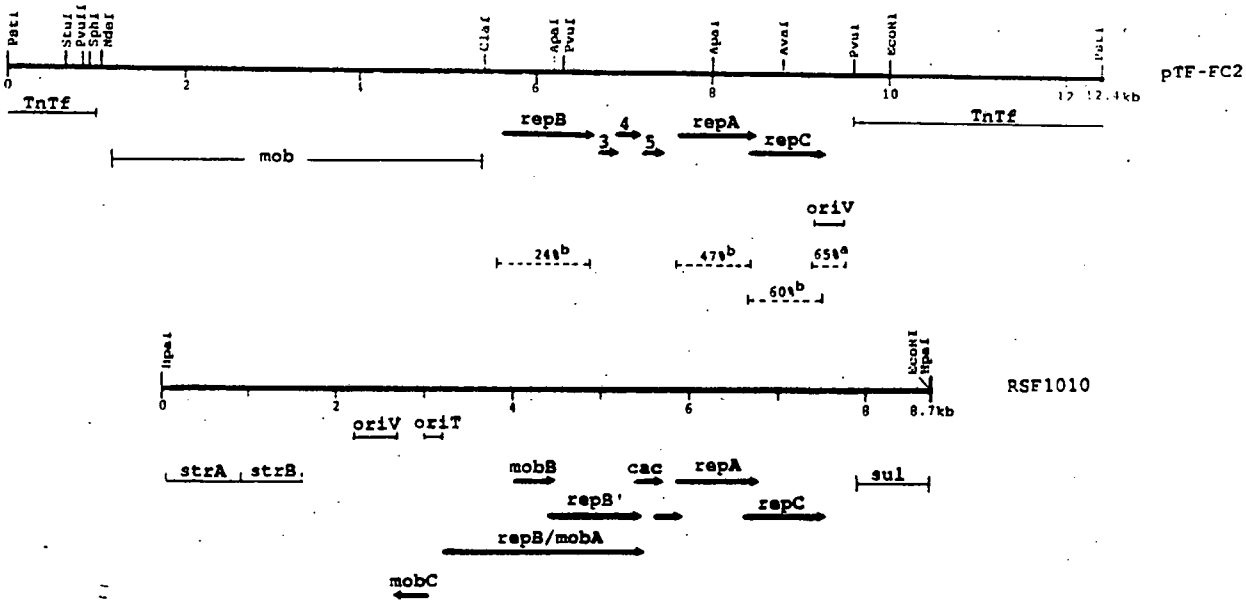

Fig. 6.1 An aligment between the replicons of pTF-FC2 and RSF1010. Thick arrows denote ORFs with thin lines showing the *mob* region of pTF-FC2, the location of the Tn*21*-like element Tn*Tf* on pTF-FC2 and the position of the Su^R and Sm^R genes on RSF1010. Dotted lines show regions of pTF-FC2 which show homology to parts of RSF1010.

An alignment of the replicons of pTF-FC2 and the IncQ plasmids shows that they are structurally similar (Fig. 6.1) A major difference, though, is that the IncQ *rep* proteins are separated from the *ori*V by the Su and Sm markers while in pTF-FC2,

the *rep* and *ori*V regions are contiguous. DNA and protein sequence comparisons have shown extensive homology between the *ori*V regions of plasmids. There is good homology between their RepC proteins, with a decrease in conservation between the RepA proteins and only limited similarity between the RepB/RepB' proteins. Despite these similarities, in particular the 65 % DNA homology between the *ori*V regions, no incompatibility could be detected between pTF-FC2 and R300B and the latter was unable to complement mutations in pTF-FC2 genes. In view of the extensive homology between the RepA and RepC proteins of these plasmids, it would be interesting to determine whether their obvious plasmid-specificity extends throughout the proteins using constructs containing chimeric proteins to complement mutations in the respective replicons. This may prove to be useful in assigning functional domains to each protein.

These results prompt speculation as to the origin of these different groups of broad-host-range plasmids. There is clear relationship between pTF-FC2 and members of the IncQ group, while there is sharp contrast with the very different IncP plasmids. This suggests that there might be at least two major groups of broad-host-range plasmids represented by pTF-FC2 and the IncQ plasmids on the one hand and by the IncP plasmids on the other.

Besides the initiator proteins (RepC), the pTF-FC2/IncQ type of replicons encode a helicase and primase which are directly involved in plasmid-specific DNA replication, while the IncP-type plasmids encode only an initiator protein (TrfA). In both types of plasmids, the ability to replicate in a wide number of bacteria appears to be as a result of modulation of the expression of the replication proteins. In the IncQ, the loss of a regulatory protein which controls *rep* expression results in changes in the host range, while the IncP plasmids have a complicated multicomponent regulatory system which regulates *trfA* expression and concomitantly determine the host range. The ability of pTF-FC2 to replicate in different bacteria appears to be associated with the presence of several small proteins. These proteins are located in a similar position to the regulatory elements of the IncQ plasmid RSF1010, but precisely what their function is remains to be determined.

It is not unusual for plasmids to carry several origins of replication, as is the case for plasmid R6K (Croza, 1980; Inuzuka *et. al.*, 1980) and pLS1 (del Solar *et. al.*, 1987). In addition, many vectors such as pSUP106 carry multiple, compaitible origins (Priefer *et. al.*, 1985). However, these origins all replicate independently of one another. Plasmid pTV400 carries two origins. Replication from the pTF-FC2 origin is absolutely dependent upon elements located on the pMB1 replicon, even in bacterial strains such as *P. aeruginosa* in which the pMB1 replicon is unable to replicate. The possiblity therefore exists that plasmids are able not only to acquire more than one functional origin, but may utilise specific elements from an origin to complement or supplement its own replicon.

The presence of the promiscuous pTF-FC2 in *T. ferrooxidans* is quite remarkable considering the fact that it carries the the inverted repeats and at least part of the transposase of a family of transposable elements which are of medical importance. *T. ferrooxidans* is an organism which grows at a low pH in an inorganic enviroment. The organism therefore occupies an extreme ecological niche which is very far removed from those occupied by bacteria which commonly contain similar promiscuous plasmids. The presence of pTF-FC2 in this organism suggests that horizontal spread of genetic material occurs across a very diverse spectrum of genera and between relatively remote ecosystems.

LITERATURE CITED

Abarzua P., Soeller W., and Marians K.J. 1984. Mutational analysis of primosome assembly: I. Distinct classes of mutants in the pBR322 *Escherichia coli* factor Y DNA effector sequences. J. Biol. Chem. 259:14286-14292.

Armstrong K., Acosta A.R., Ledner E., Machida Y., Pancetto M., McCormick M., Ohtsubo H., and Ohtsubo E. 1984. A 37 x 10^3 molecular weight plasmid-encoded protein is required for replication and copy number control in the plasmid pSC101 and its temperature-sensitive derivative pHS1. J. Mol. Biol. 175:331-347.

Austin S.J. 1988. Plasmid partition. Plasmid 20:1-9.

Bagdasarian M., Bagdasarian M.M., Lurz R., Nordheim A., Frey A., and Timmis K.N. 1982. Molecular and functional analysis of the broad host range plasmid RSF1010 and construction of vectors for gene cloning in Gram-negative bacteria. In S.Mitsubishi. (ed.). Bacterial drug resistance. Japan Sci. Society Press Tokyo. pp183-197.

Bagdasarian M., Lurz R., Ruckert B., Franklin F.C.H., Bagdasarian M., Frey J., and Timms K.N. 1981. Specific purpose cloning vectors II. broad-host-range, high copy number, RSF1010-derivedvectors and a host-vector system for gene cloning in *Pseudomonas*. Gene 16:237-247.

Bagdasarian M.M., Scholz P., Frey J., and Bagdasarian M. 1987. Regulation of the *rep* operon expression in the broad host range plasmid RSF1010. In Novick R. P. and Levy S. B. (eds). Evolution and enviromental spread of antibiotic resistance genes. 24 Banbury report-Cold Spring Harbor Laboratory. pp209-223.

Bagdasarian M., and Timmis K.N. 1982. Host-vector systems for gene cloning in *Pseudomonas*. Curr. Top. Microbiol. Immunol. 96:47-67.

Bahk J.-.D., Sakai H., and Komano T. 1988. Plasmid pACYC184 contains an *ssi*signal for initiation of single-stranded phage DNA replication. Gene 65:93-99.

Bahk J.-.D., Sakai H., and Komaru T. 1989. Mutational analysis of an *ssi* region carried by plasmid pACYC184. Gene 75:297-304.

Baker T.A., Funnell B.E., and Kornberg A. 1987. Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin *in vitro*. J. Biol. Chem. 262:6877-6885.

Balbas P., Soberon X., Merino E., Zurita M., Lomeli H., Valle F., Flores N., and Bolivar F. 1986. Plasmid vector pBR322 and its special-purpose derivatives - a review. Gene 50:3-40.

Banner D., Kokkinidis M., and Tsernoglou D. 1987. Structure of the ColE1 Rop protein at 1.7 Å resolution. J. Mol. Biol. 196:657-675.

Barth P.T., Ellis K., Bechhofer D.H., and Figurski D.H. 1984. Involvement of *kil* and *kor* genes in the phenotype of a host ragnge mutant of RP4. Mol. Gen. Genet. 197:236-243.

Barth P.T., and Grinter N.J. 1974. Comparison of the deoxyribonucleic acid molecular weight and homologies of plasmids conferring linked resistance to streptomycin and sulfonamides. J.Bacteriol. 120:618-630.

Bechhofer D.H., Kornacki J.A., Firshein W.B., Figurski D.H. 1986. Gene control in broad-host-range plasmid RK2: expression, polypeptide poroduct and multiple regulatory function of *korB*. Proc. Natl. Acad. Sci. USA. 83:394-398.

Berger D. K., Woods D. R., and Rawlings D. E. 1990. Complementation of *Escherichia coli* s^{54} (NtrA)-dependent formate hydrogenlyase activity by a cloned *Thiobacillus ferrooxidans ntrA* gene. J. Bacteriol. 172:4399-4406.

Bernardi A., and Bernardi F. 1984. Complete sequence of pSC101. Nucleic Acids Res. 12:9415-9426.

Biek D.P., and Cohen S.N. 1989a. Involvement of integration host factor (IHF) in maintenance of plasmid pSC101 in *Escherichia coli*: characterisation of pSC101 mutants that replicate in the absence of IHF. J. Bacteriol. 171:2056-2065.

Biek D.P., and Cohen S.N. 1989b. Involvement of integration host factor (IHF) in maintenance of plasmid pSC101 in *Escherichia coli*: mutations in the *topA* gene allow pSC101 replication in the absence of IHF. J. Bacteriol. 171:2066-2074.

Bramhill D., and Kornberg A. 1988. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *Escherichia coli* chromosome. Cell 52:743-755.

Brierly C. 1978. Microbiological mining. Sci. Am. 247:42-51.

Bryan L.E., Van den Elzen H.M., and Tseng T. 1972. Transferable drug resistance in *Pseudomonas aeruginosa*. Antimicrobial Ag. Chemother. 1:22-29.

Cabello F., Timmis K., and Cohen S.N. 1976. Replication control in a composite plasmid construct by *in vitro* linkage of two distinct replicons. Nature (London) 259:285-290.

Casadaban M.A., Martinez-Aras A., Shapiro S.A., and Chou J. 1983. ß-Galactosidase gene fusions for analysing expression in *E. coli* and yeast. Methods in Enzymology 100:293-308.

Castagnoli L. 1987. Characterization of a promoter up mutation in the -35 region of the promoter of the primer for ColE1 replication. Mol. Gen. Genet. **206**:180-190.

Cesarini G., Cornelissen M., Lacatena R.M., and Castagnoli L. 1984. Control of pMB1 replication: inhibition of primer formation by Rop requires RNAI. EMBO J. 3:1365-1369.

Chan P.T., Ohmori H., Tomizawa J., and Lebowitz J. 1985. Nucleotide sequence and gene organization of ColE1 DNA. J. Biol. Chem. **260**:8925-8935.

Chang A.C.Y., and Cohen S.N. 1978. Construction and characterisation of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J. Bacteriol. **134**:1141-1156.

Chikami G.K., Guiney D.G., Schmidhauser T.J., and Helinsky D.R. 1985. Comparison of 10 IncP plasmids: homology in the regions involved in plasmid replication. J. Bacteriol. **162**:656-660.

Chung C.T., and Miller R.H. 1988. A rapid and convenient method for the preparation and storage of competent bacterial cells. Nucleic Acids Res. **16**:3580.

Churchward G., Linder P., and Caro L. 1983. The nucleotide sequence of replication and maintenance functions encoded by plasmid pSC101. Nucleic Acids Res. **11**:5645-5659.

Cohen S.N., and Chang A.C.Y. 1977. Revised interpretation of the origin of the pSC101 plasmid. J. Bacteriol. **32**:734-737.

Cohen S.N., Chang A.C.Y., Boyer H.W., and Helling R.B. 1973. Construction of biologically functional plasmids *in vitro*. Proc. Natl. Acad. Sci. USA. **70**:3240-3244.

Couturier M., Bex F., Bergquist P.L., and Maas W.K. 1988. Identification and classification of bacterial plasmids. Micorbiol. Rev. **52**:375-395.

Cross M.A., Warne S.R., and Thomas C.M. 1986. Analysis of the vegetative replication origin of broad-host-range plasmid RK2 by transposon mutagenesis. Plasmid **15**:132-146.

Croza J. H. 1980. Three origins of replication are active *in vivo* in the R plasmid RSF1040. J. Biol. Chem. **255**:11075-119077.

Dasgupta S., Masukata H., and Tomizawa J. 1987. Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. Cell **51**:1113-1122.

David M., Tronchet M., and Denarie J. 1981. Transformation of *Azotobacter vinelandii* with plasmids RP4 (incP-1 group) and RSF1010 (incQ group). J. Bacteriol. **146**:1154-1157.

de Bruin J.N. and Lupski J.R. 1984. The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA sequences cloned into multicopy plasmids - a review. Gene **27**:131-149.

deGraaf J., Crosa J.H., Heffron F., and Falkow S. 1978. Replication of the nonconjugative plasmid RSF1010 in *Escherichia coli* K-12. J. Bacteriol. **134**:1117-1122.

del Solar G.H., Puyet A., and Espinosa M. 1987. Initiation signals for the conversion of single stranded to double stranded DNA forms in the streptococcal plasmid pLS1. Nucleic Acids Res. **15**:5561-5580.

del Solar G.H., de la Campa A.G., Perez-Martin J., Choli T., and Espinosa M. 1989. Purification and characterization of RepA, a protein involved in the copy number control of plasmid pLS1. Nucleic Acids Res. **17**:2405-2420.

Derbyshire K.M., Hatfull G., and Willetts N. 1987. Mobilization of the non-conjugative plasmid RSF1010: a genetic and DNA sequence analysis of the mobilization region. Mol.Gen.Genet. **206**:161-168.

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

Diaz R., and Staudenbauer W.L. 1982. Replication of the broad host range plasmid RSF1010 in cell-free extracts of *Escherichia coli* and *Pseudomonas aeruginosa*. Nucleic Acids Res. **10**:4687-4702.

Donogue D.J., and Sharp P.A. 1978. Replication of colicin E1 plasmid DNA *in vivo* requires no plasmid-encoded proteins. J. Bacteriol. **133**:1287-1294.

Dorrington R.A., and Rawlings D.E. 1989. Identification and sequence of the basic replication region of a broad-host-range plasmid isolated from *Thiobacillus ferrooxidans*. J. Bacteriol. **171**:2735-2739.

Dorrington R.A., and Rawlings D.E. 1990. Characterization of the minimum replicon of the broad-host-range plasmid pTF-FC2 and similarity between pTF-FC2 and the IncQ plasmids. J. Bacteriol. **172**:(in press).

Durland R.H., and Helinsky D.R. 1987. The sequence encoding the 43-kilodalton *trfA* protein is required for efficient replication or maintenance of minimal RK2 replicons in *Pseudomonas aeruginosa*. Plasmid **18**:164-169.

Durland R.H., and Helinsky D.R. 1990. Replication of the broad-host-range plasmid RK2: direct measurement of intracellular concentrations of the essential TrfA replication proteins and their effect on plasmid copy number. J. Bacteriol. **172**:3849-3858.

Durland R.H., Toukdarian A., Fang F., and Helinsky D.R. 1990. Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers. J. Bacteriol. **172**:3859-3867.

Ely S., and Wright A. 1985. Maintenance of plasmid pSC101 in *Escherichia coli* requires the host primase. J. Bacteriol. **164**:484-486.

Figurski D.H., and Helinski D.R. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. USA. **75**:1648-1652.

Filutowicz M., McEachern M., Greener A., Mukhopadhyay P., Uhlenhopp E., Durland R., and Helinski D.R. 1985. Role of the ' initiation protein and direct nucleotide sequence repeats in the regulation of plasmid R6K replication. In Helinski D.R., Cohen S.N., Clewell D.R., Jackson D.A. and Hollaender A. (eds). Plasmids in bacteria. Plenum, New York/London. pp125-140.

Firshein W., Strumph P., Benjamin P., Burnstein K., and Kornacki J. 1982. Replication of a low copy number plasmid by a plasmid DNA-membrane complex extracted from minicells of *Escherichia coli*. J. Bacteriol. **150**:1234-1243.

Frey J., and Bagdasarian M. 1989. The molecular biology of IncQ plasmids. In C. M. Thomas (ed.). Promiscuous Plasmids of Gram-negative bacteria. Academic Press, London. pp79-94

Frey J., Bagdasarian M., Feiss D., Franklin F.C.H., and Deshusses J. 1983. Stable cosmid vectors that enable the introduction of cloned fragments into a wide range of Gram-negative bacteria. Gene **24**:299-308.

Frieden P., Voelkel K., Sternglanz R., and Freudlick M. 1984. Reduced expression of isoleucine , valine enzymes in IHF mutants of *Escherichia coli*. J. Mol. Biol. **172**:484-489.

Fuller R.S., Funnell B.E., and Kornberg A. 1984. The *dnaA* protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. Cell **38**:889-900.

Funnell B.E., Baker T.A., and Kornberg A. 1986. Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome. J. Biol. Chem. **261**:5616-5624.

Gamas P., Burger A.C., Churchward G., Caro L., Galas D., and Chandler M. 1986. Replication of pSC101: effects of mutations in the *E. coli* DNA binding protein IHF. Mol. Gen. Genet. **204**:85-89.

Gaylo P.J., Turjman N., and Bastia D. 1987. DnaA protein is required for replication of the minimal replicon of the broad-host-range plasmid RK2 in *Escherichia coli*. J. Bacteriol. **169**:4703-4709.

Gerdes K., Rasmussen P.B., and Molin S. 1986. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. Proc. Natl. Acad. Sci. USA. **83**:3116-3120.

Grindley N.D.F., and Kelley W.S. 1976. Effects of different alleles of *Escherichia coli* K-12 *polA* gene on replication of non-transferring plasmids. Mol. Gen. Genet. **143**:311-318.

Grinter N.J. 1984. Replication control of IncP plasmids. Plasmid **11**:74-81.

Grinter N.J., and Barth P.T. 1976. Characterization of SmSu plasmids by restriction endonuclease cleavage and compatibility testing. J. Bacteriol. **128**:394-400.

Guerry P., van Embden J., and Falkow S. 1974. Molecular nature of two nonconjugative plasmids carrying drug resistance genes. J.Bacteriol. **117**:619-630.

Hansen L.B., and Yarmolinsky M.B. 1986. Host participation in plasmid maintenance: dependence upon *dnaA* of replicons derived from P1 and F. Proc. Natl. Acad. Sci. USA **83**:4423-4427.

Haring V., and Scherzinger E. 1989. Replication proteins of the IncQ plasmid RSF1010. In C.M. Thomas (ed.). Promiscuous plasmids of Gram-negative bacteria. Academic Press **London. pp95 - 124SD FG:-.**

Haring V., Scholz P., Scherzinger E., Frey J., Derbyshire K., Hatful] G., Willetts N.S., and Bagdasarian M. 1985. Protein RepC involved in copy number control of the broad host range plasmid RSF1010. Proc. Natl. Acad. USA. **82**:6090-6094.

Harrison A.P. 1984. The acidophilic thiobacilli and other acidophilic bacteria that share their habitat. Ann. Rev. Microbiol. **38**:265-292.

Hasanuma K., and Sekiguchi M. 1979. Effect of *dna* mutations on the replication of plasmid pSC101 in *Escherichia coli*. J. Bacteriol. **137**:1095-1099.

Hasunuma K., and Sekiguchi M. 1977. Replication of plasmid pSC101 in *E. coli* K-12: requirement for *dnaA* function. Mol. Gen. Genet. **154**:225-230.

Hawley F., and McClure W.R. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237-2255.

Helmer-Citterich M., Anceschi M.M., Banner D.W., and Cesareni G. 1988. Control of ColE1 replication: low affinity specific binding of Rop(Rom) to RNAI and RNAII. EMBO J. 7:557-566.

Henikoff S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-369.

Hidaka M., Akiyama M., and Horiuchi T. 1988. A consensus sequence of three DNA replication terminus sites on the *E. coli* chromosome is highly homologous to the *terR* sites of R6K plasmid. Cell 55:467-475.

Hill T.M., Pelletier A., and Tecklenburg W.D. 1988. Identification of the DNA sequence from the *Escherichia coli* terminus region that halts replication forks. Cell 55:459-466.

Hirochika H., Nakamura K., and Sakaguchi K. 1984. A linear DNA plasmid from *Streptomyces rochei* with an inverted terminal repetition of 614 base pairs. EMBO J. 3:761-766.

Holmes D.S., Lobos J.H., Bopp L.H., and Welch G.C. 1984. Cloning of a *Thiobacillus ferrooxidans* plasmid in *Escherichia coli.* J. Bacteriol. 157:324-326.

Honda Y., Sakai H., and Komano T. 1988. Two single-strand DNA initiation signals located in the *oriV* region of plasmid RSF1010. Gene 68:221-228.

Honda Y., Sakai H., Komano T., and Bagdasarian M. 1989. RepB' is required in *trans* for the two single-strand DNA initiation signals in *oriV* of plasmid RSF1010. Gene 80:155-159.

Hoyt M.A., Knight D.M., Das A., Miller H.I., and Echols H. 1982. Control of phage lambda development by stability and synthesis of cIII protein: role of viral *cII* and host *hflA, himA* and *himD* genes. Cell 31:565-573.

Inselburg J. 1974. Replication of colicin E1 plasmid DNA in minicells from a unique replication initiation site. Proc. Natl. Acad. Sci. USA. 71:2256-2259.

Inuzuku N., Inuzuka M. and Helinski D. R. Activity *in vitro* of three replication origins of the antibiotic resistance plasmid RSF1040. J. Biol. Chem. 255:11071-11074.

Ish-Horowicz D., and Burke F. 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. 9:2989-2998.

Itoh T., and Tomizawa J. 1978. Initiation of replication of plasmid ColE1 DNA by RNA polymerase, ribonuclease H and DNA polymerase I. Cold Spring Harbor Symp. Quant. 43:409-418.

Itoh T., and Tomizawa J. 1980. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc. Natl. Acad. Sci. USA. 77:2450-2454.

Itoh Y., Watson J.M., Haas D., and Leisinger T. 1984. Genetic and molecular characterization of the *Pseudomonas* plasmid pVS1. Plasmid 11:206-220.

Kaguni M., and Kornberg A. 1984. Replication initiation at the origin (*ori*C) of the *E. coli* chromosome reconstituted with purified enzymes. Cell **38**:183-190.

Kamio Y., Tabuchi A., Itoh Y., Katagiri H., and Terawaki Y. 1984. Complete nucleotide sequence of mini-Rts1 and its copy mutant. J. Bacteriol. **158**:307-312.

Kim K., and Meyer R. 1985. Copy number of the broad host-range plasmid R1162 is determined by the amounts of essential plasmid-encoded proteins. J. Mol. Biol. **185**:755-767.

Kim K., and Meyer R.J. 1986. Copy number of broad host-range plasmid R1162 is regulated by a small RNA. Nucleic Acids Res. **14**:8027-8046.

Kim Y.-J., Lin L.-S., and Meyer R.J. 1987. Two domains at the origin are required for replication and maintenance of broad-host-range plasmid R1162. J. Bacteriol. **169**:5870-5872.

Kingsbury D.T., and Helinsky D.R. 1970. DNA polymerase as a requirement for the maintenance of the bacterial plamsid colicinogenic factor. E. Biochem. Biophys. Res. Comm. **4**:1538-1544.

Kornacki J.A., Balderes P.J., and Figurski D.H. 1987. Nucleotide sequence of *korB*, a replication control gene of broad-host-range plasmid RK2. J. Mol. Biol. **198**:211-222.

Kornacki J.A., Burlage R.S., and Figurski D.H. 1990. The *kil-kor* regulon of broad-host-range plasmid RK2: nucleotide sequence, polypeptide product and expression of requlatory gene *korC*. J. Bacteriol. **172**:3040-3050.

Krishnapillai V. 1986. Genetic analysis of bacterial plasmid promiscuity. J. Genet. **65**:103-120.

Krishnapillai V., Nash J., and Lanka E. 1984. Insertion mutations in the promiscuous IncP-1 plasmid R18 which affect its host range between *Pseudomonas* species. Plasmid **12**:170-180.

Krishnapillai V., Wexler M., Nash J., and Figurski D.H. 1987. Genetic basis of a Tn*7* insertion mutation in the *trfA* region of the promiscuous IncP plasmid R18 which affects its host range. Plasmid **17**:164-166.

Kues U., and Stahl U. 1989. Replication of plasmids in Gram-negative bacteria. Microbiol. Rev. **53**:491-516.

Lacatena R.M., and Cesareni G. 1983. Interaction between RNAI and the primer precursor in the regulation of ColE1 replication. J. Mol. Biol. **170**:635-650.

Lacks S.A., Lopez P., Greenberg B., and Espinosa M. 1986. Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. J. Mol. Biol. **192**:753-765.

Lane D., Hill D., Caughey P., and Gunn P. 1984. The mini-F primary origin. J.Mol.Biol. **180**:267-282.

Lin L.-S., Kim Y.-J., and Meyer R. 1987. The 20 bp directly repeated DNA sequence of broad-host-range plasmid R1162 exerts incompatibility *in vivo* and inhibits R1162 replication *in vitro*. Mol. Gen. Genet. **208**:390-397.

Lin L.-S., and Meyer R.J. 1986. Directly repeated 20-bp sequence of plasmid R1162 DNA is required for replication, expression of incompatibility and copy-number control. Plasmid **15**:35-47.

Lin-Chao S., and Bremer H. 1986a. Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*. Mol. Gen. Genet. **203**:143-149.

Lin-Chao S., and Bremer H. 1986b. Effect of *relA* function on the replication of plasmid pBR322 in *Escherichia coli*. Mol. Gen. Genet. **203**:150-153.

Linder P., Churchward G., Guixian X., Yi-Yi Y., and Caro L. 1985. An essential replication gene, *repA*, of plasmid pSC101 is autoregulated. J. Mol. Biol. **181**:383-393.

Lodge J., Williams R., Bell A., Chan B. and Busby S. 1990. Comparison of promoter activities in *Escherichia coli* and *Pseudomonas aeruginosa*: use of a new broad-host-range promoter-probe plasmid. FEMS Microbiol. Lett. **67**:221-226.

Lopilato J., Bortner S., and Beckwith J. 1986. Mutations in a new chromosomal gene of *Escherichia coli* K-12, *pcnB*, reduce plasmid copy number of pBR322 and its derivatives. Mol. Gen. Genet. **205**:285-290.

Lui J., and Parkinson J.S. 1989. Genetic and sequence analysis of the *pcnB* locus, an *Escherichia coli* gene involved in plasmid copy number control. J. Bacteriol. **171**:1254-1261.

Lundquist P.D., and Levin B.R. 1986. Transitory derepression and the maintenance of conjugative plasmids. Genetics **113**:483-497.

Makris J.C., Nordmann P.L., and Reznikoff W.S. 1990. Integration host factor plays a role in IS*50* and Tn*5* transposition. J. Bacteriol. **172**:1368-1373.

Maniatis T., Fritsch F., and Sambrook J. 1989. Molecular cloning: a laboratory manual.. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Marians K.J., Minden J.S., and Parada C. 1986. Replication of superhelical DNAs *in vitro*. Progress in Nucleic Acid Research and Molecular Biology **33**:111-140.

Martin P.A.W., Dugan P.R., and Tuovinen O.H. 1981. Plasmid DNA in acidophilic, chemolithotrophic thiobacilli. Can. J. Microbiol. **27**:850-853.

Martin P.A.W., Dugan P.R., and Tuovinen O.H. 1983. Uranium resistance of *Thiobacillus ferrooxidans*. Eur. J. Appl. Microbiol. Biotech. **18**:392-395.

Masai H., and Arai K.-I. 1987. RepA and DnaA proteins are required for initation of R1 plasmid replication *in vitro* and interact with the *ori*R sequence. Proc. Natl. Acad. USA. **84**:4781-4785.

Masson L., and Ray D.S. 1988. Mechanism of autonomous control of the *Escherichia coli* F plasmid: purification and characterization of the *repE* gene product. Nucleic Acids Res. **16**:413-424.

Masters M., March J.B., Oliver I.R., and Collins J.F. 1990. A possible role for the *pcnB* gene product of *Escherichia coli* in modulating RNA:RNA interactions. Mol. Gen. Genet. **220**:341-344.

Masukata H., Dasgupta S., and Tomizawa J. 1987. Transcriptional activation of ColE1 DNA synthesis by displacement of the nontranscribed strand. Cell **51**:1123-1130.

Masukata H., and Tomizawa J. 1986. Control of primer formation for ColE1 plasmid replication: conformational change of the primer transcript. Cell **44**:125-136.

Masukata H., and Tomizawa J. 1987. Control of primer formation for ColE1 plasmid replication: conformational change of the primer transcript. Cell **44**:125-136.

Messer W. 1987. Initiation of replication in *Escherichia coli*. J. Bacteriol. **169**:3395-3399.

Meyer R., and Helinsky D.R. 1977. Unidirectional replication of the P-group plasmid RK2. Biochim. Biophys. Acta. **478**:109-113.

Meyer R., and Hinds M. 1982. Multiple mechanisms for expression of incompatibility by broad-host-range plasmid RK2. J. Bacteriol. **152**:1078-1090.

Meyer R., Hinds M., and Brasch M. 1982a. Properties of R1162, a broad-host-range, high-copy-number plasmid. J.Bacteriol. **150**:552-562.

Meyer R., Laux R., Boch G., Hinds M., Bayly R., and Shapiro J.A. 1982b. Broad-host-range IncP-4 plasmid R1162: effect of deletions and insertions on plasmid maintenance and host range. J. Bacteriol. **152**:140-150.

Meyer R., Lin L.-.S., Kim K., and Brasch M. 1985. Broad-host-range plasmid R1162: replication, incompatibility and copy-number control. In Helinski D.R. Cohen S.N., Clewell D.R., Jackson D.A. and Hollaender A. (eds). Plasmids in bacteria. Plenum, New York/London. pp173-188.

Miller H.I. 1984. Primary structure of the *himA* gene of *Escherichia coli*: homology with DNA binding protein HU and association with the phenylalanine tRNA synthetase operon. Cold Spring Harbor Symp. Quant. Biol. **49**:691-698.

Miller R.E., and Stadtmann E.R. 1972. Glutamate synthase from *E. coli*: an ironsulphide fluoroprotein. J. Biol.Chem. **109**:275-301.

Minden J.S., and Marians K.J. 1985. Replication of pBR322 DNA *in vitro* with purified proteins. Requirement for topoisomerase I in the maintenance of template specificity. J. Biol. Chem. **260**:9316-9325.

Morlon J., Chartier M., Bidaud M., and Ladzunski C. 1988. The complete nucleotide sequence of the colicinogenic plasmid ColA. High extent of homology with ColE1. Mol. Gen. Genet. **211**:231-243.

Mukherjee S., Patel I., and Bastia D. 1985. Conformational changes in a replication origin induced by an initiation protein. Cell **43**:189-197.

Nijkamp H.J.J., de Lang R., Stuitje A.R., van den Elzen P.J.M., Veltkamp E., and van Putten A.J. 1986. The complete nucleotide sequence of the bacteriocinogenic plasmid CloDF13. Plasmid **16**:135-160.

Nomura N., Low R.I., and Ray D.S. 1982. Identification of ColE1 DNA sequences that direct single strand-to-double strand conversion by ιX174 type mechanism. Proc. Natl. Acad. Sci. USA **79**:3153-3157.

Nordheim A., Hashimoto-Gotoh T., and Timmis K.N. 1980. Location of two relaxation nick sites in R6K and single sites in pSC101 and RSF1010 close to origins of vegetative replication: implications for conjugal transfer of plasmid deoxyribonueic acid. J. Bacteriol. **144**:023-932.

Nordström K., Molin S., and Aagard-Hansen H. 1980. Partitioning of plasmid R1 in Escherichia coli I. Kinetics of loss of plasmid derivatives deleted of the par region. Plasmid **3**:215-227.

Novick R.P. 1987. Plasmid incompatibility. Microbiol. Rev. **51**:381-395.

Ogawa T., Baker T.A., van der Ende A., and Kornberg A. 1985. Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosone: contribution of RNA polymerase and primase. Proc. Natl. Acad. Sci. USA. **82**:3562-3566.

Okazaki T., and Okazaki R. 1969. Mechanism of DNA chain growth. IV. Direction of synthesis of T4 short DNA chains as revealed by exonucleolytic degradation. Proc. Natl. Acad. Sci. USA. 64:1242-1248.

Orr E., and Staudenbauer W.L. 1981. An *Escherichia coli* mutant thermosensitive in the b-subunit of DNA gyrase: effect on the structure and replication of the colicin E1 plasmid *in vitro*. Mol. Gen. Genet. **181**:52-56.

Pabo C.O. and Sauer R.T. and Protein-DNA recognition. Ann. Rev. Biochem. 17:293-221.

Pansegrau W., and Lanka E. 1987. Conservation of a common "backbone" in the genetic organization of the IncP plasmids RP4 and R751. Nucleic Acids Res. **15**:2385-.

Persson C., and Nordstrom K. 1986. Control of replication of the broad host range plasmid RSF1010: the incompatibility determinant consists of directly repeated DNA sequences. Mol.Gen.Genet. **203**:189-192.

Pinkney M., Diaz R., Lanka E., and Thomas C.M. 1988. Replication of mini RK2 plasmid in extracts of *Escherichia coli* requires plasmid-encoded protein TrfA and host-encoded proteins DnaA, B, G, DNA gyrase and DNA polymerase III. J. Mol. Biol. 203:927-938.

Pinkney M., and Thomas C.M. 1987. Replication and maintenance of promiscuous plasmids in Gram-negative bacteria. Microbiol. Sci. **4**:186-191.

Polisky B. 1988. ColE1 replication control circuitry: sense from antisense. Cell **55**:929-932.

Pretorius I.-.M., Rawlings D.E., O' Neil E.G., Jones W.A., and Woods D.R. 1987. Nucleotide sequence of the gene encoding the nitrogenase iron protein of *Thiobacillus ferrooxidans*. J. Bacteriol **169**:367-370.

Priefer U.B., Simon R., and Puhler A. 1985. Extension of the host range of *Escherichia coli* vectors by incorporation of RSF1010 replication and mobilization functions. J. Bacteriol. **163**:324-330.

Ramesar R.S., Abratt V., Woods D.R., and Rawlings D.E. 1989. Nucleotide sequence and expression of a cloned *Thiobacillus ferrooxidans recA* gene in *Escherichia coli*. Gene **78**:1-8.

Rawlings D.E. 1988. Sequence and structural analysis of the a-and b-dinitrogenase subunits of Thiobacillus ferrooxidans. Gene **69**:337-343.

Rawlings D.E. 1990. Engineering of *Thiobacillus* for mining. In Silver S. (ed.). Biotechnology of Open Systems:Academic Press-Orlando. (in press)

Rawlings D.E., Gawith C., Petersen A., and Woods D.R. 1983. Characterization of plasmids and potential genetic markers in *Thiobacillus ferrooxidans*. In Rossi G. and Torma A.E. (eds). Recent progress in biohydrometallurgy. Associazonw\e Mineraria Sarda, Cagliari. pp 555-570.

Rawlings D.E., Jones W.A., O' Neil E.G., and Woods D.R. 1987. Nucleotde sequence of the glutamine synthetase gene and its controlling region from the acidophilic autotroph *Thiobacillus ferrooxidans*. Gene **53**:211-217.

Rawlings D.E., Pretorius I.-M., and Woods D.R. 1984. Expression of a *Thiobacillus ferrooxidans* origin of replication in *Escherichia coli*. J. Bacteriol. **158**:737-738.

Rawlings D.E., Pretorius I.-M., and Woods D.R. 1986a. Expression of *Thiobacillus ferrooixidans* plasmid functions and the development of genetic systems for the Thiobacilli. Biotechnology and Bioengineering Symp.No.29. John Wiley & Sons Inc. pp291-297.

Rawlings D.E., Sewcheran R., and Woods D.R. 1986b. Characterization of a broad-host-range mobilizable *Thiobacillus ferrooxidans* plasmid and the construction of *Thiobacillus* cloning vectors. Proc. of the International Symp. on Biohydrometallurgy Elsevier. pp419-427.

Rawlings D.E., and Woods D.R. 1985. Mobilisation of *Thibacillus ferrooxidans* plasmids among *Escherichia coli* strains. Appl.Env.Micro. **49**:1323-1325.

Rokeach L.A., Sogaard-Andersen L., and Molin S. 1985. Two functions of the E protein are key elements in the plasmid F replication control system. J.Bacteriol. **164**:1262-1270.

Salser W. 1977. Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. BIol. **42**:985-1002.

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

Schafferman A., Flashner Y., Hertman I. and Lion M. 1987. Identification and characterization of the function of the functional a origin of DNA replication of the R6K plasmid and its relatedness to the R6K ß and gamma origins. Mol. Gen. Genet. **208**:263-270.

Scherzinger E., Bagdasarian M.M., Scholz P., Lurz R., Rückert B., and Bagdasarian M. 1984. Replication of the broad host range plasmid RSF1010: requirement for three plasmid encoded proteins. Proc. Natl. Assoc. Sci. USA **81**:654-658.

Schmidhauser T.J., Filutowicz M., and Helinski D.R. 1983. Replication of derivatives of broad-host-range plasmid RK2 in two distantly related bacteria. Plasmid **9**:325-330.

Schmidhauser T.J., and Helinsky D.R. 1985. Regions of broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of Gram-negative bacteria. J. Bacteriol. **164**:446-455.

Scholz P., Haring V., Scherzinger E., Lurz R., Bagdasarian M., Schuster H., and Bagdasarian M. 1985. Replication determinants of the broad host-range plasmid RSF1010. In Helinski D.R. Cohen S.N., Clewell D.R., Jackson D.A. and Hollaender A. (eds). Plasmids in bacteria. Plenum, New York/London. pp243-259.

Scholz P., Haring V., Wittman-Liebold B., Ashman K., Bagdasarian M., and Scherzinger E. 1989. Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. Gene **75**:271-288.

Schreiner H.C., Bechhofer D.H., Pohlman R.F., Young C., Borden P.A., and Figurski D.H. 1985. Replication control in promiscuous plasmid RK2: *kil* and *kor* functions affect expression of the essential replication gene *trfA*. J. Bacteriol. **163**:228-237.

Scott J.R. 1984. Regulation of plasmid replication. Microbiol. Rev. **48**:1-23.

Selzer G., Som T., Itoh T., and Tomizawa J. 1983. The origin of replication of plasmid p15A and comparitive studies on the nucleotide sequences around the origin of related plasmids. Cell **32**:119-129.

Selzer G., and Tomizawa J. 1982. Specific cleavage of the p15A primer precursor by ribonuclease H at the origin of DNA replication. Proc. Natl. Acad. Sci. USA. **79**:7082-7086.

Seufert W., and Messer W. 1987. DnaA protein binding to the plasmid origin region can substitute for primosome assembly during replication of pBR322 *in vitro*. Cell **48**:73-78.

Shine J., and Dalgarno L. 1974. The 3' terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosomal binding sites. Proc. Natl. Acad. Sci. USA **71**:1342-1346.

Shon M., Germino J., and Bastia D. 1982. The nucleotide sequence of the replication origin ß of the plasmid R6K. J. Biol. Chem. **257**:13823-13827.

Simons R.W. 1988. Naturally occuring DNA control - a brief review. Gene **72**:35-44.

Sista P.R., Mukherjee S., Patel P., Khatri G.S., and Bastia D. 1989. A host-encoded DNA-binding protein promotes termination of plasmid replication at a sequence-specific replication terminus. Proc. Natl. Acad. Sci. USA. **86**:3026-3030.

Smith C.A., and Thomas C.M. 1989. Vegetative replication and stable maintenance of IncP plasmids. In CM Thomas (ed.). Promiscuous plasmids of Gram-negative bacteria. Academic Press:London. pp1-25.

Smith C.A., and Thomas C.M. 1985. Comparison of the nucleotide sequences of the vegetative replcation origin of broad-host-range IncP plasmids R751 and RK2 reveals conserved features of probable functional importance. Nucleic Acids Res. 13:557-572.

Som T., and Tomizawa J. 1983. Regulatory regions of ColE1 that are involved in determination of plasmid copy number. Proc. Natl. Acad. Sci. USA. 80:3232-3236.

Stalker D.M., Thomas C.M., and Helinsky D.R. 1981. Nucleotide sequence of the region of the origin of replication of the broad-host-range plasmid RK2. Mol. Gen. Genet. 181:8-12.

Stenzel T.T., Patel P., and Bastia D. 1987. The integration host factor of *Escherichia coli* binds to bent DNA at the origin of replication of the plasmid pSC101. Cell 49:709-717.

Suglio T., Domatsu C., Munakata O., Tano T., and Imai K. 1985. Role of a ferric ion-reducing system in sulfur oxidation of *Thiobacillus ferrooxidans*. Appl. Environ. Microbiol. 49:1401-1406.

Suzuki I., Takeuchi T.L., Yuthasastrakosol T.D., and Oh J.K. 1990. Ferrous iron and sulfur oxidation and ferric iron reduction activities of *Thiobacillus ferrooxidans* are affected by growth on ferrous iron sulfur, or a sulfide ore. Appl. Environ. Microbiol. 56:1620-1626.

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

te Riele H., Michel B., and Ehrlich S.D. 1986. Are single-stranded circles intermediates in plasmid DNA replication?. EMBO J. 5:631-637.

Theophilus B.D.M., Cross M.A., Smith C.A., and Thomas C.M. 1985. Regulation of the *trfA* and *trfB* promoters of broad-host-range plasmid RK2: identification of sequences essential for regulation by *trfB/korA/korD*. Nucleic Acids Res. 13:8129-8142.

Thomas C.M. 1983. Instability of a high copy number mutant of a miniplasmid derived from broad host range IncP plasmid RK2. Plasmid 10:184-195.

Thomas C.M. 1986. Evidence for the involvement of the *incC* locus of broad host range plasmid RK2 in plasmid maintenance. Plasmid 16:15-29.

Thomas C.M. 1988. Recent studies on the control of plasmid replication. Biochim. Biophys. Acta. **949**:253-263.

Thomas C.M., Cross M.M., Hussain A.A.K., and Smith C.A. 1984. Analysis of copy number control elements in the region of the vegetative replication origin of the broad host ragne plasmid RK2. EMBO J. **3**:57-63.

Thomas C.M., and Hussain A.A.K. 1984. The *korB* gene of broad-host-range plasmid RK2 is a major copy number control element which may act together with *trfB* by limiting *trfA* expression. EMBO J. **3**:1513-1519.

Thomas C.M., Hussain A.A.K., and Smith C.A. 1982. Maintenance of broad host-range plasmid RK2 replicons in *psedomonas aeruginosa*. Nature (London) **298**:674-676.

Thomas C.M., Meyer R., and Helinsky D.R. 1980. Regions of broad-host-range plasmid RK2 which are essential for replication and maintenance. J. Bacteriol. **141**:213-222.

Thomas C.M., Stalker D.M., and Helinsky D.R. 1981. Replication and incompatibility properties of segments of the origin region of replication of the broad-host range plasmid RK2. Mol. Gen. Genet. **181**:1-7.

Tokino T., Murotsu T., and Matsubara K. 1986. Purification and properties of the mini-F plasmid-encoded protein E needed for autonomous replication control of the plasmid. Proc. Natl. Acad. Sci. USA **83**:4109-4113.

Tomcsanyi T., and Apirion D. 1985. Processing enzyme ribonuclease E specifically cleaves RNAI an inhibitor of primer formation in plasmid DNA synthesis. J. Mol. Biol. **185**:713-720.

Tomizawa J. 1984. Control of ColE1 plasmid replication: the process of binding of RNAI to the primer transcript. Cell **38**:861-870.

Tomizawa J. 1985. Control of ColE1 plasmid replication: initial interaction of RNAI and the primer transcript is reversible. Cell **40**:527-535.

Tomizawa J. 1986. Control of ColE1 plasmid replication: binding of RNAI to RNAII and inhibition of primer formation. Cell **47**:89-97.

Tomizawa J. 1990a. Control of ColE1 plasmid replication: intermediates in the binding of RNAI and RNAII. J. Mol. Biol. **212**:683-694.

Tomizawa J. 1990b. Control of ColE1 plasmid replication: interaction of Rom protein with an unstable complex formed by RNAI and RNAII. J. Mol. Biol. **212**:695-708.

Tomizawa J., and Itoh T. 1981. Plasmid ColE1 incompatibility determined by interaction of RNAI with primer transcript. Proc. Natl. Acad. Sci. USA. **78**:6096-6100.

Tomizawa J., Itoh T., Selzer G., and Som T. 1981. Inhibition of ColE1 RNA primer formation by plasmid-specific small RNA. Proc. Natl. Acad. Sci. USA. **78**:1421-1425.

Tomizawa J., and Masukata H. 1987. Factor-independent termination of transcription in a strech of a deoxadenosine residues in the template DNA . Cell **51**:623-630.

Tomizawa J., Sakakibara Y., and Kakefuda T. 1974. Replication of colicin E1 plasmid DNA in cell extracts: origin and direction of replication. Proc. Natl. Acad. Sci. USA. **71**:2260-2264.

Tomizawa J., Sakakibara Y., and Kakefuda T. 1975. Replication of colicin E1 plasmid DNA added to cell extract. Proc. Natl. Acad. Sci. USA. **72**:1050-1054.

Tomizawa J., and Som T. 1984. Control of ColE1 plasmid replication: enhancement of binding of RNAI to the primer transcript by the rom protein. Cell **38**:871-878.

Twigg A.J., and Sherratt D. 1980. Trans-complementable copy number mutants of plasmid ColE1. Nature (London) **283**:216-218.

van der Ende A., Teertstra R., van der Asvoort H., and Welsbeek R.J. 1983. Initiation signals for conplementary strand DNA synthesis on single-stranded plasmid DNA. Nucleic Acids Res. **11**:4957-4975.

Vocke C., and Bastia D. 1983. DNA-protein interactions at the origin of DNA replication of the plasmid pSC101. Cell **35**:495-502.

Vocke C., and Bastia D. 1985. The replication initiator protein of plasmid pSC101 is a transcriptional repressor of its own cistron. Proc. Natl. Acad. Sci. USA. **82**:2252-2256.

Walker G.C. 1986. Plasmid biology of pKM101" role of the *muc*AB genes. Banbury Report **29**:313-321.

Ward E., and Grinsted J. 1987. The nucleotide sequence of the *tnpA* gene of Tn*21*. Nucleic Acids Res. **15**:1799-1806.

Watson J.D., Hopkins N.H., Roberts J.W., Steitz J.A., and Weiner A.M. 1987. Molecular biology of the gene. The Benjamin/Cummings Publishing Corp Inc California.

Wlodarczyk M., and Nowicka B. 1988. Preliminary evidence for the linear nature of *Thiobacillus versutus* pTAV2 plasmid. FEMS Microbiol. Lett. **55**:125-128.

Wolfson J., and Dressler D. 1972. Regions of single stranded DNA in the growing points of bacteriophage T7 chromosomes. Proc. Natl. Acad. Sci. USA. **69**:2682-2686.

Wong E.M., and Polisky B. 1985. Alternative conformations of the ColE1 replication primer modulate its interaction with RNAI. Cell **42**:959-966.

Yamaguchi K., and Masumune Y. 1985. Autogenous regulation of synthesis of the replication protein in plasmid pSC101. Mol. Gen. Genet. **200**:362-367.

Yamaguchi K., and Yamaguchi M. 1984a. The replication origin of pSC101: the nucleotide sequence and replication functions of the *ori* region. Gene **28**:211-219.

Yamaguchi K., and Yamaguchi M. 1984b. The replication origin of pSC101: replication properties of a segment capable of autonomous replication. J. Gen. Appl. Microbiol. 30:347-358.

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

Young C., Bechhofer D.H., and Figurski D.H. 1984. Gene regulation in plasmid RK2: positive control by *korA* in the expression of *korC*. J. Bacteriol. **157**:247-252.

Young C., Burlage R.S., and Figurski D.H. 1987. Control of the *kilA* gene of the broad-host-range plasmid RK2: involvement of *korA, korB* and a new gene, *korE*. J. Bacteriol. **169**:1315-1320.

Young C., Prince A.S., and Figurski D.H. 1985. *korA* function of promiscuous plasmid RK2: an autorepressor that inhibits expression of host-lethal gene *kilA* and replication gene *trfA*. Proc. Natl. Acad. Sci. USA. **82**:7374-7378.

Zabeau M., and Stanley K.K. 1982. Enhanced expression of cro-beta-galactosidase fusion products under the control of the PR promoter of bacteriophage lambda. EMBO J. **1**:1217-1224.

Zipurski S.L., and Marians K.J. 1980. Identification of two *Escherichia coli* factor Y effector sites near the origin of replication of the plasmids ColE1 and pBR322. Proc. Natl. Acad. Sci. USA 77:6521-6525.

Zyskind J.W. and Smith D.W. 1986. The bacterial origin of replication, *ori*C. Cell **46**:489-490.

APPENDIX

APPENDIX 1

Bacterials strains used during the course of this work

Bacterial Strain	Genotype	Reference
E. coli		
K514	*thr-1 leu-6 thi-1 supE44 lacY1 tonA21* r^- r^+	Zabeau and Stanley., 1982
LK111	lac^q *lacZ*ΔM15 ($LacY^+$ derivative of K514	Zabeau and Stanley., 1982
AB1157	*recA00 thr-1 leu-6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1proA2 his5 arfE3 str-31*	Maniatis *et. al.*, 1982
AB1157ts12	*polA12* derivative of AB1157	Gift from Graham Walker, MIT
GW125a	*polA1* derivative of AB1157	Dorrington and Rawlings, 1989
HMS2562	thr^- leu- lac- thy- *supE hsdR tunA- trxA*	Tabor and Richardson, 1995
P. aeruginosa		
PAO1162	*leu38 rmo1*	Bagdasarian *et. al.*, 1982

APPENDIX 2

Plasmids used during the course of this study

Plasmid	Antibiotic Resistance	Description	Reference
pBR325	Tc Cm Ap	-	(Bolivar, 1987)
pUC19	Ap	-	(Yanisch-perron *et. al.*, 1985
Bluescript SK	Ap	-	Stratagene California
pKE462	Tc As	7.5 kb EcoRI fragment from R46 carrying Tc arsenic resistance cloned into R300B	Kim Ellis, Hammersmith Hospital, London
pDER401	Cm Tc	pTF-FC2 cloned into PstI site of pBR325	(Rawlings *et. al.*, 1984)
pDER412	Cm	XhoI SalI deletion of pDER401 resulting in deletion of pBR325 origin of replication	(Rawlings *et. al.*, 1984)
pGP1-2	Km	T7 RNA polymerase inder control of the temperature sensitive Lambda P_L promoter	(Tabor and Richardson, 1985)

APPENDIX 3

Plasmid constructs generated during the course of this work

PLASMID CONTRUCT	ORIGIN OF INSERT	VECTOR	ANTIBIOTIC RESISTANCE MARKER	CHAPTER REFERENCE
pSKCm1	950 bp *AsuII* fragment from pBR325	Bluescript SK	Cm Ap^R	5.4.1
pCm2	950 bp *HindIII SalI* fragment from pSKCm1	pUC19	Cm Amp	5.4.1
pTT100	12.4 kb fragment from pDER 401	Bluescript SK	Ap^R	2.4.2b
pTV100	5 kb *ClaI-PstI* fragment from pTT100 (nts 1-4910)	Bluescript SK	Ap^R	2.4.2b
	POSITION RELATIVE TO 5 kb *ClaI-PstI* FRAGMENT FROM pTV100			
pTV101	1 - 4910	pUC19	Ap^R Cm^R	5.4.1
pTV102	312 - 4910	pUC19	Ap^R Cm^R	5.4.1
pTV109	744 - 4910	pUC19	Ap^R Cm^R	5.4.1
pTV151	1 - 1296	Bluescript SK	Ap^R	5.3.1
pTV160	1 - 965	Bluescript SK	Ap^R	5.3.1
pTV200	1 - 2485	Bluescript SK	Ap^R	2.4.2b
pTV300	2485 - 4910	Bluescript SK	Ap^R	2.4.2b
pTV400	1240 - 4441	pUC19	Ap^R	2.4.2b
pTV4100	1594 - 4441	pUC19	Ap^R	2.4.3
pTV4101	1854 - 4441	pUC19	Ap^R	2.4.2b
pTV4102	1594 - 4441	pUC19	Ap^R	4.4.5
pTV4103	1594 - 4441	pUC19	Ap^R	4.4.5
pTV4111	2160 - 4441	pUC19	Ap^R	2.4.2b
pTV4161	3779 - 4441	pUC19	Ap^R	2.4.2c
pTV4162	3779 - 3964	pUC19	Ap^R	3.4.3
pTV4163	3779 - 3913	pUC19	Ap^R	3.4.4
pTV4164	3779 - 4107	pUC19	Ap^R	3.4.4
pTV4170	3816 - 4441	pUC19	Ap^R	2.4.2c
pTV4200	1240 - 4107	pUC19	Ap^R	2.4.2b
pTV4210	1240 - 3964	pUC19	Ap^R	2.4.2b
pTV4220	1240 - 3602	pUC19	Ap^R	2.4.2b
pTV4260	1240 - 2978	pUC19	Ap^R	3.4.2a
pTV4282	1240 - 2109	pUC19	Ap^R	3.4.2a
pTV4283	1240 - 2021	pUC19	Ap^R	3.4.2a
pTV4290	1240 - 1900	pUC19	Ap^R	3.4.2a
pTF100	1240 - 4441	pUC19	Km^R Ap^R	5.4.1
pTF200	1 - 4910	-	Cm^R	5.4.1
p*PII*	1854 - 2316	pMC1403	Ap^R	3.4.2d

APPENDIX 4

GENERAL METHODS

4.1 MEDIA

4.1.1 Luria Broth / Agar

Luria	Broth	Agar
Tryptone	10 g	10 g
Yeast Extract	5 g	5 g
NaCl	5 g	5 g
Agar	-	15 g
Make up to 1 l with distilled H_2O (dH_2O)		

4.1.2 LNG Broth / LNG Agar

LNG	Broth	Agar
Tryptone	10 g	10 g
Yeast Extract	5 g	5 g
NaCl	5 g	5 g
KNO_4	4 g	4 g
Glucose	4 g	4 g
Agar	-	15 g
Make up to 1 l with distilled H_2O (dH_2O)		

4.1.3 Minimal M9 Medium (Miller *et. al.*, 1972)

10x M9 minimal salts solution

Na_2HPO_4	60 g
KH_2PO_4	30 g
NH_4Cl	10 g
NaCl	5 g
dH_2O	to 1 litre

The following reagents were autoclaved separately before mixing:

M9 salts	100 ml
1 M MgSO4	1 ml
0.1 M CaCl2	1 ml
1 M thiamine-HCl	1 ml
20% glucose	10 ml
dH_2O	to 1 litre

4.2 BACTERIAL CULTURE

4.2.1. *E. coli*

Unless otherwise stated, all strains of *E. coli* were cultured in Luria Broth (LB) or on Luria Agar (LA) plates at 37 °C. Antibiotics stocks were prepared as in Maniatis *et. al.*, (1989) and routine concentrations are shown in Table 1.

4.2.2 *P. aeruginosa*

P. aeruginosa cells were cultured in LNG broth or on LNG agar plates.and inculbated overnight at 37°C. Antibiotic concnetrations used are shown in Table 1.

Talbe 4.1 Working concentrations of antibiotics used in this study

Antibiotic		Concentration (mg/l)	
		E. coli	*P. aeruginosa*
Ampicillin	(Ap)	100	
Carbinicillin	(Cb)		1000
Chloramphenicol	(Cm)	25	500
Kanamicin	(Km)	50	
Tetracyclin	(Tc)	20	

4.3 TRANSFORMATION PROCEDURES

4.3.1 *E. coli*

The DMSO method of Chung and Miller (1988) was used to prepare competent cells from all *E. coli* strains. A 10 ml overnight culture was diluted 1\50 into 25 ml LB and the culture grown to OD_{260} 0.2 -0.6 OD units. The cells were put on ice for 5 min and pelleted by centrifugation. The pellet was resuspended in 2.5 ml chilled TSB (LB (pH 6.1) contianing 10% PEG, 5% DMSO and 20 mM Mg^{++} (10 mM $MgSO_4$ and 10 mM $MgCl_2$) and incubated on ice for 10 min. If required, 100 ul aliquots of competent cells were stored at -7) °C for 1 - 2 weeks. Five to 100 ng of DNA was added to 100 ul competent cells and incubated on ice for 5 - 30 min. Cells were expressed by adding 900 ul TSB containing 20 mM Glucose and incubation at 37°C for 1 h after which cells were plated onto LA containing the appropriate antibiotics.

4.3.2 *P. aeruginosa*

A 25 ml overnight cmlture of *P. aeruginosa* PAO1162 was centrifuged at 10 000 rpm for 15 min, the pellet resuspended in 10 ml ice cold 150 mM $MgCl_2$ and the cells incubated on ice for 5 min. The cells were centrifuged at 10 000 rpm for 15 min at 4 °C and the pellet resuspended in 2 ml cold 150 mM $MgCl_2$. DNA (100 ng - 1 mg) was added to 200 ml cells and incubated on ice for 1 h. The cells were placed at 37 °C for 5 min and then on ice for a further 5 min. Cells were expressed by adding 900 ml LNG broth and incubation at 37 °C for 1 - 2 h before plating the entire transformation mix on LNG agar plates containing the relevant antibiotics.

4.4 PREPARATION OF DNA

4.4.1 Preparation of total cellular DNA

Total DNA was extracted from *E. coli* cells using the modified method of Kirby and Wotton (1979). Cells were grown overnight in LB containing the appropriate antibiotics and harvested by centrifugation at 5 000 rpm for 10 minutes in a Sorval GSA rotor. The pellet was resuspended in 10 ml Solution I (25% sucrose; 10 mM Tris HCl, pH 8.0; 10 mM Na_2EDTA) and lysozyme was added to a final concentration of 1 mg/ml. After incubation at 37 °C, the sample was cooled on ice and 5 ml of a 0.25 M solution of Na_2EDTA was added. After incubation on ice for a further 5 minutes, 10 ml of a 2% solution of Solution II (2% SDS, 10 mM Tris HCl, pH 8.0; 10 mM Na_2EDTA) was added and the solution incubated at room temperature for 10 to 60 min. RNase was added tp a final concentration of 50 mg/ml and the sample incubated for 30 min at 50 °C after which proteinase K was added to a final concentration of 50 mg/ml. After incubation at 50 °C for a further 60 min, the sample was extracted repeatedly with phenol/chloroform/isoamylalcohol and precipitated with ethanol. The pellet was resuspended in 1 ml TE buffer.

4.4.2 Preparation of plasmid DNA

4.4.2a Minipreps

Plasmid DNA was isolated using the method of Ish-Horowicz and Burke (1981). Cells were grown overnight in LB containing the appropriate antibiotic. A 1.5 ml sample of was centrifuged in an Eppendorf microfmge tube for 1 min, the pelleted cells resuspended in 200 ml Solution I (50 mM glucose; 25 mM Tris-HCl, pH 8.0) and incubated for 5 min at room temperature. Subsequenctly, 400 ml of Solution II (0.2 M NaOH, 1% (w/v) SDS) was added, the sample was vortexed briefly and placed on ice for 5 min after which 300 ml ice-cold Solution III (5 M KOAc, pH 4.8). After thromgh mixing, and incubation for 5 min on ice, cellmlar debris and denatured chromosomal DNA were pelleted by centrifugation for 5 min. The supernatant (750 ml) was removed to a fresh tube and the plasmid DNA precipitated by adding an equal volume of isopropanol and centrifugation for 5 minutes. The pellet was washed with ice cold 70% ethanol and after drying, resuspended in 50 ml of TE buffer.

4.4.2b Maxipreps

A 200 ml cmlture was grown overnight at 37°C in LB supplemented with the appropriate antibiotics. Cells were harvested by centrifugation at 6 000 *g* for 5 min and resuspended in 4 ml Solution I. After 5 min at room temperature, 8 ml Solution II was added, and the mixture kept on ice for 5 min, before the addition of 6 ml ice cold Solution III. After a further 5 min on ice the cellmlar debris was removed by centrifugation at 12 000 *g* for 10 min. An equal volume of isopropanol was added to the supernatant and the DNA was precipitated by centrifugation at 27 000 *g* for 15 min. The pellet was washed with 70% ethanol and resuspended in 4.2 ml TE buffer, and purified by isopicnic CsCl-EtBr mltracentrifugation (Maniatis et al. 1989).

APPENDIX 5

Complete nucleotide sequence of the pTF-FC2 origin of replication

```
        ClaI
   1  ATCGATATTC ACGACTGGCG CAACCGGGAC AGCACGCTTG CCGCCTTGCA
  51  GCTATCGGCG CAGAAGTGGG GCAGCTTCAC CGTAACCGGG AACGACGAAT
 101  ACAAGGCCAT GTGCGGGAAG TTAGCAGCAG AGCACGGCTT CAAGATCACA
 151  AATGCAGAGC TTCAGGAGAG CATCCAGCAG GAGCGGCAGC GGATACAGCA
 201  GGAGAGGGCG CAGGCGATGA AATCGGAGCA GCTAAAGCAG TTCGAGCGGT
                  RepB    >
 251  ACGCGGAAGC GGTAGGCGCG GAGCGCTACC GGGTAACGTC CATCAAGATG
 301  CAGGCAGACG GAAGGAAGCA AACCTTCATC CTCGACAAGA AGGACGGCAT
 351  CACGCGGGGT TTCACACCGC AGGAGATCGA GCAGCGCACG CCGGAGATGC
 401  AGCGCCTACA GCGCCGGGGC GAAAACCTCT ACTACACGCC GCTATCGGAC
 451  AAGAAGCATC ACATCCTCAT CGACGACATG AACCGGGAGA AGCTGGAGCG
 501  GCTTATCAAA GACGGCTACC AGCCCGCCGC CGTGCTGGAA TCCAGCCCCG
 551  GCAACTATCA GGCCATCATC ACCGTGTCGA AGCTGGGGAC GGCCCACGAT
 601  AAGGACGTGG GCAACCGCCT GAGCGATGCC CTGAACCGTG AATACGGCGA
 651  CCCGAAGCTA TCGGGAGCCA TCCACCCGCA CCGCGCACCC GGCTACGAGA
 701  ACCGCAAGCC CAAGCACCAG CGGGAGGACG GCAGCTATCC AGAGGTGCGC
 751  TTGCTCAAGG CCGAGCGCCG GGAGTGCGTC AAGGCGCTGG CCTTGTCCAG
 801  CCAGATCGAC GCCGAGTATC AGCGGCAAGC GGCCTTGAAG GCGCAGCAGC
 851  CCGAGCGCAC GAAAGCCAAG CCCGCCTTGG AGCTTGCAGC GGCCAGCGGC
      AvaI
 901  AGCGCGATCG ACGCCTACCA GCGGCATTAC CGCGACGTGC TCAAGCGGCA
           PvuI
 951  GCGTGGCGGC GAGGTGGACT TGTCCCGCGT GGATTCCATG ATTGCCGTGC
1001  GTATGCGCGT CACCGGCCAC GATCAAGCGG CCATCGAGGG CGCTATCCGC
1051  CAGTGCGCAC CGGCCACCCG GCAGAAAGAC GAGGGCCGCG ATTGGAACGA
1101  CTACGCGCAG CGCACCGCCC GCTATGCCTA CAGCGCCGCA CAGCCGCAAG
1151  CCGCCGATCT TGGCAAGTAC CGGCAGCAGT GGGAGAAGCT GGAAGGGCGC
1201  GAGCCTGTAC GACAGCAGGA GCAGGCAAAG GCGCAGAAGA TCGAGCGCGA
1251  CAACTCGCCG GGAATGAGTC TCTAGCGTTG CGTGGTGGTT GTGATATACT
1301  TGTATAGCGT TTTCAGAACA GGAGCCGAAA CATGCTTGCA ATCCGACTGC
                                  ORF3    >
1351  CCGCCGAAGT GGAAACCCGC CTTGAAGCAC TGGCGCAGGC CACAGGGCGG
1401  ACCAAGACTT TCTATGCCCG CGAAGCCATC CTTGAGCACT TGGATGACCT
1451  CGAAGATTTG TACCTTGCAG AGCAACGCCT GATCGACATT CGCGCAGGCA
1501  AAACCCAAAC CGTGCCACTC GAAGAAGTGA TGAAACGCTA TGGCATGGAA
                            ORF4     >
1551  GGTTGAACTC GACCCAGCCG CCGAGCGCGA GCTAGGCAAG ATCGACCAGC
```

```
1601  AGACCGCCCG CCGCATCCTC GCTTTTTTGC ATGGCCGTGT CGCCCAGCTC
1651  GACGACCCGC GCAGCATTGG CGAAGCCCTC AAAGGCTCCA AACTGGGAGC
1701  CTTCTGGAAA TACCGCGTTG GGGATTGGCG AATCATCGCC AGCATCGAGG
1751  ACGGTGCTTT GCGCATCCTC GTTATGCGCA TCGGCAATCG TAAGGAGGTT
1801  TACCGCCAAT GATCGAATAC AGCTACCAGA TCGACCCGCG CCCCTCCGAC
        ORF5    >
1851  CTTGGCGGCG GCTGGCGGTT GCGCCTGTTG GAAAGCGGCG AGGAAGTCGG
1901  CGGCGGAGTG TTCCCGTTGT CCGAGTACGC CACAGCAGAG AACGCAGAAG
1951  AAGCGGCCAC GTACGCCTAT GAGGACGCCT TGGCCGAGGC TTCGGCGTGG
2001  CTGGCATCGA GGGGCGAAAA TTGAGCGGCG CGGCAGGGGA TTGCGGCCCC
2051  GGCAGCGCCT AACCACAACT GTCTGAAAAG GAGACAAGCA TGGCTTTAGA
                                              RepA   >
2101  CATTATGGCG GCCTTCACCA ATGAGCCGCC AGAACTTGAT TTCATCTGGC
2151  CCGGATTCTT GGCCGGAACC GTGGGCGCAC TTGTCGCACC TGGCGCAACT
2201  GGCAAGAGCT TTTTTGCTCT TGAAGCGGCC ATGTCAATCG CTTGCAGTGT
2251  GGCAGGCGGC GACCTTGTGG GACTAACCCC GGCGCACACC GGGCGCGTGG
2301  TTTATCTCGC TGGCGAAGAT CCACAGCCCG CCCTTGTGCG ACGTGTCCAC
2351  GCCATCGGCC AGCACCTCAA CCAGTCGGCC CGCGAAGCCA TCGCTGAGAA
2401  CCTGATGCTT GAGCCGATCA TGGGCAAGCG GCTAAACGTG ATGGACGACG
                                              ApaI
2451  CGCACTTGCG CCGCGTCATC GACTACAGCG CAGGGGCCCG GCTGATTGTG
2501  CTGGACACCC TGAGCCGGAT TCACATCCTC GACGAGAACA GCAATGGCGA
2551  CATGGCCCAC CTTGTTTCCG TGTTGGAACA CATCGCGGCG ACCACCGGGG
2601  CGGCTGTCCT GTACCTGCAC CACGTCAACA AGGGCAGCGC CCGCGACGGC
2651  CAGACCGACC AGCAGCAGGC AGCGCGGGGC GCGTCTGCCC TGATCGACAA
2701  CGCCAGATGG TGCGGCTATG TCGCCAAAAT GACGGAGCAG GAAGCCGAGC
2751  GCATGAGTGA CCGGGGCTTT GATCGTTCGC CATCCGGCAA CGAGCGGCGC
2801  GGCCTTTTTG TCCGCTTTGG CGTGAGCAAG CAGAACTACG ACGCGACCCC
2851  GCTAGACCGC TGGTATCAGC GGCACAGCGG CGGGGTGTTG TTGCCCGTTG
2901  AACTACAGGA GGCAATCAGC AATGGAGCAG GAAAAAAAGG GGGAAAGCGC
                          RepC   >
2951  AATGAGCTAT GACCTCACCC ATGCGCGGCA CGACCCCGCG CATTGCCTCA
3001  CGCCGGGGCT TTTCCGCAGT CTCAAGCGCG GAGAACGAAA GAGGCTCAAG
3051  CTCGATGTGA CCTACAACTA CGGAGATGAC TCAATCCGTT TTTGGGGGCC
3101  TGAACCACTT GGCGGCGATG ACTTGCGCGT ATTGCAAGGG CTGGTGGCAA
3151  TGGCTGCAAT TTCCGGAGAT AACGGGCGCG GCATCGTGCT ACGGCACGAA
3201  ACGGAATCAG AAGCAGGCCA GCAACTCCGC CTATGGCTTG ATATGCGGTG
3251  GGACGCCATA GAGAAAGATA CGATGGTAGC CAAGGGCAGC TTCCGCCAGT
        AvaI
301   TGGCCCGAGA ACTTGGCTAC GCCGAAGATG GAGGAAGTCA GTTTAAAACC
3351  ATCCGGGAAA GCATCGAACG GCTTTGGGCG GTATCGGTGA TTGTCGAAAG
3401  AGGTGGTAAG CGGCAAGGGT TCCGCATTCT GTCCGAGTAC GCGAGCGACG
3451  AGCAAGAAGG CAAGTTATTT GTTGCGCTTA ATCCCCGGCT GGCGGATGCG
3501  GTTATGGGAG AGCGCCCGCA CACCCGCATC AACATGGCAG AAGTTCGCAA
3551  GCTGGAAACA GACCCGGCAC GGCTGCTACA CCAGCGGCTA TGTGGCTGGA
3601  TTGACCCCGG AAAGTCTGGC AAAGCTGAAA TCGACACGCT GTGCGGTTAT
```

```
3651  GTATGGCCAG ACGCAGCCAA CGATGAAGCA ATGAAAAAGC GCCGCCAGAC
3701  CGCGCGCAAG GCGCTTGTCG AGCTTGCCGC CGTTGGTTGG ACGGTGAACG
3751  AGTACGCCAA GGGCAAGTGG GAAATCAGCA GGCCCAACCC CCGGCGTAAC
3801  GTTCCCCAAC CCCCGGCGTA ACGTTCCCCA ACCCCCGGCG TAACGTTCCC
3851  CGCCGAAATC TGAAAAACCT AGCAACGGCG CGGGTTTGCG GGCGATTCGC
3901  AAAATCCCTC CATGATCTAT CCAAGATCAT CCACTAGGCG CGGTACTTTT
3951  CGCCGCCCTT GAGGGCGTCA AAAGTCCTTA CCAAAAACCC CAAAAGAAAC
4001  GGCCGGGACA AGCCCGGCCA GACACACCCC GCCCCGCCCC GCTCGCCTTC
4051  ATTCTTCCAC CGGGACAATG GACACCATCA CCCGGTAGCG TTTGGCCTCT
4101  CCGGCAGGTA GCGCAGCGGC CAGCTTGGCG AGCGTTTCGG CTGGCTTGTC
4151  GGTCGTGCTT GTGGAGCACA TCGCCTCATA CCCGAACAGA AGCCATCAGA
4201  ATCGCCTACA GCGGATTTTT GGATGTTCTG GCTGCCTTGA GCTAGGGTTG
4251  GTAAAGAAAA CGCCTATGGC TGTTTGCGGG GCTTCTGCGA GCATTGCCGG
4301  GACGGTCTTG GGCTTGCTTG TGCCGTTGAG GCGAAAAACG CCACCGCCAG
4351  GACAAGCAGG GGTCGTCTCA GAAAACGGAA AATAAAGCAC GCTAAGCCGG
4401  TTGCAGCGGC GGTAGCGGCC TGAACTCGCC CGCGCCGATC
                                                  <  TnpA
```